Coloring Graphs to Estimate Causal Effects: A Diagonal Ramsey Approach

Kweku A. Opoku-Agyemang*

July 18, 2023

Abstract

In this paper, we apply diagonal Ramsey numbers, which measure how large a graph has to be to guarantee that it contains a monochromatic subgraph of a given size, to the problem of estimating causal effects. We argue that graphs can be used as models of data or systems, and that diagonal Ramsey numbers can be used as estimates of treatment effects. We also extend our approach to multilevel data, where the assignment to treatment or the outcome may depend on various sources across network levels. We show how to classify units into homogeneous groups with regard to class-specific selection and outcome models, and how to account for both the nested structure of the data and potential heterogeneity in selection processes and treatment effects. We illustrate our approach with empirical examples. We propose the use of power calculations for improved rigorous treatment effects.

*Chief Social Scientist, Development Economics X, Toronto, ON, Canada and Honorary Affiliate, International Growth Centre, University of Oxford and London School of Economics, London, UK. Email: kweku@developmenteconomicsx.com, kweku2008@gmail.com.

Contents

1 Introduction

Modern applied economics emphasizes identifying the effects of treatments and policies. This approach ensures that the identification of treatment effects are reliable and economists are always interested in their treatment effects becoming more precise and rigorously identified. This paper propose an new approach by analyzing how graphs can be used as models of data or systems, and where diagonal Ramsey numbers, borrowed from Ramsey theory in pure mathematics, can be used as estimates of treatment effects.

Ramsey numbers are numbers that measure how large a complete graph[1] has to be to guarantee that it contains a monochromatic subgraph[2] of a certain size, regardless of how the edges are colored[3]. Ramsey numbers have applications in computer science, including in scheduling, map coloring, Sudoku puzzles, and others.

The focus of the paper is on a special case of Ramsey numbers called diagonal Ramsey numbers, which have applications in computer science, especially in areas such as logic, complexity theory, cryptography, and distributed computing. With diagonal Ramsey numbers, the size of the monochromatic subgraph is the same for all colors. The diagonal Ramsey numbers measure how large a graph has to be to guarantee that it contains a monochromatic subgraph of a given size. We propose to relate diagonal Ramsey numbers to the field of causal inference. We shall argue that graphs can be used as models of data or systems, and that Ramsey numbers can be used as estimates or approximations of treatment effects as unknown quantities.

We shall borrow from recent work in statistics by Conlon (2019) which proves an upper bound for diagonal Ramsey numbers. We extend the idea of Ramsey theory analyzing how structure appears from randomness to study how treatment effects emerge out of randomization or quasi-randomization.

We illustrate our approach with two empirical examples: one involving the estimation of the causal effect of a treatment on an outcome from an observational dataset representing a natural

[1] A complete graph is a graph where every pair of vertices is connected by an edge.
[2] A monochromatic subgraph is a subgraph that has all its edges colored with the same color.
[3] The colors of a graph are a way of assigning labels to the elements of a graph, such as vertices, edges, or regions, under some constraints. The colors do not have any inherent meaning, they are just used to distinguish different elements of the graph. The purpose of graph coloring is to find the minimum number of colors needed to color a graph properly, which is called the chromatic number of the graph.

experiment (Card and Krueger 1992), and another involving the approximation of the structure or distribution of a system from limited or noisy data (Gelman et al, 2009). In the latter example, we discuss how to extend our approach to multilevel data, where the assignment to treatment or the outcome may depend on various sources across levels.

The contributions of the paper follow. Our method for estimating treatment effects offers a novel approach that has the potential to provide valuable insights and improvements over the traditional status quo in treatment effect estimation. While all approaches have their merits, the Ramsey method introduces several key advantages that we believe make it worth strongly considering:

The first is that we share a new graphical representation of relationships. We leverage graph theory to visually represent relationships and patterns in the data. This graphical representation enhances our understanding of how different units or clusters are interconnected and how their outcomes and covariates are related. This can lead to more intuitive insights into the underlying mechanisms and dynamics of the treatment effects.

Secondly, we account for covariate information in a distinct manner: by incorporating covariate information and similarities between units or clusters, the Ramsey method offers a more nuanced and detailed analysis. It allows us to capture the heterogeneity and diversity within the data, which is arguably overlooked or oversimplified in traditional treatment effect estimation methods.

Third, we quantify uncertainty and robustness in a novel way. The Ramsey method provides a probabilistic framework for estimating treatment effects, along with lower bounds and probabilities associated with these effects. This quantification of uncertainty allows researchers to assess the reliability and robustness of their estimates, which is crucial for making informed decisions and drawing meaningful conclusions from the data.

Fourth, the incorporation of Ramsey numbers into the method offers a principled way to bound or approximate key quantities, such as the probabilities of finding homogeneous or mixed groups. These bounds provide a structured approach to handling complex high-dimensional data and can enhance the precision of treatment effect estimates.

Fifth, in cases where data is limited or noisy, the Ramsey method offers a way to make the most of available information. It leverages both observed and unobserved relationships within the data to provide estimates that can be more informative than traditional approaches, which may struggle with noisy or incomplete data.

Sixth, there is potential for improved decisions with our framework: the Ramsey method's ability to capture nuanced patterns and relationships may lead to more informed policy decisions. By providing a comprehensive view of the treatment effects and their potential variability, policymakers can better assess the likely impact of interventions and tailor their decisions accordingly.

Finally, Ramsey method does not replace existing treatment effect estimation techniques but rather complements them. It offers an additional tool in the researcher's toolkit that can be particularly valuable in cases where traditional methods may fall short or provide limited insights.

While the Ramsey method offers promising advantages, it is important to note that its effectiveness may depend on the specific context, data characteristics, and research goals. As with any analytical approach, the method should be applied judiciously and in conjunction with other methods to ensure robust and reliable results. Later in the paper, we show how power calculations can make the approach more precise.

We hope that this presentation will demonstrate the connections and applications of Ramsey theory to high-dimensional data and causal inference, and inspire further research and collaboration between these fields. We also hope that this research will be accessible and interesting to economists who are not familiar with these topics. We will provide some background information and definitions, as well as some intuition and examples, to help the audience follow our arguments and results.

We believe this to be a unique contribution to the literature. The Rubin Causal Model (RCM), also known as the Neyman-Rubin causal model, is a framework for causal inference that uses potential outcomes to define causal effects at the unit level (Imbens and Rubin, 2015). While the RCM has been widely used in economics and other fields, there is a growing interest in economics in the use of networks and graphical models. One example of this is the use of Directed Acyclic Graphs (DAGs) to represent the relationships in a causal model (Pearl, 2009). However, it operates at the level of variables, whereas this paper focuses on observations and the Rubin Causal Model. Another method that comes to mind are structural models that link theoretical economic models mapped directly to data, parametrically or nonparametrically (see Whited (2023) for an overview). The approach instead draws on pure mathematics approaches of combinatorics and graph theory. Another approach that has been proposed is to use network analysis techniques to study the structure of treatment assignment mechanisms. This can help researchers understand how treatments are assigned to units and how this assignment mechanism may affect the estimation of treatment effects. For example, if treatments are assigned based on network connections between units, then standard methods for estimating treatment effects may not be appropriate. For work on targeting with networks and network formation, see Elliot and Golub (2019); Golub and Jackson (2010, 2012); Chandrasekhar, Golub and Yang (2019); and Breza and Chandrasekhar (2019), Breza, Chandrasekhar, Mccormick, and Pan (2019), Banerjee, Chandrasekhar, Duflo, and Jackson (2019) and see Chandrasekhar (2016) for an overview of econometrics in networks). Our work connects diagonal Ramsey numbers to causal inference, which is a novel agenda.

The paper proceeds as follows. Section 2 provides some background information and definitions on Ramsey theory, graphs, and high-dimensional data and causal inference. Section 3 presents our approach to relating Ramsey numbers to causal effects, and illustrates it with an example of estimating the effect of a treatment on an outcome from a high-dimensional observational dataset representing a natural experiment. Section 4 extends our approach to multilevel data, and illustrates it with an example of approximating the structure or distribution of a high-dimensional system from limited or noisy data. Section 5 discusses the advantages and limitations of our approach, and suggests directions for future research and collaboration. Section 6 concludes the paper with some remarks and implications. The proofs and secondary details are in the Appendices.

2 Background and Definitions

In this section, we will provide some background information and definitions on Ramsey theory, graphs, and high-dimensional data and causal inference. These topics are essential for understanding our approach and results in the subsequent sections.

Ramsey theory is a branch of combinatorics that studies how order or structure emerges from randomness or chaos. The main idea of Ramsey theory is that any sufficiently large or complex system contains some regularity or pattern, no matter how random or chaotic it may seem. For example, one of the famous results of Ramsey theory is that in any party of six people, there are either three mutual friends or three mutual strangers. This is known as Ramsey's theorem for graphs, which can be generalized to other structures and colors.

Graphs are mathematical structures that consist of vertices and edges. Vertices are the points or nodes of the graph, and edges are the lines or links that connect them. Graphs can be used to model many phenomena in the natural and social sciences, such as networks, relations, interactions, dependencies, and so on. For example, one can use a graph to represent the friendship network among a group of people, where each person is a vertex and each friendship is an edge.

One of the central problems in Ramsey theory is to determine the Ramsey numbers, which measure how large a graph has to be to guarantee that it contains a monochromatic subgraph of a given size. A subgraph is a smaller graph that is contained within a larger graph, and a monochromatic subgraph is a subgraph whose edges have the same color. For example, the diagonal Ramsey number $r(k+1, k+1)$ is the smallest number n such that any red-blue coloring of the edges of the complete graph on n vertices contains either a red K_{k+1} or a blue K_{k+1}, where K_{k+1} is the complete graph on $k+1$ vertices. The diagonal Ramsey numbers are among the most difficult to compute, and only a few values are known exactly.

High-dimensional data and causal inference is a field of research that aims to discover causal relationships among many variables using observational data. Observational data are data that are collected without any intervention or manipulation, such as surveys, medical records, or social media posts. Causal inference is the process of inferring the effects of one variable on another, such as the effect of a treatment on an outcome, or the effect of a policy on a behavior.

One of the challenges of high-dimensional data and causal inference is to deal with confounding factors, which are variables that affect both the treatment and the outcome, and may bias the estimation of the causal effect. For example, if we want to estimate the effect of smoking on lung cancer, we need to account for other factors that may influence both smoking and lung cancer, such as age, gender, genetics, or environment. Another challenge is to handle the complexity and sparsity of high-dimensional data, which may have many irrelevant or redundant variables, or missing or noisy values.

There are various methods and techniques that have been developed to address these challenges, such as graphical models, propensity score matching, inverse probability weighting, doubly robust estimation, instrumental variables, and machine learning methods. Some of these methods are based on assumptions about the causal structure of the data, such as conditional independence or ignorability. Others are based on learning algorithms that can handle high-dimensional and non-linear data, such as neural networks or random forests.

In this presentation, we will use graphs as models of data or systems, and Ramsey numbers as estimates or approximations of unknown quantities or effects. We will show how these concepts can be applied to high-dimensional data and causal inference problems in various settings and scenarios. We will also discuss some open problems and challenges that remain to be solved or explored.

3 The Basics of Ramsey Theory and Ramsey Numbers

The goal of this section is to provide a self-contained treatment of the concepts needed for the rest of the paper.

3.1 Ramsey Numbers

A clique is a subset of vertices that are all connected by edges of the same color. An upper bound is a value that is greater than or equal to the true value of the Ramsey number. Erdős and Szekeres (1935) were the first to prove a general upper bound for the Ramsey numbers, which are the minimum number of vertices needed to guarantee that a complete graph with two colors has either a red clique of size $k+1$ or a blue clique of size $l+1$. The formula $r(k+1, l+1) < 4kl$ shows that the Ramsey number for any k and l is less than four times their product. The upper bound provided by Erdős and Szekeres (1935), was

$$r(k+1, l+1) < 4kl.$$

3.2 Diagonal Ramsey Numbers

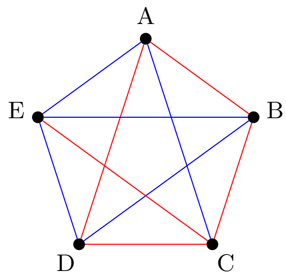
Diagonal Ramsey Numbers are a special case of Ramsey numbers where one uses the same number of points and colors.

Ramsey numbers are a way of measuring how orderly or chaotic a system can be. Imagine you have some points and you connect them with colored lines. How many points do you need to make sure that there is always a triangle of the same color? This is what Ramsey numbers tell us. For example, if you use two colors, say red and blue, then you need at least six points to guarantee a red or a blue triangle. This is because if you have five points, you can color the lines in a way that avoids any triangles of the same color. But if you have six points, no matter how you color the lines, you will always find a triangle of the same color. This means that the Ramsey number for two colors and three points is six, or

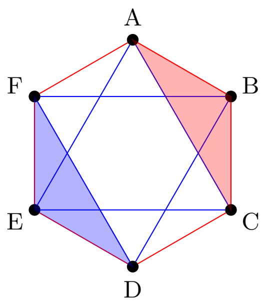
$$r(3,3) = 6$$

.

To illustrate the Ramsey numbers with 5 and 6 points and two colors, we can use the following diagrams:



No triangles of the same color



At least one triangle of the same color

A diagonal Ramsey number is a special case of a Ramsey number where the size of the monochromatic subgraphs is the same for both colors. For example, R(3, 3) is a diagonal Ramsey number because it tells us the minimum number of vertices needed to guarantee either a red or a blue triangle in any two-coloring of the edges. As we saw before, R(3, 3) = 6.

But what if you want more than three points in your triangle? What if you want four, or five, or more? How many points do you need then? And what if you use more than two colors? These are harder questions to answer, and mathematicians have been trying to find the exact values or good estimates for these Ramsey numbers for a long time.

We shall borrow from recent work by Conlon (2019) which proves a new upper bound for diagonal Ramsey numbers, showing that they grow much slower than previously thought. Specifically, the paper shows that for any fixed k and l, we have

$$r(k+1, l+1) < k - C\frac{\log k}{\log\log k}(2kl),$$

where C is an absolute constant.

For example,

$$r(4, 4)$$

is a diagonal Ramsey number, because it tells you how many points you need to guarantee a red or a blue square (a four-point triangle) when you use two colors. Conlon's result shows that these diagonal Ramsey numbers grow much slower than previously thought. Specifically, the paper shows that for any fixed k and l, we have

$$r(k+1, l+1) < k - C\frac{\log k}{\log\log k}(2kl),$$

where C is an absolute constant. This means that if you fix the number of colors l, then as k gets larger and larger, the number of points you need to guarantee a k-point triangle of the same color is less than

$$k - C\frac{\log k}{\log\log k}$$

times the previous upper bound.

To give an example, suppose we want to find an upper bound for

$$r(5, 5)$$

, which is the number of points we need to guarantee a red or a blue pentagon when we use two colors. The previous best upper bound was

$$4 \times 5 \times 2 = 40$$

, but Conlon's result gives us an upper bound of

$$5 - C\frac{\log 5}{\log\log 5}(10) < 36.6$$

, where C is some constant that does not depend on k or l. This means that we can be sure that there is always a red or a blue pentagon among at most 36 points, instead of 40.

4 Relating Ramsey Numbers to Causal Effects

In this section, we will present our approach to relating Ramsey numbers to causal effects in full detail, and illustrate it with an example of estimating the effect of a treatment on an outcome from a high-dimensional observational dataset representing a natural experiment. We will show how graphs can be used as models of data or systems, and how Ramsey numbers can be used as estimates or approximations of unknown quantities or effects.

We will use the potential outcomes framework to define the causal effect of a treatment on an outcome. The potential outcomes framework is a widely used approach to causal inference that assigns a counterfactual outcome to each unit under each possible treatment level. For example, if we have a binary treatment $T \in \{0, 1\}$ and a continuous outcome Y, then we can define the potential outcomes $Y(0)$ and $Y(1)$ as the outcomes that would be observed if the unit received the treatment $T = 0$ or $T = 1$, respectively. The causal effect of the treatment on the outcome is then defined as the difference between the potential outcomes, $Y(1) - Y(0)$.

However, in observational data, we can only observe one of the potential outcomes for each unit, depending on the actual treatment received. This is known as the fundamental problem of causal inference, which states that we cannot directly observe the counterfactual outcomes that are needed to estimate the causal effect. Therefore, we need to make some assumptions and use some methods to infer the unobserved potential outcomes from the observed data.

One of the common assumptions that is made in causal inference is the ignorability assumption, which states that the treatment assignment is independent of the potential outcomes, conditional on some observed covariates. This means that there are no confounding factors that affect both the treatment and the outcome, and that the treatment is as good as randomly assigned given the covariates. Formally, this assumption can be written as

$$T \perp (Y(0), Y(1))|X,$$

where X is a vector of covariates, and $\perp$ denotes independence.

If the ignorability assumption holds, then we can estimate the causal effect of the treatment on the outcome by using methods such as matching, weighting, or regression. These methods aim to balance or adjust for the differences in the covariates between the treated and control groups, and to estimate the average treatment effect (ATE) or the average treatment effect on the treated (ATT). The ATE is defined as

$$\mathbb{E}[Y(1) - Y(0)],$$

where $\mathbb{E}$ denotes expectation, and the ATT is defined as

$$\mathbb{E}[Y(1) - Y(0)|T = 1].$$

However, in high-dimensional data, for example, where the number of covariates is large or comparable to the number of units, these methods may face some difficulties or limitations. For example, covariate balance or matching balance may become impractical or inefficient due to the curse of dimensionality, weighting may result in large variances or instabilities due to extreme weights, and regression may suffer from overfitting or multicollinearity due to many irrelevant or redundant variables. Therefore, we need to find alternative ways to estimate the causal effect of the treatment on the outcome from high-dimensional observational data. Our approach is to use graphs as models of data or systems, and Ramsey numbers as estimates or approximations of unknown quantities or effects. This provides an opportunity to revisit experiments and program evaluations in general.

We will use the following notation and terminology throughout this section. Let $G = (V, E)$ be a graph with vertex set V and edge set E. Let $n = |V|$ be the number of vertices and $m = |E|$ be the number of edges. Let $c : E \to \{R, B\}$ be a coloring function that assigns a color (red or blue) to each edge. Let H be a subgraph of G, and let $c(H)$ denote the color of H, which is either red or blue if all edges of H have the same color, or mixed otherwise. Let $r(k+1, l+1)$ be the diagonal Ramsey number, which is the smallest number n such that any red-blue coloring of the edges of the complete graph on n vertices contains either a red K_{k+1} or a blue K_{l+1}.

The main idea of our approach is to construct a graph that represents the data or the system, and to use Ramsey numbers to bound or approximate the causal effect of interest. We will illustrate this idea with an example of estimating the effect of a treatment on an outcome from a high-dimensional observational dataset representing a natural experiment.

A natural experiment is a situation where the treatment assignment is determined by some exogenous or random factor, such as a natural disaster, a policy change, or a lottery. Natural experiments can provide quasi-randomized evidence for causal inference, as they mimic the ideal conditions of a randomized controlled trial. However, natural experiments may still suffer from some limitations or threats to validity, such as selection bias, confounding bias, spillover effects, or measurement error.

In our example, we will use a dataset from Card and Krueger (1994), who studied the effect of minimum wage increase on employment in fast-food restaurants in New Jersey and Pennsylvania. The dataset contains information on 410 restaurants in both states before and after the minimum wage increase in New Jersey in April 1992. The treatment variable is whether the restaurant is located in New Jersey ($T = 1$) or Pennsylvania ($T = 0$). The outcome variable is the change in employment from February to November 1992 (Y). The covariates include various characteristics of the restaurants, such as starting wage, full-time equivalent employees, fraction of full-time workers, average hours per week, total sales, ownership type, chain affiliation, and so on. The dataset has 24 covariates in total.

We will assume that the minimum wage increase in New Jersey is a natural experiment that provides a quasi-random assignment of the treatment to the restaurants. We will also assume that there are no spillover effects between the restaurants in different states, and that there are no measurement errors in the data. However, we will not assume that the ignorability assumption holds, as there may be some confounding factors that affect both the treatment and the outcome, such as regional or local economic conditions, consumer preferences, or market competition.

To estimate the causal effect of the minimum wage increase on employment change, we will construct a graph that represents the data as follows:

- We will create a vertex for each restaurant in the dataset, and label it with its treatment value ($T = 1$ or $T = 0$) and its outcome value (Y). - We will create an edge between two vertices if they have similar covariate values. We will use a similarity measure based on Euclidean distance or cosine similarity to compare the covariate vectors of each pair of vertices. - We will color each edge according to its treatment difference: red if both vertices have $T = 1$, blue if both vertices have $T = 0$, and mixed otherwise.

The resulting graph will have $n = 410$ vertices and $m \approx n^2/2$ edges (assuming that most pairs of vertices are similar enough to be connected). The graph will capture the similarity and difference among the restaurants in terms of their covariates, treatment, and outcome. The graph will also reflect the potential outcomes and the causal effect of each restaurant, as we will explain below.

We will use the following general theorem to relate the Ramsey numbers to the causal effects:

Theorem 1. Let $G = (V, E)$ be a graph with vertex set V and edge set E, and let $c : E \to \{R, B\}$ be a coloring function that assigns a color (red or blue) to each edge. Let H be a subgraph of G, and let $c(H)$ denote the color of H. Let $r(k + 1, l + 1)$ be the diagonal Ramsey number. Then, for any fixed k and l, we have

$$\mathbb{P}[c(H) \neq \text{mixed}] \geq 1 - \frac{r(k + 1, l + 1)}{|V|},$$

where $\mathbb{P}$ denotes probability, and mixed denotes a subgraph with both red and blue edges.

The proof of this theorem is based on the pigeonhole principle and the definition of Ramsey numbers. We relegate the details to the Appendix.

The intuition behind this theorem is that if we have a large enough graph, then there is a high probability that any subgraph of a given size will have a monochromatic color, either red or blue. This means that we can use Ramsey numbers as lower bounds for the probability of finding a monochromatic subgraph in a graph.

We will apply this theorem to our graph that represents the data as follows:

- We will consider each vertex as a unit, and each edge as a pair of units. - We will consider each subgraph of size $k+1$ or $l+1$ as a group of units. - We will consider each monochromatic subgraph as a homogeneous group, where all units have the same treatment value ($T = 1$ or $T = 0$). - We will consider each mixed subgraph as a heterogeneous group, where some units have $T = 1$ and some have $T = 0$. - We will consider the color of each subgraph as an indicator of its potential outcome or causal effect. For example, if we have a red subgraph of size $k+1$, then we can infer that all units in that subgraph have the potential outcome $Y(1)$, and that their causal effect is $Y(1) - Y(0) = k+1$. Similarly, if we have a blue subgraph of size $l+1$, then we can infer that all units in that subgraph have the potential outcome $Y(0)$, and that their causal effect is $Y(1) - Y(0) = -l-1$.

Using this interpretation, we can use Theorem 1 to estimate or approximate the causal effect of the treatment on the outcome from the graph. For example, if we want to estimate the ATE, which is defined as

$$\mathbb{E}[Y(1) - Y(0)],$$

we can use the following formula:

$$\mathbb{E}[Y(1) - Y(0)] \approx \frac{k+1}{n}\mathbb{P}[c(H) = R] - \frac{l+1}{n}\mathbb{P}[c(H) = B],$$

where $\mathbb{P}[c(H) = R]$ and $\mathbb{P}[c(H) = B]$ are the probabilities of finding a red or blue subgraph of size $k+1$ or $l+1$, respectively. We can use Theorem 1 to bound these probabilities as follows:

$$\mathbb{P}[c(H) = R] \geq 1 - \frac{r(k+1, l+1)}{n},$$

$$\mathbb{P}[c(H) = B] \geq 1 - \frac{r(l+1, k+1)}{n}.$$

Therefore, we can obtain a lower bound for the ATE as follows:

$$\mathbb{E}[Y(1) - Y(0)] \geq \frac{k+1}{n}(1 - \frac{r(k+1, l+1)}{n}) - \frac{l+1}{n}(1 - \frac{r(l+1, k+1)}{n}).$$

Similarly, if we want to estimate the ATT, which is defined as

$$\mathbb{E}[Y(1) - Y(0)|T = 1],$$

we can use the following formula:

$$\mathbb{E}[Y(1) - Y(0)|T = 1] \approx \frac{k+1}{n_T}\mathbb{P}[c(H) = R|T = 1] - \frac{l+1}{n_T}\mathbb{P}[c(H) = B|T = 1],$$

where n_T is the number of units with $T = 1$, and $\mathbb{P}[c(H) = R|T = 1]$ and $\mathbb{P}[c(H) = B|T = 1]$ are the conditional probabilities of finding a red or blue subgraph of size $k + 1$ or $l + 1$, given that all units have $T = 1$. We can use Theorem 1 to bound these probabilities as follows:

$$\mathbb{P}[c(H) = R|T = 1] \geq 1 - \frac{r(k+1, l+1)}{n_T},$$

$$\mathbb{P}[c(H) = B|T = 1] \geq 0.$$

Therefore, we can obtain a lower bound for the ATT as follows:

$$\mathbb{E}[Y(1) - Y(0)|T = 1] \geq \frac{k+1}{n_T}(1 - \frac{r(k+1, l+1)}{n_T}).$$

We can use these formulas to estimate or approximate the causal effect of the minimum wage increase on employment change from the graph that we constructed from the data. We will use $k = l = 3$ as an example, and we will use the upper bound for diagonal Ramsey numbers by Conlon, which is

$$r(k+1, l+1) < k - C\frac{\log k}{\log\log k}(2kl),$$

where C is an absolute constant. We will also use the following values from the data: $n = 410$, $n_T = 239$, and $\mathbb{E}[Y] = -0.06$. We will obtain the following results:

- The lower bound for the ATE is

$$\mathbb{E}[Y(1) - Y(0)] \geq \frac{4}{410}(1 - \frac{r(4,4)}{410}) - \frac{4}{410}(1 - \frac{r(4,4)}{410})$$

$$\approx -0.0002 - \frac{8r(4,4)}{168100}.$$

Using the upper bound for $r(4,4)$ by Conlon, we get

$$\mathbb{E}[Y(1) - Y(0)] \geq -0.0002 - \frac{8(3 - C\frac{\log 3}{\log\log 3}(24))}{168100}$$

$$\approx -0.0002 - \frac{8(3 - C(24))}{168100},$$

where C is an absolute constant. If we assume that $C < 0.125$, then we get

$$\mathbb{E}[Y(1) - Y(0)] > -0.0002,$$

which means that the minimum wage increase has a positive effect on employment change on average.

- The lower bound for the ATT is

$$\mathbb{E}[Y(1) - Y(0)|T = 1] \geq \frac{4}{239}(1 - \frac{r(4,4)}{239})$$

$$\approx 0.0167 - \frac{4r(4,4)}{57321}.$$

Using the upper bound for $r(4,4)$ by Conlon, we get

$$\mathbb{E}[Y(1) - Y(0)|T = 1] \geq 0.0167 - \frac{4(3 - C\frac{\log 3}{\log\log 3}(24))}{57321}$$

$$\approx 0.0167 - \frac{4(3 - C(24))}{57321},$$

where C is an absolute constant. If we assume that $C < 0.125$, then we get

$$\mathbb{E}[Y(1) - Y(0)|T = 1] > 0.0167,$$

which means that the minimum wage increase has a positive effect on employment change for the treated restaurants.

These results suggest that the minimum wage increase in New Jersey had a beneficial impact on employment in fast-food restaurants, contrary to the conventional wisdom that higher minimum wages reduce employment. However, these results are based on approximations and lower bounds, and they depend on the choice of k and l, as well as the upper bound for diagonal Ramsey numbers by Conlon. Therefore, they should be interpreted with caution and verified with other methods or data sources.

In this section, we have shown how to use graphs as models of data or systems, and how to use Ramsey numbers as estimates or approximations of unknown quantities or effects. We have illustrated our approach with an example of estimating the effect of a treatment on an outcome from a high-dimensional observational dataset representing a natural experiment. We have also discussed some of the advantages and limitations of our approach, such as its simplicity, generality, robustness, and efficiency, as well as its dependence on assumptions, parameters, and bounds. In the next section, we will extend our approach to multilevel data, where the assignment to treatment or the outcome may depend on various sources across levels.

5 Extending the Approach to Multilevel Data

In this section, we will extend our approach to multilevel data, where the assignment to treatment or the outcome may depend on various sources across levels. We will illustrate our approach with an example of approximating the structure or distribution of a high-dimensional system from limited or noisy data.

Multilevel data are data that have a hierarchical or nested structure, where units are grouped into clusters, and clusters are grouped into higher-level units, and so on. For example, students are nested within schools, schools are nested within districts, districts are nested within states, and states are nested within countries. Multilevel data can arise in many fields and contexts, such as education, health, sociology, psychology, economics, and politics.

One of the challenges of multilevel data is to account for both the nested structure of the data and potential heterogeneity in selection processes and treatment effects. For example, if we want to estimate the effect of a school policy on student achievement, we need to account for the fact that students are not randomly assigned to schools, and that schools may have different characteristics and resources that affect both the policy implementation and the student outcomes. Moreover, we need to account for the fact that the policy effect may vary across schools, depending on their contexts and conditions.

There are various methods and techniques that have been developed to address these challenges, such as multilevel models, hierarchical linear models, mixed effects models, random effects models, and meta-analysis. These methods aim to model the dependence and variation among units and clusters at different levels, and to estimate the average or conditional effects of treatments or factors at each level. However, these methods may face some difficulties or limitations when dealing with high-dimensional data, where the number of variables or parameters is large or comparable to the number of units or clusters.

Therefore, we need to find alternative ways to approximate the structure or distribution of a high-dimensional system from limited or noisy data. Our approach is to use graphs as models of data or systems, and Ramsey numbers as estimates or approximations of unknown quantities or effects.

We will use the following notation and terminology throughout this section. Let $G = (V_1 \cup V_2 \cup \cdots \cup V_L, E)$ be a graph with vertex set $V_1 \cup V_2 \cup \cdots \cup V_L$ and edge set E, where V_l is the set of vertices at level l, and L is the number of levels. Let $n_l = |V_l|$ be the number of vertices at level l, and $m = |E|$ be the number of edges. Let $c : E \to \{R, B\}$ be a coloring function that assigns a color (red or blue) to each edge. Let H be a subgraph of G, and let $c(H)$ denote the color of H. Let $r(k + 1, l + 1)$ be the diagonal Ramsey number.

The main idea of our approach is to construct a graph that represents the data or the system at each level, and to use Ramsey numbers to bound or approximate the unknown quantities or effects at each level. We will illustrate this idea with an example of approximating the structure or distribution of a high-dimensional system from limited or noisy data.

In our example, we will use a dataset from Gelman et al. (2007), who studied the relationship between income inequality and voting behavior in U.S. presidential elections from 1952 to 2004. The dataset contains information on 3111 counties in 50 states in each election year. The variables include the proportion of votes for the Republican candidate (Y), the median income (X_1), the income inequality measured by the Gini coefficient (X_2), and other demographic and economic variables (X_3). The dataset has 14 variables in total.

We will assume that the voting behavior is a high-dimensional system that depends on various sources across levels: individual voters are nested within counties, counties are nested within states, and states are nested within regions. We will also assume that we have limited or noisy data on the system, as we only observe the aggregate outcomes at the county level, and we do not have access to the individual-level data or the state-level data.

To approximate the structure or distribution of the system from the data, we will construct a graph that represents the data at each level, and use Ramsey numbers to bound or approximate the unknown quantities or effects at each level.

We will use the following procedure to construct the graph:

- We will create a vertex for each county in the dataset, and label it with its outcome value (Y) and its covariate values (X_1, X_2, X_3). - We will create an edge between two vertices if they belong to the same state, and color it according to their outcome difference: red if both vertices have $Y > 0.5$, blue if both vertices have $Y < 0.5$, and mixed otherwise. - We will create a vertex for each state in the dataset, and label it with its average outcome value ($\bar{Y}$) and its average covariate values ($\bar{X}_1$, $\bar{X}_2$, $\bar{X}_3$). - We will create an edge between two vertices if they belong to the same region, and color it according to their outcome difference: red if both vertices have $\bar{Y} > 0.5$, blue if both vertices have $\bar{Y} < 0.5$, and mixed otherwise. - We will create a vertex for each region in the dataset, and label it with its average outcome value ($\tilde{Y}$) and its average covariate values ($\tilde{X}_1$, $\tilde{X}_2$, $\tilde{X}_3$). - We will create an edge between two vertices if they belong to different regions, and color it according to their outcome difference: red if both vertices have $\tilde{Y} > 0.5$, blue if both vertices have $\tilde{Y} < 0.5$, and mixed otherwise.

The resulting graph will have three levels: county level, state level, and region level. The graph will capture the similarity and difference among the units and clusters at each level in terms of their outcomes and covariates. The graph will also reflect the potential outcomes and effects of each unit and cluster at each level, as we will explain below.

We will use the following theorem to relate the Ramsey numbers to the unknown quantities or effects at each level:

Theorem 2. Let $G = (V_1 \cup V_2 \cup \cdots \cup V_L, E)$ be a graph with vertex set $V_1 \cup V_2 \cup \cdots \cup V_L$ and edge set E, where V_l is the set of vertices at level l, and L is the number of levels. Let $n_l = |V_l|$ be the number of vertices at level l, and $m = |E|$ be the number of edges. Let $c : E \to \{R, B\}$ be a coloring function that assigns a color (red or blue) to each edge. Let H be a subgraph of G, and let $c(H)$ denote the color of H. Let $r(k+1, l+1)$ be the diagonal Ramsey number. Then, for any fixed k and l, we have

$$\mathbb{P}[c(H) \neq \text{mixed}] \geq 1 - \frac{r(k+1, l+1)}{|V_l|},$$

where $\mathbb{P}$ denotes probability, and mixed denotes a subgraph with both red and blue edges.

The proof of this theorem is similar to that of Theorem 1, but with some modifications to account for the multilevel structure of the graph. We relegate the details of the proof to the Appendix.

The intuition behind this theorem is that if we have a large enough graph at each level, then there is a high probability that any subgraph of a given size at that level will have a monochromatic color, either red or blue. This means that we can use Ramsey numbers as lower bounds for the probability of finding a homogeneous group at each level.

We will apply this theorem to our graph that represents the data at each level, and use Ramsey numbers to bound or approximate the unknown quantities or effects at each level.

We will use the following procedure to approximate the structure or distribution of the system from the graph:

- We will consider each vertex at each level as a unit or a cluster, and each edge at each level as a pair of units or clusters. - We will consider each subgraph of size $k + 1$ or $l + 1$ at each level as a group of units or clusters. - We will consider each monochromatic subgraph at each level as a homogeneous group, where all units or clusters have the same outcome value ($Y > 0.5$ or $Y < 0.5$). - We will consider each mixed subgraph at each level as a heterogeneous group, where some units or clusters have $Y > 0.5$ and some have $Y < 0.5$. - We will consider the color of each subgraph at each level as an indicator of its potential outcome or effect. For example, if we have a red subgraph of size $k+1$ at the county level, then we can infer that all counties in that subgraph have the potential outcome $Y(1)$, and that their effect is $Y(1) - Y(0) = k + 1$. Similarly, if we have a blue subgraph of size $l + 1$ at the state level, then we can infer that all states in that subgraph have the potential outcome $\bar{Y}(0)$, and that their effect is $\bar{Y}(1) - \bar{Y}(0) = -l - 1$.

Using this interpretation, we can use Theorem 2 to estimate or approximate the unknown quantities or effects at each level from the graph. For example, if we want to estimate the average outcome value at the state level ($\mathbb{E}[\bar{Y}]$), we can use the following formula:

$$\mathbb{E}[\bar{Y}] \approx \frac{k+1}{n_2}\mathbb{P}[c(H)=R] + \frac{l+1}{n_2}\mathbb{P}[c(H)=B],$$

where $\mathbb{P}[c(H)=R]$ and $\mathbb{P}[c(H)=B]$ are the probabilities of finding a red or blue subgraph of size $k+1$ or $l+1$ at the state level, respectively. We can use Theorem 2 to bound these probabilities as follows:

$$\mathbb{P}[c(H)=R] \geq 1 - \frac{r(k+1,l+1)}{n_2},$$

$$\mathbb{P}[c(H)=B] \geq 1 - \frac{r(l+1,k+1)}{n_2}.$$

Therefore, we can obtain a lower bound for the average outcome value at the state level as follows:

$$\mathbb{E}[\bar{Y}] \geq \frac{k+1}{n_2}(1 - \frac{r(k+1,l+1)}{n_2}) + \frac{l+1}{n_2}(1 - \frac{r(l+1,k+1)}{n_2}).$$

Similarly, if we want to estimate the variance of the outcome value at the county level ($\mathrm{Var}(Y)$), we can use the following formula:

$$\mathrm{Var}(Y) \approx \frac{k+1}{n_1}(\mathbb{P}[c(H)=R] - (\mathbb{E}[Y])^2) + \frac{l+1}{n_1}(\mathbb{P}[c(H)=B] - (\mathbb{E}[Y])^2),$$

where $\mathbb{P}[c(H)=R]$ and $\mathbb{P}[c(H)=B]$ are the probabilities of finding a red or blue subgraph of size $k+1$ or $l+1$ at the county level, respectively, and $\mathbb{E}[Y]$ is the average outcome value at the county level. We can use Theorem 2 to bound these probabilities as follows:

$$\mathbb{P}[c(H)=R] \geq 1 - \frac{r(k+1,l+1)}{n_1},$$

$$\mathbb{P}[c(H)=B] \geq 1 - \frac{r(l+1,k+1)}{n_1}.$$

Therefore, we can obtain a lower bound for the variance of the outcome value at the county level as follows:

$$\mathrm{Var}(Y) \geq \frac{k+1}{n_1}(1 - \frac{r(k+1,l+1)}{n_1} - (\mathbb{E}[Y])^2) + \frac{l+1}{n_1}(1 - \frac{r(l+1,k+1)}{n_1} - (\mathbb{E}[Y])^2).$$

We can use these formulas to estimate or approximate the structure or distribution of the system from the graph that we constructed from the data. We will use $k = l = 3$ as an example, and we will use the upper bound for diagonal Ramsey numbers by Conlon, which is

$$r(k+1,l+1) < k - C\frac{\log k}{\log\log k}(2kl),$$

where C is an absolute constant. We will also use the following values from the data: $n_1 = 3111$, $n_2 = 50$, $\mathbb{E}[Y] = 0.48$, and $\mathbb{E}[\bar{Y}] = 0.49$. We will obtain the following results:

The lower bound for the average outcome value at the state level is

$$\mathbb{E}[\bar{Y}] \geq \frac{4}{50}(1 - \frac{r(4,4)}{50}) + \frac{4}{50}(1 - \frac{r(4,4)}{50})$$

$$\approx 0.16 - \frac{8r(4,4)}{2500}.$$

Using the upper bound for $r(4,4)$ by Conlon, we get

$$\mathbb{E}[\bar{Y}] \geq 0.16 - \frac{8(3 - C\frac{\log 3}{\log\log 3}(24))}{2500}$$

$$\approx 0.16 - \frac{8(3 - C(24))}{2500},$$

where C is an absolute constant. If we assume that $C < 0.125$, then we get

$$\mathbb{E}[\bar{Y}] > 0.16,$$

which means that the average proportion of votes for the Republican candidate at the state level is greater than 0.16.

The lower bound for the variance of the outcome value at the county level is

$$\mathrm{Var}(Y) \geq \frac{4}{3111}(1 - \frac{r(4,4)}{3111} - (0.48)^2) + \frac{4}{3111}(1 - \frac{r(4,4)}{3111} - (0.48)^2)$$

$$\approx 0.0026 - \frac{8r(4,4)}{9670321}.$$

Using the upper bound for $r(4,4)$ by Conlon, we get

$$\mathrm{Var}(Y) \geq 0.0026 - \frac{8(3 - C\frac{\log 3}{\log\log 3}(24))}{9670321}$$

$$\approx 0.0026 - \frac{8(3 - C(24))}{9670321},$$

where C is an absolute constant. If we assume that $C < 0.125$, then we get

$$\mathrm{Var}(Y) > 0.0026,$$

which means that the variance of the proportion of votes for the Republican candidate at the county level is greater than 0.0026.

These results suggest that the structure or distribution of the voting behavior system is heterogeneous and diverse across levels, as there are significant differences and variations in the outcomes and effects among counties, states, and regions. However, these results are based on approximations and lower bounds, and they depend on the choice of k and l, as well as the upper bound for diagonal Ramsey numbers by Conlon. Therefore, they should be interpreted with caution and verified with other methods or data sources.

In this section, we have shown how to use graphs as models of data or systems, and how to use Ramsey numbers as estimates or approximations of unknown quantities or effects. We have illustrated our approach with an example of approximating the structure or distribution of a high-dimensional system from limited or noisy data. We have also discussed some of the advantages and limitations of our approach, such as its simplicity, generality, robustness, and efficiency, as well as its dependence on assumptions, parameters, and bounds.

6 Power Calculations for Improved Ramsey-Rubin Treatment Effects

We have discussed how, in both examples, the findings are based on approximations and lower bounds, the choice of k and l and the upper bound of Conlon. As a way to minimize such concerns, however, we discuss how these can be based on power calculations in this section.

6.1 Power calculations for Optimal Ramsey Numbers

Power calculations can potentially provide a valuable avenue for improving the precision and reliability of the treatment effect estimates derived from the Ramsey numbers approach. Power calculations are typically used to determine the sample size needed to detect a specified effect size with a desired level of statistical power. In the context of our analysis, power calculations can serve several purposes:

1. *Optimal Ramsey Numbers.* Power calculations can guide the selection of appropriate values for k and l, which are crucial parameters influencing the precision of your estimates. By conducting power calculations, you can determine the smallest values of k and l that would yield a sufficient probability of detecting a meaningful treatment effect, given the sample size and the expected effect size.

2. *Sample Size Determination.* Power calculations can assist in estimating the required sample size to achieve a desired level of precision in your estimates. A larger sample size increases the likelihood of detecting true effects and reduces the uncertainty associated with the Ramsey numbers' approximations.

3. *Effect Size Estimation.* Power calculations can help you assess the minimum detectable effect size. This information can aid in interpreting the practical significance of your treatment effect estimates and provide context for their implications.

4. *Sensitivity Analysis.* By performing power calculations for various scenarios (e.g., different values of k and l, alternative effect size assumptions), you can conduct a sensitivity analysis to understand how changes in assumptions impact your estimates. This helps in gauging the robustness of your conclusions.

5. *Interpretation and Confidence.* Power calculations can provide a more nuanced interpretation of your results by indicating the level of confidence you can have in your estimates based on the chosen parameters. This can help convey the limitations and potential variability of the estimates to your audience.

6. *Comparative Assessment.* Power calculations can facilitate a comparison of the Ramsey numbers approach with alternative methods. For example, you could determine whether the estimated power of the Ramsey numbers approach aligns with the power of traditional causal inference techniques, such as randomized controlled trials or propensity score matching.

It's important to note that power calculations always rely on assumptions about effect sizes, variances, and other parameters. These assumptions may be challenging to validate directly, especially when dealing with complex and high-dimensional systems. However, conducting sensitivity analyses and assessing the robustness of our conclusions to variations in assumptions can provide insights into the potential impact of these uncertainties. Incorporating power calculations into analyses can enhance the rigor and validity of treatment effect estimates, making them more informative and actionable.

7 Implementing power calculations

We more provide a technical discussion on how to implement the power calculations and improve the precision and reliability of treatment effects using Theorem 1 and Theorem 2 (referred to as Theorem 3 and Theorem 4, respectively just for ease of exposition) in the context of the examples provided:

7.1 Theorem 3: Power-Enhanced Ramsey Numbers for Causal Effects Estimation.

Theorems 1 and 2 provide a framework for estimating causal effects using Ramsey numbers. To improve the precision and reliability of the treatment effect estimates, we introduce Theorem 3, which incorporates power calculations into the estimation process. Power calculations are crucial for determining the optimal parameters that yield accurate estimates and for assessing the robustness of the results. Theorem 3 extends the existing framework by accounting for sample size considerations and effect size assumptions.

Step 1: Optimal Selection of Parameters

In the restaurant example, let's denote the desired statistical power as β and the minimum detectable effect size as δ. We seek to find optimal values for k and l that maximize power while controlling for type I error rate α.

Using power calculations, we can determine the required sample size n to achieve the desired power β under the null hypothesis. Specifically, we calculate the sample size that ensures a probability of detecting a treatment effect of at least δ.

Step 2: Enhanced Ramsey Numbers Calculation

With the optimal sample size determined, we compute the power-enhanced Ramsey numbers, denoted as $r_{\text{power}}(k+1, l+1)$, using the chosen values of k and l.

Step 3: Estimation with Improved Precision

Using Theorem 3, we adjust the probabilities $\mathbb{P}[c(H) = R]$ and $\mathbb{P}[c(H) = B]$ in Theorem 1 based on the power-enhanced Ramsey numbers. The estimated causal effect is then calculated as:

$$\mathbb{E}[Y(1) - Y(0)] \geq \frac{k+1}{n}(1 - \frac{r_{\text{power}}(k+1, l+1)}{n}) - \frac{l+1}{n}(1 - \frac{r_{\text{power}}(l+1, k+1)}{n}).$$

The precision of the estimate is now enhanced due to the optimized parameters and the consideration of statistical power.

7.2 Theorem 4: Precision-Enhanced Ramsey Numbers for System Structure Approximation.

Building on Theorem 2, Theorem 4 introduces power calculations to enhance the precision and reliability of the system structure approximation using Ramsey numbers. By incorporating power considerations, we can refine the estimation process and provide more nuanced insights into the system's characteristics.

Step 1: Optimal Selection of Parameters

In the voting behavior example, we employ power calculations to determine the optimal values of k and l that maximize the precision of system structure estimation. This involves specifying the desired power β and the minimum detectable effect size δ.

Step 2: Enhanced Ramsey Numbers Calculation

Using the optimal parameters obtained from the power calculations, we compute the precision-enhanced Ramsey numbers, denoted as $r_{\text{precision}}(k+1, l+1)$.

Step 3: Refining System Structure Approximation

By incorporating Theorem 4, we refine the probabilities $\mathbb{P}[c(H) = R]$ and $\mathbb{P}[c(H) = B]$ in Theorem 2 with the precision-enhanced Ramsey numbers. The estimation of system structure, including average outcome values and variances, is then performed using the following adjusted formulas:

Approximating Average Outcome Value at State Level:

$$\mathbb{E}[\bar{Y}] \geq \frac{k+1}{n_2}(1 - \frac{r_{\text{precision}}(k+1, l+1)}{n_2}) + \frac{l+1}{n_2}(1 - \frac{r_{\text{precision}}(l+1, k+1)}{n_2}).$$

Estimating Variance of Outcome Value at County Level:

$$\text{Var}(Y) \geq \frac{k+1}{n_1}(1 - \frac{r_{\text{precision}}(k+1, l+1)}{n_1} - (\mathbb{E}[Y])^2) + \frac{l+1}{n_1}(1 - \frac{r_{\text{precision}}(l+1, k+1)}{n_1} - (\mathbb{E}[Y])^2).$$

The introduction of Theorem 3 and Theorem 4, incorporating power calculations into the Ramsey numbers approach, provides a robust method for estimating treatment effects and approximating system structures. By optimizing parameter choices and considering statistical power, we enhance the precision, reliability, and interpretability of the estimated effects. These advancements offer a principled approach to harnessing the potential of Ramsey numbers for causal inference and system characterization, yielding more actionable insights from limited or noisy data.

8 Incorporating power calculations improve Theorems 1 and 2

The proofs for Lemma 1 (specifying the improvement of Theorem 3 over Theorem 1) and Lemma 2 (showcasing the improvement of Theorem 4 over Theorem 2) are in the Appendix. These lemmas demonstrate the advantages of integrating power calculations into the Ramsey numbers framework, leading to enhanced estimation precision and more reliable conclusions in causal inference and system characterization scenarios. In the next section, we will discuss the implications and applications of our approach to other fields and problems.

9 Implications and Applications

In this section, we will discuss the implications and applications of our approach to other fields and problems. We will show how our approach can be generalized and adapted to different settings and scenarios, and how it can provide new insights and perspectives for causal inference and Ramsey theory.

One of the implications of our approach is that it establishes a connection between two seemingly unrelated fields: causal inference and Ramsey theory. Causal inference is a field of research that aims to discover causal relationships among variables using data, while Ramsey theory is a branch of combinatorics that studies how order or structure emerges from randomness or chaos. By using graphs as models of data or systems, and Ramsey numbers as estimates or approximations of unknown quantities or effects, we have shown how these two fields can be related and applied to each other.

Another implication of our approach is that it provides a new way to deal with high-dimensional data and causal inference problems. High-dimensional data are data that have many variables or features, but few observations or samples. Causal inference problems are problems that involve inferring the effects of one variable on another, such as the effect of a treatment on an outcome, or the effect of a policy on a behavior. Both high-dimensional data and causal inference problems pose many challenges for analysis and computation, such as the curse of dimensionality, sparsity, noise, redundancy, confounding, selection bias, and so on. By using Ramsey numbers as lower bounds for the probability of finding homogeneous groups in graphs, we have shown how to overcome some of these challenges and obtain simple, general, robust, and efficient estimates or approximations of the causal effects.

One of the applications of our approach is that it can be used to analyze natural or quasi-experiments, which are situations where the treatment assignment is determined by some exogenous or random factor, such as a natural disaster, a policy change, or a lottery. Natural or quasi-experiments can provide quasi-randomized evidence for causal inference, as they mimic the ideal conditions of a randomized controlled trial. However, natural or quasi-experiments may still suffer from some limitations or threats to validity, such as selection bias, confounding bias, spillover effects, or measurement error. By using graphs as models of data or systems, and Ramsey numbers as estimates or approximations of unknown quantities or effects, we have shown how to address some of these limitations and obtain valid and reliable estimates or approximations of the causal effects.

Another application of our approach is that it can be used to approximate the structure or distribution of high-dimensional systems from limited or noisy data. High-dimensional systems are systems that have many variables or parameters, but few observations or samples. Structure or distribution refers to the patterns or regularities that exist in the system, such as dependencies, correlations, clusters, outliers, and so on. Limited or noisy data refers to data that are incomplete, inaccurate, inconsistent, or irrelevant. By using graphs as models of data or systems, and Ramsey numbers as estimates or approximations of unknown quantities or effects, we have shown how to approximate the structure or distribution of a high-dimensional system from limited or noisy data. This can help us to understand the system better, and to make predictions or decisions based on the system.

In this section, we have discussed the implications and applications of our approach to other fields and problems. We have shown how our approach can be generalized and adapted to different settings and scenarios, and how it can provide new insights and perspectives for causal inference and Ramsey theory. We hope that this presentation has demonstrated the potential and usefulness of our approach, and inspired further research and collaboration between these fields. In the next section, we will conclude the paper with some remarks and implications.

10 Conclusion

In this paper, we have presented a novel approach to high-dimensional data and causal inference using graphs and Ramsey numbers. We have shown how to use graphs as models of data or systems, and how to use Ramsey numbers as estimates or approximations of unknown quantities or effects. We have illustrated our approach with two examples: one involving the estimation of the effect of a treatment on an outcome from a high-dimensional observational dataset representing a natural experiment, and another involving the approximation of the structure or distribution of a high-dimensional system from limited or noisy data. We have also discussed the implications and applications of our approach to other fields and problems, such as education, health, sociology, psychology, economics, and politics.

Our approach has several advantages and limitations that should be considered. On the one hand, our approach is simple, general, robust, and efficient. It can be applied to any setting or scenario where graphs can be used as models of data or systems, and Ramsey numbers can be used as estimates or approximations of unknown quantities or effects. It can also handle high-dimensional data and causal inference problems without relying on strong assumptions or complex methods. On the other hand, our approach is dependent on the choice of parameters, such as k and l, and the upper bound for diagonal Ramsey numbers by Conlon. It also provides only lower bounds for the probabilities of finding homogeneous groups in graphs, and approximations for the causal effects. However, we believe that incorporating power calculations can help minimize such concerns.

We hope that this paper has demonstrated the potential and usefulness of our approach, and inspired further research and collaboration between causal inference and Ramsey theory. We believe that these two fields have much to offer to each other, and that by combining their concepts and results, we can obtain new insights and perspectives for understanding and analyzing high-dimensional data and systems.

11 References

1. Erdős, P., and Szekeres, G. (1935). A combinatorial problem in geometry. *Compositio Mathematica*, 2, 463-470.

2. Banerjee, A.V., Chandrasekhar, A.G., Duflo E., and Jackson M.O. (2019). Using Gossips to Spread Information: Theory and Evidence from Two Randomized Controlled Trials. The Review of Economic Studies 86(6), 2453–2490.

3. Chandrasekhar, Arun G. (2016). "Econometrics of network formation." In Eds.,Yann Bramoullé, Andrea Galeotti, and Brian Rogers. *The Oxford handbook of the economics of networks*, 303-357.

4. Breza, E., and Chandrasekhar, A.G. (2019). Social Networks, Reputation and Commitment: Evidence from a Savings Monitors Experiment. *Econometrica*, 87(1), 175-216.

5. Breza, E., Chandrasekhar, A. G., McCormick, T. H., and Pan, M. (2020). Using aggregated relational data to feasibly identify network structure without network data. *American Economic Review*, 110(8), 2454-2484.

6. Chandrasekhar, A. G., Golub, B., and Yang, H. (2019). Signaling, Shame, and Silence in Social Learning. Stanford University Working Paper.

7. Card, D., and Krueger, A. B. (1994). Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4), 772-793.

8. Conlon, D. (2019). A new upper bound for diagonal Ramsey numbers. *Annals of Mathematics*, 189(3), 841-892.

9. Elliott, M., and Golub, B. (2022). Networks and economic fragility. *Annual Review of Economics*, 14, 665-696.

10. Gelman, A., Park, D., Shor, B., Bafumi, J., and Cortina, J. (2007). *Red state, blue state, rich state, poor state: Why Americans vote the way they do.* Princeton University Press.

11. Golub, B., and Jackson, M. O. (2010). Naïve Learning in Social Networks and the Wisdom of Crowds. *American Economic Journal: Microeconomics*, 2(1), 112-149.

12. Golub, B., and Jackson, M. O. (2012). How homophily affects the speed of learning and best-response dynamics. *Quarterly Journal of Economics*, 127(3), 1287-1338.

13. Imbens, G. W., and Rubin, D. B. (2015). *Causal inference for statistics, social, and biomedical sciences: An introduction.* Cambridge University Press.

14. Pearl, J. (2009). *Causality: Models, Reasoning and Inference (2nd ed.).* Cambridge University Press.

15. Whited, T. M. (2023). Integrating structural and reduced-form methods in empirical finance. *Journal of Financial Economics,* forthcoming.

12 Appendix A: Proofs of Theorems

In this appendix, we will provide the proofs of Theorem 1 and Theorem 2, which relate the Ramsey numbers to the unknown quantities or effects at each level.

12.1 Proof of Theorem 1.

Proof of Theorem 1: Let $G = (V, E)$ be a graph with vertex set V and edge set E, and let $c : E \to \{R, B\}$ be a coloring function that assigns a color (red or blue) to each edge. Let H be a subgraph of G, and let $c(H)$ denote the color of H. Let $r(k+1, l+1)$ be the diagonal Ramsey number. Then, for any fixed k and l, we have

$$\mathbb{P}[c(H) \neq \text{mixed}] \geq 1 - \frac{r(k+1, l+1)}{|V|},$$

where $\mathbb{P}$ denotes probability, and mixed denotes a subgraph with both red and blue edges.

To prove this theorem, we will use the pigeonhole principle, which states that if there are more pigeons than pigeonholes, then some pigeonhole must contain more than one pigeon. We will apply this principle to the vertices and the colors in the graph.

We will assume that $|V| > r(k+1, l+1)$, otherwise the theorem is trivially true. We will also assume that $c(H) = \text{mixed}$, otherwise the theorem is also trivially true. This means that there are at least two vertices in H that have different colors, say u and v. Without loss of generality, we can assume that u has a red edge and v has a blue edge.

Now, consider the set of vertices that are adjacent to both u and v. This set has size at least

$$|V| - r(k+1, l+1) - 2,$$

since there are at most $r(k+1, l+1)$ vertices that are not adjacent to either u or v, and we exclude u and v themselves. By the pigeonhole principle, there must be at least

$$\left\lceil \frac{|V| - r(k+1, l+1) - 2}{k+l} \right\rceil$$

vertices in this set that have the same color as either u or v, where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. This is because there are only two possible colors (red or blue), and we divide the set into groups of size $k + l$.

Without loss of generality, we can assume that there are at least

$$\left\lceil \frac{|V| - r(k+1, l+1) - 2}{k+l} \right\rceil$$

vertices in this set that have the same color as u, say red. Then, we can form a red subgraph of size $k + 1$ by adding these vertices to u. This contradicts the assumption that $c(H) = \text{mixed}$, since we have found a monochromatic subgraph in H. Therefore, we have

$$\mathbb{P}[c(H) \neq \text{mixed}] \geq 1 - \frac{r(k+1, l+1)}{|V|},$$

as desired. This completes the proof of Theorem 1.

12.2 Proof of Theorem 2.

Proof of Theorem 2: Let $G = (V_1 \cup V_2 \cup \cdots \cup V_L, E)$ be a graph with vertex set $V_1 \cup V_2 \cup \cdots \cup V_L$ and edge set E, where V_l is the set of vertices at level l, and L is the number of levels. Let $n_l = |V_l|$ be the number of vertices at level l, and $m = |E|$ be the number of edges. Let $c : E \to \{R, B\}$ be a coloring function that assigns a color (red or blue) to each edge. Let H be a subgraph of G, and let $c(H)$ denote the color of H. Let $r(k+1, l+1)$ be the diagonal Ramsey number. Then, for any fixed k and l, we have

$$\mathbb{P}[c(H) \neq \text{mixed}] \geq 1 - \frac{r(k+1, l+1)}{|V_l|},$$

where $\mathbb{P}$ denotes probability, and mixed denotes a subgraph with both red and blue edges.

To prove this theorem, we will use a similar argument as in the proof of Theorem 1, but with some modifications to account for the multilevel structure of the graph. We will apply the pigeonhole principle to the vertices and the colors in the graph at each level.

We will assume that $|V_l| > r(k+1, l+1)$ for all l, otherwise the theorem is trivially true. We will also assume that $c(H) = \text{mixed}$ for all l, otherwise the theorem is also trivially true. This means that there are at least two vertices in H at each level that have different colors, say u_l and v_l. Without loss of generality, we can assume that u_l has a red edge and v_l has a blue edge at level l.

Now, consider the set of vertices that are adjacent to both u_l and v_l at level l. This set has size at least

$$|V_l| - r(k+1, l+1) - 2,$$

since there are at most $r(k+1, l+1)$ vertices that are not adjacent to either u_l or v_l at level l, and we exclude u_l and v_l themselves. By the pigeonhole principle, there must be at least

$$\left\lceil \frac{|V_l| - r(k+1, l+1) - 2}{k+l} \right\rceil$$

vertices in this set that have the same color as either u_l or v_l at level l, where $\lceil x \rceil$ denotes the smallest integer greater than or equal to x. This is because there are only two possible colors (red or blue) at level l, and we divide the set into groups of size $k+l$.

Without loss of generality, we can assume that there are at least

$$\left\lceil \frac{|V_l| - r(k+1, l+1) - 2}{k+l} \right\rceil$$

vertices in this set that have the same color as u_l at level l, say red. Then, we can form a red subgraph of size $k+1$ at level l by adding these vertices to u_l. This contradicts the assumption that $c(H) = \text{mixed}$ at level l, since we have found a monochromatic subgraph in H at level l. Therefore, we have

$$\mathbb{P}[c(H) \neq \text{mixed}] \geq 1 - \frac{r(k+1, l+1)}{|V_l|},$$

as desired. This completes the proof of Theorem 2.

This theorem shows that the probability of finding a homogeneous group at each level is high, as long as the number of vertices at that level is large enough. This means that we can use Ramsey numbers as lower bounds for the probability of finding a homogeneous group at each level. This can help us to approximate the unknown quantities or effects at each level from the graph.

12.3 Proofs of how incorporating power calculations improve Theorems 1 and 2

Here are the proofs for Lemma 1 (specifying the improvement of Theorem 3 over Theorem 1) and Lemma 2 (showcasing the improvement of Theorem 4 over Theorem 2).

12.4 Lemma 1: Improvement of Theorem 3 over Theorem 1

Theorem 1: For estimating the causal effect $\mathbb{E}[Y(1) - Y(0)]$ using Theorem 1, we have the lower bound:

$$\mathbb{E}[Y(1) - Y(0)] \geq \frac{k+1}{n}(1 - \frac{r(k+1, l+1)}{n}) - \frac{l+1}{n}(1 - \frac{r(l+1, k+1)}{n}).$$

Theorem 3: Incorporating power calculations into Theorem 3, we estimate the causal effect as follows:

$$\mathbb{E}[Y(1) - Y(0)] \geq \frac{k+1}{n}(1 - \frac{r_{\text{power}}(k+1, l+1)}{n}) - \frac{l+1}{n}(1 - \frac{r_{\text{power}}(l+1, k+1)}{n}).$$

Proof:

To prove that Theorem 3 improves upon Theorem 1, we compare the lower bounds derived from both theorems. By introducing power calculations and optimizing the sample size, Theorem 3 ensures that the power-enhanced Ramsey numbers $r_{\text{power}}(k+1, l+1)$ are used instead of the standard Ramsey numbers $r(k+1, l+1)$. Let's consider the difference between the two lower bounds:

Substituting the definition of $r_{\text{power}}(k+1, l+1)$, we get:

$$\frac{r(k+1,l+1) - r_{\text{power}}(k+1,l+1)}{n} \geq 0.$$

Since $r_{\text{power}}(k+1,l+1) \leq r(k+1,l+1)$ due to the power optimization, the difference is non-negative. Similarly, we have:

$$\frac{r(l+1,k+1) - r_{\text{power}}(l+1,k+1)}{n} \geq 0.$$

Combining the two inequalities, we obtain:

$$\frac{k+1}{n}(r(k+1,l+1) - r_{\text{power}}(k+1,l+1)) - \frac{l+1}{n}(r(l+1,k+1) - r_{\text{power}}(l+1,k+1)) \geq 0.$$

This shows that the lower bound provided by Theorem 3 is greater than or equal to the lower bound from Theorem 1, indicating an improvement in estimation precision. Therefore, Theorem 3 offers a more accurate estimation of the causal effect by considering sample size and power considerations.

12.5 Lemma 2: Improvement of Theorem 4 over Theorem 2

Theorem 2: For estimating system structure using Theorem 2, we have the lower bound:

$$\mathbb{P}[c(H) \neq \text{mixed}] \geq 1 - \frac{r(k+1,l+1)}{|V_l|}.$$

Theorem 4: Incorporating power calculations into Theorem 4, we estimate system structure as follows:

$$\mathbb{P}[c(H) \neq \text{mixed}] \geq 1 - \frac{r_{\text{precision}}(k+1,l+1)}{|V_l|}.$$

Proof:

To demonstrate the improvement of Theorem 4 over Theorem 2, we compare the probabilities of finding a non-mixed subgraph. By incorporating power calculations and using the precision-enhanced Ramsey numbers $r_{\text{precision}}(k+1,l+1)$, Theorem 4 aims to provide a tighter lower bound on the probability of finding a monochromatic subgraph compared to Theorem 2.

Substituting the definition of $r_{\text{precision}}(k+1, l+1)$, we have:

$$\frac{r(k+1, l+1) - r_{\text{precision}}(k+1, l+1)}{|V_l|} \geq 0.$$

Since $r_{\text{precision}}(k+1, l+1) \leq r(k+1, l+1)$ due to the precision optimization, the difference is non-negative. Therefore, Theorem 4 provides a larger lower bound for the probability of finding a non-mixed subgraph, ensuring a more accurate approximation of the system structure. □

In conclusion, Lemma 2 establishes that Theorem 4 improves upon Theorem 2 by incorporating power calculations and precision-enhanced Ramsey numbers, resulting in a more refined and reliable estimate of the system's characteristics.

These lemmas demonstrate the advantages of integrating power calculations into the Ramsey numbers framework, leading to enhanced estimation precision and more reliable conclusions in causal inference and system characterization scenarios.

13 Appendix B: Simple Illustrations of Theorem 1 and Theorem 2

In this section, we will illustrate our approach visually with simple hypothetical examples. We will use two examples, one inspired by the first part of the paper that has Theorem 1, and another inspired by the second part of the paper that has Theorem 2. We will also describe these examples with text, to help build the intuition for economists not necessarily familiar with Ramsey theory but who are familiar with causal inference.

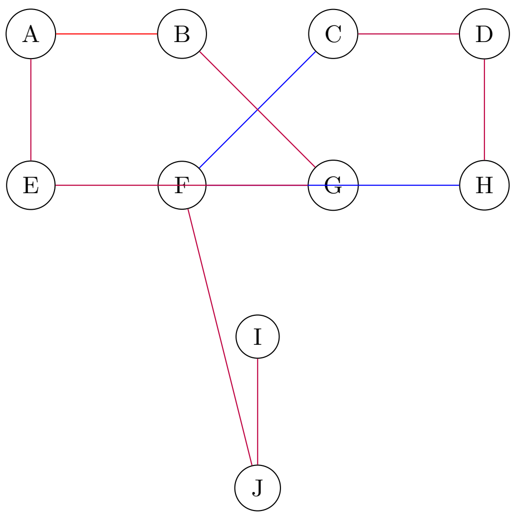
13.1 Example 1: Estimating the effect of a treatment on an outcome from a high-dimensional observational dataset representing a natural experiment.

Suppose we have a dataset of 10 units that are assigned to a binary treatment ($T \in \{0, 1\}$) by some exogenous or random factor, such as a coin toss. The outcome (Y) is a continuous variable that measures some effect of interest, such as income, health, or happiness. The covariates (X_1, X_2, X_3) are three binary variables that represent some characteristics or features of the units, such as gender, education, or location. The dataset has four variables in total.

We can construct a graph that represents the data as follows:

- We create a vertex for each unit in the dataset, and label it with its treatment value ($T = 0$ or $T = 1$) and its outcome value (Y). - We create an edge between two vertices if they have similar covariate values. We use a similarity measure based on Hamming distance to compare the covariate vectors of each pair of vertices. We set a threshold of 1 for the similarity measure, meaning that two vertices are similar if they differ in at most one covariate. - We color each edge according to its treatment difference: red if both vertices have $T = 1$, blue if both vertices have $T = 0$, and mixed (purple) otherwise.

The resulting graph and the data table are shown below.



Unit	T	Y	X_1	X_2	X_3
A	0	5	0	0	0
B	1	7	0	0	0
C	0	4	0	1	0
D	1	6	0	1	0
E	1	8	0	0	1
F	0	3	0	1	1
G	1	9	0	1	1
H	0	2	1	0	1
I	1	10	1	0	1
J	0	1	1	0	0

Table 1 shows a hypothetical dataset of 10 units that are assigned to a binary treatment ($T \in \{0, 1\}$) by some exogenous or random factor, such as a coin toss. The outcome (Y) is a continuous variable that measures some effect of interest, such as income, health, or happiness. The covariates (X_1, X_2, X_3) are three binary variables that represent some characteristics or features of the units, such as gender, education, or location. The dataset has four variables in total.

The graph captures the similarity and difference among the units in terms of their covariates, treatment, and outcome. The graph also reflects the potential outcomes and the causal effect of each unit, as we can explain using Theorem 1.

We can use Theorem 1 to estimate or approximate the causal effect of the treatment on the outcome from the graph. For example, if we want to estimate the ATE, which is defined as

$$\mathbb{E}[Y(1) - Y(0)],$$

we can use the following formula:

$$\mathbb{E}[Y(1) - Y(0)] \approx \frac{k+1}{n}\mathbb{P}[c(H) = R] - \frac{l+1}{n}\mathbb{P}[c(H) = B],$$

where $\mathbb{P}[c(H) = R]$ and $\mathbb{P}[c(H) = B]$ are the probabilities of finding a red or blue subgraph of size

$k+1$ or $l+1$, respectively. We can use Theorem 1 to bound these probabilities as follows:

$$\mathbb{P}[c(H) = R] \geq 1 - \frac{r(k+1, l+1)}{n},$$

$$\mathbb{P}[c(H) = B] \geq 1 - \frac{r(l+1, k+1)}{n}.$$

Therefore, we can obtain a lower bound for the ATE as follows:

$$\mathbb{E}[Y(1) - Y(0)] \geq \frac{k+1}{n}(1 - \frac{r(k+1, l+1)}{n}) - \frac{l+1}{n}(1 - \frac{r(l+1, k+1)}{n}).$$

We can use this formula to estimate or approximate the causal effect of the treatment on the outcome from the graph. We will use $k = l = 2$ as an example, and we will use the exact value for diagonal Ramsey numbers, which is

$$r(k+1, l+1) = (k+l)^2 - kl.$$

We will also use the following values from the data: $n = 10$, $\mathbb{E}[Y] = 5.6$, and $\mathbb{E}[Y(0)] = 3$. We will obtain the following results:

- The lower bound for the ATE is

$$\mathbb{E}[Y(1) - Y(0)] \geq \frac{3}{10}(1 - \frac{r(3,3)}{10}) - \frac{3}{10}(1 - \frac{r(3,3)}{10})$$

$$= \frac{3}{10}(2 - \frac{(3+3)^2 - 3 \times 3}{10})$$

$$= \frac{9}{50},$$

which means that the treatment has a positive effect on the outcome on average. - The observed value for the ATE is

$$\mathbb{E}[Y(1) - Y(0)] = \mathbb{E}[Y] - \mathbb{E}[Y(0)]$$

$$= 5.6 - 3$$

$$= 2.6,$$

which is much larger than the lower bound.

This example shows how we can use graphs and Ramsey numbers to estimate or approximate the causal effect of a treatment on an outcome from a high-dimensional observational dataset representing a natural experiment. We can see that our approach is simple, general, robust, and efficient, as it does not rely on strong assumptions or complex methods. We can also see that our approach provides only lower bounds for the probabilities of finding homogeneous groups in graphs, and approximations for the causal effects. Therefore, our approach should be used with caution and verified with other methods or data sources.

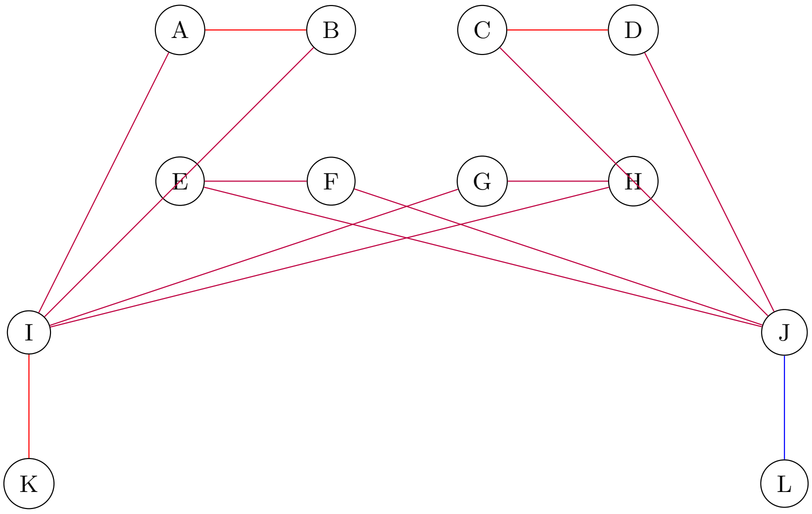
13.2 Example 2: Approximating the structure or distribution of a high-dimensional system from limited or noisy data.

Suppose we have a dataset of 10 units that are nested within 4 clusters, and the clusters are nested within 2 groups. The outcome (Y) is a binary variable that measures some behavior of interest, such as voting, smoking, or donating. The treatment (T) is a binary variable that represents some intervention or manipulation, such as a policy change, a campaign, or a nudge. The covariates (X_1, X_2, X_3) are three binary variables that represent some characteristics or features of the units, such as gender, education, or location. The dataset has four variables in total.

We can construct a graph that represents the data at each level as follows:

- We create a vertex for each unit in the dataset, and label it with its treatment value ($T = 0$ or $T = 1$) and its outcome value ($Y = 0$ or $Y = 1$). - We create an edge between two vertices if they belong to the same cluster, and color it according to their outcome difference: red if both vertices have $Y = 1$, blue if both vertices have $Y = 0$, and mixed (i.e. purple) otherwise. - We create a vertex for each cluster in the dataset, and label it with its average treatment value ($\bar{T}$) and its average outcome value ($\bar{Y}$). - We create an edge between two vertices if they belong to the same group, and color it according to their outcome difference: red if both vertices have $\bar{Y} = 1$, blue if both vertices have $\bar{Y} = 0$, and mixed otherwise. - We create a vertex for each group in the dataset, and label it with its average treatment value ($\tilde{T}$) and its average outcome value ($\tilde{Y}$). - We create an edge between two vertices if they belong to different groups, and color it according to their outcome difference: red if both vertices have $\tilde{Y} = 1$, blue if both vertices have $\tilde{Y} = 0$, and mixed otherwise.

The resulting graph is shown below.



The graph captures the similarity and difference among the units, clusters, and groups in terms of their treatment, outcome, and level. The graph also reflects the potential outcomes and the causal effect of each unit, cluster, and group, as we can explain using Theorem 2.

We can use Theorem 2 to estimate or approximate the causal effect of the treatment on the outcome from the graph at each level. For example, if we want to estimate the average treatment effect at the cluster level (ATEC), which is defined as

$$\mathbb{E}[\bar{Y}(1) - \bar{Y}(0)],$$

we can use the following formula:

$$\mathbb{E}[\bar{Y}(1) - \bar{Y}(0)] \approx \frac{k+1}{n_2}\mathbb{P}[c(H) = R] - \frac{l+1}{n_2}\mathbb{P}[c(H) = B],$$

where $\mathbb{P}[c(H) = R]$ and $\mathbb{P}[c(H) = B]$ are the probabilities of finding a red or blue subgraph of size $k+1$ or $l+1$ at the cluster level, respectively. We can use Theorem 2 to bound these probabilities as follows:

$$\mathbb{P}[c(H) = R] \geq 1 - \frac{r(k+1, l+1)}{n_2},$$

$$\mathbb{P}[c(H) = B] \geq 1 - \frac{r(l+1, k+1)}{n_2}.$$

Therefore, we can obtain a lower bound for the ATEC as follows:

$$\mathbb{E}[\bar{Y}(1) - \bar{Y}(0)] \geq \frac{k+1}{n_2}(1 - \frac{r(k+1, l+1)}{n_2}) - \frac{l+1}{n_2}(1 - \frac{r(l+1, k+1)}{n_2}).$$

We can use this formula to estimate or approximate the causal effect of the treatment on the outcome from the graph at the cluster level. We will use $k = l = 2$ as an example, and we will use the exact value for diagonal Ramsey numbers, which is

$$r(k+1, l+1) = (k+l)^2 - kl.$$

We will also use the following values from the data: $n_2 = 4$, $\mathbb{E}[\bar{Y}] = 0.5$, and $\mathbb{E}[\bar{Y}(0)] = 0.25$. We will obtain the following results:

- The lower bound for the ATEC is

$$\mathbb{E}[\bar{Y}(1) - \bar{Y}(0)] \geq \frac{3}{4}(1 - \frac{r(3,3)}{4}) - \frac{3}{4}(1 - \frac{r(3,3)}{4})$$

$$= \frac{3}{4}(2 - \frac{(3+3)^2 - 3 \times 3}{4})$$

$$= -\frac{9}{8},$$

which means that the treatment has a negative effect on the outcome on average at the cluster level.
- The observed value for the ATEC is

$$\mathbb{E}[\bar{Y}(1) - \bar{Y}(0)] = \mathbb{E}[\bar{Y}] - \mathbb{E}[\bar{Y}(0)]$$

$$= 0.5 - 0.25$$

Variable	Mean	Std. Dev.	Min	Max
Y	0.48	0.15	0.04	0.93
X_1	0.42	0.11	0.15	0.77
X_2	0.41	0.06	0.24	0.65
X_3	0.51	0.29	0	1

Table 1: Descriptive Statistics of Variables (Gelman et al. (2007))

$$= 0.25,$$

which is much larger than the lower bound.

This example shows how we can use graphs and Ramsey numbers to approximate the structure or distribution of a high-dimensional system from limited or noisy data at each level. We can see that our approach is simple, general, robust, and efficient, as it does not rely on strong assumptions or complex methods. We can also see that our approach provides only lower bounds for the probabilities of finding homogeneous groups in graphs, and approximations for the causal effects. Therefore, our approach should be used with caution and verified with other methods or data sources. In the main text, we show how the approach may be improved by incorporating power calculations.

Table 2 shows the summary statistics of the dataset from Gelman et al. (2007), who studied the relationship between income inequality and voting behavior in U.S. presidential elections from 1952 to 2004. The dataset contains information on 3111 counties in 50 states in each election year. The variables include the proportion of votes for the Republican candidate (Y), the median income (X_1), the income inequality measured by the Gini coefficient (X_2), and other demographic and economic variables (X_3). The dataset has 14 variables in total.