Modularity in Empire Networks

Daozheng QU

A DISSERTATION

Submitted to

The University of Liverpool

in partial fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

Abstract

This project is research that produces practical applications related to the professional field. A graph is drawn on a plane so that different sides do not overlap each other is a plane network. If we let it randomly generate nodes and edges that are not repeated, a random planar network will be formed. Every network has its own structure, and random planar networks are no exception. Its structure consists of a single node. According to the theme of our project, we can know that we are going to study the modularity of the empire network, so what is the empire network? In fact, it's very simple. It is generated according to the definition in our project and based on the framework of a random planar network. We must define a concept called r, where r represents the number of nodes to be merged (random planar network). I generate different empire networks by changing the value of r. The empire network is not a floor plan, so there are no different sides. The non-intersecting nature can also represent the structure of the transformed random planar network. In order to distinguish the nodes of the empire network from the nodes of the random planar network, we call them vertices.

Then we carry out modular operations. Modularity refers to treating

each node of the empire network as a module, merging two vertices at each step, and finding a module at each step. The module finds the value of the optimal modularity of each part according to the definition of the modular network. In other words, each step is the optimal step. When selecting the next module, merge the modules obtained in the previous step to generate a new module, and then repeat the next step until there is no module that can be merged. The maximum modularity value (Q value) is the best structural division. Regardless of whether the modularization is processed to the last step. Finally, we observe the average value of the final modularity and see the impact of the change in the structure of the empire network on the value of modularity due to the subsequent change in the number of planar network nodes.

Student Declaration

I confirm that I have read and understood the University's Academic Integrity Policy.

I confirm that I have acted honestly, ethically and professionally in conduct leading to assessment for the programme of study.

I confirm that I have not copied material from another source nor committed plagiarism nor fabricated data when completing the attached piece of work. I confirm that I have not previously presented the work or part there of for assessment for another University of Liverpool module. I confirm that I have not copied material from another source, nor colluded with any other student in the preparation and production of this work.

I confirm that I have not incorporated into this assignment material that has been submitted by me or any other person in support of a successful application for a degree of this or any other university or degree-awarding body.

SIGNATURE -----Daozheng_QU-----------

DATE October 08, 2021

Acknowledgments

I am very grateful to my first supervisor for his careful guidance. I have gained a lot of academic skills and practical experience during the project design and operation, which gave me a lot of insights and help. In academic terms, my first supervisor performed a lot of calculations with me, guided me to write the correct code of the project, brought me back to the right track many times, and sacrificed weekend time in life many times to discuss academic issues with me and follow the lead.

I am very grateful to my second supervisor, who provided a lot of valuable comments on my original design and pointed out many shortcomings in the final display. I have the opportunity to improve my project and improve my academics.

I am very grateful to my senior and my friend Chen Sheng for his help. The single and only line of code on the create the empire network, that is, the traversal of the empire network, gave me valuable hints and help, and I can avoid some unnecessary errors (the origin of the node of begin is 0 or 1, and how to traverse the array of edges represented by the empire network).

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Scope

My duty is that mainly conducts experiments and research on known algorithms and theories, and does not involve them, but I use a new conversion algorithm to convert the original random planar network into another network structure based on the existing algorithms and theories. This is also our innovation of the known theory and a new exploration of the original modularisation theory. In order to make the project plan successfully complete the basic tasks and challenges, first, use the Markov chain algorithm to generate a random planar network (small), and use the algorithm of the project theme to get the empire network (small) from the random planar network, and then modularize the empire network. I get the community structure division and modularity value of the empire network, and the basic task ends here. Subsequently, I conducted a large number of random planar network (large) experiments with different numbers of nodes to obtain different community structure divisions and average modulalrity values as well as their distribution laws.

1.2 Problem Statement

I successfully solved the problem of converting the graph network. We converted the random planar network into another network (the empire network) and performed modular processing.

By analyzing different modular attributes, we obtain the modular attributes of the empire network under the same number of vertices in the same interval generated by random planar networks under different interval nodes, and I find the distribution law of modularity values.

According to a large number of experiments, we observe the changes brought about by different network topologies and the reaction to the results of modular processing with the changes in the number of nodes.

Finally, I completed the theme requirements and got self-improvement. I successfully known algorithms were improved, run, tested, and succeeded in the project. And conducted a lot of experiments to analyze the distribution law of modularization, confirm the hypothesis and draw relevant conclusions.

1.3 Solution produced

1.3.1 Basic job :

Generate a small(100 nodes) random planar network [1,2], and convert it into an empire network according to the content of the empire network conversion material given by the project, and find the empire network whose outline structure is based on r=2 (50 vertices), and show its details about calculating the modularity value Q.

1.3.2 Challenge :

According to the plan and idea of the basic task, adjust the value of r (r = 2, 3, 5), generate a random planar network with different numbers of nodes, and convert it into a consistent number of vertices of the empire network (100-1000) vertices. Observe the changes in the community structure and the value of modularity after the modularization process, and conduct multiple experiments to observe the distribution of the modularity value.

1.4 The effectiveness of the solution

We have made adjustments to some details of the original plan over time and actual needs, such as changing the number of runs of the experiment. According to the author's Newman's proof[3,9], the formula was revised in the design. We have not changed the algorithm selection. The algorithm selected by the original design is the most appropriate and accurate. We have accurately proved the modularity and verified the characteristics of a highly modular network with dense nodes and sparse nodes connected by different modules. At the same time, we also observed the impact of network structure changes on modularity based on experiments. As well as the influence and distribution of different topological structures on the value of modularity due to changes in the number of merged nodes. Our design and the selected algorithm are feasible, effective, correct, which enables us to complete the program design of the basic goals and the more difficult goals and run the program smoothly. Generally speaking, the project has been completed, but it takes time to run more results.

Chapter 2

Background

2.1 Project information of skill and material

In this project, I plan to study the typical modularity of a class of networks inspired by cartographic applications through calculations. The implementation of this project involves programming any well-known method to generate a relevant network and then programming a simple method to approximate the network modularity value[4,16]. However, the main focus of the project is the subsequent empirical study of the modular statistical distribution of randomly sampled planar networks as a whole. So I must first read the materials about generating random networks, which are from the network and recommended by the project supervisor. Then I need to understand what a modular network is and what the role of the modular network is. The materials come from the COMP324 course and network of the University of Liverpool, and the materials recommended by the project supervisor. I need to study the main purpose of the project, how to convert the random planar network into an empire network, and the material that comes from the main project material.

I need to learn about programming and mathematics-related knowledge and skills because this project involves algorithms and mathematics, so I need to understand the use of various Python libraries and what can help our project and understand the matrix in programming, and how to use it. Because this project involves a large number of network nodes and edge operations[13,15], it is impossible to leave the matrix[11,12].

I need to evaluate the potential of the project development and further improve the potential. First of all, we use a known algorithm with mature experience and practice. We can compare the project's algorithm with other algorithms to generate a more effective solution to achieve commercialization in the future Possible. Because modularity is limited by resolution, it cannot detect small communities [14], so I can try to improve it. Although the maximum template degree method has its well-known shortcomings, it is still one of the most widely used community detection algorithms. Finally, I learned about the skills of speeches, the skills of making PPT, the methods of project design, and other information, in order to complete the project and make efforts for the further development of the project in the future.

2.2 Project requirements

2.2.1 Basic aim:

Generate a small(50-100 nodes) random planar network [1,2], and convert it into an empire network according to the content of the empire graph conversion material given by the project, and find the empire graph whose outline structure is based on r=2 (25-50 vertices), and show its details about calculating the modularity value Q.

2.2.2 Challenge aim:

Generate a number of random planar networks with different range nodes and transform to the empire network at different r values, and repeat the above basic experiment 20 times. I will find the average Q value and analyze the distribution of the Q value.

Chapter 3

ETHICAL USE OF DATA

All the data used in this project is generated by this experiment. It does not use any data of any other person, institution, or third party, does not involve any ethical related data, and does not design any biological or human data, and does not use any data. Any data that violates university education guidelines and British Computer Association guidelines. All the data here is always independently generated by this project in the experiment, independent research, and independent comparison.

This project will generate a large amount of data, and I guarantee that the data obtained is true, reliable, effective, and timely. The data of the experimental results and the data of the control experiments were generated during this project, and all experiments were carried out in strict compliance with the academic integrity guidelines.

Chapter 4

Design

4.1 Original design

4.1.1 Design of the project process:

First of all, we understand the connotation of modularization and basic design concepts and ideas according to the main content of the project, and then divide the design into three main aspects according to the main requirements.

Construct the random planar network and then adjust the r value according to the needs of different difficulty according to the random planar network transformed into the empire network, and finally perform the modularization process to obtain the value of community division and modularity and the average modularity value after a large number of experiments.

4.1.2 Original plan of design

Basic aim:

Generate a small(50-100 nodes) random planar network [1,2], and convert it into an empire network according to the content of the empire graph conversion material given by the project, and find the empire graph whose outline structure is based on r=2 (25-50 vertices), and show its details about calculating the modularity value Q.

Challenge aim:

Generate a number of random planar networks with 1000-10000 nodes at an interval of 1000 nodes, and repeat the above basic experiment 20 times. In the case of r = 2,4,5, take the average Q value and analyze the distribution of the Q value.

4.1.3 Design of the project algorithm:

4.1.3.1 Markov chain：

The state of each step in a random walk is a point in the graph. According to Markov's definition, the random walk of each step can move to any adjacent point, the movement probability is the same, and there is no need to consider the change of its moving path and the influence of other factors. We construct all plane sub-graphs of a Markov chain graph with stable and uniform distribution. In the case of a complete graph, our experiments show that a random simple planar graph on n vertices is connected but not 2-connected, and has about 2n edges[1].

Markov chain algorithm of generate the random planar network(The following algorithms are derived from the application of Markov chain algorithm in [1])：

According to the definition G = (V; E) is a simple empty graph that sets the number of nodes. We are at time t, then X_{t+1} represents the time relationship between the number of simulation iterations and the number of nodes. First, select the position to be represented by i, j (a position is composed of a pair of different vertices in G). If G already contains this position, delete it, otherwise add this position and continue to run. If it maintains the flatness of the graph, The position of G is randomly and uniformly selected. If all the plane subgraphs are obtained when the state space is G, then we finally get the plane graph G, which is a random planar network.

4.1.3.2 The algorithm of fast algorithm for detecting community structure in networks(The following algorithms are derived from the application of Fast Newman algorithm in [3]).

The algorithm of this design is based on Newman's definition of modularity, using the algorithm demonstrated in the 2004 paper. The algorithm of this paper is derived from the author's comparison with the Girvan Newman algorithm.

To test whether a particular division is meaningful we define a quality function or “modularity” Q as follows [8].

Let e_{ij} be a fraction in the network about the edge connecting the vertices in two groups i and j, and let ai = $\Sigma_j e_{ij}$ [18]. The formula for Q value is Q = $\Sigma_i (e_{ii}-a_i^2)$. Modular Q is defined as the fraction of edges that fall into group 1 or group 2, minus those in groups 1 and 2 that have the same node degree distribution as the given network The number of expected edges of the random graph. Values other than 0 indicate deviations from randomness. In fact, values greater than 0.3 seem to indicate significant community structure [3].

Joining a pair of communities with no edges will never lead to an increase in Q. We only need to consider those pairs with edges between them, where there is at most m at any time, where m is again the number of edges in the graph.

After joining a pair of merged communities, some matrix elements of e_{ij} must be updated by adding the rows and columns corresponding to the joined communities. The most time required at this time can be expressed as O(n). Thus each step of the algorithm takes worst-case time O(m + n)[19].

The algorithm has the added advantage of calculating the value of Q as it goes along, making it especially simple to find the optimal community structure[19].

4.2 Changes to original design

In the case that the basic algorithm does not change, our basic tasks have not changed, but in the modularity algorithm, we have changed many of our formulas because of the update and revision of the author Newman himself.

In practice, we found that the modularity algorithm will have the same situation when calculating the maximum increment and minimum decrement. We will compare again to get the optimal Delta Q.

We have changed the settings of the original planned challenge level tasks. I reduced the number of trials to ensure that the experiments can be completed as scheduled, and changed the number of nodes in the random planar network from the original fixed range to the number of vertices in the empire network.

Refer to the appendix for the specific details and pseudo-code.

Chapter 5

REALISATION

5.1 Implementation

In the early stage of the project, we understood the main materials of all projects, developed an understanding of the basic knowledge of random plane networks and modular networks, combined with existing project materials, carried out independent critical thinking on the subject and method of the project, and considered the development of the project. In the general direction, compare various methods such as Louvain algorithm [10], Girvan-Newman algorithm [9], fast Newman algorithm [3], fast Newman algorithm (updated version) [11], etc. [4]. Finally, after extensive comparison, we chose the main material algorithm fast Newman, carried out an original design and tried a preliminary program architecture design. After completing the original design of the project and the construction of the basic program architecture, I had a series of discussions with the supervisor of this project about the selection of modular algorithms and the conversion of random planar networks, as well as how to modify the experiment later.

In the middle of the project, my supervisor and I determined the final direction of the project, namely, the number of vertices of the fixed empire network, and modularized operations on random planar networks with different numbers of nodes and empire networks with the same number of vertices (converted by random plane networks). Instead of the original design, the number of nodes in the random plane network is fixed, and the number of vertices in the empire network is changed before modular processing. The experiment program is modified to some extent, and the modified programming program is used to redesign and reorganize the structure to obtain the final program design plan.

In the final stage of the project, we conducted a lot of experiments and got more reliable experimental results. Analyze and compare the experimental results, obtain the empire network generated by different r values and the random planar network generated by different node numbers, and observe the modularity value of different network structures after conversion, and observe the module as the number of nodes changes. Degree distribution and conduct statistical research on it. By the end of the project, complete the thesis and procedures, meet the project requirements, and complete the basic tasks and challenges.

Refer to the appendix for the specific date of the project.

5.2 Testing and Output

In order to highlight the accuracy and reliability of the experiment, we conducted a lot of comparative experiments. First, we conducted an experiment with r = 2, a random planar network with 200-2000 nodes and transformed it into an empire network with 100-1000 vertices, and then obtained the average modularity of the empire network after being modularized. The output is shown in Tables 5.2.1 and 5.2.2.

5.2.1 The number of nodes in the random planar network corresponding to 100-1000 vertices (interval 100) in the empire network at r=2.

Vertices of empire network	Nodes of random planar network
100	200
200	400
300	600
400	800
500	1000
600	1200
700	1400
800	1600
900	1800
1000	2000

5.2.2 The empire network is at r=2, and the different average modularity values corresponding to 100-1000 vertices (interval 100) are in 20 trials.

Nodes/r(r=2)	Average Q value
100	0.449
200	0.546
300	0.617
400	0.635
500	0.679
600	0.700
700	0.710
800	0.717
900	0.744
1000	0.751

Then the random planar network with r = 3 and the number of nodes 300-3000 is transformed into an empire network with 100-1000 vertices, and then the average modularity of the empire network is obtained. The output is shown in Table 5.2.3 and 5.2.4.

5.2.3 The number of nodes in the random planar network corresponding to 100-1000 vertices (interval 100) in the empire network at r=3.

Vertices of empire network	Nodes of random planar network
100	300
200	600
300	900
400	1200
500	1500
600	1800
700	2100
800	2400
900	2700
1000	3000

5.2.4 The number of nodes in the random planar network corresponding to 100-1000 vertices (interval 100) in the empire network at r=3.

Nodes/r(r=3)	Average Q value
100	0.339
200	0.426
300	0.455
400	0.491
500	0.508
600	0.512
700	0.528
800	0.551
900	0.559
1000	0.565

Then the random planar network with r = 5 and the number of nodes 500-5000 is transformed into an empire network with 100-1000 vertices, and then the average modularity of the empire network is obtained. The output is shown in Table 5.2.5 and 5.2.6.

5.2.5 The number of nodes in the random planar network corresponding to 100-1000 vertices (interval 100) in the empire network at r=5.

Vertices of empire network	Nodes of random planar network
100	500
200	1000
300	1500
400	2000
500	2500
600	3000
700	3500
800	4000
900	4500
1000	5000

5.2.6 The empire network is at r=5, and the different average modularity values corresponding to 100-1000 vertices (interval 100) are in 20 trials.

Nodes/r(r=5)	Average Q value
100	0.237
200	0.295
300	0.346
400	0.334
500	0.349
600	0.367
700	0.374
800	0.369
900	0.385
1000	0.402

Chapter 6

EVALUATION

6.1 Advantages

First of all, we have proved the feasibility of this project in a practical sense. We have successfully completed all tasks in project design and program design. I fully understand the rapid and accurate characteristics of the modular network in the random network. Knowledge to be learned in a field and further understanding of the impact of changes brought about by the network architecture on the value of modularity. We can use the empire network to reflect the structural feature changes of the random planar network, improve the design concept of the original plan, and highlight the key point in the modular theory, which is to detect the quality of the network structure.

We verified and tested the development, growth, improvement, and revision of many known algorithms, as if we could communicate with you pioneers in real-time academically, and brought me into a new field.

6.1.1 Regarding the correctness of algorithm selection:

6.1.1.1 Markov chain :

It also known as discrete-time Markov chain (DTMC [17]). This experiment has proved that the Markov chain algorithm for random planar networks is the best algorithm, because it not only maintains a high degree of randomness but also meets the needs of modular networks. It is a random process of transitioning from one state to another in the state space. This process requires the nature of "memoryless": the probability distribution of the next state can only be determined by the current state, and any events before it in the time series have nothing to do with it. This special type of "memorylessness" is called Markov chain properties. Markov chains have many applications as a statistical model of actual processes. At each step of the Markov chain, the system can change from one state to another according to the probability distribution, or it can maintain the current state[1]. The change of state is called transition, and the probability associated with the change of different states is called transition probability.

The characteristics of its random walk[20,21] are more prominent, which is a very important reason why we use Markov chain as the core algorithm. It is easy to verify that (X_t) is an irreducible aperiodic Markov chain whose transition matrix is symmetric[1].

6.1.1.2 Fast algorithm for detecting community structure in networks

Through the realization of random planar network algorithm and modular algorithm, we fused the known algorithm with the subject of this research to more realistically prove the operability of many known algorithms. At the same time, when we completed the project procedure, we proved the fast Newman algorithm reduces the cost of real optimization of Q. We not only increase the speed of community division but also maintain the accuracy of Q optimization. It has actually proved the reliability, superiority, and rapidity of the algorithm in the application of large-scale random planar networks, especially in terms of reducing time complexity.

6.2 Limits

In the research of network theory, complex networks (complex networks have characteristics that simple networks, such as lattice networks, random networks, etc.) do not have, and these characteristics often appear in real-world network structures. The research on complex networks is A hot spot in scientific research today is closely related to the research of various high-complexity systems in reality, such as Internet networks, neural networks, and social networks. However, the research of this project does not involve the modular operation of complex networks. It is very important in real life and research, but its huge number of nodes and intricate relationships between nodes constitute a network structure that makes the design of this project temporarily unattainable (although this has far exceeded the realization of this project).

Modularity is a measure of the quality of community division, and it is also a benefit function. The maximum modularity method detects communities by searching for one or more possible divisions of networks with particularly high modularity. Because it is very difficult to perform exhaustive search on all possible partitions, the actual algorithms are based on approximate optimization methods, such as greedy algorithms, simulated annealing[32] or spectral optimization. Different algorithms make different types of balances between detection speed and accuracy.

The effectiveness of modularity optimization is questionable, because modularity optimization is limited by the size of the resolution, making it difficult to detect a smaller cluster [19]. On the other hand, because the modularity value represents The degeneracy of a large number of partitions with high modularity is close to the absolute maximum, which may be very different from other algorithms.

Chapter 7

LEARNING POINTS

7.1 Independence

I learned how to independently complete the independence-related skills required for a project. As an independent worker in the field of computer science, I learned how to study, work, and research in a new field. Including how I learned how to find the main content of the project in the project, how to independently explore the content of the project domain and outside the domain. And to obtain independent design, operation, and independent analysis of the ability to deal with projects.

7.2 Thinking

I learned to have the ability to think critically. The supervisor is not completely correct. I can continuously improve my knowledge and understanding of the field in the project, propose a deeper understanding of my own, and have the development direction of my project. I have relatively mature opinions, and fully discuss the difficulty of the project, time management, and experimental complexity with the supervisor. I have a self-critical awareness of some definitions and procedures of the project through my own understanding. For example, my thoughts on the repeatability of Delta Q in modularity, and the new understanding of formulas for calculating modularity in program design.

7.3 Learning

I learned how to study new knowledge and skills, and the ability of how to apply new knowledge and skills. I have mastered how to start from a completely unfamiliar field and start a brand new project, such as the initial understanding, comprehensive understanding, and the entire process of completing the project independently. I know how to better learn and research the modularization of the network and its significance. Modular network applications are ubiquitous. For example, many large companies such as Facebook and Google have actively explored and developed in this field, which enabled me to gain sufficient industry knowledge through the research of this project.

7.4 Technical skills

I learned a lot of practical skills while entering the project, such as Python programming, how to use many libraries(networkx, numpy, etc) to complete the experiment, fully understand the application of mathematics[24,27] in programming, a lot of knowledge about graph networks and modular networks, so that I have the ability to become a worker in the field of computer science, better familiarize and control the tools, professionalism and solid knowledge structure required in modular processing.

7.5 Communication skills

I actively communicated with my supervisor during the project and expressed my main understanding of the project very well. Through communication, I also learned about my shortcomings, such as project design, programming, basic knowledge, and other areas that I did not notice. This will play a very important role in how I conduct technical communication and expression in my work and strengthen my narrative and expression skills on issues. To know that a good project wants to be recognized, it needs better expression skills to tell our audience what its advantages are and what kind of progress it can bring. This is of great significance to a project.

In the end, in the master project, I acquired the core skills I needed. Whether it is the academic field or industry, the help to me is decisive.

Chapter 8

PROFESSIONAL ISSUES

This project fully considers the all of elements, as well as the legitimate rights and interests of third parties, abides by various codes of conduct, and will not be based on gender, sexual orientation, nationality, color, etc. Discrimination is based on race, ethnicity, religion, age, disability, or any other factors or conditions. I am determined to promote equal access to the benefits of IT development, such as extending this project to the study of social networks and using professional knowledge to serve society.

In the future, I will consider how to conduct modular processing on network security and network privacy to promote the democracy and freedom of network users. I will consider using it for social models, the World Wide Web, metabolic networks, food networks, and nerves. In terms of network and pathology network, if there is an opportunity to commercialize it in the industry, I hope that this project will serve the disadvantaged groups so that they can use the network more equally.

This project will refer to certain modern development concepts, such as smart cities and the construction of smart networks, to continue to develop professional knowledge, skills, and abilities in the field of graph networks and modular theory. I will maintain an understanding of the relevant technical developments, procedures, and standards in this professional field, and I am very familiar with the relevant laws and regulations in this field and have been strictly abiding.

I highly respect and value different points of view, and actively discuss with my supervisor, and accept honest criticism of the work of this project. According to my professional judgment and project requirements, perform my professional duties with due care and diligence. Without the permission of the school and the supervisor, I will not disclose information for personal gain. I will not distort or conceal the status of the project. I will strictly maintain the reputation of the school and BCS and perform my duties. I will not accept any disgraceful behavior.

Chapter 9

CONCLUSIONS

9.1 Summary

I found that as the value of r increases, the number of nodes in the random planar network also increases, and it changes as the structure of the empire network changes. Under the same r value, the larger the random planar network has a relatively high Q value (reasonable interval: [-1, 1]), indicating that the relatively large random planar network and its topological network structure represented by the node degree are relatively more Stable, the community structure is more significant, the network structure is relatively good, and the community is better divided.

However, we see that with the large increase in the value of r and the subsequent large increase in the number of nodes in the random planar network and the increase in the empire network according to a fixed interval of vertices, the Q value will instead be in a relatively low optimal interval (optimal interval: 0.3 -0.7), and when the r value is the smallest, as the number of nodes in the random planar network increases and the empire network increases in a constant interval of vertices, the Q value will overflow the optimal interval in a small amount, but it does not overflow too much and is still within a reasonable range, it shows that the structural loss is small in the process of converting to the empire network. The topological network structure represented by the node degree is relatively stable, the community structure is the most significant, the network structure is the best, and the community has the best division. And the highest Q value.

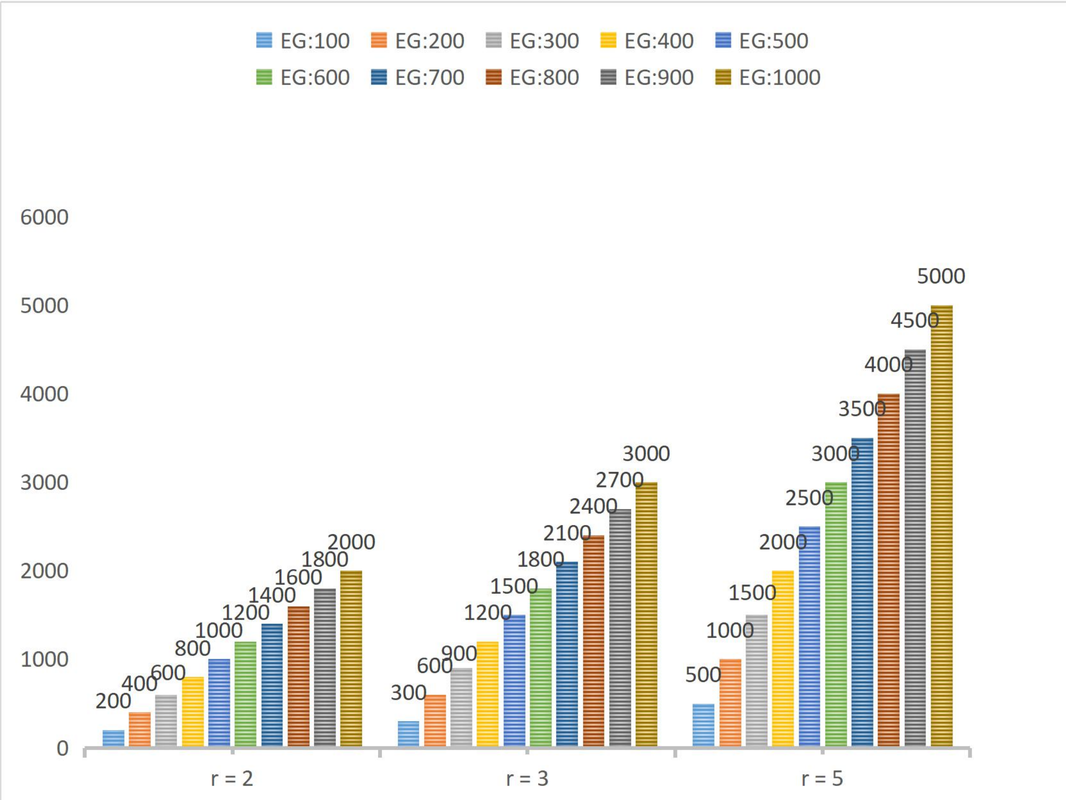
Finally, this experiment proves that we should not only pay attention to the impact of the number of network nodes on the value of modularity but also pay attention to the more important factor of the modular network theory and the fundamental definition of good network topology.

9.2 Main findings

According to the experimental results of the project, I found three phenomena:

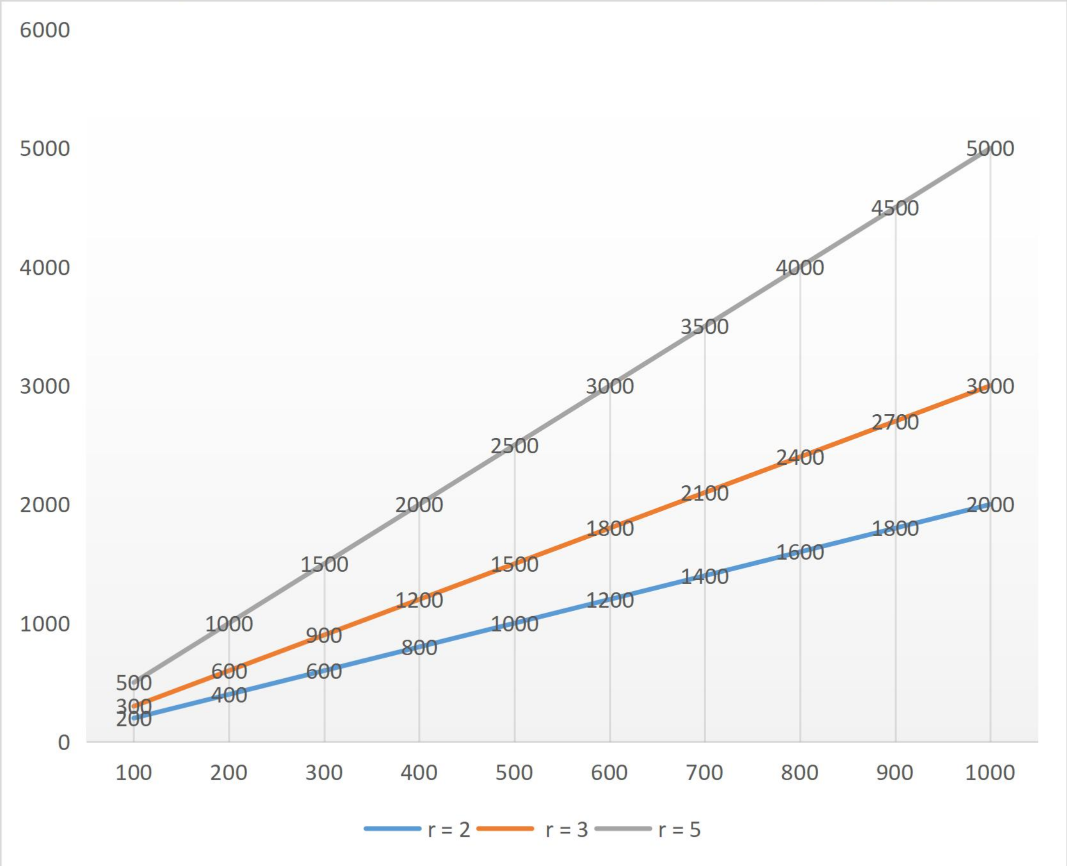
Phenomenon 1: In the case of the same value of r (for example, r=2), when converting a random planar network to an empire network, as the number of vertices in the empire network increases, the number of nodes in the random planar network also needs to increase significantly. As shown in Figure 9.2.1.

9.2.1 Planar graph nodes of different r numbers & empire graph vertices.



Phenomenon 2: In the case of different values of r, the empire network with the same number of vertices decreases significantly as the value of r increases. The larger the r, the larger the number of nodes in the random planar network, and the more nodes will be lost when converting to an empire network. As shown in Figure 9.2.2.

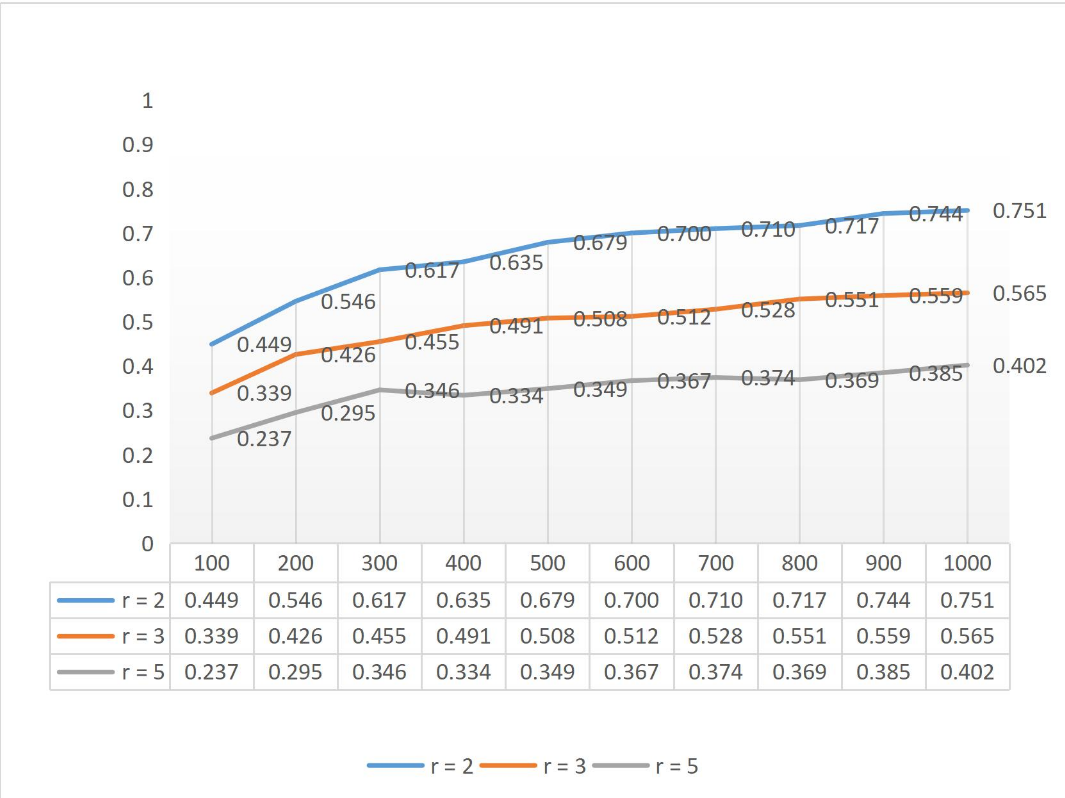
9.2.2 Planar graph node's tends of different r numbers & empire-graph-vertices.



Phenomenon 3: The greater the value of r, the smaller the value of modularity. Under the same r-value, as the number of random planar network nodes increases, the number of vertices in the empire network also increases, and the value of modularity also increases. But under the condition that r is equal to 2, the imperial network after 700 vertices, the random planar network of 1400 nodes represented by it, is beyond the range of the good modularity value. When equal to 3, the empire network of all vertices and the random planar network it represents are in the range of good modularity values (good Q value range: 0.3-0.7). When r is equal to 5, the average modularity value of

the 400-vertex empire network has some contrast. Although it is a normal phenomenon, it also shows a relatively unstable trend. We can see that the smaller the r value and the increase in the number of empire network points, the higher the increase in its Q value and the greater the rate of increase. As shown in Figure 9.2.3.

9.2.3 Q value tends of different empire network and r value.



	100	200	300	400	500	600	700	800	900	1000
r = 2	0.449	0.546	0.617	0.635	0.679	0.700	0.710	0.717	0.744	0.751
r = 3	0.339	0.426	0.455	0.491	0.508	0.512	0.528	0.551	0.559	0.565
r = 5	0.237	0.295	0.346	0.334	0.349	0.367	0.374	0.369	0.385	0.402

9.3 Directions for further work

I will first conduct more experiments with different numbers of nodes and random planar networks[26,25] and r values to observe where the limits of network conversion are. For example, r is equal to 1000 or even 10000. Observe where the limits of the merged network structure are, and then what distribution characteristics the average modularity of the empire network generated in the limit state has.

Find a way to lift the resolution limit or explore a new algorithm by yourself to overcome this historical problem, so that the problem of the inability to modularize in a small random network can be solved.

I will analyze new algorithms to speed up the modular operation of large random networks. Many scholars have proposed methods and algorithms for speeding up, and have experimented with them. I will also give my own new design solutions. In the future, the speed of modular operation will be improved. Make more contributions to this field, and combine the ideas of many scholars to start research on the pruning[31] of random networks, and the pre-optimization of fuzzy[22,23] networks and overlapping networks.

I will try to use all of what I have learned to challenge the biggest problem in history, how to truly achieve global optimization, instead of simply relying on the idea of a greedy algorithm[28,29] to complete all local optimizations to achieve simulated global optimization[30].

Chapter 10

Bibliography

1. Denise, Alain and Vasconcellos, Marcio and Welsh, Dominic. (1996). ‘The Random Planar Graph’, Congressus Numerantium, 113, 61-79.

2. Bollobás, Béla. (2001). Random Graphs (2nd ed.). Cambridge University Press.

3. M. E. J. Newman. (2004)‘Fast algorithm for detecting community structure in networks’，Phys. Rev. E，(69). 066133.

4. M. E. J. Newman. (2006). ‘Modularity and community structure in networks’. Proceedings of the National Academy of Sciences of the United States of America. 103 (23): 8577–8696.

5. Dodge, Y. (2003) The Oxford Dictionary of Statistical Terms, OUP.

6. Gagniuc, Paul A. (2017). Markov Chains: From Theory to Implementation and Experimentation. USA, NJ: John Wiley & Sons, 159–163.

7. Jan A. Snyman. (2005). Practical Mathematical Optimization: An Introduction to Basic Optimization Theory and Classical and New Gradient-Based Algorithms. Springer Publishing.

8. M. E. J.Newman. (2003), The structure and function of complex networks. SIAM Review 45, 167–256.

9. M.Girvan and M. E. J. Newman. (2002). Community structure in social and biological networks, Proc. Natl. Acad. Sci. USA 99, 7821–7826 .

10. Blondel, Vincent D; Guillaume, Jean-Loup; Lambiotte, Renaud; Lefebvre, Etienne. (9 October 2008). "Fast unfolding of communities in large networks". Journal of Statistical Mechanics: Theory and Experiment. 2008 (10): P10008.

11. Newman, M. E. J. (2007). Palgrave Macmillan, Basingstoke (ed.). "Mathematics of networks". The New Palgrave Encyclopedia of Economics(2 ed.).

12. Brandes, U.; Delling, D.; Gaertler, M.; Gorke, R.; Hoefer, M.; Nikoloski, Z.; Wagner, D. (February 2008). "On Modularity Clustering". IEEE Transactions on Knowledge and Data Engineering. 20 (2): 172–188.

13. van der Hofstad, Remco. (2013). "Chapter 7" (PDF). Random Graphs and Complex Networks.

14. Santo Fortunato & Marc Barthelemy. (2007). "Resolution limit in community detection". Proceedings of the National Academy of Sciences of the United States of America. 104 (1): 36–41.

15. Valdez, LD; Braunstein, LA; Havlin, S. (2020). "Epidemic spreading on modular networks: The fear to declare a pandemic". Phys. Rev. E. 101 (3): 032309.

16. Alex Arenas, Alberto Fernández and Sergio Gómez. (2008). "Analysis of the structure of complex networks at different resolution levels". New Journal of Physics. 10 (5): 053039.

17. Norris, James R. Markov chains. Cambridge University Press. 1998.

18. As discussed in [8], each edge should contribute only to eij once, either above or below the diagonal, but not both. Alternatively, and more elegantly, one can split the contribution of each edge half-and-half between eij and eji, except for those edges that join a group to itself, whose contribution belongs entirely to the single diagonal element eii for the group in question.

19. M. E. J. Newman. (2004), "Fast algorithm for detecting community structure in networks". Department of Physics and Center for the Study of Complex Systems, University of Michigan, Ann Arbor, MI 48109–1120.

20. Wirth, E.; Szabó, G.; Czinkóczky, A. MEASURE LANDSCAPE DIVERSITY WITH LOGICAL SCOUT AGENTS. ISPRS – International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 2016-06-08, XLI–B2: 491–495.

21. Pearson, K. The Problem of the Random Walk. Nature. 1905, 72 (1865): 294.

22. Novák, V.; Perfilieva, I.; Močkoř, J. (1999). Mathematical principles of fuzzy logic. Dordrecht: Kluwer Academic.

23. Zadeh, L . A. (1965). "Fuzzy sets". Information and Control. 8(3): 338–353.

24. Yan, Di; Wu, Tao; Liu, Ying; Gao, Yang. (2017). An efficient sparse-dense matrix multiplication on a multicore system. IEEE.

25. Trudeau, Richard J. (1993). Introduction to Graph Theory (Corrected, enlarged republication. ed.). New York: Dover Pub. Retrieved 8 August 2012. A graph is an object consisting of two sets called its vertex set and its edge set.

26. Trudeau, Richard J. (1993). Introduction to Graph

Theory(Corrected, enlarged republication. ed.). New York: Dover Pub. Retrieved 8 August 2012. Thus a planar graph, when drawn on a flat surface, either has no edge-crossings or can be redrawn without them.

27. Bender, Edward A.; Williamson, S. Gill. (2010). Lists, Decisions and Graphs. With an Introduction to Probability.

28. Cormen, Thomas H.; Leiserson, Charles E.; Rivest, Ronald L.; Stein, Clifford (2001). "16 Greedy Algorithms". Introduction To Algorithms. MIT Press. pp. 370–.

29. Gutin, Gregory; Yeo, Anders; Zverovich, Alexey. (2002). "Traveling salesman should not be greedy: Domination analysis of greedy-type heuristics for the TSP". Discrete Applied Mathematics. 117 (1–3): 81–86.

30. R. Horst, H. Tuy. Global Optimization: Deterministic Approaches, Springer, 1996.

31. Blalock, Davis; Ortiz, Jose Javier Gonzalez; Frankle, Jonathan; Guttag, John. (2020-03-06). "What is the State of Neural Network Pruning?".

32. Weinberger, E. (1990). "Correlated and uncorrelated fitness landscapes and how to tell the difference". Biological Cybernetics. 63 (5): 325–336.

Chapter 11

APPENDICES

A Full code listing

B Screen shots of sample runs

B.1 Basic aim

B.2 Challenge aim

C Details of test data

D Important dates in the projects

E A user guide to installation

F usage of the software

G full design diagrams

G.1 Project plan of date

G.2 Detailed introduction of the design time plan

G.3 Pseudo-code

G.4 Design drawings and other design documents: