

Masterarbeiten am Lehrstuhl 2

Prof. Dr. Amin Coja-Oghlan
amin.coja-oghlan@tu-dortmund.de

28. September 2021

Effiziente Algorithmen bilden das Rückgrat aller IT-Systeme und gute Algorithmiker sind dementsprechend heiß begehrt. Am LS2 befassen wir uns mit der Entwicklung und Analyse von Algorithmen, mit Anwendungen in den verschiedensten Bereichen, und mit den kombinatorischen Grundlagen der Algorithmik. Wenn Sie sich für eine Masterarbeit bei uns interessieren, sollten Sie Interesse am Tüfteln und an einer analytischen Arbeitsweise mitbringen. Einige Themenvorschläge aus den Bereichen

- *Probabilistische Kombinatorik,*
- *Theorie des maschinellen Lernen und Data Science,*
- *Epidemiologie*

sind im folgenden aufgelistet. Sprechen Sie uns aber auch gern persönlich an oder kontaktieren Sie uns per email, denn viele Masterthemen sind auf die jeweiligen individuellen Interessen zugeschnitten.

Probabilistische Kombinatorik und Algorithmen

Algorithmen nutzen immer strukturelle Eigenschaften des in Frage stehenden Problems aus. Die Algorithmik ist daher eng mit der Mathematik, insbesondere der Kombinatorik verbunden, die sich mit diskreten Objekten wie Graphen, Netzwerken oder Codes befaßt. In jüngster Zeit spielen dabei Techniken, die den Zufall nutzen, eine entscheidende Rolle. Dies führt auf das Studium zufälliger diskreter Strukturen wie zufälliger Graphen, Matrizen oder Codes.

Thema: Das Rekonstruktionsproblem

In dem bekannten Spiel “Stille Post” flüstert ein Spieler dem nächsten eine Nachricht zu. Die Frage ist, wieviel die Nachricht, die beim letzten Spieler ankommt, noch mit der ursprünglichen Nachricht zu tun hat. Dieses Spiel ist ein einfacher Spezialfall des Rekonstruktionsproblems. Anstelle einer einfachen Kette von Spielern haben wir es im allgemeinen Rekonstruktionsproblem mit einer Baumstruktur zu tun, in dem beispielsweise jeder Spieler zwei verschiedenen weiteren Spielern, seinen Nachkommen in dem Baum, die Nachricht zuflüstert. Die Aufgabe besteht dann darin, aus *allen* Nachrichten, die die Blätter des Baums empfangen, die ursprüngliche Nachricht zu rekonstruieren. Das Rekonstruktionsproblem spielt eine extrem wichtige Rolle in der statistischen Physik, der Kombinatorik aber auch der Bioinformatik (Rekonstruktion von Erbfolgen, “phylogenetische Bäume”). Ziel dieses Projektes ist es, anhand konkreten Nachrichtenmodelle das Rekonstruktionsproblem theoretisch zu untersuchen.

Referenzen:

- E. Mossel: Reconstruction on trees: beating the second eigenvalue. [\[link\]](#)
- E. Mossel, Y. Peres: Information flow on trees. [\[link\]](#)

Thema: Algorithmische Regularität

Das Regularitätslemma ist ein zentrales Ergebnis der Kombinatorik. Es sagt aus, daß auch sehr große, komplizierte kombinatorische Strukturen immer als Überlagerung einer kleinen Zahl sehr regelmäßig strukturierter Strukturen beschrieben werden können. Aus algorithmischer Sicht stellt sich die Frage, wie eine solche reguläre Zerlegung möglichst effizient bestimmt werden kann und wie diese Zerlegung algorithmisch genutzt werden kann. In diesem Projekt geht es um die Entwicklung entsprechender Algorithmen für verschiedene kombinatorische Strukturen wie Graphen, Hypergraphen oder Wahrscheinlichkeitsverteilungen.

Referenzen:

- Abelpreis E. Szemerédi. [\[link\]](#)
- V. Rödl, M. Schacht: Regularity lemmas for graphs. [\[link\]](#)

Thema: Hashing, Codes und Datenkompression

Der Zufall kann als wichtiges Hilfsmittel bei der Konstruktion fehlerkorrigierender Codes herangezogen werden. Derartige Codes spielen eine wichtige Rolle in modernen Kommunikationsstandards ("low-density parity check codes"). Derartige Konstruktionen stellen uns vor zwei Herausforderungen, nämlich einerseits die informationstheoretische Frage, was probabilistische Konstruktion prinzipiell erreichen können, und andererseits die algorithmische Frage, inwiefern sich dieses Potential auch mit effizienten Algorithmen (z.B. Decodierungsalgorithmen) heben läßt. Neben Codes im engeren Sinne sind auch Anwendungen im Bereich Hashing und Datenkompression von Interesse.

Referenzen:

- D. MacKay: Information theory, Inference and Learning Algorithms. [\[link\]](#)
- M. Dietzfelbinger, A. Goerdt, M. Mitzenmacher, A. Montanari, R. Pagh, M. Rink: Tight thresholds for cuckoo hashing via XORSAT. [\[link\]](#)

Thema: Spektralalgorithmen

Eine wichtige Klasse von Algorithmen, sog. Spektralalgorithmen, verwenden die Eigenwerte und Eigenvektoren von geschickt konstruierten Matrizen, um kombinatorische Probleme zu lösen. Eine neue Konstruktion, mit der ein Netzwerk durch eine Matrix dargestellt werden kann, ist die *Hashimoto-Matrix*. In diesem Projekt geht es darum, den Nutzen dieser Matrix für verschiedene kombinatorische Aufgabenstellungen zu erforschen. Ein besonderer Fokus liegt dabei auf der Verbindung zwischen der Hashimoto-Matrix und dem Belief Propagation-Algorithmus.

Referenzen:

- F. Krzakala, C. Moore, E. Mossel, J. Neeman, A. Sly, L. Zdeborova, P. Zhang: Spectral redemption: clustering sparse networks. [\[link\]](#)

- C. Bordenave, M. Lelarge, L. Massoulié: Non-backtracking spectrum of random graphs: community detection and non-regular Ramanujan graphs. [\[link\]](#)

Thema: Algorithmen für zufällige Erfüllbarkeitsprobleme

Das Erfüllbarkeitsproblem ('SAT') ist eines der wichtigsten algorithmischen Probleme überhaupt, weil es als Unterproblem in einer Vielzahl von Anwendungen begegnet (z.B. Datenbankabfragen, Korrektheit von Programmen). Dabei geht es darum, für eine gegebene aussagenlogische Formel eine Belegung der Variablen zu finden, die die Formel erfüllt. Dieses Problem ist NP-vollständig, d.h. es gibt sehr wahrscheinlich keinen effizienten Algorithmus, der dieses Problem immer löst. Durch diese Erkenntnis verschwindet das SAT-Problem aber natürlich nicht von der Bildfläche. Stattdessen stellt sich die Aufgabe, dem SAT-Problem mit möglichst guten Heuristiken beizukommen. Zur Evaluation solcher Heuristiken werden häufig zufällig erzeugte Instanzen des Erfüllbarkeitsproblems herangezogen. Um solche zufällige Instanzen und ihre algorithmische Lösbarkeit geht es in diesem Projekt.

Referenzen:

- A. Coja-Oghlan: A better algorithm for random k -SAT. [\[link\]](#)
- A. Braunstein, M. Mezard, R. Zecchina: Survey propagation: an algorithm for satisfiability. [\[link\]](#).

Thema: Lokale Grenzwertsätze

Eine der wichtigsten Erkenntnisse in der Wahrscheinlichkeitstheorie ist, daß Summen unabhängiger Zufallsvariablen und sehr milden Annahmen gegen die Normalverteilung konvergieren. Diese Aussage ist bekannt als der "zentrale Grenzwertsatz". In der Kombinatorik hat man es dabei häufig mit Zufallsvariablen zu tun, die irgendetwas zählen. Für solche Zufallsvariablen gilt manchmal eine noch genauere Aussage, nämlich ein "lokaler Grenzwertsatz". In diesem Projekt soll es darum gehen, eine möglichst allgemeine Version des lokalen Grenzwertsatzes herzuleiten, die auf Anwendungen in der Kombinatorik zugeschnitten ist.

Referenzen:

- Wikipedia-Artikel. [\[link\]](#)
- Encyclopedia of Mathematics. [\[link\]](#)

Thema: Das Potts-Modell

Das Potts-Modell ist ein einfaches Modell aus der Physik zur Beschreibung des Ferromagnetismus. Die Idee ist, daß Magnetismus aus dem Zusammenspiel vieler Elementarteilchen entsteht, von denen jedes eine endliche Zahl von magnetischen Ausrichtungen annehmen kann. Benachbarte Teilchen wechselwirken, so daß es energetisch günstiger ist, wenn sie die gleich magnetische Ausrichtung besitzen. In Abhängigkeit von der Temperatur ergibt sich somit ein Wechselspiel aus Entropie (Anzahl der Möglichkeiten) und Energie. In diesem Projekt geht es darum, das Potts-Modell sowohl algorithmisch als auch strukturell zu analysieren. Hilfsmittel sind dabei u.a. der *Belief Propagation*-Algorithmus sowie Algorithmen, die auf Markovketten beruhen.

Referenzen:

- D. Welsh, C. Merino: The Potts model and the Tutte polynomial. [\[link\]](#)
- M. Bordewich, C. Greenhill, V. Patel: Mixing of the Glauber dynamics for the ferromagnetic Potts model. [\[link\]](#)

Maschinelles Lernen und Data Science

Aufgrund spektakulärer Erfolge wie AlphaGo ist in den letzten Jahren das maschinelle Lernen wieder ins Zentrum der Informatikforschung gerückt. Auch der Umgang mit großen Datenmengen spielt eine wichtige Rolle an der Schnittstelle zwischen Statistik und Informatik. Algorithmen spielen dabei natürlich eine fundamentale Rolle. Ziel der folgenden Themen ist eine analytische Auseinandersetzung mit Algorithmen aus dem maschinellen Lernen und aus dem Gebiet Data Science.

Thema: Analyse von neuronalen Netzwerkmodellen

Selbst relative einfache Modelle neuronaler Netzwerke wie z.B. das Hopfield-Netzwerk oder *restricted Boltzmann machines* stellen uns in Hinblick auf eine exakte Analyse vor Herausforderungen. Dabei geht es zum einen um das Verständnis informationstheoretischer Eigenschaften (z.B. wieviel Trainingsdaten notwendig sind, wie hoch die Speicherkapazität des Netzwerks ist), aber auch um algorithmische Probleme wie z.B. die Analyse von Lernalgorithmen. In diesem Projekt geht es darum, Fortschritte beim analytischen Verständnis solcher Netzwerkmodelle zu machen.

Referenzen:

- G. Hinton: A practical guide to training restricted Boltzmann machines. [\[link\]](#)
- M. Gabriele, E. Tramel, F. Krzakala: Training restricted Boltzmann machines via the Thouless-Anderson-Palmer free energy. [\[link\]](#)

Thema: Das stochastische Blockmodell

Das stochastische Blockmodell ist ein einfaches aber wichtiges Modell zur Beschreibung strukturierter Netzwerkdaten. Die gegebenen Daten beschreiben dabei einfach ein Netzwerk, ohne weitere Zusatzinformation. Die Aufgabenstellung besteht darin, die versteckte Struktur des Netzwerkes, z.B. aus welchen "communities" das Netzwerk besteht, zu lernen. Obwohl es in den letzten Jahren hat es zahlreiche Forschungsbeiträge zu diesem Modell gegeben hat, sind einige der wichtigsten Fragestellungen noch nicht genau beantwortet. In diesem Projekt geht es darum, verschiedene informationstheoretische und algorithmische Ansätze auf diesen Modell anzuwenden und zu verbessern.

Referenzen:

- C. Moore: The computer science and physics of community detection: landscapes, phase transitions, and hardness. [\[link\]](#)
- E. Abbe: Community detection and stochastic block models: recent developments. [\[link\]](#)

Thema: Message Passing-Algorithmen

In jüngster Zeit hat eine neue Klasse von Algorithmen Aufsehen erregt, die auf der Idee des Nachrichtenaustausches beruht. Dabei werden Daten durch ein Netzwerk modelliert, dessen Knoten

entlang der Kanten des Netzwerkes Nachrichten austauschen, bis ein Fixpunkt erreicht ist. Aus den Nachrichten, gegen die das Verfahren konvergiert, wird dann die Lösung des Lernproblems gewonnen. Beispiele derartiger Algorithmen sind *Belief Propagation* oder *Approximate Message Passing*. In diesem Projekt geht es um ein verbessertes analytisches Verständnis solcher Algorithmen.

Referenzen:

- Wikipedia-Artikel [\[link\]](#)
- O. Feng, R. Venkataramanan, C. Rush, R. Samworth: A unifying tutorial on approximate message passing. [\[link\]](#)

Epidemiologie

In der Covid-Pandemie boten computerbasierte (und damit algorithmenbasierte) Modelle der Ausbreitung des Virus eine wichtige Entscheidungsgrundlage für Eindämmungsmaßnahmen. Bei den folgenden Themen geht es um eine kritische Auseinandersetzung mit solchen Modellen und ihren mathematischen Grundlagen. Welche Vorgänge können die verschiedenen Modelle abbilden? Erlauben sie eine nachvollziehbare Erklärung der Vorhersagen? Wo gibt es Verbesserungsmöglichkeiten?

Thema: Das SIR-Modell auf Netzwerken

Einige wichtige epidemiologische Modelle beruhen auf Kontaktnetzwerken. Zur vereinfachten Modellierung wird dabei manchmal angenommen, daß das zugrundeliegende Netzwerk durch ein geeignetes Zufallsmodell beschrieben werden kann. Dies führt dann natürlich auf die Herausforderung, einen epidemischen Prozeß auf einem solchen Zufallsnetzwerk auch zu analysieren. Das bekannteste Beispiel eines solchen Prozesses ist das SIR-Modell. Dabei ist jeder Netzwerkknoten entweder im Zustand “susceptible” (empfindlich für Ansteckung), “infected” (bereits angesteckt) oder “recovered” (genesen). Die Aufgabenstellung besteht darin, die zeitliche Evolution dieses Prozesses auf dem zufälligen Netzwerk und die treibenden Kräfte dahinter zu verstehen.

Referenzen:

- S. Janson, M. Luczak, P. Windridge: Law of large numbers for the SIR epidemic on a random graph with given degrees. [\[link\]](#)
- C. Fransson, P. Trapman: SIR epidemics and vaccination on random graphs with clustering. [\[link\]](#)

Thema: First passage percolation

Ein weiteres vereinfachtes Modell von Ansteckungsprozessen auf Netzwerken ist der sog. “first passage percolation”-Prozeß. Hier geht es, grob gesprochen, darum, wie sich eine erste Welle einer Ansteckung in einem Netzwerk völlig ungeschützter Knoten ausbreitet. Ziel dieses Projektes ist es, diesen Prozeß auf einfachen aber aussagekräftigen Netzwerkmodellen zu untersuchen.

Referenzen:

- S. Dereich, M. Ortgiese: Local neighbourhoods for first passage percolation on the configuration model. [\[link\]](#)
- S. Bhamidi, R. van der Hofstad, G. Hooghiemstra: First passage percolation on random graphs with finite mean degrees. [\[link\]](#)