NTNU

DEPARTMENT OF COMPUTER SCIENCE

TDT4900 — MASTER'S THESIS

Exploring Matroids in Fair Allocation: Building the Matroids.jl Library

Author:
Andreas Aaberge Eide

Supervisor:
Magnus Lie Hetland

June 1, 2023

In matroid theory, we explore
A structure with properties galore
From bases to circuits
It never discourages
Mathematicians always wanting more

The axioms it holds are so grand
And its applications vast and unplanned
From optimization to graphs
It can solve so many tasks
Matroid theory, truly a wonderland

<div align="right">ChatGPT</div>

Acknowledgements

I would like to extend a heartfelt thanks to my friends and foes at Radio Revolt, without whom I would have completed this degree years earlier, with far better grades and in significantly better health.

Contents

Introduction

Imagine you are a dean at a large university, responsible for allocating seats in courses to students for the coming semester. Each student has applied for some subset of the courses available, and each course has a limit on the number of students it can accept. In addition, there are scheduling conflicts between courses, and the students have hard limits on the number of courses they can take in a semester[1]. As a good dean, you wish to allocate fairly, ensuring that the students are happy with their courses and do not feel disfavored compared to the other students. At the same time, you strive for efficiency, so that every student gets enough courses to make the expected progress on their grade. How would you go about solving this problem? Over lunch, you discuss your quandary with a colleague from the computer science department who immediately assuages your fears by informing you that your situation is in fact an instance of indivisible fair allocation with matroid-rank valuations, a well-studied problem for which several algorithms exist! Eager to learn more, you ask your colleague to explain her cryptic remark.

What is fair allocation?

Fair allocation is the problem of fairly partitioning a set of resources (in this case, the courses) among agents (the students) with different preferences or valuations

---

[1] This example is due to Benabbou et al [1].

over these resources. This has been a hot topic of interest since antiquity (a 2000-year old allocation strategy can be found in the Talmud [2]), and remains so today. The mathematical study of fair allocation started with a seminal work by Steinhaus in 1948 [3], and for decades the focus was largely on the *divisible* case, in which the resources can be divided into arbitrary small pieces. In the divisible case, fair allocations always exist, and they can be computed efficiently [4]. In the dean's scenario, however, the course seats are *indivisible goods*. A fair allocation of indivisible goods is, depending on the measure of fairness, not always achievable; consider for example allocating a course with one seat between two students who both applied for it – there is no way of allocating the seat without one agent being unhappy.

Generally speaking, an allocation is measured against two justice criteria: *fairness* and *efficiency*. Fairness relates to the degree to which agents preceive the allocation as favoring other agents over themselves. One common way to describe the *fairness* of an allocation is with the concept of *envy-freeness*. Envy is defined as the degree to which an agent values another agent's received bundle of resources higher than their own. An allocation is envy-free if no agent envies another agent. In the trivial example above, the only envy-free allocation is the one in which no student receives the seat; while this is technically speaking fair, it is highly inefficient. Efficiency deals with maximizing some notion of resource utilization, or, equivalently, reducing waste. The perfectly fair allocation in which no one receives anything is rarely desirable for reasons of efficiency. Conversely, while the allocation in which one agent receives everything might be highly efficient in terms of the total sum of bundle values, it is obviously unfair. The task of the fair allocation algorithm, then, is to find some balance between these criteria.

How do matroids enter into this?

What the colleague from the computer science department noticed about the dean's problem, was that it was well-structured, in fact it is a textbook example of *matroid-rank valuations* in practice. Matroid rank functions (MRFs) are are a class of functions with properties that make them both easy to reason about and practically applicable in a setting such as fair allocation, and can equivalently be referred to as *binary submodular functions*. A submodular function is a set function that obeys the law of *diminishing returns* – as the size of the input set increases, the marginal value of a single additional good decreases. MRFs are the class of submodular functions with *binary marginals*, meaning that the

value of any single good is either 0 or 1.

In practice, these properties make MRFs a compelling framework for modelling user preferences in a setting such as the dean's allocation scenario. The binary marginals reflect a student's willingness (value of 1) or unwillingness (value of 0) to enroll in a course. The diminishing returns property allow us to implement what are known in economics terms as *supplementary goods* and *fixed demand*. A student might be interested in two similar courses, but not wish to enroll in both, so given one, the marginal value of the other drops to 0 (the courses are supplementary goods). In addition, a student needs only at most one seat per course, so the many available seats for the same course are also supplementary goods. A student has limited time and energy, and so for each course seat received, the marginal value of the other courses can only decrease – after some threshold is reached in the number of enrolled courses, all remaining courses have value 0 (there is a fixed demand for courses).

Matroids are extensively studied mathematical structures that generalize concepts from a variety of different fields. A number of interesting algorithms have been developed for fair allocation with matroid-rank valuations [1, 5–8] that make use of deep results from matroid theory in their analysis, and deliver well on a range of justice criteria which might be computationally intractable to achieve under general valuations.

What does this thesis contribute?

Perhaps because matroids are so well-understood and pleasant to work with theoretically, there is a dearth of tooling available for generating and working with them programmatically. In an effort to fill this gap, this thesis proposes Matroids.jl, a library for the Julia programming language [9], which extends the existing Allocations.jl library [10] with the functionality required to enable the empirical study of matroidal fair allocation algorithms.

This thesis describes the work that has been done to design and build a working, proof-of-concept version of Matroids.jl. It is structured as follows. In the next chapter, I establish the concepts from matroid theory and fair allocation necessary to understand the rest of the thesis. In Chapter 3, I describe the design and implementation of the Matroids.jl API, which includes various classic matroid algorithms that have found use in fair allocation algorithms. In Chapter 4, I show how this API can be used to implement Viswanathan and Zick's Yankee Swap algorithm [8] and some other algorithms for matroid-rank-valued fair allocation. In Chapter 5, I show how Matroids.jl implements the

random generation of a range of matroid types. Of particular interest here is Knuth's classic method for generating arbitrary matroids [11], the successful implementation of which was a significant sub-goal of the project. In Chapter 6, I provide some experimental results for the algorithms over different matroid types. Finally, in Chapter 7, I give a summary discussion on the limitations of Matroids.jl and suggests a few possible avenues of future work.

Chapter 2

Preliminaries

> For simplicity, we also assume that every point in a geometry is a closed set. Without this additional assumption, the resulting structure is often described by the ineffably cacaphonic term "matroid", which we prefer to avoid in favor of the term "pregeometry".
>
> Gian-Carlo Rota [12]

Matroids were first introduced by Hassler Whitney in 1935 [13], in a seminal paper where he described two axioms for independence in the columns of a matrix, and defined any system obeying these axioms to be a "matroid" (which unfortunately for Rota is the term that has stuck). Whitney's key insight was that this abstraction of "independence" is applicable to both matrices and graphs. Matroids have also received attention from researchers in fair allocation, as their properties make them useful for modeling user preferences. We have already seen this with the course allocation problem described in the previous chapter; other use cases include the assignment of kindergarten slots or public housing estates among people of different ethnicities [7].

2.1 Fair allocation

To ease readability, I abuse notation a bit and use $S + e$ and $S - e$ to refer to $S \cup \{e\}$ and $S \setminus \{e\}$, respectively.

An instance of a fair allocation problem consists of a set of agents $N = \{1, 2, \ldots, n\}$ and a set of m goods $E = \{e_1, e_2, \ldots, e_m\}$. Each agent has a valuation function $v_i : 2^E \to \mathbb{R}^+$; $v_i(S)$ is the value agent i ascribes to the bundle of goods S. The marginal value of agent i for the good g, given that she already owns the bundle S, is given by $\Delta_i(S, g) := v_i(S+g) - v_i(S)$. Throughout most of this thesis, we assume that v_i is a matroid rank function, or, equivalently, a binary submodular function. To formalize the description given in Chapter 1, this means that

  (a) $v_i(\emptyset) = 0$,

  (b) v_i has binary marginals: $\Delta_i(A, g) \in \{0, 1\}$ for every $A \subset E$ and $g \in E$,

  (c) v_i is submodular: for every $A \subseteq B \subseteq E$ and $g \in E \setminus B$, we have that $\Delta_i(A, g) \geq \Delta_i(B, g)$.

Any function v_i adhering to these properties is a valid characterization of exactly one matroid in terms of its rank function [14]. There are many other ways to characterize matroids, some of which are given in Section 2.2.

The output of an algorithm for fair allocation is an allocation of the goods to the agents. An allocation A is an n-partition of E, $A = (A_1, A_2, \ldots, A_n)$, where each A_i is the bundle of goods allocated to agent i. Sharing is not allowed, so we require that $A_i \cap A_j = \emptyset$ for all $i \neq j$. We say that an allocation is *clean* (also known as *non-redundant* in the literature) if no agent has received any good they value at 0. An allocation is *complete* if all goods are allocated. It might not be possible to guarantee both cleanness and completeness; for instance in a case where a good is 0-valued by all agents.

2.1.1 Envy-freeness

We are interested in producing *fair* allocations. One of the most popular notions of fairness in the literature is envy-freeness (EF), which states that no agent should prefer another agent's bundle over her own (in fair allocation, an agent is *envious* of another agent if she prefers that agent's bundle). An allocation A is EF iff, for all agents $i, j \in N$,

$$v_i(A_i) \geq v_i(A_j). \tag{EF}$$

Because, as mentioned in the introduction, EF is not always achievable when the goods are indivisible, the literature has focused on relaxations thereof. The most prominent such relaxation, which can be guaranteed, is *envy-freeness up to one good* (EF1) [15], which allows for the envy of up to the value of one (highest-valued) good. This is equivalent to saying that any envy can be eliminated by dropping one good from the envied bundle. A is an EF1 allocation iff, for all agents $i, j \in N$ where $|A_j| > 0$, there exists a $e \in A_j$ such that

$$v_i(A_i) \geq v_i(A_j - e). \tag{EF1}$$

Envy-freeness up to any good (EFX) is an even stronger version of EF. While EF1 allows that agent i envies agent j up to their highest valued good, EFX requires that the envy can be removed by dropping agent j's least valued good. There are two slightly different definitions of EFX in use in the literature. I follow the naming scheme used by Benabbou et al. [7] and refer to these as EFX_+ and EFX_0. Caragiannis et al. [16] requires that this least valued good be positively valued. We call this fairness objective EFX_+. A is an EFX_+ allocation iff, for all agents $i, j \in N$,

$$v_i(A_i) \geq v_i(A_j - e), \ \forall e \in A_j \text{ st. } v_i(A_j - e) < v_i(A_j). \tag{EFX$_+$}$$

Plaut and Roughgarden [17], on the other hand, allow for 0-valued goods in the envy check – we call this version EFX_0. It is stronger requirement than EFX_+. A is an EFX_0 allocation iff, for all agents $i, j \in N$,

$$v_i(A_i) \geq v_i(A_j - e), \ \forall e \in A_j. \tag{EFX$_0$}$$

In the general, additive case, the existence of EFX_0 allocations is an open question for instances with $n \geq 4$ [4].

2.1.2 Proportionality

Proportionality is a fairness objective that is fundamentally different from envy-freeness, in that it checks each bundle value against some threshold, instead of comparing bundle values against each other. An allocation A is proportional (PROP) iff each agent $i \in N$ receives at least her proportional share PROP_i, which is the $\frac{1}{n}$ fraction of the value she puts on the whole set of goods, ie.

$$v_i(A_i) \geq \text{PROP}_i := \frac{v_i(E)}{n}. \tag{PROP}$$

Proportionality might not be achievable in the indivisible case (again, consider two agents and one positively valued good), and so relaxations in the same vein as EF1 and EFX have been introduced – these are called PROP1 and PROPX [4]. An allocation is PROP1 iff there for each agent $i \in N$ exists some good $g \in E \setminus A_i$ that, if given to i, would ensure that agent i received her proportional share; that is,

$$\exists e \in E \setminus A_i \text{ st. } v_i(A_i + e) \geq \text{PROP}_i \qquad \text{(PROP1)}$$

PROPX is a stronger fairness objective than PROP1, and has, as in the case of EFX, two slightly different definitions in the literature. I follow the naming scheme established for EFX above, and refer to these as PROPX_+ and PROPX_0. The logic is similar to that of EFX. An allocation is PROPX_0 iff each agent can achieve her proportional share by receiving one additional, least-valued good from the goods not allocated to her. This good might be zero-valued.

$$\min_{e \in E \setminus A_i} v_i(A_i + e) \geq \text{PROP}_i \qquad \text{(PROPX}_0)$$

If we disallow zero-valued items, we arrive at the slightly weaker criteria PROPX_+, given by:

$$\min_{e \in E \setminus A_i, \Delta_i(A_i, e) > 0} v_i(A_i + e) \geq \text{PROP}_i \qquad \text{(PROPX}_+)$$

Maximin share fairness

Budish introduces a relaxation of proportionality known as *maximin share fairness* [18], in which the threshold for each agent i is her *maximin share* (MMS). The MMS of agent i, denoted by μ_i, is defined as the maximum value she could receive if she partitioned E among all agents and then picked the worst bundle. Let $\Pi_n(E)$ be the family of all possible allocations of the goods in E to the agents in N. Then,

$$\mu_i := \max_{A \in \Pi_n(E)} \min_{A_j \in A} v_i(A_j).$$

An allocation is MMS-fair if all agents receive at least as much as their maximin share:

$$v_i(A_i) \geq \mu_i, \ \forall i \in N. \qquad \text{(MMS)}$$

In the general, additive case, MMS-fair allocations do not always exist, and even computing the MMS of an agent is an NP-hard problem [4]. In a setting with matroid-rank valuations, however, Barman and Verma showed that MMS-fair allocations always exist, and can be computed in polynomial time [6].

2.1.3 Efficiency

Fairness is usually coupled with some efficiency criterion, to prevent the perfectly fair solution in which the whole set of goods is thrown away. The efficiency of an allocation can be measured with some *welfare function* on the values of the agents. There are three welfare functions commonly used in the literature:

1. **Egalitarian social welfare (ESW):** The ESW of an allocation A is given by the minimum value of an agent. $\text{ESW}(A) = \min_{i \in N} v_i(A_i)$.

2. **Utilitarian social welfare (USW):** The USW of an allocation is the total value received by all agents. $\text{USW}(A) = \sum_{i \in N} v_i(A_i)$.

3. **Nash welfare (NW):** The Nash welfare of an allocation is a compromise between the utilitarian and egalitarian approaches, given by the product of agent utilities. $\text{NW}(A) = \prod_{i \in N} v_i(A_i)$.

Allocations that maximize one of these welfare functions are referred to as MAX-ESW, MAX-USW and MNW, respectively.

Leximin

An obvious drawback of the egalitarian approach is that, in the situation where there are multiple possible allocations that maximize the minimum bundle value, it is indifferent to which of these bundles it prefers. Consider for instance two possible allocations of goods to three agents, where their bundle values are given as (A,B,C):

- Allocation 1: $(5, 7, 10)$

- Allocation 2: $(5, 8, 9)$

Both of these allocations are MAX-ESW. A stricter version of the egalitarian rule is *leximin*: an allocation is leximin if it maximizes the smallest value; subject to that, it maximizes the second-smallest value; subject to that, it maximizes the next-smallest value, and so on. The leximin rule prefers Allocation 2 over Allocation 1.

Pareto Optimality

2.2 Matroid theory

If a mathematical structure can be defined or axiomatized in multiple different, but not obviously equivalent, ways, the different definitions or axiomatizations of that structure make up a cryptomorphism. The many obtusely equivalent definitions of a matroid are a classic example of cryptomorphism, and belie the fact that the matroid is a generalization of concepts in many, seemingly disparate areas of mathematics. As a result, the terms used in matroid theory are borrowed from analogous concepts in both graph theory and linear algebra.

The most common way to characterize a matroid is as an *independence system*. An independence system is a pair $(E, \mathcal{I})$, where E is the ground set of elements, $E \neq \emptyset$, and $\mathcal{I}$ is the set of independent sets, $\mathcal{I} \subseteq 2^E$. The *dependent sets* of a matroid are $2^E \setminus \mathcal{I}$.

In practice, the ground set E represents the universe of elements in play, and the independent sets of typically represent the legal combinations of these items. In the context of fair allocation, the independent sets represent the legal (in the case of matroid constraints) or desired (in the case of matroid valuations) bundles of items.

A matroid is an independence system with the following properties [13]:

(1) If $S \subseteq T$ and $T \in \mathcal{I}$, then $S \in \mathcal{I}$.

(2) If $S, T \in \mathcal{I}$ and $|S| > |T|$, then there exists $e \in S \setminus T$ such that $S + e \in S$.

(2') If $S \subseteq E$, then the maximal independent subsets of S are equal in size.

Property (1) is called the *hereditary property* and (2) the *exchange property*. Properties (2) and (2') are equivalent. To see that (2) $\implies$ (2'), consider two maximal subsets of S. If they differ in size, (2) tells us that there are elements we can add from one to the other until they have equal cardinality. We get (2') $\implies$ (2) by considering $S = A \cup B$. Since $|A| > |B|$, they cannot both be

maximal, and some $e \in A \setminus B$ can be added to B to obtain another independent set.

When $S = E$, (2') gives us that the maximal independent sets of a matroid are all of the same size. A maximal independent subset of E is known as a *basis*. The size of the bases is the *rank* of the matroid as a whole. A matroid can be exactly determined by $\mathcal{B}$, its collection of bases, since a set is independent if and only if it is contained in a basis (this follows from (1)). A theorem by Whitney [13] gives the axiom system characterizing a collection of bases of a matroid:

1. No proper subset of a basis is a basis.

2. If $B, B' \in \mathcal{B}$ and $e \in B$, then for some $e' \in B'$, $B - e + e' \in \mathcal{B}$.

The rank function of a matroid is a function $v : 2^E \rightarrow \mathbb{Z}^+$ which, given a subset $B \subseteq E$, returns the size of the largest independent set contained in B. That is,

$$v(B) = \max_{A \subseteq B, A \in \mathcal{I}} |A|.$$

2.2.1 A graphic example

Different types of matroids exist, arising from different sources of "independence"; one well-known subclass of matroids, arising from notions of independence in graphs, is the class of *graphic matroids*.

A *tree* is a connected acyclic graph, and a *forest* is a disconnected graph consisting of some number of trees. A *spanning tree* of G is a subgraph with a unique simple path between all pairs of vertices of G. A *spanning forest* of G is a collection of spanning trees, one for each component. A graph will have some number of different spanning trees. Figure 2.1 shows two spanning trees of the same graphs (or alternatively, a spanning forest over one graph with two components).

Given a graph $G = (V, E)$, let $\mathcal{I} \subseteq 2^E$ be the family of subsets of the edges E such that, for each $I \in \mathcal{I}$, (V, I) is a forest. It is a classic result of matroid theory that $\mathfrak{M} = (E, \mathcal{I})$ (the ground set of the matroid being the edges of the graph) is a matroid [14, p. 657]. To understand how, we will show that it adhers to axioms (1) and (2'), as given in above. By investigating the highlighted spanning trees in Figure 2.1, it is easy to convince oneself that all subsets of a spanning tree are trees, since no subset of an acyclic set of edges will contain a cycle. Thus axiom (1) – the hereditary property – holds.
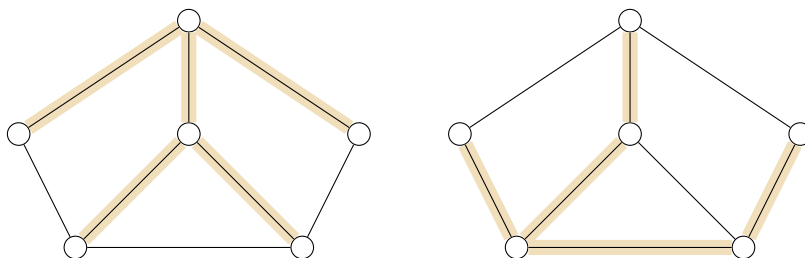
Figure 2.1: Two spanning trees of a graph with 8 edges.

To see that (2') holds, consider the set of bases of the matroid, $\mathcal{B} \subseteq \mathcal{I}$. Figure 2.1 shows two bases of the matroid described by the graph. By definition, each basis $B \in \mathcal{B}$ is a maximal forest over G. Since a spanning tree of a graph with n nodes must needs have $n - 1$ edges (I recommend drawing trees and counting their edges until one is convinced that this must be the case), we have $|B| = |V| - k$, where k is the number of components of G. This is the same for every $B \in \mathcal{B}$, which proves property (2'). Any matroid given by a graph G, denoted by $\mathfrak{M}(G)$, is called a graphic matroid.

2.2.2 Other characterizations of a matroid

We have already seen the independent sets characterization of a matroid, but the other properties of a matroid have their own axiom systems that can equivalently be used to characterize a matroid.

Characterization via rank function.

Figure 2.2: Two overlapping circuits on a graph.

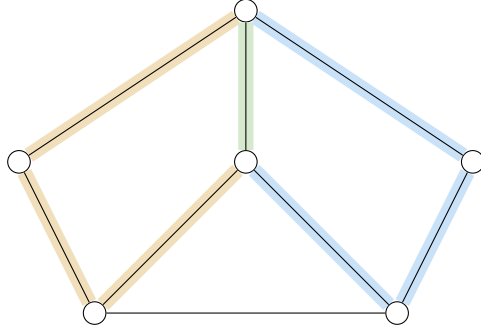**Characterization via circuits.** A *circuit* is a minimal dependent set of a matroid – it is an independent set plus one redundant element. Equivalently, the collection of circuits of a matroid is given by

$$\mathcal{C} = \big\{C : |C| = v(C) + 1, C \subseteq E\big\}.$$

A set is independent if and only if it contains no circuit [14], and so a matroid is uniquely determined by the collection of its circuits. The following conditions characterize $\mathcal{C}$ [13]:

(1) No proper subset of a circuit is a circuit.

(2) If $C, C' \in \mathcal{C}$, $x \in C \cap C'$ and $y \in C \setminus C'$, then $C \cup C'$ contains a circuit containing y but not x.

These properties are easily grokked by studying an example. In Figure 2.2, we see two circuits of a graph, highlighted in yellow and blue, overlapping at the edge highlighted in green. Obviously, no circuit in this figure contains a circuit. We can also verify that the union of the yellow edge set and the blue edge set, minus the green edge, is in fact a third, bigger circuit, as promised by property (2).

Characterization via closed sets. We also need to establish the concept of the *closed sets* (sometimes referred to as *flats* [14]) of a matroid. A closed set is a set whose cardinality is maximal for its rank. Equivalently to the definition given above, we can define a matroid as $\mathfrak{M} = (E, \mathcal{F})$, where $\mathcal{F}$ is the set of closed sets of $\mathfrak{M}$, satisfying the following properties [11]:
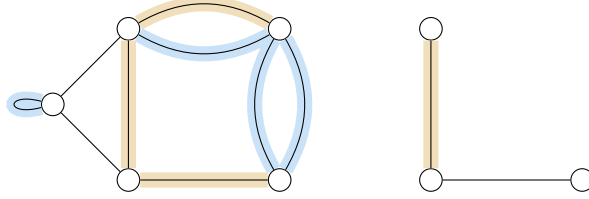
Figure 2.3: An independent subset of edges in yellow, with its closure in blue.

1. The set of all elements is closed: $E \in \mathcal{F}$

2. The intersection of two closed sets is a closed set: If $A, B \in \mathcal{F}$, then $A \cap B \in \mathcal{F}$

3. If $A \in \mathcal{F}$ and $a, b \in E \setminus A$, then b is a member of all sets in $\mathcal{F}$ containing $A \cup \{a\}$ if and only if a is a member of all sets in $\mathcal{F}$ containing $A \cup \{b\}$

The *closure function* is the function $cl : 2^E \to 2^E$, such that

$$cl(S) = \big\{ x \in E : v(S) = v(S \cup \{x\}) \big\}.$$

That is to say, the closure function, when given a set $S \subseteq E$, returns the set of elements in $x \in E$ such that x can be added to S with no increase in rank. It returns the closed set of the same rank as S, that contains S. The *nullity* of a subset S is the difference $|S| - v(S)$, ie. the number of elements that must be removed from S to obtain an independent set.

Figure 2.3 shows an independent (acyclic) set S of edges in highlighted yellow, along with its closure $cl(S)$ in blue. The closure is the set of edges e such that $v(S + e) = v(S)$, or, equivalently, such that the spanning tree of $S + e$ has the same size as that of S.

2.2.3 Matroid union

The matroid union is an operation that allows us to produce a new matroid by combining the independent sets of some number of existing matroids. Given n matroids $\mathfrak{M}_i = (E, \mathcal{I}_i)$, their union is given by

$$\widehat{\mathfrak{M}} = (E, \{I_1 \cup \ldots \cup I_n : I_i \in \mathcal{I}_i, \ \forall i \in [n]\}).$$

$\widehat{\mathfrak{M}}$ is in fact a matroid [14, Ch. 42], whose independent sets are the the subsets $S \subseteq E$ that allow an n-partition $S_1, \ldots, S_n$ such that $S_i \in \mathcal{I}_i$, for all $i \in [n]$. In

fair allocation jargon, if each $\mathfrak{M}_i$ is the matroid described by agent i's valuation function v_i, a set of goods S is independent in $\widehat{\mathfrak{M}}$ if and only if we can allocate it among the agents and produce utilitarian social welfare equal to $|S|$. Hence, each basis in $\widehat{\mathfrak{M}}$ corresponds to a clean (but not necessarily complete), MAX-USW allocation of the goods in E [6]. It is a classic result of Edmonds [19] that a basis in $\widehat{\mathfrak{M}}$ can be computed in polynomial time, using the *matroid union algorithm*. This is achieved by making use of two additional concepts, the *exchange graph* and *transfer paths*.

The exchange graph

Given n matroids $\mathfrak{M}_1 = (E, \mathcal{I}_1), \ldots, \mathfrak{M}_n = (E, \mathcal{I}_n)$, let A be a collection of n sets $A_1, \ldots, A_n$ such that (a) $A_i \cap A_j = \emptyset$ when $i \neq j$, and (b) for each i, $A_i \subseteq E$ and $A_i \in \mathcal{I}_i$. In the context of a matroid-rank-valued fair allocation problem, A is a clean allocation of the goods in E to the agents in N.

We follow the example of Schrijver [14] and define the exchange graph of A as the directed graph $D(A) = (E, x(A))$, where each node corresponds to a good in E, and the edges are given by

$$x(A) = \{(p, q) : p \in A_i, q \in E \setminus A_i, v_i(A_i - p + q) = v_i(A_i)\},$$

where v_i is the rank function of the matroid $\mathfrak{M}_i$. In other words, an edge exists between goods $p \in A_i$ and $q \in E \setminus A_i$ iff we can replace p with q for no decrease in the rank of the set containing p. Intuitively, we can understand D as representing for each good p, which other good q the current owner of p can replace p with and be just as happy. This intuitive explanation lets us begin to see why the matroid union algorithm and the concept of the exchange graph has found widespread use in fair allocation with matroid-rank valuations; they allow us to model equitable transfers of goods between agents.

Transfer paths and path augmentation

Let $P = (e_1, \ldots, e_t)$ be a path in the exchange graph $D(A)$. The *transfer* of goods along P is the operation in which e_t is given to the agent who owns e_{t-1}, e_{t-1} is given to the agent who owns e_{t-2}, and so on until e_1 is discarded. This transfer is called *path augmentation*; we use the notation established by Viswanathan and Zick [8] and denote the bundle A_i after augmentation with the path P by $A_i \Lambda P$.

For some $i \in N$, we define $F_i = \{e \in E \setminus A_i : A_i + e \in \mathcal{I}_i\}$ as the set of elements whose addition to A_i yields another, larger independent set (remember

that A is clean, so $A_i \in \mathcal{I}_i$ for each i). In a paper on the matroid union algorithm, Knuth [20] shows that, by augmenting along a shortest path $P = (e_1, \ldots, e_t)$ from F_i to A_j for some $j \in N - i$, we get

(a) $v_i(A_i \Lambda P + e_1) = v_i(A_i) + 1$,

(b) $v_k(A_k \Lambda P) = v_k(A_k), \ \forall k \in N - i - j$, and

(c) $v_j(A_j \Lambda P) = v_j(A_j) - 1$.

By greedily growing clean bundles in this manner, we can find a maximal independent set over $\widehat{\mathfrak{M}}$ in polynomial time [14]. Chapter 3 discusses Matroids.jl implementation of Knuth's matroid union algorithm. Viswanathan and Zick's Yankee Swap algorithm [8], discussed in Chapter 4, uses these concepts as well.

2.2.4   Matroid erection

Knuth's general matroid construction, the implementation of which is discussed in Section 5.4, requires us to establish a few more matroid concepts. The *rank-k truncation* of a rank-r matroid $\mathfrak{M} = (E, \mathcal{I})$, is the rank-$k$ matroid $\mathfrak{M}^{(k)} = (E, \mathcal{I}^{(k)})$, where

$$\mathcal{I}^{(k)} = \{I \in \mathcal{I} : |I| \leq k\}.$$

The *truncation* of $\mathfrak{M}$ is given as $T(\mathfrak{M}) = \mathfrak{M}^{(r-1)}$. As a simple example, we have that the uniform matroid $U_n^{n-1} = T(F_n)$, where F_n is the free matroid with n elements. The *erection* (I defer to Crapo for the choice of term [21]) of the matroid $\mathfrak{M}$ is the matroid $\mathfrak{N}$ of rank $\leq r + 1$ such that $\mathfrak{M} = T(\mathfrak{N})$. In other words, when taking the truncation of a matroid, we produce a new matroid by disregarding all independent sets of cardinality higher than some k. When taking the erection of a rank-k matroid, we produce a new matroid by declaring some number of subsets $S \subseteq E$ such that $|S| = r + 1$ to be independent, or equivalently, declaring some number of them to be closed sets of rank r. By declaring all of them to be dependent, we get the *trivial erection* $\mathfrak{N} = \mathfrak{M}$. By declaring none of them dependent (ie. they are all independent sets $r(S) = |S| = r + 1$), $\mathfrak{N}$ is the *free erection* of $\mathfrak{M}$ [22].

A matroid can have many erections. Figure 2.4 shows the Hasse diagram for the matroid of 4 elements, where every subset of two or fewer elements is independent (denoted in yellow). Each subset is encoded as a binary string, where a 1 designates membership in the set (a syntax we will become very familiar with in Chapter 5). This matroid is known as the uniform matroid U_4^2. The uniform matroid U_4^3 is the free erection of U_4^2, where we have simply
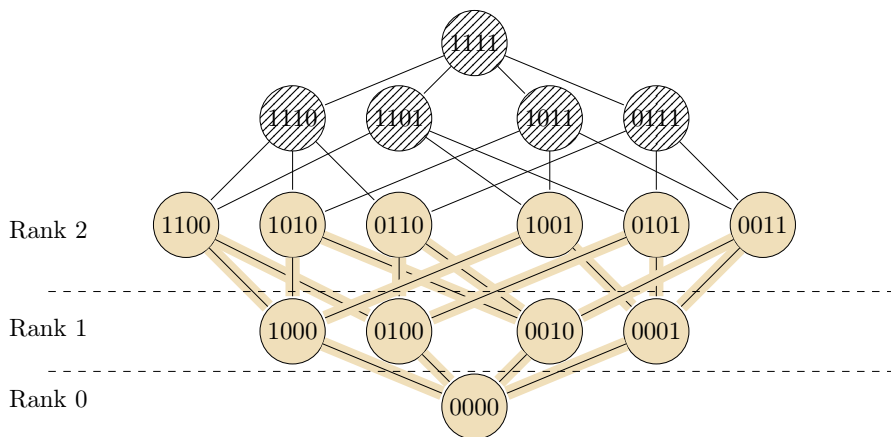
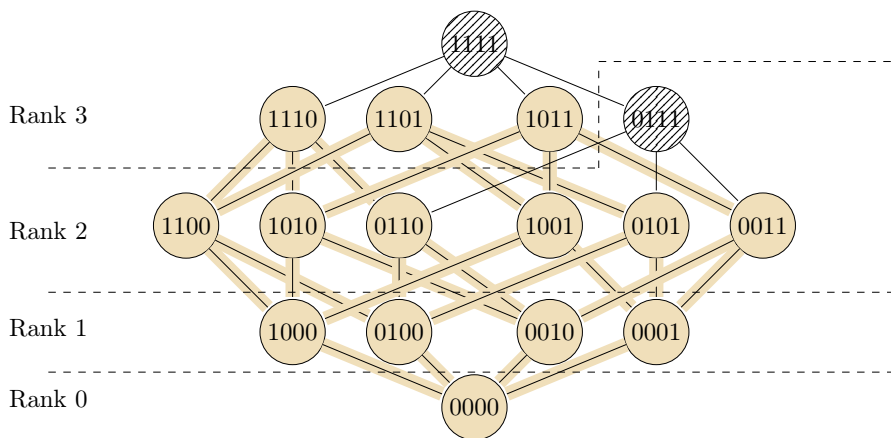Figure 2.4: The uniform matroid U_4^2.

Figure 2.5: An erection of U_4^2.

21

designated all sets of rank 3 as independent. Figure 2.5 shows another erection of U_4^2, one in which only three of the subsets with three elements are designated as independent. The final set, 0111, remains a dependent set, and has therefore in fact been designated as a closed set of rank 2. By starting from the rank-0 matroid $\mathfrak{M}^{(0)}$, one can iteratively erect any matroid $\mathfrak{M}$, at each iteration i designating the sets to be closed sets of rank i (all the while ensuring that the axioms for the closed sets of a matroid are obeyed). This is the approach taken by Knuth in his matroid erection algorithm, which is discussed in Chapter 5.4.

2.3 Matroids in fair allocation

It should be clear at this point that matroids are compelling structures to work with in the context of fair allocations. There are two main use cases for matroids in fair allocation: matroid-rank valuation functions (as in the example scenario from the introduction) and matroid constraints. Matroids.jl will be developed with the empirical study of algorithms for these two scenarios in mind.

2.3.1 Matroid-rank valuations

In a fair allocation instance with matroid-rank valuations, each agent i has a corresponding matroid $\mathfrak{M}_i = (\mathcal{M}, \mathcal{I}_i)$, where $\mathcal{M}$ (the set of goods) is the ground set of elements common to all agents' matroids.

Babaioff 2021 – PE mechanism – non-redundant Lorenz dominating allocation always exists and is MNW, MAX-USW, leximin, EFX (EF1) and .5-MMS. With a randomized prioritization it is ex ante EF and ex ante proportional as well. 3Yankee Swap computes prioritized non-redundant Lorenz dominating allocations.

Limitation: MRFs cannot model complementary goods!

2.3.2 Matroid constraints

Another usage found for matroids in fair allocation is that of *matroid constraints*. The majority of work on fair division assumes that any allocation is feasible, and the sole concern is finding an allocation that aligns well with the agents' valuation profiles. In many practical applications, however, there will be allocations that are not legal or desirable. Suksompong [23] gives the example of a museum with multiple branches distributing exhibits of different categories (sculpture, paintings, et cetera) among the branches. For each category, it wants to create a

balanced distribution among the branches, so that the difference in the number of exhibits of a given category differ by at most one between any branch. This is an example of a *cardinality constraint*, which can be modeled with a partition matroid, and is a subset of the broader class of matroid constraints.

When matroid constraints are enforced on a fair allocation instance, we require that all bundles be independent sets on some supplied matroid, common to all agents.

Chapter 3

The Matroids.jl API

Matroids.jl exists to enable the empirical study of matroidal fair allocation. In this chapter, I consider what fair allocation-specific methods the Matroids.jl API should expose to achieve this goal, and how they might be implemented. While doing so, I keep track of which properties are required from the matroids being used. Chapter 5 describes how Matroids.jl generates a number of different matroid types, and how the getter functions for the properties we need are implemented.

The implementation will draw inspiration from, and be designed to integrate with, Hummel and Hetland's well-organized Allocations.jl library [10], which provides a range of algorithms for fair allocation of indivisible items. Allocations.jl currently supports additive and submodular valuations and a number of constraint types, including conflict constraints and cardinality constraints (which, incidentally, are partition matroid constraints [5]). Our goal is to extend Allocations.jl with support for matroid-rank valuations and matroidal constraints. As such, Matroids.jl should be structured in such a manner as to be familiar to those acquainted with Allocations.jl.

This chapter is structured as follows:

1. I first extend Allocation.jl's existing functionality to support representing and evaluating fair allocation instances with matroid-rank valuations and matroidal constraints

2. I then present the requirements of the algorithms bla bla bla

3. I implement some stuff (matroid partition??)

In the next chapter, I describe an implementation, using Matroids.jl, of the Yankee Swap algorithm for fair allocation with matroidal valuations [8] and present some experimental results.

3.1 Fairness under matroid-rank valuations

If Matroids.jl is to be of use in the empirical study of matroidal fair allocation algorithms, we need to be able to evaluate the fairness of an allocation. In this chapter, I show how Matroids.jl implements the fairness criteria given in Chapter 2. The valuation profile of matroid-rank-valued allocation problem instance gives the valuation function of each agent. This is represented as a struct containing the matroid $\mathfrak{M}_i$ for each agent i. Agent i's value for the set of goods S, $v_i(S)$, is the rank of S in $\mathfrak{M}_i$.

```
"""
    struct MatroidRank <: Profile

A matroid rank valuation profile, representing how each agent values all
    possible bundles. The profile is constructed from `n` matroids, one for
    each agents, each matroid over the set of goods [m].
"""
struct MatroidRank <: Profile
    matroids::Vector{Matroid}
    m::Int
end

value(V::MatroidRank, i, S) = rank(V.matroids[i], S)
value(V::MatroidRank, i, g::Int) = value(V, i, Set(g))
```

Envy-freeness

Checking if an allocation is EF is the same for matroid-rank valuations as for additive valuations – simply compare each agent's own bundle value with that agent's subjective valuation of each other agent's bundle. This is already implemented in Allocations.jl. In this section, I give the functions value_1, value_x and value_x0, which are used for computing EF1, EFX_+ and EFX_0, respectively. These functions take in a valuation profile, an agent i and a bundle S, and return the agent i's value for S, up to some item.

To check efficiently if an allocation is EF1, we make use of the fact that a matroid rank function has binary marginals; in other words, the highest valued

good in a bundle will always have value 1, unless the bundle value is 0. This gives us a simple way of checking for EF1. Similarly, since the least positively-valued good also has value 1, EFX_+ is the same as EF1.

```
value_1(V::MatroidRank, i, S) = max(value(V, i, S)-1, 0)
value_x(V::MatroidRank, i, S) = value_1(V, i, S)
```

The value of the least valued good overall (including 0-values) depend on whether the bundle is independent. An independent set contains by definition no redundant elements, so if the bundle is independent, the least-valued good has value 1. If the bundle is dependent, it contains at least one 0-valued good, or, equivalently, a good whose removal does not affect the bundle value. This gives us EFX_0.

```
value_x0(V::MatroidRank, i, A) =
    is_indep(V.matroids[i], A) ? value_1(V, i, A) : value(V, i, A)
```

Proportionality

To check whether an allocation A is PROP or some relaxation thereof, we compare $v_i(A_i)$ against some threshold for every agent i. In this section, we give the functions for computing the threshold for PROP and its relaxations.

PROP_i is simply the rank of $\mathfrak{M}_i$, $v_i(\mathcal{M})$, as this is the maximum value achievable for agent i, divided by the number of agents in the problem instance.

```
prop(V::MatroidRank, i, _) = rank(V.matroids[i])/na(V)
```

To check for PROP1, we need to figure out if there exists some $g \in \mathcal{M}$ such that $v_i(A_i + g) \geq \frac{1}{n}v_i(\mathcal{M})$. We know, due to the hereditary property (as given in Section 2.2.2) that unless $v_i(A_i) = v_i(\mathcal{M})$ already (in which case we have trivial PROP1), there exists $g \in \mathcal{M} \setminus A_i$ such that $\Delta_i(A_i, g) = 1$. To figure out if A is PROP1, then, we need to check whether $v_i(A_i) + 1 \geq \frac{1}{n}v_i(\mathcal{M})$, or equivalently, whether $v_i(A_i) \geq \frac{1}{n}v_i(\mathcal{M}) - 1 = \text{PROP}_i - 1$. This is our PROP1 threshold. Since the least positively-valued element will also have a marginal value of 1, PROPX_+ is the same as PROP1.

```
prop_1(V::MatroidRank, i, A) = prop(V, i, A) - 1
prop_x(V::MatroidRank, i, A) = prop_1(V, i, A)
```

When checking for PROPX$_0$, we want the $g \in E \setminus A_i$ whose addition would increase the value of A_i the least. The question, then, is whether there exists an element $g \in E \setminus A_i$ such that $\Delta_i(A_i, g) = 0$. If A_i is a closed set (ie. maximal for its rank), then any additional good will increase the rank by 1, otherwise there exists some such g.

```
prop_x(V::MatroidRank, i, A) =
    is_closed(V.matroids[i], A) ? prop_1(V, i, A) : prop(V, i, A)
```

Maximin share

Matroids.jl implements Barman and Verma's [6] method for computing agent i's maximin share μ_i in polynomial time. It starts off with a partitioning of $\mathcal{M}$ into n sets that are independent in $\mathfrak{M}_i$, which can be produced in polynomial time using the matroid partioning algorithm [14, p. 731]. This

3.2 The matroid partitioning algorithm

Chapter 4

Implementing Yankee Swap

4.1 The matroid union algorithm

4.2   Some experimental results

Chapter 5

Generating matroids

The overarching goal for this project is to make Matroids.jl, a proof-of-concept library for working programmatically with matroids, specifically in the context of fair allocation. This chapter covers how Matroids.jl enables the creation of specific matroids and the generation of random ones, as well as how to access important properties such as independent sets, closed sets, circuits, bases, the rank function and the closure function. The first part of the chapter focuses on implementing these features for various types of matroids, including uniform, linear (vector), graphic, and partition matroids. The final part of the chapter is a significant portion of the thesis as a whole, and describes the implementation of Knuth's interesting algorithm [11] for the erection of arbitrary rank-r matroids.

1. `rank`

2. `closure`

3. `bases`

4. `is_indep`

5. `is_circuit`

6. `is_basis`

5.1 Uniform matroids and partition matroids

A uniform matroid U_n^r is the matroid over n elements where the independent sets are exactly the sets of cardinality at most r. The free matroid $U_n^n = (E, 2^E)$ is a special case of the uniform matroid and is the simplest, biggest and least interesting type of matroid, being the trivial case in which every subset of E is an independent set. In Matroids.jl, we represent uniform matroids with a simple struct.

```
struct UniformMatroid
  n::Integer
  r::Integer
end

FreeMatroid(n) = UniformMatroid(n, n)
```

5.2 Linear matroids

5.3 Graphic matroids

We begin with defining the graph theory terms used in this section. An undirected graph $G = (V, E)$ is said to be *connected* if there exists at least one path between each pair of nodes in the graph; otherwise it is *disconnected*. A disconnected graph consists of at least two connected subsets of nodes. These connected subgraphs are called *components*. The *degree* of a node v is the number of edges for which v is an endpoint. A *regular graph* is a graph in which all nodes have the same degree. An *induced subgraph* $G[S]$, where S is either a subset of the nodes of G (in which case $G[S]$ is a *node-induced subgraph*) or of the edges of G (*edge-induced*).

5.3.1 Random graphs

In order to generate random graphic matroid, we will need to generate random graphs, let us take a look at some of the options available to us for this. Luckily for us, random graphs has been an area of extensive study for more than sixty years, and several models with different properties exist.

The Erdős-Rényi (ER) model (also known as Erdős-Rényi-Gilbert [24]) picks uniformly at random a graph from among the $\binom{\binom{n}{2}}{M}$ possible graphs with n nodes

and M edges, or, alternatively, constructs a graph with n nodes where each edge is present with some probability p [25, 26]. This model produces mostly disconnected graphs, and the size distribution of its components with respect to the number of edges has been studied extensively. With n nodes and fewer than $\frac{n}{2}$ edges, the resulting graph will almost always consist of components that are small trees or contain at most one cycle. As the number of edges exceeds $\frac{n}{2}$, however, the so-called "giant" component of size $\mathcal{O}(n)$ emerges, and starts to absorb the smaller components [27]. The ER model is the oldest and most basic random graph model, and is often referred to simply as the random graph, denoted by $G(n, p)$.

Variations of the ER model have been developed by physicists and network scientists to produce phenomena commonly seen in real-world networks [24]. These variations include the Barabási-Albert model, which grows an initial connected graph using preferential attachment (a mechanism colloquially known as "the rich get richer"), in which more connected nodes are more likely to receive new connections. This results in graphs in which a small number of nodes ("hubs") have a significantly higher degree than the rest, creating a power-law distribution of node degrees. This property is known as scale-freeness and is thought to be a characteristic of the Internet [28].

Another approach is the Watts-Strogatz model, which starts with a ring lattice, a regular graph with n nodes, each with degree k, and then rewires each edge with some probability p. By changing p, one is able to 'tune' the graph between regularity (p=0) and disorder (p=1). For intermediate values of p, Watts-Strogatz produces so-called "small-world" graphs, which exhibit both a high degree of clustering (how likely two nodes with a common neighbor are to be adjacent), and short average distance between nodes. This phenomenon is found in many real-world networks, such as social systems or power grids [29].

5.3.2 Properties of random graphic matroids

We will use the Graphs.jl library [30] for handling graphs in Matroids.jl. This library has built-in methods for the random graph models described in the previous chapter[1].

When constructing matroids, we want to be able to specify the size of the ground set, and perhaps also the rank of the matroid. Let us see how we can achieve this with the random graph models we have discussed. The method `barabasi_albert(n,k)` generates a Barabási-Albert model random graph with

---

[1]https://docs.juliahub.com/Graphs/VJ6vx/1.4.1/generators/

n nodes. It starts with an initial graph of k nodes, and adds the remaining $n-k$ nodes one at a time, each new node receiving k edges via preferential attachment. Thus, the final graph has $|E| = (n-k)k$ edges. To specify a matroid with m edges, we pick some $k|m$ and solve for n. Remember that the rank of a graphic matroid is the size of a spanning tree over the graph, which is $n-1$ when the graph is connected. If we select a smaller k from among the factors of $|E|$, we get a larger final rank, and vice versa. We can generate a Watts-Strogatz model random graph with the method `watts_strogatz(n, k, β)`, where n is the number of nodes, k the node degree and β the probability of rewiring. The number of edges of a regular graph with n nodes and degree k (and thus the size of the ground set of the induced graphic matroid) is given by $\frac{nk}{2}$, so nk must be even. Erdős-Rényi is the simplest model for our purposes, as the method `erdos_renyi(nv, ne)` simply takes in the desired number of nodes and edges. However, since the resulting graph has a large number of components for $|E| < \frac{n}{2}$, we have less fine-grained control over the final rank of the induced graphic matroid.

In Matroids.jl, we "generate" a graphic matroid by simply accepting some graph, and figure out the rank of the matroid using Kruskal's algorithm for maximal spanning forests, which runs in $\mathcal{O}(|E|\lg|E|)$ time [31]. Implementing the methods for finding the properties of our graphic matroids is simple, as they reduce to well-known algorithms (implemented by Graphs.jl) for finding the properties of the graphs they are derived from.

```julia
using Graphs

struct GraphicMatroid
  g::Graph
  n::Integer
  r::Integer
  GraphicMatroid(g::Graph) = new(g, ne(g), length(kruskal_mst(g)))
end
```

The rank function returns the size of a spanning forest of the subgraph induced by some subset of the edges. This is the rank of that subset. Thus, the rank of a subset $S \subseteq E$ can be found in $\mathcal{O}(|S|\lg|S|)$ time (when the MST is found using Kruskal's algorithm).

```julia
function rank(m::GraphicMatroid, S)
  edgelist = [e for (i, e) in enumerate(edges(g)) if i in S]
  subgraph, _vmap = induced_subgraph(m.g, edgelist)
  return length(kruskal_mst(subgraph))
```

```
end
```

The indepence oracle returns whether the subgraph induced by a supplied subset of edges is acyclic. While independence can also be determined with the rank function, by checking whether the cardinality of a set equals its rank, this uses a DFS behind the scenes[2], which runs in linear time [31].

```
function is_indep(m::GraphicMatroid, S)
  edgelist = [e for (i, e) in enumerate(edges(g)) if i in S]
  subgraph, _vmap = induced_subgraph(m.g, edgelist)
  return !is_cyclic(subgraph)
end
```

The circuit oracle

```
function is_circuit(m::GraphicMatroid, S)
  #TODO
end
```

The closure function accepts a set of elements S, and returns the largest set of elements $cl(S)$ such that $S \subseteq cl(S) \subseteq E, r(S) = r(cl(S))$. In a graph context, given a graph $G = (V, E)$ and an edge-induced subgraph $G[S] = (V, S), S \subseteq E$, this is the same as finding the largest edge-induced subgraph $G[T], S \subseteq T \subseteq E$, in which a spanning tree has the same number of edges as one in $G[S]$. Since the size of a spanning tree in $G[S]$ is given by $|V| - 1$, $G[T]$ cannot contain any edges to nodes not in V, as this would increase the rank of $G[T]$. Therefore, we get that the closure of S is the largest set T of edges between nodes that are present in the edge-induced subgraph $G[S]$. The method `closure` below returns the set of all edges whose endpoints are both located in the subgraph induced by S.

```
function closure(m::GraphicMatroid, S)
  edgelist = [e for (i, e) in enumerate(edges(m.g)) if i in S]
  _sg, vmap = induced_subgraph(m.g, edgelist)
  return [e for e in edges(m.g) if [e.src, e.dst] ⊆ vmap]
end
```

---

[2]https://docs.juliahub.com/Graphs/VJ6vx/1.4.1/pathing/#Graphs.is_cyclic

5.4 Knuth's matroid construction

In the preparatory project delivered fall of 2022, I implemented Knuth's 1974 algorithm for the random generation of arbitrary matroids via the erection of closed sets [11]. With this, I was able to randomly generate matroids of universe sizes $n \leq 12$, but for larger values of n my implementation was unbearably slow. In this section, Knuth's method for random matroid construction will be described, along with the steps I have taken to speed up my initial, naïve implementation.

KNUTH-MATROID (given in Algorithm 1) accepts the ground set E and a list X such that $X[i] \subseteq 2^E$, and produces the rank-r matroid $\mathfrak{M}$ such that $\text{rank}(X) = k$ for each $X \in X[k]$. This is done in a bottom-up manner through r sequential erections starting from the empty rank-0 matroid, $\mathfrak{M}^{(0)}$, each iteration i producing the erection $\mathfrak{M}^{(i+1)}$ from $\mathfrak{M}^{(i)}$ and $X[i]$. The algorithm outputs the tuple (E, F), where $F = [F_0, \ldots, F_r]$, r being the final rank of $\mathfrak{M}$ and F_i the family of closed sets of rank i. In the paper, Knuth shows that $\bigcup_{i=0}^{r} F[r] = \mathcal{F}$, where $\mathcal{F}$ is the set of closed sets of a matroid, and so the algorithm produces a valid matroid represented by its closed sets.

Algorithm 1 KNUTH-MATROID(E, X)

Input: The ground set of elements E, and a list of enlargements X.

Output: The list of closed sets of the resulting matroid grouped by rank,
          $F = [F_0, \ldots, F_r]$, where F_i is the set of closed sets of rank i.

---

1 $r = 0, F = [\{\emptyset\}]$
2 **while** TRUE
3   PUSH!$(F, \text{GENERATE-COVERS}(F, r, E))$
4      $F[r+1] = F[r+1] \cup X[r+1]$
5   SUPERPOSE!$(F[r+1], F[r])$
6   **if** $E \notin F[r+1]$
7          $r \leftarrow r + 1$
8   **else**
9   **return** (E, F)

To understand the procedure, let us investigate what Algorithm 1 does at iteration $1 < i < r$, where r is the final rank $\mathfrak{M}$, the matroid under construction.

At iteration i, we produce a rank-$(i + 1)$ erection $\mathfrak{M}^{(i+1)}$ of $\mathfrak{M}^{(i)}$, which is represented by its closed sets $F = [F_0, F_1, ..., F_i]$, where F_i is the set of closed sets of rank i. We want to produce the set F_{i+1} of rank-r closed sets of an erection of $\mathfrak{M}^{(i)}$ such that each $X \in X[i]$ is contained in some rank-r closed set. First we find the "covers" of each closed set in F_i. The covers of a closed set A of rank r are the sets obtained by adding one more element from E to A. The covers are generated with GENERATE-COVERS(F, r, E).

GENERATE-COVERS(F, r, E)
1 return $\{A \cup \{a\} : A \in F[r], a \in E \setminus A\}$

Given no enlargements ($X[i] = \emptyset$), the resulting matroid $\mathfrak{M}^{(i+1)}$ is the *free erection* of $\mathfrak{M}^{(i)}$, and there are no essential closed sets in F_{i+1}. Arbitrary matroids can be generated by supplying different lists X. When enlarging, the sets in $X[r+1]$ are simply added to $F[r+1]$, before SUPERPOSE! is run to ensure that the newly enlarged family of closed sets of rank $r + 1$ is valid (ie. in accordance with the closed set axioms given in Section 2.2). If F_{r+1} contains two sets A, B whose intersection $A \cap B \nsubseteq C$ for any $C \in F_r$ (in other words, their intersection is not a closed set), replace A, B with $A \cup B$. Repeat until no two sets exist in F_{r+1} whose intersection is not contained within some set $C \in F_r$.

SUPERPOSE!(F_{r+1}, F_r)
 1 **for** $A \in F_{r+1}$
 2   **for** $B \in F_{r+1}$
 3     $flag \leftarrow$ TRUE
 4   **for** $C \in F_r$
 5   **if** $A \cap B \subseteq C$
 6       $flag \leftarrow$ FALSE
 7
 8   **if** $flag =$ TRUE
 9      $F_{r+1} \leftarrow F_{r+1} \setminus \{A, B\}$
10      $F_{r+1} \leftarrow F_{r+1} \cup \{A \cup B\}$

5.4.1 Randomized KMC

In the randomized version of KNUTH-MATROID, we generate matroids by applying a supplied number of random coarsening steps, instead of enlarging with

supplied sets. This is done by applying SUPERPOSE! immediately after adding the covers, then choosing a random member A of F$[r+1]$ and a random element $a \in E \setminus A$, replacing A with $A \cup \{a\}$ and finally reapplying SUPERPOSE!. The parameter $p = (p_1, p_2, \ldots)$ gives the number of such coarsening steps to be applied at each iteration of the algorithm.

The pseudocode given up to this point corresponds closely to the initial Julia implementation, which can be found in Appendix **??**. It should already be clear that this brute force implementation leads to poor performance – for instance, the SUPERPOSE! method uses a triply nested for loop, which should is a candidate for significant improvement. Section 5.4.2 describes the engineering work done to create a more performant implementation.

Algorithm 2 RANDOMIZED-KNUTH-MATROID(E, p)

Input: The ground set of elements E, and a list $p = [p_1, p_2, ...]$, where p_r is the number of coarsening steps to apply at rank r in the construction.

Output: The list of closed sets of the resulting matroid grouped by rank, F $= [F_0, \ldots, F_r]$, where F_i is the set of closed sets of rank i.

1   $r = 0, \mathrm{F} = [\{\emptyset\}]$
2   **while** TRUE
3   PUSH!(F, GENERATE-COVERS(F, r, E))
4   SUPERPOSE!(F$[r+1]$, F$[r]$)
5   **if** $E \in \mathrm{F}[r+1]$ **return** (E, F)
6   **while** $p[r] > 0$
7          $A \leftarrow$ a random set in F$[r+1]$
8          $a \leftarrow$ a random element in $E \setminus A$
9   **replace** A with $A \cup \{a\}$ in F$[r+1]$
10   SUPERPOSE!(F$[r+1]$, F$[r]$)
11   **if** $E \in \mathrm{F}[r+1]$ **return** (E, F)
12        $p[r] = p[r] - 1$
13      $r = r + 1$

5.4.2 Improving performance

When recreating Knuth's table of observed mean values for the randomly generated matroids, some of the latter configurations of n and $(p_1, p_2, \ldots)$ was unworkably slow, presumably due to my naïve implementation of the algorithm. Table 5.1 shows the performance of this first implementation.

Table 5.1: Performance of `random_kmc_v1`.

n	$(p_1, p_2, \ldots)$	Trials	Time	GC Time	Bytes allocated
10	(0, 6, 0)	100	0.0689663	0.0106786	147.237 MiB
10	(0, 5, 1)	100	0.1197194	0.0170734	251.144 MiB
10	(0, 5, 2)	100	0.0931822	0.0144022	203.831 MiB
10	(0, 6, 1)	100	0.0597314	0.0094902	132.460 MiB
10	(0, 4, 2)	100	0.1924601	0.0284532	406.131 MiB
10	(0, 3, 3)	100	0.3196838	0.0463972	678.206 MiB
10	(0, 0, 6)	100	1.1420602	0.1671325	2.356 GiB
10	(0, 1, 1, 1)	100	2.9283978	0.3569357	5.250 GiB
13	(0, 6, 0)	10	104.0171128	9.9214449	161.523 GiB
13	(0, 6, 2)	10	11.4881308	1.3777947	20.888 GiB
16	(6, 0, 0)	1	-	-	-

The performance was measured using Julia's `@timed`[3] macro, which returns the time it takes to execute a function call, how much of that time was spent in garbage collection and the number of bytes allocated. As is evident from the data, larger matroids are computationally quite demanding to compute with the current approach, and the time and space requirements scales exponentially with n. Can we do better? As it turns out, we can; after the improvements outlined in this section, we will be able to generate matroids over universes as large as $n = 128$ in a manner of seconds and megabytes.

---

[3]https://docs.julialang.org/en/v1/base/base/#Base.@timed

Representing sets as binary numbers

The first improvement we will attempt is to represent our closed sets using one of Julia's `Integer` types of bit width at least n, instead of as a `Set`[4] of elements of E. Appendix B contains all the code referenced in this chapter; the Julia implementation at this point can be found in **??**.

1

The idea is to define a family of closed sets of the same rank as `Set{UInt16}`. Using `UInt16` we can support ground sets of size up to 16. Each 16-bit number represents a set in the family. For example, the set $\{2, 5, 7\}$ is represented by

$$164 = \text{0x00a4} = \text{0b0000000010100100} = 2^7 + 2^5 + 2^2.$$

At either end we have $\emptyset \equiv \text{0x0000}$ and $E \equiv \text{0xffff}$ (if $n = 16$). The elementary set operations we will need have simple implementations using bitwise operations.

Set operation	Bitwise operation
$A \cap B$	A AND B
$A \cup B$	A OR B
$A \setminus B$	A AND NOT B
$A \subseteq B$	A AND $B = A$

We can now describe the bitwise versions of the required methods. The bitwise implementation of GENERATE-COVERS finds all elements in $E \setminus A$ by finding each value $0 \leq i < n$ for which `A & 1 << i === 0`, meaning that the set represented by `1 << i` is not a subset of A. The bitwise implementation of SUPERPOSE! is unchanged apart from using the bitwise set operations described above.

---

[4]https://docs.julialang.org/en/v1/base/collections/#Base.Set

Table 5.2: Performance of `random_kmc_v2`.

n	$(p_1, p_2, \ldots)$	Trials	Time	GC Time	Bytes allocated
10	[0, 6, 0]	100	0.0010723	0.0001252	1.998 MiB
10	[0, 5, 1]	100	0.0017543	0.0001431	3.074 MiB
10	[0, 5, 2]	100	0.0008836	0.0001075	2.072 MiB
10	[0, 6, 1]	100	0.0007294	6.73e-5	1.700 MiB
10	[0, 4, 2]	100	0.0020909	0.0001558	3.889 MiB
10	[0, 3, 3]	100	0.0024636	0.0002139	4.530 MiB
10	[0, 0, 6]	100	0.007082	0.0004801	9.314 MiB
10	[0, 1, 1, 1]	100	0.0132477	0.0008307	17.806 MiB
13	[0, 6, 0]	10	0.042543	0.0014988	31.964 MiB
13	[0, 6, 2]	10	0.0183313	0.0012176	21.062 MiB
16	[0, 6, 0]	10	1.2102877	0.0146129	450.052 MiB

The performance of `random_kmc_v2` is shown in Table 5.2. It is clear that representing closed sets using binary numbers represents a substantial improvement – we are looking at performance increases of 100x-1000x across the board. Great stuff!

Sorted superpose

Can we improve the running time of our implementation further? It is clear that SUPERPOSE! takes up a large portion of the compute time. In the worst case, when no enlargements have been made, F_{r+1} is the set of all $r + 1$-sized subsets of E, $|F_{r+1}| = \binom{n}{r+1}$. Comparing each $A, B \in F_{r+1}$ with each $C \in F_r$ in a triply nested for loop requires $\mathcal{O}(\binom{n}{r+1}^2 \binom{n}{r})$ operations. In the worst case, no enlargements are made at all, and we build the free matroid in $\mathcal{O}(2^{3n})$ time (considering only the superpose step).

After larger closed sets have been added to F[$r + 1$], SUPERPOSE! will cause sets to merge, so that only maximal dependent sets remain. Some sets will even simply disappear. In the case where $X = \{1, 2\}$ was added by GENERATE-COVERS, and the $Y = \{1, 2, 3\}$ was added manually as an enlargement, the smaller set will be fully subsumed in the bigger set, as $\{1, 2\} \cap \{1, 2, 3\} = \{1, 2\}$ (which is not a subset of any set in F[r]) and $\{1, 2\} \cup \{1, 2, 3\} = \{1, 2, 3\}$. In

this situation, Y would "eat" the covers $\{1,3\}$ and $\{2,3\}$ as well. This fact is reflected in the performance data – compare the memory allocation differences between the 10-element matroid with $p = [0, 0, 6]$ and the one with $p = [0, 6, 0]$ in any of the performance tables in this section. Making enlargements at earlier ranks result in smaller matroids as more sets get absorbed.

```
function sorted_bitwise_superpose!(F, F_prev)
  As = sort!(collect(F), by = s -> length(bits_to_set(s)))
  while length(As) !== 0
    A = popfirst!(As)

    for B in setdiff(F, A)
      if should_merge(A, B, F_prev)
        insert!(As, 1, A | B)
        setdiff!(F, [A, B])
        push!(F, A | B)
        break
      end
    end
  end

  return F
end
```

Since the larger sets will absorb so many of the smaller sets (around $\binom{p}{r+1}$, where p is the size of the larger set and $r + 1$ is the size of the smallest sets allowed to be added in a given iteration), might it be an idea to perform the superpose operation in descending order based on the size of the sets? This should result in fewer calls to SUPERPOSE!, as the bigger sets will remove the smaller sets that fully overlap with them in the early iterations, however, the repeated sorting of the sets might negate this performance gain. This is the idea behind `sorted_bitwise_superpose!`, which was used in `random_kmc_v3`. The full code can be found in Appendix **??**.

Unfortunately, as Table 5.3 shows, this implementation is a few times slower and more space demanding than the previous implementation. This is might be due to the fact that an ordered list is more space inefficient than the hashmap-based `Set`.

Iterative superpose

The worst-case $\mathcal{O}(\binom{n}{r+1}^2 \binom{n}{r})$ runtime of SUPERPOSE! at step r is due to the fact that it takes in F after all covers and enlargements have been indiscriminately added to F[$r + 1$] and then loops through to perform the superposition. Might

Table 5.3: Performance of `random_kmc_v3`.

n	$(p_1, p_2, \ldots)$	Trials	Time	GC Time	Bytes allocated
10	[0, 6, 0]	100	0.0023382	0.0001494	4.042 MiB
10	[0, 5, 1]	100	0.001853	0.0001433	4.383 MiB
10	[0, 5, 2]	100	0.0017845	0.0001341	4.043 MiB
10	[0, 6, 1]	100	0.0015145	0.0001117	3.397 MiB
10	[0, 4, 2]	100	0.0030704	0.0002125	6.385 MiB
10	[0, 3, 3]	100	0.0037838	0.0002514	7.018 MiB
10	[0, 0, 6]	100	0.008903	0.000557	14.159 MiB
10	[0, 1, 1, 1]	100	0.0142828	0.0008823	21.838 MiB
13	[0, 6, 0]	10	0.0627633	0.002094	51.492 MiB
13	[0, 6, 2]	10	0.0106478	0.0007704	20.774 MiB
16	[0, 6, 0]	10	0.6070136	0.0095656	310.183 MiB

there be something to gain by inserting new closed sets into the current family one at a time, and superposing on the fly?

```
# Superpose (random_kmc_v4)
push!(F, Set()) # Add F[r+1].
while length(to_insert) > 0
  A = pop!(to_insert)
  push!(F[r+1], A)

  for B in setdiff(F[r+1], A)
    if should_merge(A, B, F[r])
      push!(to_insert, A | B)
      setdiff!(F[r+1], [A, B])
      push!(F[r+1], A | B)
    end
  end
end
```

In `random_kmc_v4`, the full code of which can be found in Appendix **??**, the covers and enlargements are not added directly to $F[r + 1]$, but to a temporary array `to_insert`. Each set A is then popped from `to_insert` one at a time, added to $F[r + 1]$ and compared with the other sets $B \in F[r + 1] \setminus \{A\}$ and $C \in F[r]$ in the usual SUPERPOSE! manner. This results in fewer comparisons,

as each set is only compared with the sets added before it; the first set is compared with no other sets, the second set with one other and the sets in F[r], and so on. The number of such comparisons is therefore given by the triangular number $T_{\binom{n}{r+1}}$, and so we should have roughly halved the runtime at step r. It is worth noting that this implementation of SUPERPOSE! uses a subroutine `should_merge` that returns early when it finds one set $C \in \text{F}[r]$ such that $C \supseteq A \cap B$, so in practice it usually does not require $\binom{n}{r}$ comparisons in the innermost loop.

Table 5.4 shows that the iterative superpose was a meaningful improvement. For most input configurations, it is a few times faster and a few times less space demanding than `random_kmc_v2`.

Table 5.4: Performance of `random_kmc_v4`.

n	$(p_1, p_2, \ldots)$	Trials	Time	GC Time	Bytes allocated
10	[0, 6, 0]	100	0.0014585	3.94e-5	724.635 KiB
10	[0, 5, 1]	100	0.0007192	9.39e-5	659.729 KiB
10	[0, 5, 2]	100	0.0005943	3.53e-5	617.668 KiB
10	[0, 6, 1]	100	0.0003502	2.88e-5	408.666 KiB
10	[0, 4, 2]	100	0.001013	5.36e-5	887.618 KiB
10	[0, 3, 3]	100	0.0011847	5.03e-5	1.003 MiB
10	[0, 0, 6]	100	0.0015756	9.7e-5	1.066 MiB
10	[0, 1, 1, 1]	100	0.0046692	0.0001385	2.455 MiB
13	[0, 6, 0]	10	0.0118201	0.0005486	6.289 MiB
13	[0, 6, 2]	10	0.0075668	0.0002458	4.666 MiB
16	[0, 6, 0]	10	0.2819294	0.0040792	81.317 MiB
16	[0, 6, 1]	10	0.8268207	0.0070206	154.451 MiB
16	[0, 0, 6]	10	95.1959596	0.0290183	553.597 MiB

Rank table

While SUPERPOSE! is getting more efficient, it is still performing the same comparisons over and over again. Let's consider what we are really trying to achieve with this function, to see if we can't find a smarter way to go about it.

After adding the closed sets for a rank, Superpose! is run to maintain the closed set properties of the matroid (given in Section 2.2). These are maintained by ensuring that, for any two newly added sets $A, B \in \mathrm{F}[r + 1]$, there exists $C \in \mathrm{F}[r]$ such that $A \cap B \subseteq C$. Up to this point, this has been done by checking if the intersection of each such A, B is contained in a set C of rank r. We remember that one of the properties of the closed sets of a matroid is that the intersection of two closed sets is itself a closed set. Therefore, we do not need to find a closed set C of rank r that *contains* $A \cap B$, since if A and B are indeed closed sets, their intersection will be *equal* to some closed set C of any rank $\leq r$. This insight leads us to the next improvement: if we keep track of all added closed sets in a rank table, then we can memoize Superpose! and replace the innermost loop with a constant time dictionary lookup.

```
# The rank table maps from the representation of a set to its assigned rank.
rank = Dict{T, UInt8}(0=>0)

[...]

# Superpose.
push!(F, Set()) # Add F[r+1].
while length(to_insert) > 0
  A = pop!(to_insert)
  push!(F[r+1], A)
  rank[A] = r


  for B in setdiff(F[r+1], A)
    if !haskey(rank, A&B) || rank[A&B] >= r
      # Update insert queue.
      push!(to_insert, A | B)

      # Update F[r+1].
      setdiff!(F[r+1], [A, B])
      push!(F[r+1], A | B)

      # Update rank table.
      rank[A|B] = r
      break
    end
  end
end
```

The full code for `random_kmc_v5` can be found in Appendix **??**. Table 5.5 shows that implementing a rank table was an extremely significant improvement. For smaller matroids, it is around 5-10x faster, however it is for larger matroids that it truly outshines its predecessors – `random_kmc_v5` is a whopping 13 000 times faster than `random_kmc_v4` with $n = 16, p = [0, 0, 6]$ as input.

Table 5.5: Performance of `random_kmc_v5`.

n	$(p_1, p_2, \ldots)$	Trials	Time	GC Time	Bytes allocated
10	[0, 6, 0]	100	0.0001335	0.0	138.966 KiB
10	[0, 5, 1]	100	0.0001436	0.0	158.691 KiB
10	[0, 5, 2]	100	0.0001928	0.0	167.487 KiB
10	[0, 6, 1]	100	0.0002204	0.0	148.812 KiB
10	[0, 4, 2]	100	0.0001578	0.0	173.455 KiB
10	[0, 3, 3]	100	0.0001743	0.0	202.566 KiB
10	[0, 0, 6]	100	0.0003433	0.0	431.089 KiB
10	[0, 1, 1, 1]	100	0.0004987	0.0	439.511 KiB
13	[0, 6, 0]	100	0.0004776	0.0	422.431 KiB
13	[0, 6, 2]	100	0.0003469	0.0	441.621 KiB
16	[0, 6, 0]	100	0.0009073	0.0	1010.452 KiB
16	[0, 6, 1]	100	0.0007939	0.0	997.022 KiB
16	[0, 0, 6]	100	0.0066951	0.0	8.564 MiB
20	[0, 6, 0]	100	0.0030797	0.0	4.042 MiB
20	[0, 6, 2]	10	0.0022849	0.0	4.547 MiB
32	[0, 6, 2, 1]	10	0.0269912	0.0	63.082 MiB

Non-redundant cover generation

Up to this point, our cover generation routine has not taken into account that any two sets of rank r will have at least one cover in common. To see this, consider a matroid-under-construction with $n = 10$ where $A = \{1, 2\}$ and $B = \{1, 3\}$ are closed sets of rank 2. Currently, GENERATE-COVERS will happily generate the cover $C = \{1, 2, 3\}$ twice, once as the cover of A and subsequently as the cover of B. Throughout this analysis, we will assume the worst case scenario of no enlargements, as any enlargements will strictly lower the number of sets in play at a given rank. In this case, $|\mathrm{F}[r]| = \binom{n}{r}$, and for each closed set A of rank r we are generating $|E \setminus A| = (n - r)$ covers, giving us a total of $\binom{n}{r}(n - r)$ covers generated at each rank r, including the duplicates. With no enlargements, we know that there are $\binom{n}{r+1}$ covers, and

$$(n-r)\binom{n}{r} = \frac{n!(n-r)}{r!(n-r)!}$$

$$= \frac{n!}{r!(n-r-1)!}$$

$$= (r+1)\frac{n!}{(r+1)!(n-r-1)!}$$

$$= (r+1)\binom{n}{r+1}.$$

For each step r, we are generating $r + 1$ times as many covers as we need to. Over the course of all steps $0 \le r \le n$, we are generating

$$\sum_{r=0}^{n}(r+1) = \sum_{r=1}^{n+1} r = T_{n+1}$$

times the actual number of covers, where $T_{n+1} = \frac{(n+1)(n+2)}{2}$ is the triangular number. In other words, if we find a way to generate each cover only once, we will have shaved off an n^2 factor from the asymptotic complexity of our implementation.

When generating covers, `random_kmc_v6` improves upon the brute force cover generation described above by only adding the covers

$$\Big\{A \cup \{a\} : A \in \mathrm{F}[r], a \in E \setminus A, a \notin \bigcup\{B : B \in \mathrm{F}[r+1], A \subseteq B\}\Big\}.$$

In other words, we find the covers of A, that is, the sets obtained by adding one more element a from E to A, but we do not include any a that is to be found in another, already added, cover B that contains A. This solves the problem described above; the cover $\{1, 2, 3\} = B \cup \{2\}$ will not be generated, as $2 \in C$ and $B \subseteq C$. This is implemented in the following manner:

```
# Generate minimal closed sets for rank r+1 (random_kmc_v6)
for y in F[r] # y is a closed set of rank r.
  t = E - y # The set of elements not in y.
  # Find all sets in F[r+1] that already contain y and remove excess elements
      from t.
  for x in F[r+1]
    if (x & y == y) t &= ~x end
    if t == 0 break end
  end
  # Insert y ∪ a for all a ∈ t.
```

```
    while t > 0
      x = y|(t&-t)
      add_set!(x, F, r, rank)
      t &= ~x
    end
  end
```

We have extracted the iterative superpose logic described above into its own function to allow it to be performed on a cover-per-cover basis:

```
function add_set!(x, F, r, rank)
  if x in F[r+1] return end
  for y in F[r+1]
    if haskey(rank, x&y) && rank[x&y]<r
    continue
    end

    # x ∩ y has rank > r, replace with x ∪ y.
    setdiff!(F[r+1], y)
    return add_set!(x|y, F, r, rank)
  end

  push!(F[r+1], x)
  rank[x] = r
end
```

As such, $F[r+1]$ is empty when the first cover $y \in F$ is generated, and all covers $\{y \cup \{a\} : a \in E \setminus y\}$ are added. For later sets y, we are comparing with the previously added covers, and dropping any element to be found in a cover x that fully includes y. This way, we avoid re-generating the cover x.

The full code for `random_kmc_v6` can be found in Appendix **??**.

Table 5.6: Performance of `random_kmc_v6`.

n	$(p_1, p_2, \ldots)$	Trials	Time	GC Time	Bytes allocated
10	[0, 6, 0]	100	0.000157	0.0	11.306 KiB
10	[0, 5, 1]	100	0.0001427	0.0	12.257 KiB
10	[0, 5, 2]	100	0.000121	0.0	11.568 KiB
10	[0, 6, 1]	100	8.61e-5	0.0	10.447 KiB
10	[0, 4, 2]	100	0.0001237	0.0	13.597 KiB
10	[0, 3, 3]	100	0.0001233	0.0	14.029 KiB
10	[0, 0, 6]	100	0.0002856	0.0	15.414 KiB
10	[0, 1, 1, 1]	100	0.0001942	0.0	14.446 KiB
13	[0, 6, 0]	100	0.0004483	0.0	19.117 KiB
13	[0, 6, 2]	100	0.0004541	0.0	18.957 KiB
16	[0, 6, 0]	10	0.0014919	0.0	34.531 KiB
16	[0, 6, 1]	10	0.0014731	0.0	36.016 KiB
16	[0, 0, 6]	10	0.0168858	0.0	127.652 KiB
20	[0, 6, 0]	10	0.0061574	0.0	81.573 KiB
20	[0, 6, 2]	10	0.0059717	0.0	82.323 KiB
32	[0, 6, 2, 1]	10	0.1599507	0.0	279.531 KiB
63	[0, 6, 4, 2, 1]	1	11.138914	0.0	4.912 MiB
64	[0, 6, 4, 4, 2, 1]	1	12.508729	0.0	4.912 MiB
128	[0, 6, 6, 4, 4, 2, 1]	1	1232.8570	0.0114583	102.159 MiB

5.4.3   What do the generated matroids look like?

Observations

1. The average cardinality of the closed sets of a given rank is usually not very much higher than the rank. If the average cardinality were to stray much, all the sets merge instead and the sole closed set of that rank would become E. This might be a useful heuristic when finding the rank of a set.
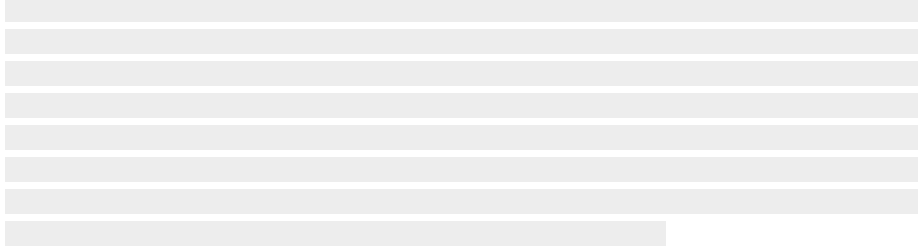
5.4.4 Finding the properties of arbitrary matroids

To build a general matroid library, we want to be able to access all properties of a generated matroid $\mathfrak{M}$, viz. its bases $\mathcal{B}$, independent sets $\mathcal{I}$, circuits $\mathcal{C}$, and rank function $rank : 2^E \to \mathbb{Z}^*$.

In this section, I will first describe an extension of KNUTH-MATROID that is also fully enumerates $\mathcal{I}$ and $\mathcal{C}$ for $\mathfrak{M}$ when n is small enough. However, this approach does not scale well for larger values of n. For values of n up to 128, we will therefore restrict our attention to independent sets and the rank function, as these are the matroid properties that are relevant to our usecase of fair allocation.

Up-front enumeration of circuits and independent sets for smaller matroids

In his 1974 paper [11], Knuth includes an ALGOL W [32] implementation that also enumerates all circuits and independent sets for the generated matroid. A later implementation in C called ERECTION.W can be found at Knuth's home page [33]. `random_erect` is an extension of `random_kmc_v6` that finds $\mathcal{I}$ and $\mathcal{C}$ by pre-populating the rank table with all subsets of E. The full source code for `random_erect` can be found in Appendix XXX.

```
# Populate rank table with 100+cardinality for all subsets of E.
k=1; rank[0]=100;
while (k<=mask)
  for i in 0:k-1 rank[k+i] = rank[i]+1 end
  k=k+k;
end
```

Covers are generated and sets inserted in the same manner as in `random_kmc_v6`. After all covers and enlargements have been inserted and

superposed (meaning F[$r + 1$] contains the closed sets of rank $r + 1$), a new operation, `mark_independent_subsets!` is called on each closed set.

Determining matroid properties after erection

In his 1989 paper, Greene introduces the concept of *descriptive sufficiency*. A subcollection of closed sets of a matroid is descriptively sufficient if it can be used to identify the fundamental properties of the matroid using certain easily applied conditions [22].

```julia
"""
  function is_indep(M::ClosedSetsMatroid, S::Integer)

Determines whether a given set S is independent in the matroid M, given by the
    closed sets of M grouped by rank. Uses (I.1) in Greene (1989).
"""
function is_indep(M::ClosedSetsMatroid, S::Integer)
  t = Base.count_ones(S)

  for F in M.F[t]
    if Base.count_ones(S&F) > t-1 return false end
  end

  return true
end
```

Indepence oracle.

50

```
"""
    function is_circuit(M::ClosedSetsMatroid, S::Integer)

Determines whether a given set S is a circuit in the matroid M, given by the
    closed sets of M. Uses (C.1) and (C.2) in Greene (1989).
"""
function is_circuit(M::ClosedSetsMatroid, S::Integer)
  t = Base.count_ones(S)

  for F in M.F[t] # (C.1) S ⊆ F for some F ∈ F_{t-1}.
    if S&F==S @goto C2 end
  end
  return false

  @label C2
  for F in M.F[t-1] # (C.2) |S ∩ F| ≤ r(F) for all F ∈ F_{t-2}.
    if Base.count_ones(S&F) > t-2 return false end
  end

  return true
end
```

Circuit oracle. ▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨▨

```
"""
    function minimal_spanning_subsets(M::ClosedSetsMatroid, A::Integer)

A modification of Algorithm 3.1 from Greene (1989) that finds all minimal
    spanning subsets of A ⊆ E, given a matroid M = (E, F). If A = E, this finds
    the bases of M.
"""
minimal_spanning_subsets(M::ClosedSetsMatroid, A::Integer) = _mss_all(M, 0, A)

function _mss_all(M::ClosedSetsMatroid, j::Integer, Ā::Integer)
  B = [Ā&F for F in M.F[j+1] if Base.count_ones(Ā&F) > j]

  while length(B) == 0
    if j >= Base.count_ones(Ā)-1 return Ā end

    j += 1
    B = [Ā&F for F in M.F[j+1] if Base.count_ones(Ā&F) > j]
  end

  bases = Set()
  t = reduce(|, B)
  while t > 0
    x = t&-t
    bases = bases ∪ _mss_all(M, j, Ā&~x)
    t &= ~x
  end
  return bases
end
```

Minimal spanning subsets and bases.

Chapter 6

Results

Chapter 7

Conclusions

7.1 Limitations

What types of user preferences are we unable to represent using MRFs? Ref. Halpern's introduction; MRFs can model substitutes, but not complementaries (left shoe worthless without right shoe).

7.2 Future work

1. Barman and Verma, in their 2020 paper on MMS-fair allocations under matroid-rank valuations, give a

Notes

1. Skrive mer om hvordan Set{Set{Integer}} lagres i minnet og fordelene med å gå over til Set{Integer}.

2. @Benchmarking. Histogrammer. Beskrive variansen i matroide-størrelse ifht input.

Bibliography

[1] N. Benabbou, M. Chakraborty, A. Igarashi, and Y. Zick, "Finding fair and efficient allocations when valuations don't add up," in *Algorithmic Game Theory*, Springer International Publishing, 2020, pp. 32–46. DOI: `10.1007/978-3-030-57980-7_3`. [Online]. Available: `https://doi.org/10.1007%2F978-3-030-57980-7_3`.

[2] R. J. Aumann and M. Maschler, "Game theoretic analysis of a bankruptcy problem from the Talmud," *Journal of Economic Theory*, vol. 36, no. 2, pp. 195–213, 1985, ISSN: 0022-0531. DOI: `https://doi.org/10.1016/0022-0531(85)90102-4`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/0022053185901024`.

[3] H. Steinhaus, "The problem of fair division," *Econometrica*, no. 16, pp. 101–104, 1948.

[4] G. Amanatidis *et al.*, *Fair division of indivisible goods: A survey*, 2022. arXiv: `2208.08782 [cs.GT]`.

[5] S. Barman and A. Biswas, *Fair division under cardinality constraints*, 2020. arXiv: `1804.09521 [cs.GT]`.

[6] S. Barman and P. Verma, *Existence and computation of maximin fair allocations under matroid-rank valuations*, 2021. arXiv: `2012.12710 [cs.GT]`.

[7] N. Benabbou, M. Chakraborty, A. Igarashi, and Y. Zick, "Finding fair and efficient allocations for matroid rank valuations," *ACM Trans. Econ. Comput.*, vol. 9, no. 4, Oct. 2021, ISSN: 2167-8375. DOI: `10.1145/3485006`. [Online]. Available: `https://doi.org/10.1145/3485006`.

[8] V. Viswanathan and Y. Zick, *Yankee swap: A fast and simple fair allocation mechanism for matroid rank valuations*, 2023. arXiv: `2206.08495 [cs.DS]`.

[9] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A fresh approach to numerical computing," *SIAM review*, vol. 59, no. 1, pp. 65–98, 2017. [Online]. Available: `https://doi.org/10.1137/141000671`.

[10] M. L. Hetland and H. Hummel, *Allocations.jl*, version 0.1, Nov. 2022. [Online]. Available: `https://github.com/mlhetland/Allocations.jl`.

[11] D. E. Knuth, "Random matroids," *Discrete Mathematics*, vol. 12, pp. 341–358, 4 1975.

[12] H. H. Crapo and G.-C. Rota, *On the foundations of combinatorial theory: Combinatorial geometries*. M.I.T. Press, 1970.

[13] H. Whitney, "On the abstract properties of linear dependence," *American Journal of Mathematics*, vol. 57, pp. 509–533, 3 Jul. 1935.

[14] A. Schrijver, *Combinatorial Optimization: Polyhedra and Efficiency*. Springer, 2003.

[15] R. J. Lipton, E. Markakis, E. Mossel, and A. Saberi, "On approximately fair allocations of indivisible goods," in *Proceedings of the 5th ACM Conference on Electronic Commerce*, ser. EC '04, New York, NY, USA: Association for Computing Machinery, 2004, pp. 125–131, ISBN: 1581137710. DOI: `10.1145/988772.988792`. [Online]. Available: `https://doi.org/10.1145/988772.988792`.

[16] I. Caragiannis, D. Kurokawa, H. Moulin, A. D. Procaccia, N. Shah, and J. Wang, "The unreasonable fairness of maximum nash welfare," *ACM Trans. Econ. Comput.*, vol. 7, no. 3, Sep. 2019, ISSN: 2167-8375. DOI: `10.1145/3355902`. [Online]. Available: `https://doi.org/10.1145/3355902`.

[17] B. Plaut and T. Roughgarden, *Almost envy-freeness with general valuations*, 2017. arXiv: `1707.04769 [cs.GT]`.

[18] E. Budish, "The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes," *Journal of Political Economy*, vol. 119, no. 6, pp. 1061–1103, Dec. 2011. DOI: `10.1086/664613`. [Online]. Available: `https://doi.org/10.1086/664613`.

[19] J. Edmonds, "Matroid partition," in *50 Years of Integer Programming 1958-2008*, Springer Berlin Heidelberg, Nov. 2009, pp. 199–217. DOI: `10.1007/978-3-540-68279-0_7`. [Online]. Available: `https://doi.org/10.1007/978-3-540-68279-0_7`.

[20] D. Knuth, *Matroid partitioning* (Report (Stanford University. Computer Science Department)). Computer Science Department, Stanford University, 1973.

[21] H. H. Crapo, "Erecting geometries," *Annals of the New York Academy of Sciences*, vol. 175, no. 1, pp. 89–92, Jul. 1970. DOI: `10.1111/j.1749-6632.1970.tb56458.x`. [Online]. Available: `https://doi.org/10.1111/j.1749-6632.1970.tb56458.x`.

[22] T. Greene, "Descriptively sufficient subcollections of flats in matroids," *Discrete Mathematics*, vol. 87, no. 2, pp. 149–161, 1991, ISSN: 0012-365X. DOI: `https://doi.org/10.1016/0012-365X(91)90044-3`. [Online]. Available: `https://www.sciencedirect.com/science/article/pii/0012365X91900443`.

[23] W. Suksompong, "Constraints in fair division," *SIGecom Exch.*, vol. 19, no. 2, pp. 46–61, Dec. 2021. DOI: `10.1145/3505156.3505162`. [Online]. Available: `https://doi.org/10.1145/3505156.3505162`.

[24] S. E. Fienberg, "A brief history of statistical models for network analysis and open challenges," *Journal of Computational and Graphical Statistics*, vol. 21, no. 4, pp. 825–839, 2012. DOI: `10.1080/10618600.2012.738106`. eprint: `https://doi.org/10.1080/10618600.2012.738106`. [Online]. Available: `https://doi.org/10.1080/10618600.2012.738106`.

[25] P. Erdős and A. Rényi, "On random graphs I.," *Publicationes Mathematicae Debrecen*, vol. 6, no. 3-4, pp. 290–297, Jul. 1959. DOI: `10.5486/pmd.1959.6.3-4.12`. [Online]. Available: `https://doi.org/10.5486/pmd.1959.6.3-4.12`.

[26] E. N. Gilbert, "Random Graphs," *The Annals of Mathematical Statistics*, vol. 30, no. 4, pp. 1141–1144, 1959. DOI: `10.1214/aoms/1177706098`. [Online]. Available: `https://doi.org/10.1214/aoms/1177706098`.

[27] S. Janson, D. E. Knuth, T. Łuczak, and B. Pittel, *The birth of the giant component*, 1993. arXiv: `math/9310236 [math.PR]`.

[28] R. Albert and A.-L. Barabási, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74, pp. 47–97, 1 Jan. 2002. DOI: `10.1103/RevModPhys.74.47`. [Online]. Available: `https://link.aps.org/doi/10.1103/RevModPhys.74.47`.

[29] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, no. 6684, pp. 440–442, Jun. 1998. DOI: `10.1038/30918`. [Online]. Available: `https://doi.org/10.1038/30918`.

[30] J. Fairbanks, M. Besançon, S. Simon, J. Hoffiman, N. Eubank, and S. Karpinski, *JuliaGraphs/Graphs.jl: An optimized graphs package for the julia programming language*, 2021. [Online]. Available: `https://github.com/JuliaGraphs/Graphs.jl/`.

[31] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to Algorithms* (The MIT Press), 3rd ed. London, England: MIT Press, Jul. 2009.

[32] N. Wirth and C. A. R. Hoare, "A contribution to the development of ALGOL," *Commun. ACM*, vol. 9, no. 6, pp. 413–432, Jun. 1966, ISSN: 0001-0782. DOI: `10.1145/365696.365702`. [Online]. Available: `https://doi.org/10.1145/365696.365702`.

[33] D. E. Knuth, *ERECTION.W*, `https://www-cs-faculty.stanford.edu/~knuth/programs/erection.w`, Mar. 2003. [Online]. Available: `https://www-cs-faculty.stanford.edu/~knuth/programs/erection.w`.

[34] R. Fourquet, *BitIntegers.jl*, https://github.com/rfourquet/BitIntegers.jl, Accessed: 2023-05-20.

Appendices

Appendix A

Sets as numbers – some useful tricks

Section* 5.4.2 details a number of steps taken in order to build a performant Julia implementation of Knuth-Matroid. Perhaps chief among these steps in terms of sheer performance gain compared to the initial, naïve implementation, was the transition from representing subsets of E as a `Set` of integers (or whatever type the elements of E might have), to representing them as a single integer, whose 1-bits denote which elements are in the set. This is possible as long as n is less than the widest available integer type (in off-the-shelf Julia, 128 bits, though one can go wider with the help of libraries [34]). We reiterate the bitwise equivalents of the basic set operations in Table A.1. These bitwise

Table A.1: Set operations and their equivalent bitwise operations

Set operation	Bitwise equivalent
$A \cap B$	$A \wedge B$
$A \cup B$	$A \vee B$
$A \setminus B$	$A \wedge \neg B$
$A \subseteq B$	$A \wedge B = A$

equivalents allow us to perform the set operations in constant time (right???),

resulting in significant performance increases. In the code snippets included throughout Section 5.4.2 and Appendix B, a number of "tricks" are performed with bitwise operations whose workings and purpose might be a bit obtuse. This appendix came to be as I got to grips with working with sets in this manner.

How do I...

...create a singleton set?

The left-shift operator ($<<$) can be used to set the ith bit to 1 and the others to 0. In general, $\{a\} = 1 << a$. This is used in an early version of GENERATE-COVERS:

```
function generate_covers_v2(F_r, n)
  Set([A | 1 << i for A ∈ F_r for i in 0:n-1 if A & 1 << i === 0])
end
```

...find the smallest element of a set?

Using the two's complement of a set T, denoted by $-T = \neg T + 1$, we can find the smallest element with the operation $T \wedge -T$. This is used in the next trick.

...enumerate all elements of a set one by one?

Using the previous trick, we can repeatedly pop the smallest element in the following manner:

```
t = 0b11111111
while t > 0
  x = t&-t  # x is the singleton set consisting of the smallest element of t
  output(x)
  t &= ~x   # t = t setminus x
end
```

This outputs all numbers from 1 to 0xff with a Hamming weight of 1.

...get a random element from a set?

We find all the positions at which the reversed bitstring of the set has a '1'
character, and choose a random one.

```
function rand_el(S::Integer)
  x = rand([2^(i-1) for (i,c) in enumerate(reverse(bitstring(S))) if c == '1'])
  return convert(typeof(S), x)
end
```

Appendix B

Code snippets

```
"""
Knuth's 1973 Matroid Partitioning algorithm for partitioning a set into subsets
independent in various given matroids.

Knuth's description: Given k matroids Ms = [M1, ..., Mk] on the same ground set
E, the algorithm finds a k-partitioning [S1, ..., Sk] of the elements of E such
that Sj is independent in matroid Mj, and nj <= |Sj| <= nj', for given limits
nj and nj'.

This implementation drops the upper limit nj' for each element j (implicitly it
is infinity for all matroids). Supply nj in array lims (lims[j] = nj).
"""
function knuth_partition(Ms, lims=nothing)
  n = Ms[1].n; k = length(Ms)
  S0 = Set(1:n) # The unallocated items.
  S = [Set() for _ in 1:k] # The partition-to-be.
  color = Dict(x=>0 for x in 1:n) # color[x] = j iff x ∈ S[j].
  for y in 1:k color[-y] = y end # -y is the 'standard' element of color y.
  succ = [0 for _ in 1:n]

  lims = lims === nothing ? [0 for _ in 1:k] : lims

  function augment(r)
    for x in 1:n succ[x] = 0 end

    A = Set(1:n)
    B = r > 0 ? Set(-r) : Set(-j for j in 1:k)

    while B != Set()
      C = Set()
      for y ∈ B for x ∈ A
```

```
      j = color[y]

      if is_indep(Ms[j], x ∪ setdiff(S[j], y))
        succ[x] = y
        A = setdiff(A, x)
        C = C ∪ x
        if color[x] == 0 repaint(x); return end
      end
    end end
    B = C
  end

  println("$A violates the condition of Theorem 3")
end

function repaint(x)
  while x ∈ 1:n
    y = succ[x]
    j = color[x]

    if j == 0 setdiff!(S0, x) else setdiff!(S[j], x) end

    j = color[y]
    S[j] = S[j] ∪ x
    color[x] = j
    x = y
  end
end

# Ensure every part gets at least its lower limit.
for j in 1:k for _ in 1:lims[j]
  augment(j)
end end

# Allocate the rest.
while S0 != Set()
  augment(0)
end

return S
end
```