Project Review-I

on

Ramsey Numbers using Social Media Analytics

by

Swastik Banerjee (RA1711008010239)

Under the guidance of

Dr. D. Hemavathi

Professor Department of IT

Introduction

In this Combinatorial Graph Theory and Social Media Analytics research project, we aim to improve the upper bounds and lowers bounds of Ramsey Numbers $R(r,s)$ with $r,s > 5$ using social media analytics.

What are Ramsey Numbers?

- The numbers R(r, s) in Ramsey's theorem (and their extensions to more than two colours) are known as Ramsey numbers. The Ramsey number, R(m, n), gives the solution to the party problem, which asks the minimum number of guests, R(m, n), that must be invited so that at least m will know each other or at least n will not know each other. In the language of graph theory, the Ramsey number is the minimum number of vertices, v = R(m, n), such that all undirected simple graphs of order v, contain a clique of order m, or an independent set of order n. Ramsey's theorem states that such a number exists for all m and n.

- By symmetry, it is true that R(m, n) = R(n, m). An upper bound for R(r, s) can be extracted from the proof of the theorem, and other arguments give lower bounds. (The first exponential lower bound was obtained by Paul Erdős using the probabilistic method.) However, there is a vast gap between the tightest lower bounds and the tightest upper bounds. There are also very few numbers r and s for which we know the exact value of R(r, s).

- Computing a lower bound L for R(r, s) usually requires exhibiting a blue/red colouring of the graph KL−1 with no blue Kr subgraph and no red Ks subgraph. Such a counterexample is called a Ramsey graph. Brendan McKay maintains a list of known Ramsey graphs.[6] Upper bounds are often considerably more difficult to establish: one either has to check all possible colourings to confirm the absence of a counterexample, or to present a mathematical argument for its absence.

Our Background of Research and Objective

The requirement for this research project is to build two-coloured graphs using social media networks . Each node in the graph will represent a person/user of the social media, connected to another node using either of the two colours, e.g. blue and red, one colour representing both of them are connected to each other using the social media, and the other colour representing the two of them not being connected to each other via the social media. Our aim is to find, through statistical experiments, the minimum and maximum number of nodes needed, i.e, the interval with highest probability to find monochromatic cliques each of edges 6,7,8 etc respectively.

Motivation

The computational complexity of determining the values of Ramsey Numbers more than 5 is so difficult that Joel Spencer had once said, "Erdős asks us to imagine an alien force, vastly more powerful than us, landing on Earth and demanding the value of R(5, 5) or they will destroy our planet. In that case, he claims, we should marshal all our computers and all our mathematicians and attempt to find the value. But suppose, instead, that they ask for R(6, 6). In that case, he believes, we should attempt to destroy the aliens." A sophisticated computer program does not need to look at all colourings individually in order to eliminate all of them; nevertheless it is a very difficult computational task that existing software can only manage on small sizes. Each complete graph Kn has 1/2n(n − 1) edges, so there would be a total of c^(n(n − 1)/2) graphs to search through (for c colours) if brute force is used. Therefore, the complexity for searching all possible graphs (via brute force) is O(c^n^2) for c colourings and at most n nodes. The situation is unlikely to improve with the advent of quantum computers. The best known algorithm exhibits only a quadratic speedup (c.f. Grover's algorithm) relative to classical computers, so that the computation time is still exponential in the number of colours.

Literature Abstract

The computational complexity for finding Ramsey Numbers more than 5 with the existing algorithms of classical and quantum computers are exponential in the number of colours.

This research project aims at finding the same probabilistically using an alternative approach, with the help of social media analytics and repeated statistical experiments.

Project Plan/Methodology

Collect data from social media analytics website and form a complete two-coloured graph using social media networks.

Choose a number N and run the experiment a large number of times, say 100, and note in each case the min. no. of nodes needed to find a clique of N edges

We get a distribution of the sample modes. By seeing the histogram, determine the underlying distribution.

Find the confidence interval of the mean of the distribution using R.

This interval will contain the number of nodes needed for N-edged clique with highest probability.

Impact of our method

- Using our experiments, we can move one step forward into conjecturing the exact values of R(6,6), R(7,7) etc.

- For R(r, s) with r, s > 5, only weak bounds are available. Lower bounds for R(6, 6) and R(8, 8) have not been improved since 1965 and 1972, respectively. Using our research experiments, we can improve the lower bounds.

The Dataset

The dataset that we are trying for this experiment is being gathered from various social media analytics website like http://networkrepository.com/socfb.php and https://snap.stanford.edu/data/egonets-Facebook.html .
A constant research for better and more appropriate data is still being done by the person involved with this project.

Present Known values

- As described above, R(3, 3) = 6. It is easy to prove that R(4, 2) = 4, and, more generally, that R(s, 2) = s for all s: a graph on s − 1 nodes with all edges coloured red serves as a counterexample and proves that R(s, 2) ≥ s; among colourings of a graph on s nodes, the colouring with all edges coloured red contains a s-node red subgraph, and all other colourings contain a 2-node blue subgraph (that is, a pair of nodes connected with a blue edge.)
- Using induction inequalities, it can be concluded that R(4, 3) ≤ R(4, 2) + R(3, 3) − 1 = 9, and therefore R(4, 4) ≤ R(4, 3) + R(3, 4) ≤ 18. There are only two (4, 4, 16) graphs (that is, 2-colourings of a complete graph on 16 nodes without 4-node red or blue complete subgraphs) among 6.4 × 1022 different 2-colourings of 16-node graphs, and only one (4, 4, 17) graph (the Paley graph of order 17) among 2.46 × 1026 colourings.[6] (This was proven by Evans, Pulham and Sheehan in 1979.) It follows that R(4, 4) = 18.
- The fact that R(4, 5) = 25 was first established by Brendan McKay and Stanisław Radziszowski in 1995.
- The exact value of R(5, 5) is unknown, although it is known to lie between 43 (Geoffrey Exoo (1989) and 48 (Angeltveit and McKay (2017) (inclusive).
- In 1997, McKay, Radziszowski and Exoo employed computer-assisted graph generation methods to conjecture that R(5, 5) = 43. They were able to construct exactly 656 (5, 5, 42) graphs, arriving at the same set of graphs through different routes. None of the 656 graphs can be extended to a (5, 5, 43) graph.
- For R(r, s) with r, s > 5, only weak bounds are available. Lower bounds for R(6, 6) and R(8, 8) have not been improved since 1965 and 1972, respectively.

Research Aim and Responsibilities

The standard survey on the development of Ramsey number research is the Dynamic Survey 1 of the Electronic Journal of Combinatorics, by Stanisław Radziszowski, which is periodically updated.

A fruitful research can result in publication of our results in the above reputed journal.

References/Related Work

1.https://www.combinatorics.org/ojs/index.php/eljc/article/view/DS1

2.https://journals.aps.org/pra/abstract/10.1103/PhysRevA.93.032301

3.https://onlinelibrary.wiley.com/doi/abs/10.1002/jgt.3190190304

4.https://onlinelibrary.wiley.com/doi/abs/10.1002/jgt.3190130113

5. https://arxiv.org/abs/1703.08768v2

6. https://www.sciencedirect.com/science/article/pii/S0095895696917414?via%3Dihub

THANK YOU