Title: Ramsey Numbers Analysis using Social Network Analytics

Guide: Dr. D. Hemavathi

Members:

Swastik Banerjee [RA1711008010239]

Introduction

In this Combinatorial Graph Theory and Social Media Analytics research project, we aim to analyze already existing anonymous graph networks taken from social media datas like Facebook and find out how much they tally with Ramsey’s theorem.

Objective: What is Ramsey’s Theorem?

`

In combinatorics, Ramsey Theorem states that one will always find monochromatic cliques when the edges are coloured in a sufficiently large complete graph. For example, say we have two positive integers r and s. Ramsey-theorem guarantees that there has to be one least positive integer for which every blue-red (in the case of two-coloring) edge coloring of the complete graph of R(r,s) vertices contains a blue r-clique or a red s-clique.

Purpose of project

In our research analysis project, we will survey on social media graphs taken from Facebook users of MIT and see how much the results are close to the Ramsey Theorem in practice.

That is, Ramsey’s theorem gives a minimum number of edges to guarantee the existence of a monochromatic clique of either red color or blue color.

In our experiment, we have designed red colored edge as knowing each other in Facebook, and blue colored edge as not-knowing each other in Facebook, where the complete graph is formed by vertices where each vertex is an user of Facebook.

Existing System

A social friendship network extracted from Facebook consisting of people (nodes) with edges representing friendship ties was used from MIT was used in our research analysis.

A complete graph consisting of undirected edges were created using MIT Facebook Analytics. As stated above, the data had 6.4K nodes, where a connected to b meant that node a was connected to node b in Facebook. These edges were coloured red. The nodes that did not know each other were also joined using blue edges to make it a complete graph.
The above manipulation was done using the software Mathematica licensed under Wolfram Research, Inc, a very good tool to do these mathematical manipulations of graphs, etc.

Limitations of existing models: Currently Known Values of Ramsey Numbers -

R(3, 3) = 6 as previously mentioned. Ramsey for 4,2 and Ramsey for any s,2 is always equal to s for any s. It is extremely simple to prove: As a counterexample, let us take a graph with s1 nodes and the edges when all coloured red proves that Ramsey of s,2 is always equal to s. A s-node red subgraph exists in the colouring of all edges coloured red, and a 2-node blue subgraph exists in all other colourings of a graph on s nodes (that is, a pair of nodes connected with a blue edge.)

The values of Ramsey of 4,3 can be calculated using values of Ramsey for 4,2 and 3,3. Similarly, it is easy to deduce the values of Ramsey of 4,4 for 2-coloruing. Various colourings of graph were constructed to find lower bounds of these numbers, and finally Ramsey of 4,4 has been deduced to be 18.

The exact values of Ramsey of 5,5 is unknown, but it known to lie between 43 and 48. The exact value of Ramsey of 4,5 is however known to be exactly 25.

Literature Survey

Some related works are:

▷“Small Ramsey Numbers
Stanisław Radziszowski
DS1: Jan 15, 2021
2011-08-22 ”

▷“Ramsey Theory Applications
Vera Rosta
DS13: Dec 7, 2004
2004-12-07 ”

▷“Ramsey Numbers of Ordered Graphs
Martin Balko, Josef Cibulka, Karel Král, Jan Kynčl
P1.16
2020-01-10 ”

▷“Constructive Lower Bounds on Classical Multicolor Ramsey Numbers
Xu Xiaodong, Xie Zheng, Geoffrey Exoo, Stanisław P. Radziszowski
R35

Proposed System Analysis

A complete graph consisting of undirected edges were created using MIT Facebook Analytics.

As stated above, the data had 6.4K nodes, where a connected to b meant that node a was connected to node b in Facebook. These edges were coloured red. The nodes that did not know each other were also joined using blue edges to make it a complete graph.

The above manipulation was done using the software Mathematica licensed under Wolfram Research, Inc, a very good tool to do these mathematical manipulations of graphs, etc.

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.

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.

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.

Tools Used

Mathematica software

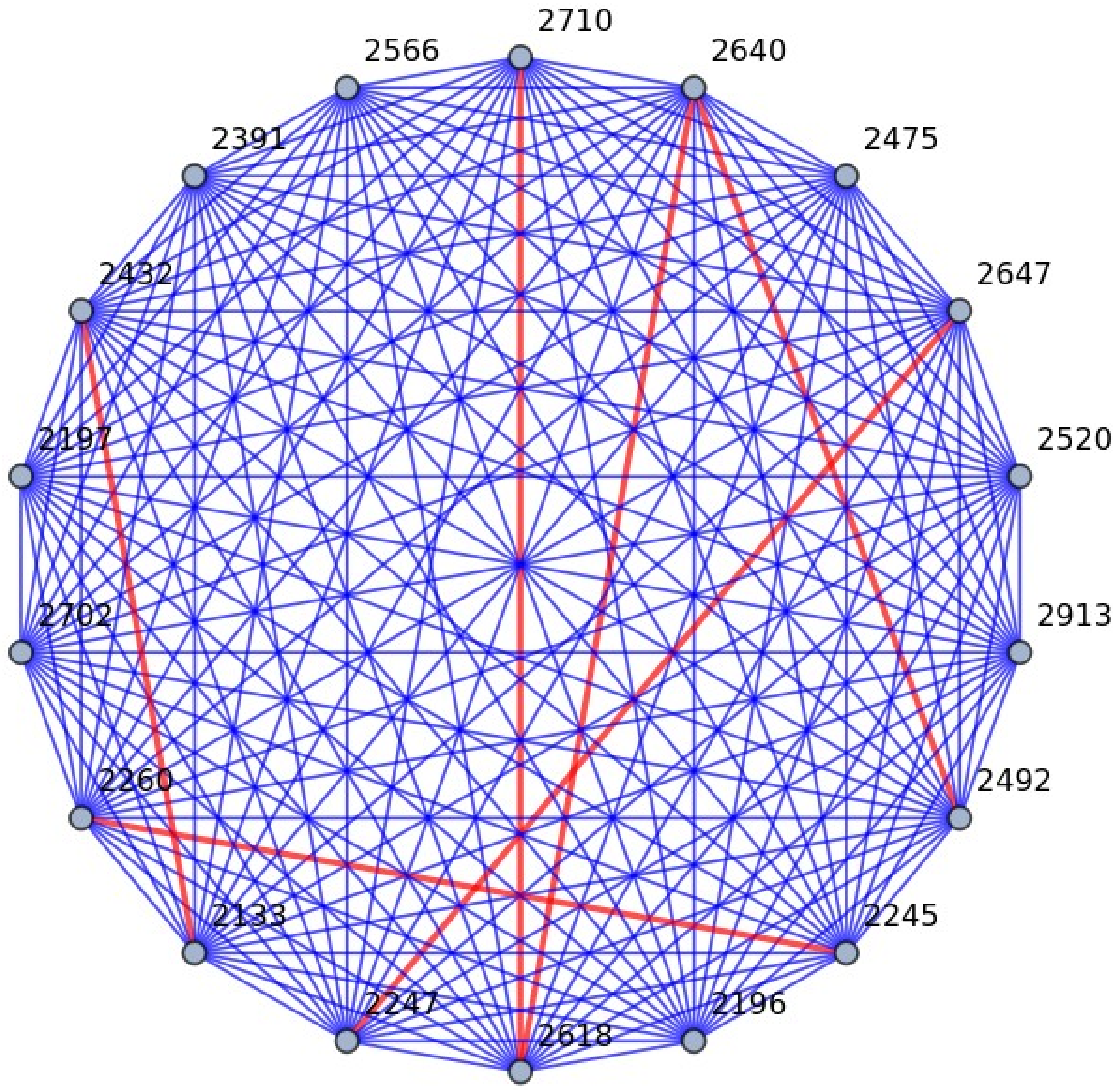
Stanford Dataset Social Network Analytics

Dataset

Network Data Statistics

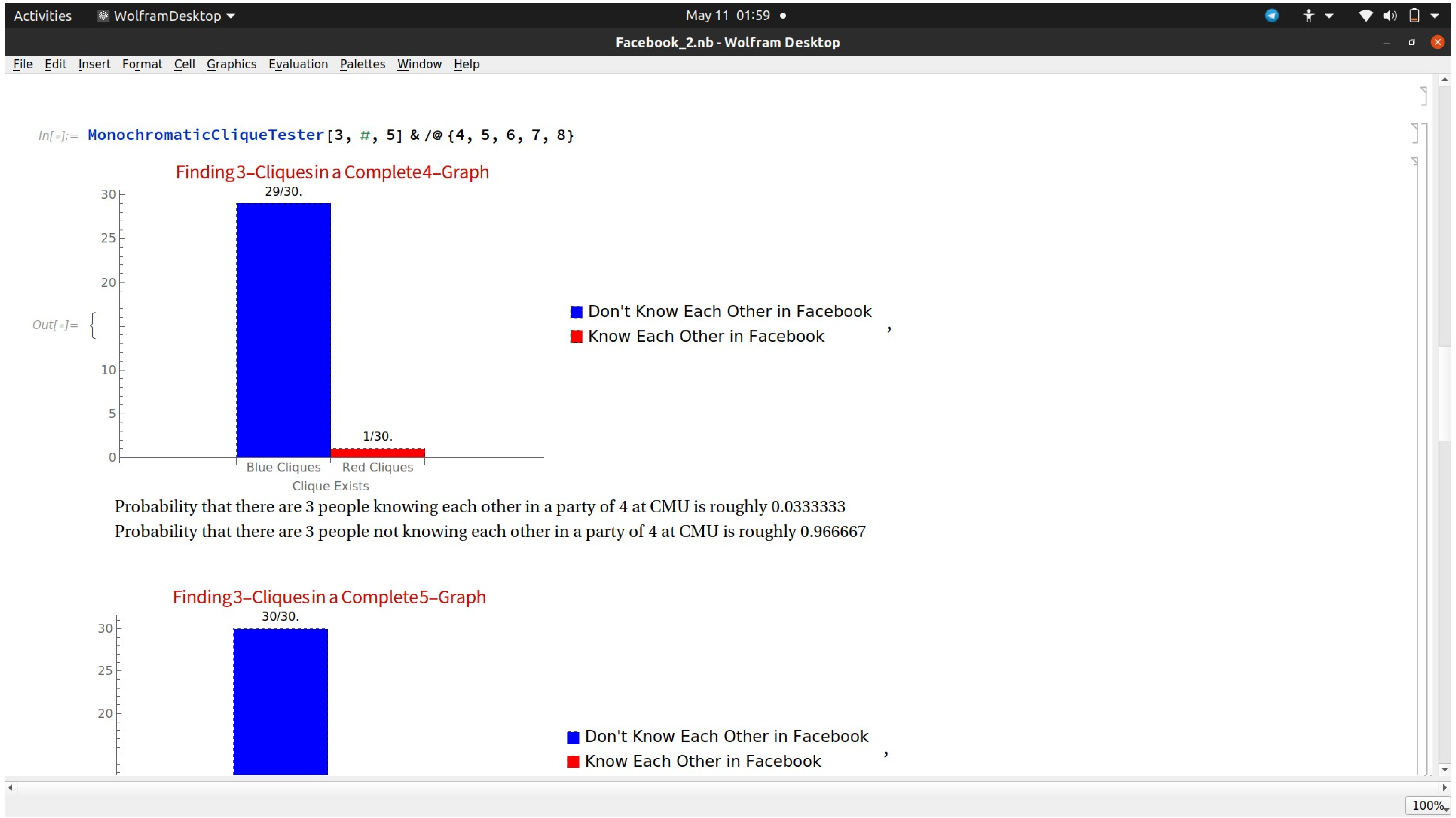
Nodes	6.4K
Edges	251.2K
Density	0.0122613
Maximum degree	708
Minimum degree	1
Average degree	78
Assortativity	0.119896
Number of triangles	7.1M
Average number of triangles	1.1K
Maximum number of triangles	27.8K
Average clustering coefficient	0.27236
Fraction of closed triangles	0.180288
Maximum k-core	73
Lower bound of Maximum Clique	9

Input



2710
2640
2475
2647
2520
2913
2492
2245
2196
2618
2247
2133
2260
2702
2197
2432
2391
2566

Output


Activities
WolframDesktop
May 11 01:59
Facebook_2.nb - Wolfram Desktop
File Edit Insert Format Cell Graphics Evaluation Palettes Window Help
In[∘]:= MonochromaticCliqueTester[3, #, 5] & /@ {4, 5, 6, 7, 8}
Out[∘]=
Finding 3–Cliques in a Complete 4–Graph
29/30.
1/30.
Blue Cliques
Red Cliques
Clique Exists
Don't Know Each Other in Facebook
Know Each Other in Facebook
Probability that there are 3 people knowing each other in a party of 4 at CMU is roughly 0.0333333
Probability that there are 3 people not knowing each other in a party of 4 at CMU is roughly 0.966667
Finding 3–Cliques in a Complete 5–Graph
30/30.
Don't Know Each Other in Facebook
Know Each Other in Facebook
100%

Code Sample

Activities   WolframDesktop   May 11 01:58

MIT.nb - Wolfram Desktop

File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

```
    coloredSubgraphs = Quiet[GroupBy[MapAt[Lookup[LineColor], AnnotationValue[var, EdgeStyle], {All, 2}], Last → First]];
    cliques = Length@First@FindClique[#] ≥ n & /@ coloredSubgraphs;
    AppendTo[list[[1]], cliques[[1]]];
    If[Length@cliques > 1, AppendTo[list[[2]], cliques[[2]]]];
   ]
  ];


Labeled[BarChart[{{Length@Cases[list[[1]], True], Length@Cases[list[[2]], True]}}, ChartBaseStyle → EdgeForm[Dashed],
  ChartStyle → {Blue, Red}, ChartLegends → {"Don't Know Each Other in Facebook", "Know Each Other in Facebook"},
  BarSpacing → None, ChartLabels -> {{"Clique Exists"}, {"Blue Cliques", "Red Cliques"}},
  LabelingFunction → (Placed[Row[{#, "/", N[x * i]}], Above] &),
  PlotLabel → Style["Finding " <> ToString[n] <> "-Cliques in a Complete " <> ToString[e] <> "-Graph", "Title", 16],
  ImageSize → Medium, PerformanceGoal → "Speed"],
 Column[{"Probability that there are " <> ToString[n] <> " people knowing each other in a party of " <> ToString[e] <>
    " at MIT is roughly " <> ToString[N[Length@Cases[list[[2]], True] / N[x * i]]],
   "Probability that there are " <> ToString[n] <> " people not knowing each other in a party of " <> ToString[e] <>
    " at MIT is roughly " <> ToString[N[Length@Cases[list[[1]], True] / N[x * i]]]}],
 LabelStyle → Directive[FontSize → 13.5, FontFamily → "Times"]]
] (*Gives a bar chart of how many times monochromatic n-
cliques of colors blue and red were found resp. in a complete graph of e vertices
 (taken randomly from the entire imported social-media graph in batches of successive 1000 vertices at a time till
   all the vertices are taken into consideration) when the experiment is run a total of x times*)

In[*]:= MonochromaticCliqueTester[5, #, 100] & /@ Range[43, 48]
```

100%

Activities WolframDesktop May 11 01:58

MIT.nb - Wolfram Desktop

File Edit Insert Format Cell Graphics Evaluation Palettes Window Help

```
  {i, Length[intersectpos]}]];

In[]:= MonochromaticCliqueTester[n_, e_, x_] := Module[{subedges, newEdges, graph, i, j, var, coloredSubgraphs, cliques, list = {{}, {}}},


  For[i = 0, i < Floor@N@(Max[g1] / 1000), i++,
   subedges = Select[g2, (1000 * i < #[[1]] < 1000 * (i + 1)) && (1000 * i < #[[2]] < 1000 * (i + 1)) &];
   newEdges = Replace[subedges, {DirectedEdge[a_, b_] :> Style[UndirectedEdge[a, b], Red, Thick],
      UndirectedEdge[a_, b_] :> Style[UndirectedEdge[a, b], Blue]}, {1}];
   graph = Graph[newEdges];

   For[j = 1, j <= x, j++,

    var = Subgraph[graph, {RandomInteger[{1000 * i, 1000 * (i + 1)}, e]}, VertexLabels → Automatic];

    Export["/home/justanotherlad/Documents/BTech Degree Project/MIT/Pictures/5-Clique/Ramsey_for_" <> ToString[n] <>
      "_in_" <> ToString[e] <> "_" <> ToString[i] <> "_" <> ToString[j] <> ".png", var];


    coloredSubgraphs = Quiet[GroupBy[MapAt[Lookup[LineColor], AnnotationValue[var, EdgeStyle], {All, 2}], Last → First]];
    cliques = Length@First@FindClique[#] ≥ n & /@ coloredSubgraphs;
    AppendTo[list[[1]], cliques[[1]]];
    If[Length@cliques > 1, AppendTo[list[[2]], cliques[[2]]]];
   ]
  ];


  Labeled[BarChart[{{Length@Cases[list[[1]], True], Length@Cases[list[[2]], True]}}, ChartBaseStyle → EdgeForm[Dashed],
```

100%

Conclusion

Our experiment determines that practically, in the famous party problem, the probability of people not knowing each other is far more than people knowing each other, althought theoretically we assume both are equally likely.

References

Bohman, Tom; Keevash, Peter (2010), "The early evolution of the H-free process", Invent. Math., ***181*** *(2): 291–336, arXiv:0908.0429, Bibcode:2010InMat.181..291B, doi:10.1007/s00222-010-0247-x*

Brualdi, Richard A. (2010), Introductory Combinatorics (5th ed.), Prentice-Hall, pp. 77–82, ISBN

Conlon, David (2009), "A new upper bound for diagonal Ramsey numbers", Annals of Mathematics, ***170*** *(2): 941–960, arXiv:math/0607788v1, doi:10.4007/annals.2009.170.941, MR 2552114.*

Erdős, Paul (1947), "Some remarks on the theory of graphs", Bull. Amer. Math. Soc., ***53*** *(4): 292–294, doi:10.1090/S0002-9904-1947-08785-1.*

Erdős, P.; Moser, L. (1964), "On the representation of directed graphs as unions of orderings" (PDF), A Magyar Tudományos Akadémia, Matematikai Kutató Intézetének Közleményei, ***9****: 125–132, MR 0168494*

*Exoo, G. (1989), "A lower bound for R(5,5)", Journal of Graph Theory, **13**: 97–98, doi: 10.1002/jgt.3190130113.*

Graham, R.*; Rothschild, B.;* Spencer, J. H. *(1990), Ramsey Theory, New York: John Wiley and Sons.*

Gross, Jonathan L. (2008), Combinatorial Methods with Computer Applications, CRC Press, p. 458, ISBN

Harary, Frank (1972), Graph Theory, Addison-Wesley, pp. 16–17, ISBN

Ramsey, F. P. *(1930), "On a problem of formal logic", Proceedings of the London Mathematical Society, **30**: 264–286, doi:10.1112/plms/s2-30.1.264.*

Spencer, J. *(1975), "Ramsey's theorem – a new lower bound", J. Combin. Theory Ser. A, **18**: 108–115, doi:10.1016/0097-3165(75)90071-0.*

FUTURE WORK :

1. This experiment can be used to determine lower bounds for larger Ramsey Numbers
2. Similar Social Media Analysis can be done to obtain results in the fields of Combinatorics or Graph Theory.