

The Clique Problem and Covid-19:

How a theoretical problem can inform the spread of a deadly disease

Robert Moseley, Andrew Cantrell

Department of Computer Science, California Baptist University

CSC414: Theory of Computation

Professor Benjamin Sanders

December 6, 2021

### **Abstract**

The maximal clique problem is a popular theoretical problem in computer science that has never been solved. Despite this, optimizations of the clique problem have incredibly interesting real-world applications that could even help save the lives of other people. One such cause is contact tracing in the context of the Covid-19 virus which has plagued the world for much of the last 2 years. By analyzing existing algorithms and graph variations, a new model for contact tracing is proposed that builds off of previous work and combines the tools in a unique way. While the maximal clique problem is still unsolved, the implementation proposed in this paper is an optimization of the problem that is particularly suited to perform well in contact tracing and to even theoretically improve the accuracy of contact tracing as it currently stands. In addition to this, a current epidemic model is used and applied to the chosen algorithm as a means of attempting to understand contact tracing in a new way. Even though this solution is theoretical, current research in the domain of contact tracing suggests that these improvements are needed and could have a significant impact on the field if instituted correctly.

*Keywords:* clique problem, maximal clique, Covid-19, contact tracing, social distancing

### **Introduction to the Clique Problem and NP-completeness**

Since computers and the theory of computation were first developed, there have been hundreds—if not thousands—of complex and difficult problems that needed to be solved in order to further develop the field of computing. Problems such as integer factorization in polynomial time, the dynamic optimality conjecture, solving 3SUM in under the quadratic time barrier, and finding clustered planar drawings in polynomial time are all examples of problems in the field of computer science that are yet to be solved. One such problem that has never been solved, that may have interesting applications to the way that modern diseases are studied, is the clique problem. According to Bomze et al. “The maximum clique problem is a classical problem in combinatorial optimization which finds important applications in different domains” (p. 2). The clique problem is considered effectively “unsolved” because it is what is known as an NP-complete problem. NP-complete problems are problems that are accepted as being computationally “intractable” (Garey et al., p. 2). This means, in brief, that it is impossible given current knowledge of the problem to assess or solve the problem in a deterministic manner. NP-complete, definitionally, stands for nondeterministic polynomial-time complete—which references the idea that even applying standard brute force algorithms to problems in this domain produces results that are nondeterministic and thus unreliable for producing consistent results (Garey et al., p. 3).

Given the above explanation of NP-completeness, it is now important to understand the clique problem itself and the different ways in which it can be assessed. First and foremost, cliques are a definition associated with a subdomain of computation and mathematics known as graph theory. Knowing this, it is possible to define the first two primary components to understanding the clique problem: 1) complete graphs, and 2) cliques. Graphs, in their most basic

form, are simply distributions of nodes connected by edges which will be referred to as vertices. Using this definition of graphs, it is now possible to give a definition of complete graphs and cliques:

1. **Complete Graphs:** a graph that features pairs of vertices that are all adjacent (Orlin, p. 406)
2. **Clique:** A subgraph of a complete graph (Orlin, p. 406)

Now that two of the quintessential components of the clique problem have been clearly defined, it is now possible to begin understanding the clique problem and why it is NP-complete.

Before giving mathematical definitions of the components, it is necessary to understand that the clique problem is also often referred to as the maximal clique problem. The goal of “solving” or assessing the clique problem is to find an algorithm that deterministically finds the maximum-sized clique within any given graph. Due to the fact that graphs come in many sizes, shapes, and varieties it can be difficult to apply one algorithm or approach to solving for maximum clique size. Thus, even though it is possible to find the maximum size of a clique within a graph, this problem is considered a problem because of how unreliable existing methods are for finding this elusive number which, in computer science terms, is known as the cardinality—a term that will be used throughout this analysis. If it was possible to determine a maximum clique within a graph with both mathematical certainty and a deterministic algorithm, the clique problem would no longer be a problem that needed to be solved.

Mathematically, a graph can be defined as  $G = (V, E)$  while a complete graph can be defined as  $\forall i, j \in V, (i, j) \in E$  (Pardalos & Xue, p. 2). This definition of a graph will also be applied to subgraphs (which will be considered independent graphs when applicable) to simplify the discussion of cliques and how they can be visualized. A clique given as  $C$ , using

this notation, “is a subset of  $V$  such that  $G(C)$  is complete” (Pardalos & Xue, p. 2). An example of a maximal clique is given by Orlin (p. 409) which is “Let  $C_\alpha$  be the unique maximal clique that contains edge  $\alpha$ . Then it follows that if  $C$  is any clique containing edge  $\alpha$ ,  $C \subseteq C_\alpha$ .” This simply states that any graph containing a maximal clique, also contains non-maximal cliques and that sub-cliques can be contained within maximal cliques and that sub-cliques can contain the same—or some of the same—vertices as their parent cliques. This idea will contribute to the way that cliques are discussed, and how they can be considered in the application of cliques to real-world problems.

### **Analysis of Covid-19 Spread**

Throughout 2020 and 2021, the world has watched as a virus known as Covid-19 has spread around the world and taken the lives of millions of people. During this difficult time, many mechanisms have been utilized to attempt to slow down the spread of the virus and the deadly effect that it can produce in human hosts. These mechanisms include medical interventions such as vaccines and specific medications, wearing masks, and social distancing. One of these mechanisms, however, is of particular importance as it relates to the clique problem stated above: social distancing. Social distancing, for those unaware, is the process of putting physical distance in between people in order to prevent infected people from accidentally or intentionally spreading the virus to other people. Social distancing, in many places all over the world, has been instituted as a long-term measure in places like restaurants, amusement parks, movie theaters, and any other places that tend to become congested with people or put people into close contact with one another. Matrajt and Leung (2020) established through their research that social distancing can be incredibly beneficial in when used with other anti-Covid protocols, but that social distancing alone is not enough and “must be used in conjunction with testing and

contact tracing of all suspected cases to mitigate virus transmission” (p. 1). In this instance, the authors of this study used a common definition of social distancing which dictates that a minimum of six feet be maintained between people at all times (Matrajt & Leung, pp. 2-3)

While it is true, as given above, that social distancing needs to be used with other mechanisms in order to be truly successful—depending of course on the definition of social distancing used—other studies have shown promising results about the effectiveness of social distancing that suggests it should not be ignored as a way of intervening in the spread of the virus. In another study which included isolation and quarantine within the definition of social distancing, it was found that almost nothing was more effective in preventing the spread of the virus (Milne & Xie, 2020). Intuitively, this makes perfect sense. If a virus requires humans as a carrier and there are no humans present or available, the virus will not be able to spread. This is a convenient solution in theory, but in practice is not as easily attainable because it is incredibly difficult to prevent all human-to-human contact. According to Moosa (p. 6295):

Naturally, new cases will continue to emerge during periods of social distancing for two reasons. The first is that people need to go shopping as and when they need food and other essential household goods, while workers with essential jobs continue to go to work. The second is that the benefit of social distancing will not be felt until these policies have been in place for a period of time that is equal to the incubation period.

Many individuals who contracted Covid-19 in the weeks before the start of social distancing will not be diagnosed until they begin to show symptoms, which can take up to 14 days.

Given this information, it is clear that social distancing will always be limited by the society in which it is enforced, the number of people that have essential jobs, and the ability to quickly

detect symptoms when they are present. Despite this, one useful tool for when social distancing fails can be found in the practice of contact tracing. While rudimentary, contact tracing has been shown to aid in the overall suppression of the Covid virus and to at best prevent the spread between people and at worst slow the spread (Matrajt & Leung, p. 7). With this information in mind about how Covid can spread between people and how it can be controlled, it is now possible to present a recommendation for how these protocols can be improved with the goal of protecting as many lives as possible.

### **The Intersection of the Clique Problem and Covid-19**

It might not be obvious, at first, how the clique problem and Covid-19 could ever be possibly related or how one could inform the other, but the intersection between the two is actually rather striking and surprisingly useful. In order to present the intersection of the two, however, there must first be a general understanding of some of the algorithms that are currently used to look for maximal cliques within complete graphs and subgraphs. One such algorithm, and one of the first of its kind, was presented by Carraghan & Pardalos (1990) which was a partially enumerative algorithm and was shown to be efficient for up to 3000 vertices and over a million edges (p. 1). This result is particularly striking 1) because it was one of the first applicable efficient algorithms, and 2) because of its efficiency despite the very dramatic hardware limitations of computers in the early 90s. Despite this, however, this algorithm does not scale in the necessary ways and should not be relied upon now—especially for something as sensitive as Covid-19 tracking. An algorithm that might be interesting as applied to Covid-19, however, is an algorithm known as the Reactive Local Search algorithm, which was originally published in 2001 (Battiti & Protasi, p. 1). One of the favorable aspects of this algorithm is its heightened performance within dense graphs—an area in which many algorithms can fail. Worst

case complexity of this algorithm is given as  $O(\max\{n, m\})$  with  $n$  representing the number of nodes in the graph and  $m$  representing the numbers of edges present (p. 2). This result, while not solving the NP-completeness issue, does present a more reasonable time complexity to a problem that rarely approaches polynomial time even in best case scenarios (Bomze et al., pp. 19-25).

Beyond the above two algorithms, one final and important algorithm needs to be discussed. This algorithm is known as the Bron-Kerbosch algorithm for temporal graphs (Himmel et al., p. 6). The reason that this algorithm is so important is because it is adaptable to temporal graphs, which are just graphs that change over time. Given how complicated the Covid-19 problem is, it is necessary to implement the most advanced algorithm possible that can account for changes in graphs—no matter how dramatic they are. This algorithm has been demonstrated in practice to achieve dramatic speedups over previous enumerative algorithms, and the ability to be applied to temporal graphs also gives it a distinct advantage over static algorithms (Himmel et al., p. 15). Given this, the Bron-Kerbosch algorithm will be the algorithm that will be referenced moving forward and the one discussed in the context of Covid-19.

In the section above that introduced the topic of Covid-19 transmission and how current research indicates how it can be mitigated, the context of contact tracing was mentioned as a tool that can assist the mechanism of social distancing. One of the difficult things about contact tracing, of course, is accurately assessing who is sick, for how long, who made them sick, and who was exposed to them. Current research on contact tracing suggests that not enough effort has been put into understanding how to adequately use the tool and what an “optimal level of investment” could look like (Armbruster & Brandeau, p. 342). Despite this, graphs remain an excellent way to accurately model populations and the links and interactions between people within the population group. Specifically, within the context of contact tracing, graphs can be an



incredibly useful tool for quickly establishing the types of exposures that are occurring which can then be modeled on the graph to prevent further unnecessary exposures. One of the difficulties in using graphs, however, is choosing which algorithms to implement and what type of graph to use—not all graphs are created equal. In the study mentioned previously, one of the limitations of the research is that the graphs implemented in the study are not dynamic, but instead use a cyclic graph where each node has, on average, degree 4 (Armbruster & Brandeua, p. 343). Knowing the limitations of static graphs and how they can fail to be representative of real-world data, it should be possible to implement an improved system based on concepts discussed above.

### Theoretical Implementation and Results

The implementation for the proposed improvement has two primary components: 1) the chosen algorithm and graph methodology, and 2) the model chosen for simulating epidemic conditions. For the first part, as discussed prior, the Bron-Kerbosch algorithm represents a dramatic improvement over previous enumerative algorithms, thus the chosen algorithm for this will be the Bron-Kerbosch Pivot algorithm from the work of Himmel et al. (2017).

---

#### Algorithm 2 Enumerating all Maximal Cliques in a Graph with Pivoting

---

```

1: function BRONKERBOSCHPIVOT( $P, R, X$ )
    $\triangleright R$  : a clique
    $\triangleright P \cup X$  : set of all vertices  $v$  such that  $R \cup \{v\}$  is a clique and where
     – vertices in  $P$  have not yet been considered as additions to  $R$  and
     – vertices in  $X$  already have been considered in earlier steps
2:   if  $P \cup X = \emptyset$  then
3:     add  $R$  to the solution
4:   end if
5:   choose pivot vertex  $u \in P \cup X$  with  $|P \cap N(u)| = \max_{v \in P \cup X} |P \cap N(v)|$ 
6:   for  $v \in P \setminus N(u)$  do
7:     BRONKERBOSCHPIVOT( $P \cap N(v), R \cup \{v\}, X \cap N(v)$ )
8:      $P \leftarrow P \setminus \{v\}$ 
9:      $X \leftarrow X \cup \{v\}$ 
10:  end for
11: end function

```

---

(Himmel et al., 2017)

While many past algorithms have been utilized in conjunction with static unweighted graphs, the graph implemented with Bron-Kerbosch should be a temporal weighted graph. The benefits of using a temporal weighted graph are 1) the temporal nature of the graph will allow it to reflect a changing population size and makeup, and 2) the weighted component—implemented as a numeric score indicating how susceptible a person is to becoming ill—will add additional detail and realism to the simulation, hopefully producing more valuable data. As for the epidemic model used, there is one model that stands above the others in terms of both accuracy and frequency of use. The SIRS model, implemented by Armbruster and Brandeau (2007), requires that all infected individuals be removed from the simulation when they are being “treated,” and added back as susceptible individuals once treated (p. 343). Importantly, this model is a current industry standard for contact tracing and is also used in similar models and simulations—which is also why it has been chosen here.

### **Conclusion**

Using the Bron-Kerbosch implementation and SIRS epidemic model above, it should be possible to create an incredibly powerful tool for simulating improved contact tracing. Not only does the Bron-Kerbosch model allow for more dynamic simulations, but the weighted nature of the graph also takes into consideration the likelihood of someone actually becoming ill. While this model is still theoretical in nature, current research, as outlined above, indicates that the combination of these tools should represent a dramatic improvement over current processes and models used for pandemic contact tracing.

While it is true, additionally, that the clique problem is still unsolved as a function of determinism, the algorithms and methods chosen above represent an optimization for maximal clique finding in this specific problem and performance issues should not prevent the

implementation from succeeding. One possible limitation that would have to be addressed in a true implementation would be the lack of data on immune system strength in individuals—but solving for this in a simulation would be as simple as randomizing the data set within normal bounds and adding the possibility of people falling outside the normal deviation in order to account for those with especially poor health conditions. Having established all of this, this proposed system, if properly utilized, could present an improvement over current contact tracing solutions—which, in a global pandemic, would not only be useful, but could be used to improve the lives of many people.

### References

- Armbruster, B., Brandeau, M. L. (2007). Contact tracing to control infectious disease: when enough is enough. *Health Care Management Science*, 10, 341-355. DOI: <https://doi.org/10.1007/s10729-007-9027-6>
- Bomze I.M., Budinich M., Pardalos P.M. et al. (1999) The Maximum Clique Problem. In: Du DZ., Pardalos P.M. (eds) *Handbook of Combinatorial Optimization*. Springer, Boston, MA. [https://doi.org/10.1007/978-1-4757-3023-4\\_1](https://doi.org/10.1007/978-1-4757-3023-4_1)
- Battiti, R., Protasi, M. (2001). Reactive Local Search for the Maximum Clique Problem<sup>1</sup>. *Algorithmica*, 29, 610–637. <https://doi.org/10.1007/s004530010074>
- Carraghan R., Pardalos P. (1990). An exact algorithm for the maximum clique problem. *Operations Research Letters*, 9(6), 375-382. [https://doi.org/10.1016/0167-6377\(90\)90057-C](https://doi.org/10.1016/0167-6377(90)90057-C)
- Garey, M.R., Johnson, D. S., Stockmeyer, L. (1974). Some simplified NP-complete problems. *Proceedings of the Sixth Annual ACM Symposium on Theory of Computing*, 47-63. *Association for Computing Machinery*, New York, NY. DOI: <https://doi.org/10.1145/800119.803884>
- Gendreau, M., Soriano, P. & Salvail, L. (1993). Solving the maximum clique problem using a tabu search approach. *Ann Oper Res* 41, 385–403. <https://doi.org/10.1007/BF02023002>
- Himmel, AS., Molter, H. Niedermeier, R. et al. (2017). Adapting the Bron-Kerbosch algorithm for enumerating maximal cliques in temporal graphs. *Social Network Analysis and Mining*. 7, 1-16. DOI: <https://doi.org/10.1007/s13278-017-0455-0>

- Matrajt, L., & Leung, T. (2020). Evaluating the Effectiveness of Social Distancing Interventions to Delay or Flatten the Epidemic Curve of Coronavirus Disease. *Emerging infectious diseases*, 26(8), 1740–1748. <https://doi.org/10.3201/eid2608.201093>
- Milne, J. G., Xie, S. (2020). The effectiveness of social distancing in mitigating COVID-19 spread: a modelling analysis. *Medrxiv*. DOI: <https://doi.org/10.1101/2020.03.20.20040055>
- Moosa, I. (2020). The effectiveness of social distancing in containing Covid-19. *Applied Economics*, 52(58), 6292-6305. DOI: <https://doi.org/10.1080/00036846.2020.1789061>
- Orlin, J. (1977). Contentment in Graph Theory: Covering graphs with cliques. *Indagationes Mathematicae*, 80(5), 406-424. DOI: [https://doi.org/10.1016/1385-7258\(77\)90055-5](https://doi.org/10.1016/1385-7258(77)90055-5)
- Pardalos, P.M., Xue, J. (1994). The maximum clique problem. *J Glob Optim* 4, 301–328. <https://doi.org/10.1007/BF01098364>
- Patric R. J. Östergård. (2001). A new algorithm for the maximum-weight clique problem. *Nordic J. of Computing*, 8(4), 424–436.
- P. M. Pardalos & A. T. Phillips (1990) A global optimization approach for solving the maximum clique problem. *International Journal of Computer Mathematics*, 33:3-4, 209-216, DOI: [10.1080/00207169008803851](https://doi.org/10.1080/00207169008803851)
- Solnon, C., Fenet, S. (2006). A study of ACO capabilities for solving the maximum clique problem. *J Heuristics* 12, 155–180. <https://doi.org/10.1007/s10732-006-4295-8>