Sampling Problems Associated with Matroids

CEng 461

Jakob Roberts & Jim Galloway

Matrix? Matroid? - An Introduction

matroid
/ˈmeɪtrɔɪd/

Matroid can be defined as $\mathcal{M}$ comprised of the ground set G and the collection of I subsets $\mathcal{M} = (\mathcal{E}, \mathcal{I})$. The concept of a Matroid was first described in 1935 by Hassler Whitney (mathematician) as a combinatorial generalization of linear independence of vectors.

A mathematical entity consisting of a finite set $\mathcal{E}$ together with a collection $\mathcal{I}$ of subsets of $\mathcal{E}$ such that: the empty set is a member of $\mathcal{I}$; any subset of a member of $\mathcal{I}$ is also a member of $\mathcal{I}$; and if two subsets $\mathcal{I}1$ and $\mathcal{I}2$ are in $\mathcal{I}$, where the cardinality of $\mathcal{I}1$ is less than that of $\mathcal{I}2$, then there exists an element e that is in $\mathcal{I}2$ but not $\mathcal{I}1$ such that the union of $\mathcal{I}1$ with $\{e\}$ is also an element of $\mathcal{I}$.

These three properties are the essence of a matroid and are described in further detail on the following slide.


MATROIDS
HISTORY.COM

Properties of Matroids

For a matroid $\mathcal{M} = (\mathcal{E}, \mathcal{I})$, there exists a finite collection of independent, finite sets ($\mathcal{I}$) that satisfies three axioms:

Non-emptiness property: $\varnothing \in I$

The empty set is in $\mathcal{I}$ (thus, $\mathcal{I}$ is not itself empty).

Heredity property: $X \in I \Rightarrow \wp(X) \in I$

If a set $\mathcal{X}$ is an element of $\mathcal{I}$, then every subset of $\mathcal{X}$ is also in $\mathcal{I}$.

Exchange property: $X \in I,\ Y \in I,\ |X| > |Y| \Rightarrow \exists\ x \in X \setminus Y \mid Y \cup \{x\} \in I$

If $\mathcal{X}$ and $\mathcal{Y}$ are two sets in $\mathcal{I}$ where the cardinality of $\mathcal{X}$ is greater than that of $\mathcal{Y}$, then there is an element x in $\mathcal{X}$ minus $\mathcal{Y}$ such that the union of $\mathcal{Y}$ and $\{x\}$ is in $\mathcal{I}$.

Sound Familiar Yet?

Matroid theory borrows much of its terminology from set theory, linear algebra, and graph theory.

i. The union of all sets in $\mathcal{M}$ is called the **ground set**.
set theory

ii. An independent set is called a **basis** if it is not a proper subset of another independent set.
linear algebra

iii. The **rank** of a subset $\mathcal{X}$ of the ground set is the size of the largest independent subset of X.
linear algebra

iv. A dependent set is called a **circuit** if every proper subset is independent.
graph theory

Yearning for Independence

Essentially, Matroids generalize the idea of independence. Whitney provided two axioms for his definition of independence, and stated that any structure adhering to these axioms is a matroid. These two axioms provided his abstract version of independence that can be applied to both graph theory and linear algebra

Whitney used two of the previous properties as his axioms. Both the heredity property and the exchange property were used to describe a matroid's independence. Whitney failed to include the 3rd axiom of non-emptiness as he simply assumed that the set of independent sets must at least contain one set that is the empty set.

Graphic Matroids

A graphic matroid (also called a cycle matroid or polygon matroid) generalizes the notion of an undirected graph.

Consider an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ and set of independent sets $\mathcal{I}$ such that $\mathcal{I} = \{ I : I \subseteq \mathcal{E}, I \text{ induces a forest in } \mathcal{G}\}$.

Then, $\mathcal{M} = (\mathcal{E}, \mathcal{I})$ is a matroid, and the set of independent sets $\mathcal{I}$ are the forests of $\mathcal{G}$. Some analogous properties of $\mathcal{M}$ and $\mathcal{G}$ are:

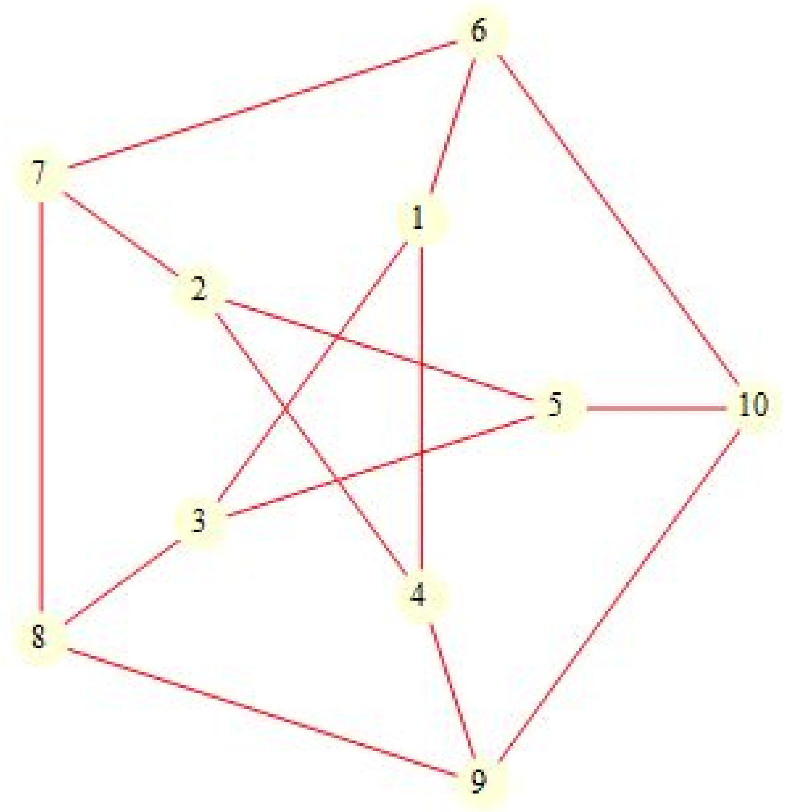
- the bases of a graphic matroid $\mathcal{M}(\mathcal{G})$ are the spanning forests of $\mathcal{G}$
- the circuits of $\mathcal{M}(\mathcal{G})$ are the simple cycles of $\mathcal{G}$
- the rank in $\mathcal{M}(\mathcal{G})$ of a set $\mathcal{X}$ of edges of a graph $\mathcal{G}$ is $r(\mathcal{X}) = n - c$ where n is the number of vertices in the subgraph formed by the edges in $\mathcal{X}$ and c is the number of connected components of the same subgraph

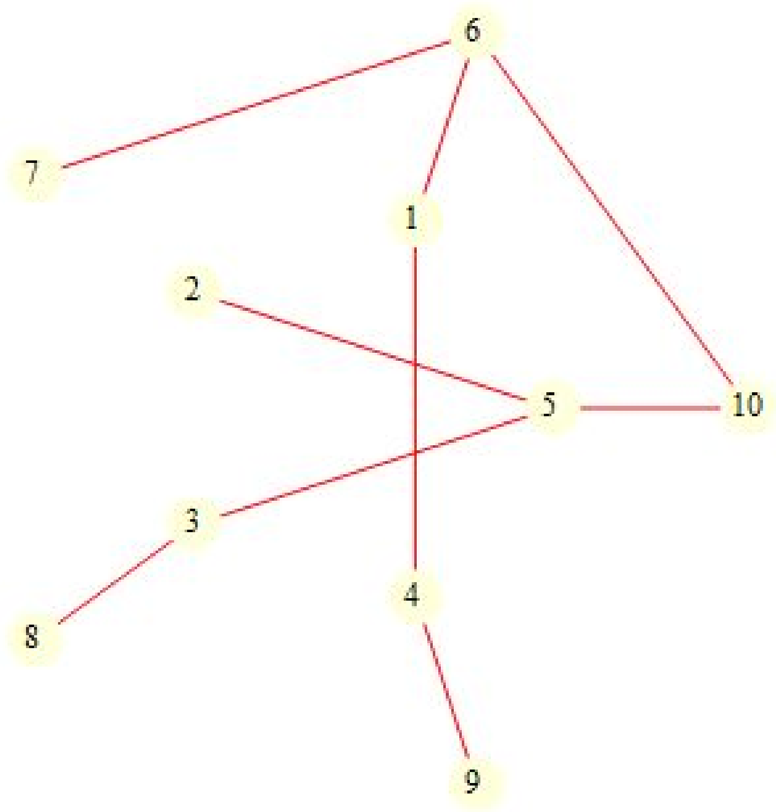
More Examples of Matroid Types & Algorithms

Vector Matroid
Graphic Matroid
Uniform Matroid
Partition Matroid
Laminar Matroid
Transversal Matroid
Matching Matroid
Uniform Matroids

Graphic/cycle matroid
Cographic/cocycle matroid
Disjoint path matroid
Matroids from Field Extensions
Infinite Matroids
The Greedy Algorithm
Matroid Partitioning
Matroid Intersection

Petersen Graph (and a possible spanning tree)





Random Sampling's Complications

An important part of statistics is determining a proper representative sample of the data to examine a small portion in order to gather data about a larger set.

Random sampling is a powerful tool to gather a small representative sample.

- The chief complication is determining an appropriate representative sample that will not skew the results of the overall sample space.

For some problems involving matroids and using a random sample, we will be talking **specifically about a matroid in a graphical representation where the edges connecting the graph are weighted**. This weighted quality allows for these matroids to be *optimized*.

Random Sampling and Matroid Optimization

The objective is to find the optimum basis of the set $\mathcal{S}$: denoted **OPT**($\mathcal{S}$), where $\mathcal{S} \subseteq \mathcal{M}$, and $\mathcal{M}$ is an m-element matroid with rank r. Through assuming that all element weights in $\mathcal{M}$ are unique, we can say that the optimum base of $\mathcal{M}$ is also unique.

The concept to optimize is that if an element is added to an independent set that decreases the cost or increase the size of the set, it is considered an improvement. If $\mathcal{T}$ is an independent set of $\mathcal{G}$, adding element e improves $\mathcal{T}$ if and only if the elements of $\mathcal{T}$ weighing less than e do not span e.

Taking into account sampling, if we take a random submatroid of $\mathcal{M}$ that is constructed by including each element of $\mathcal{M}$ independently with a given probability, we can determine that the optimum of the submatroid is the inverse of the probability of the inclusion of the independent elements in the submatroid.

Random Sampling to Find a Packing Number

The packing number of a matroid, $\mathbf{P}(\mathcal{M}) = \boldsymbol{k}$, describes the number of disjoint bases in $\mathcal{M}$.

Random sampling significantly decreases the running time of algorithms that find $\mathbf{P}(\mathcal{M})$ given $\mathcal{M}$ as input.

The improvement is drastic:

Knuth algorithm: $O(mr^2k^2)$

Random sampling: $O(r^3 \ln(r))$

m = matroid size

r = matroid rank

Summary & Conclusion

Matroids are defined as a set of elements, and a set of subsets of the set of elements.

Matroids are mathematical structures that generalize to other mathematical structures.

Problems on matroids can be applied to problems on other domains.

Random sampling improves optimization algorithms on weighted matroids.

Random sampling improves algorithms that find disjoint bases in a matroid.

References

http://jeffe.cs.illinois.edu/teaching/algorithms/notes/08-matroids.pdf
https://www.math.lsu.edu/~oxley/survey4.pdf
https://en.oxforddictionaries.com/definition/matroid
https://www.math.washington.edu/~morrow/336_09/papers/Will.pdf
https://courses.engr.illinois.edu/cs598csc/sp2010/Lectures/Lecture14.pdf
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.63.7731&rep=rep1&type=pdf
http://www.cs.tau.ac.il/~azar/matr.pdf
https://www.computer.org/csdl/proceedings/focs/1993/4370/00/0366879.pdf
http://people.csail.mit.edu/karger/Papers/matroid.pdf