Färben von zufälligen Schnittgraphen

Diplomarbeit

Michael Ueckerdt

Betreuung: Prof. Dr. Anusch Taraz, Michael Behrisch

Humboldt Universität zu Berlin
Institut für Informatik
Arbeitsgruppe Algorithmen

24. Juli 2005

Erklärung

Ich erkläre hiermit, die vorliegende Arbeit selbst verfasst und keine anderen Hilfsmittel als die angegebenen verwendet zu haben. Ich bin damit einverstanden, dass die vorliegende Arbeit in der Bibliothek des Instituts für Informatik der Humboldt Universität zu Berlin öffentlich ausgelegt wird.

Berlin, den 24. Juli 2005

Michael Ueckerdt

Danksagung

Ich möchte meinen beiden Betreuern Prof. Dr. Anusch Taraz und Michael Behrisch für die vielen wertvollen Hinweise und Denkanstöße sowie die konstruktive Kritik danken.

Weiterer Dank gebührt meinen Eltern, die mich unermüdlich zur Fertigstellung meiner Arbeit anhielten und auch das Korrekturlesen der Arbeit übernahmen.

Inhaltsverzeichnis

1 Einleitung

In der vorliegenden Arbeit wird das Verhalten von Färbungsalgorithmen auf komplexen Netzwerken auf Grundlage des Modells der zufälligen Schnittgraphen untersucht. Für das verwendete Zufallsmodell werden Optimalität der Algorithmen bzw. Aussagen zu ihrer Approximationsgüte bewiesen.

Eine Zuordnung von Farben zu den Knoten eines Graphen heißt Färbung, wenn je zwei durch eine Kante verbundenen Knoten unterschiedliche Farben zugeordnet werden. In der Graphentheorie werden diese Farben üblicherweise durch natürliche Zahlen repräsentiert. Eine solche Färbung ist sehr leicht zu konstruieren indem beispielsweise jedem Knoten eine eigene Farbe zugeordnet wird. In der Graphentheorie besteht jedoch das Ziel darin, mit möglichst wenig Farben auszukommen. Die Anzahl der minimal benötigten Farben wird chromatische Zahl eines Graphen genannt. Das Bestimmen einer optimalen Färbung, d.h. einer Färbung, die mit der minimal möglichen Anzahl von Farben auskommt, wird als das Färbeproblem bezeichnet.

Eine klassische Anwendung besteht im Einfärben einer politischen Weltkarte, so dass benachbarte Länder unterschiedliche Farben bekommen. Dieses Problem kann auf das Färben von Graphen zurückgeführt werden. Dazu fasst man die Karte als Graphen auf, dessen Knoten die Länder repräsentieren und in dem zwei Knoten verbunden sind, wenn die dazugehörigen Länder eine gemeinsame Grenze besitzen. Im Rahmen der Graphentheorie konnte gezeigt werden, dass solche Karten immer mit vier Farben gefärbt werden können [1].

Neben dem Einfärben von Karten gibt es eine Reihe weiterer Anwendungen für Färbungen. Ein populäres Beispiel ist die Zuordnung von Frequenzen zu Sendern (z.B. Radiostationen), so dass Sender mit sich überlappendem Einzugsbereich unterschiedliche Frequenzen bekommen. Dazu betrachtet man die Sender als Knoten und verbindet je zwei Knoten/Sender wenn sich ihr Sendebereich überschneidet. Eine Färbung eines solchen Graphen liefert eine Frequenzzuordnung, wenn die Farben als Frequenzen interpretiert werden.

Ein weiteres Beispiel ist die Zuordnung von Variablen zu Hardware-Registern (einer CPU) z.B. im Rahmen der Compiler-Optimierung. Der Zugriff auf Variablen in Registern ist normalerweise deutlich schneller als der Zugriff auf andere Variablen. Da jedoch die Anzahl der verfügbaren Register häufig sehr viel kleiner als die Anzahl der verwendeten Variablen ist, muss ein Register verschiedene Variablen nacheinander aufnehmen. Zwei Variablen sollen außerdem wenn möglich gleichzeitig in verschiedenen Registern vorliegen, wenn die eine kurz vor und nach der anderen benötigt wird. Wieder lässt sich ein Graph konstruieren, dessen Färbung eine Lösung für das Problem liefert. Dazu interpretiert man die Variablen als Knoten eines Graphen und verbindet zwei Knoten/Variablen durch eine Kante, wenn die eine kurz vor und nach der anderen benötigt wird. Eine Färbung kann nun als Zuordnung von Registern/Farben zu den Variablen/Knoten aufgefasst werden.

Als letztes Beispiel sei das Testen von Leiterplatten auf Kurzschlüsse durch Ungenauigkeiten in den Leiterbahnen angeführt. Dazu können die Leiterbahnen in einzelne Schaltkreise eingeteilt werden, die dann die Knoten eines Graphen bilden. Zwei Schaltkreise/Knoten werden durch eine Kante verbunden, wenn sie nicht unabhängig voneinander getestet werden können (d.h. gemeinsame Leiterbahnen besitzen oder nah genug zusammenliegen, dass ein Kurzschluss möglich ist). Ziel ist es, den Testprozess durch

Parallelisierung der Tests zu beschleunigen. Über eine Färbung lässt sich bestimmen, welche Schaltkreise gleichzeitig auf Kurzschlüsse überprüft werden können. Die Anzahl der Farben entspricht dann der Anzahl der nacheinander auszuführenden Tests, wobei in einem Test alle Schaltkreise der entsprechenden Farbe parallel getestet werden können [2].

1.1 Resultate zum Färben aus der Graphentheorie

Das Finden einer optimalen Färbung für Graphen ist ein NP-vollständiges Problem [3]. D.h. wenn $P \neq NP$ gilt, gibt es keinen polynomiellen Algorithmus der dieses Problem für beliebige Graphen löst. Ein Algorithmus heißt polynomiell, wenn seine Worst-Case Laufzeit durch ein Polynom über der Größe der Probleminstanz abgeschätzt werden kann. Beim Färbeproblem wird üblicherweise die Anzahl der Knoten des zu färbenden Graphen als Größe angesetzt. Die Suche nach schnellen Algorithmen zum Finden einer optimalen Färbung erscheint damit wenig erfolgversprechend.

Wie kann also das Färben von Graphen in der Praxis angegangen werden? Ein erster Ansatz ist der Umstieg auf Approximationsalgorithmen. Man sucht nach einer guten aber nicht mehr unbedingt optimalen Färbung. Allerdings konnten Lund und Yannakakis 1993 zeigen, dass es ein $\epsilon > 0$ gibt, so dass auch das Berechnen einer Färbung mit nicht mehr als $n^\epsilon \chi$ Farben NP-vollständig ist, wobei n die Anzahl der Knoten des Graphen und χ dessen chromatische Zahl ist [3]. Der beste bekannte Approximationsalgorithmus garantiert eine Approximationsgüte von $O\left(n\frac{(\log\log n)^2}{(\log n)^3}\right)$, d.h. es gibt eine Konstante c, so dass maximal $cn\frac{(\log\log n)^2}{(\log n)^3}\chi$ Farben benötigt werden [4]. Gute Approximationen von Färbungen sind daher genauso schwierig zu finden wie eine optimale Färbung.

Es gibt jedoch auch Graphenklassen, auf denen das Färbeproblem in polynomieller Zeit lösbar ist. So lässt sich beispielsweise sehr einfach zeigen, dass alle Graphen ohne Kreise ungerader Länge in linearer Zeit mit maximal zwei Farben gefärbt werden können [3]. Diese Färbung ist optimal, da jeder Graph mit mindestens einer Kante, eine chromatische Zahl größer oder gleich zwei hat. Für $\chi \leq 2$ ist das Färbeproblem folglich in linearer Zeit lösbar. Umgekehrt lässt sich ebenfalls zeigen, dass zum Färben eines Graphen mit einem Kreis ungerader Länge mindestens drei Farben benötigt werden [3].

Planare Graphen bilden eine weitere Klasse von Graphen, zu denen interessante Aussagen bezüglich des Färbens bekannt sind. Ein Graph heißt planar, wenn er in die Ebene eingebettet werden kann. D.h. für planare Graphen gibt es immer eine Darstellung in der Ebene, in der sich seine Kanten nicht kreuzen. Die oben angesprochene Graphendarstellung einer politischen Karte liefert immer einen planaren Graphen. Im Jahre 1976 wurde bewiesen, dass alle planaren Graphen mit vier Farben gefärbt werden können [1]. Dennoch bleibt das Konstruieren einer solchen Färbung schwierig (NP-vollständig) [3]. Bereits Ende des 19. Jahrhundert wurde allerdings ein Algorithmus mit quadratischer Laufzeit gefunden, der planare Graphen mit maximal fünf Farben färbt [3]. Das Färbeproblem kann damit auf planaren Graphen deutlich besser approximiert werden als im allgemeinen Fall.

Abschließend soll noch die Klasse der perfekten Graphen erwähnt werden. Ein Graph ist perfekt, wenn für den Graphen selbst und alle Graphen, die durch das Herausnehmen von Knoten (und der mit ihnen verbundenen Kanten) entstehen, die Cliquenzahl gleich der chromatischen Zahl ist. Dabei ist eine Clique eine Menge von Knoten, bei der jeder Knoten mit jedem anderen durch eine Kante verbunden ist. Die Cliquenzahl eines Graphen beschreibt die Größe der größten Clique des Graphen. Offensichtlich ist die Cliquenzahl eine untere Schranke für die chromatische Zahl, da zum Färben einer Clique jeder ihrer Knoten eine andere Farbe bekommen muss. Diese Schranke kann jedoch beliebig schlecht werden. So gab unter anderem Mycielsky ein Verfahren zum Konstruieren von Graphen mit Cliquenzahl zwei aber einer beliebigen chromatischen Zahl größer als zwei an [3]. Im Falle der perfekten Graphen ist diese Schranke jedoch scharf. In der Klasse der perfekten Graphen werden das Färbeproblem und einige andere NP-vollständige Optimierungsprobleme zugänglich, d.h in polynomieller Zeit lösbar [3].

1.2 Zufällige Graphen

Das Konstruieren einer optimalen Färbung oder einer guten Approximation für einen beliebigen Graphen ist schwierig. Eine Möglichkeit dennoch optimale oder gute Färbungen zu erzeugen, liegt in der Einschränkung der Betrachtung auf spezielle Graphenklassen. Eine andere Möglichkeit an das Färbeproblem heranzugehen ist die Betrachtung zufälliger Graphen und der erwarteten Laufzeit, oder auch der erwarteten Approximationsgüte von Algorithmen darauf. Möglicherweise sind Algorithmen mit exponentieller Worst-Case Laufzeit bzw. sehr schlechter Approximationsgüte nur für einen kleinen Teil der Probleminstanzen wirklich so langsam oder so schlecht und damit trotzdem in der Praxis brauchbar. Um Algorithmen daraufhin zu untersuchen darf die Analyse jedoch nicht auf die Betrachtung des Worst-Case beschränkt sein.

Das vermutlich einfachste und am besten verstandene Modell zufälliger Graphen wurde von P. Erdös und A. Renyi entwickelt. Dabei wird ausgehend von einem kantenlosen Graphen auf n Knoten für jedes Knotenpaar mit Wahrscheinlichkeit p eine Kante zwischen diesen Knoten eingefügt. Im Folgenden wird dieses Modell als das Modell der klassischen zufälligen Graphen bezeichnet und als $G_{n,p}$ notiert. Für $p = \frac{1}{2}$ entsprechen die klassischen zufälligen Graphen in einem gewissen Sinne der Gleichverteilung auf der Menge der Graphen, da jeder mögliche Graph mit derselben Wahrscheinlichkeit erzeugt wird, wenn man Graphen, die durch Vertauschen bzw. Umbenennen der Knoten ineinander überführt werden können, als verschiedene Graphen versteht.

Für dieses Modell zufälliger Graphen gibt es einen einfachen polynomiellen Färbealgorithmus, der für ein großes Spektrum von Kantenwahrscheinlichkeiten p fast immer weniger als $(2+\epsilon)\chi$ Farben benötigt (für ein beliebiges $\epsilon > 0$). Das genaue Resultat inklusive des verwendeten Algorithmus ist in Kapitel 5 angegeben. Während also die oben aufgeführten Worst-Case-Resultate nahelegen, dass das Färbeproblem in der Praxis nur für sehr kleine Graphen bzw. spezielle Graphenklassen gelöst oder wenigstens gut approximiert werden kann, liefert die Untersuchung auf zufälligen Graphen ein anderes Bild.

Die praktische Relevanz dieses Resultats hängt davon ab, wie stark sich die Verteilung der echten Eingabegraphen von der Verteilung der klassischen zufälligen Graphen unterscheidet.

1.3 Zufällige Schnittgraphen

Lange Zeit war das $G_{n,p}$-Modell das einzig verfügbare um komplexe Graphen zu untersuchen [5]. Analysen von verschiedensten Netzwerken aus der realen Welt zeigten jedoch häufig strukturelle Unterschiede zwischen typischen klassischen Zufallsgraphen und den untersuchten realen Graphen auf. Eine Zusammenfassung zur Analyse verschiedener realer Graphen ist in [6] zu finden. Bei der Suche nach gemeinsamen Eigenschaften dieser realen Graphen kristallisierten sich vor allem die folgenden drei Konzepte heraus (siehe auch [6]):

1. kurze durchschnittliche Pfadlänge (englisch: small worlds)
   Dieses Merkmal beschreibt die Tatsache, dass reale Graphen trotz ihrer Größe häufig vergleichsweise kurze Verbindungen zwischen zwei beliebig gewählten Knoten aufweisen. Beispielsweise wurde 1999 für die Graphenrepräsentation des WWW ausgehend von der Analyse eines Ausschnitts die durchschnittliche Pfadlänge zwischen zwei Knoten auf 19 geschätzt (bei ca. 800 Millionen Knoten). Dabei bilden die Web-Seiten die Knoten des Graphen und zwei Knoten des Graphen sind durch eine Kante verbunden, wenn eine der beiden Seiten auf die andere via Hyperlink verweist.

2. Zusammenballung von Kanten (englisch: clustering)
   In realen Graphen wird häufig eine Tendenz zur Zusammenballung von Kanten beobachtet. D.h. die Verteilung der Kanten in realen Graphen ist nur selten gleichförmig. Vielmehr liegt ein Großteil der Kanten in lokalen Anhäufungen vor. Für den WWW-Graphen lässt sich diese Tendenz erklären, wenn man die Ursache der Kanten, d.h. den Grund für Links berücksichtigt. Häufig verweisen Seiten aufeinander, die sich mit dem selben Thema beschäftigen oder einen gemeinsamen Autor haben. Beim Aufbau neuer Seiten ist nun die Wahrscheinlichkeit wesentlich höher, dass auf eine andere Seiten desselben Themas verwiesen wird, als die Wahrscheinlichkeit eines Verweises auf eine beliebig gewählte Seite. Auf diese Weise entstehen lokale Kantenanhäufungen zwischen Knoten/Webseiten mit gleichen oder verwandten Themen. Die Stärke der Zusammenballung kann mit Hilfe des Clustering-Koeffizienten (englisch: clustering coefficient) ermittelt werden. Eine Definition sowie genaue Betrachtung dieser Größe findet sich in [6]. Hier genügt die Aussage, dass ein höherer Clustering-Koeffizient eine stärkere Tendenz zur Cluster-Bildung beschreibt.

3. exponentielle Gradverteilung für große Knotengrade (englisch: scale-free degree distribution)
   Der Grad eines Knotens ist die Anzahl der von diesem Knoten ausgehenden Kanten. Die Verteilung des Knotengrads gibt an, welcher Grad wie häufig auftritt. In realen Graphen kann die Verteilung des Knotengrads für Grade oberhalb einer gewissen Grenze häufig durch eine Funktion der Form

   $$P(k) \sim k^{-\gamma}$$

   beschrieben werden, wobei $P(k)$ die Wahrscheinlichkeit beschreibt, dass ein beliebiger Knoten den Grad k hat, und γ eine positive Konstante ist [6]. Das Spektrum der Knotengrade realer Graphen umfasst also normalerweise mehrere Größenordnungen, wobei die Anzahl der Knoten mit dem Grad k ab irgendeinem k exponentiell abnimmt. Bei der Analyse des WWW konnte für ein geeignet gewähltes γ eine sehr gute Übereinstimmung mit der realen Gradverteilung erzielt werden.

Bei fester Knotenzahl n gibt es im $G_{n,p}$-Modell mit der Kantenwahrscheinlichkeit p genau einen Parameter, mit dem das Aussehen der erzeugten Graphen kontrolliert werden kann. Dieser Parameter wird benötigt, um die Anzahl der Kanten zu justieren.

Damit sind die Einstellmöglichkeiten dieses Modells erschöpft und es bleibt nur noch zu untersuchen, wie gut die so erzeugten Graphen in den drei aufgeführten Merkmalen durchschnittliche Pfadlänge, Kantenverteilung und Verteilung der Knotengrade mit realen Graphen übereinstimmen. Für $G_{n,p}$-Graphen konnte gezeigt werden, dass die durchschnittliche Pfadlänge logarithmisch in der Anzahl der Knoten wächst [6]. Die Graphen weisen folglich eine kurze durchschnittliche Pfadlänge auf, ebenso wie man sie häufig in realen Graphen vorfindet. Die Verteilung der Kanten ist jedoch vergleichsweise homogen, so dass der Clustering-Koeffizient aussschließlich abhängig von der Kantenwahrscheinlichkeit p ist und um mehrere Größenordnungen unter denen der untersuchten realen Graphen liegt [6]. Die Knotengrade klassischer Zufallsgraphen sind poissonverteilt, und weichen damit signifikant von der in realen Graphen beobachteten Verteilung ab [6].

Alles in allem können die Untersuchungen von Performanz und Approximationsgüte von Algorithmen auf den klassischen zufälligen Graphen nicht ohne Weiteres zur Vorhersage für das Verhalten bei realen Graphen benutzt werden. In der Literatur wurde daher eine Vielzahl von alternativen Modellen zufälliger Graphen vorgeschlagen, die die Verteilung der realen Graphen besser nachbilden. Die verschiedenen Modelle liefern unterschiedlich gute Annäherungen an reale Graphen, wobei die Güte der Annäherung auch von der Art der realen Graphen abhängt. Außerdem unterscheiden sich die Modelle auch in ihrer Zugänglichkeit bzw. Handhabbarkeit, d.h. in der Schwierigkeit der Untersuchung von erzeugten Graphen auf bestimmte Eigenschaften oder das Verhalten von Algorithmen darauf.

In dieser Arbeit wird das Modell der zufälligen Schnittgraphen benutzt. Nach Meinung des Autors ist dieses Modell ein guter Kompromiss zwischen Zugänglichkeit und Annäherung an reale Graphen. Das Modell wird im Folgenden als $\mathcal{G}_{n,m,p}$ notiert. Man geht von einer Knotenmenge mit n Knoten und einer disjunkten Menge von m Features aus. In einem ersten Schritt werden den Knoten unabhängig von einander Features zugeordnet, so dass die Wahrscheinlichkeit dafür, dass ein beliebiger Knoten ein bestimmtes Feature hat, exakt p ist. Der zufällige Schnittgraph auf der Knotenmenge entsteht durch Einfügen aller Kanten zwischen Knoten mit gemeinsamen Features. In der hier benutzten Form wurde diese Modell in [7] eingeführt. Im Gegensatz zu den klassischen zufälligen Graphen sind die Kanten in zufälligen Schnittgraphen nicht mehr unabhängig voneinander. Betrachtet man beispielsweise drei Knoten a, b und c, so erhöht sich die Wahrscheinlichkeit einer Kante zwischen a und b, wenn sowohl a als auch b einen gemeinsamen Nachbar c besitzen. Diese Form der Transitivität kann in vielen realen Graphen beobachtet werden und ist eine Ursache für die ungleichmäßige Verteilung der Kanten. Der Clustering-Koeffizient von zufälligen Schnittgraphen ist bei geeigneter Wahl von m wesentlich größer als in klassischen Zufallsgraphen mit gleicher Knoten- und Kantenzahl [6].

Anders als bei $G_{n,p}$ gibt es für $\mathcal{G}_{n,m,p}$ bei fester Knotenzahl n zwei Parameter, mit denen das Aussehen der Graphen kontrolliert werden kann. Es gibt daher verschiedene Möglichkeiten die gewünschte Kantenzahl zu erzielen und somit auch die Chance die Schnittgraphen besser an reale Eingabegraphen anzupassen.

Die Gestalt der zufälligen Schnittgraphen hängt entscheidend von der Anzahl der verwendeten Features m ab. Ist m klein, so können nicht alle Graphen erzeugt werden. Erst ab $m \geq \frac{1}{4}n^2$ können alle Graphen als Ergebnis des $\mathcal{G}_{n,m,p}$-Zufallsexperiments auftreten [7]. Mit steigendem m nimmt außerdem der Einfluss der oben angesprochenen Transitivität ab. In [7] wurde gezeigt, dass bei $m = n^\alpha$ für $\alpha > 6$ die zufälligen Schnittgraphen genauso aussehen wie klassische zufällige Graphen (bei gleicher Knoten- und Kantenzahl). Die Abhängigkeit zwischen den einzelnen Kanten ist hier so gering, dass sie keinen messbaren Einfluss mehr auf die Gestalt der Graphen hat. Die Äquivalenz kann sogar für kleinere m nachgewiesen werden, wenn die Featurewahrscheinlichkeit p hinreichend klein ist. Details dazu sind in Abschnitt 2.4 aufgeführt. Da das Modell der klassischen Zufallsgraphen bereits umfassend untersucht worden ist, wird sich diese Arbeit auf zufällige Schnittgraphen mit kleinen Featuremengen konzentrieren.

1.4 Zielstellung

Die vorliegende Arbeit beschäftigt sich mit dem Färben von zufälligen Schnittgraphen. Wie die vorangegangenen Abschnitte zeigen, gibt es eine Reihe praktischer Probleme, die auf das Färbeproblem zurückgeführt und somit durch effiziente Färbeverfahren gelöst werden können. Da das optimale Färben von Graphen im Allgemeinen sehr aufwendig (NP-vollständig) ist, wird die Untersuchung auf zufälligen Graphen durchgeführt. Das Modell der klassischen zufälligen Graphen ist nur begrenzt geeignet reale Grapheneingaben, die ja bei der Lösung praktischer Probleme auftreten werden, nachzubilden. Daher benutzt der Autor das Modell der zufälligen Schnittgraphen, dessen Ausgaben oft deutlich näher an realen Graphen liegen.

Natürlich ist nicht zu erwarten, dass sich ein effizienter Algorithmus zum optimalen Färben aller zufälligen Schnittgraphen, d.h. für die komplette Bandbreite von Featurezahlen und -wahrscheinlichkeiten, finden lässt. Im nicht abgedeckten Parameterbereich wird daher eine Schranke für die Approximationsgüte des verwendeten Färbealgorithmus abgeleitet. Abschließend wird das untersuchte Verfahren auf einige reale Graphen angewandt. Die dabei erhaltenen Färbungen sind optimal bzw. sehr nah am Optimum und demonstrieren somit einerseits die Praxistauglichkeit des Verfahrens und liefern andererseits einen Hinweis darauf, an welchen Stellen eine Verbesserung der hier erhaltenen Resultate möglich sein könnte. Da zur Untersuchung der Approximationsgüte eine Aussage über die Stabilitätszahl benötigt wird (die Größe der größten Knotenteilmenge, deren Knoten nicht durch Kanten verbunden sind), beschäftigt sich ein Kapitel ausführlich mit der Analyse dieser Größe. Dabei werden für fast den kompletten Parameterraum obere und untere Schranken für die Stabilitätszahl aufgestellt.

Dem Autor sind weder zum Färbeproblem noch zur Stabilitätszahl zufälliger Schnittgraphen vergleichbare Arbeiten bekannt. Die vorliegende Arbeit erweitert daher das Wissen über zufällige Schnittgraphen und trägt wesentlich zum Verständnis dieses Modells zufälliger Graphen bei. Unter der Annahme dass dieses Graphenmodell gute Approximationen einiger realer Graphen liefert, sind die hier erhaltenen Ergebnisse auch für entsprechende reale Graphen relevant.

2 Grundlagen

Nachfolgend werden die nötigen Definitionen und verwendeten Abschätzungen bereitgestellt, sowie gängige Beweisverfahren und Aussagen aufgeführt, die in dieser Arbeit vorausgesetzt werden. Dieses Kapitel eignet sich daher zum Nachschlagen von Begriffen und Notation und wird häufig von den darauffolgenden Kapiteln referenziert.

2.1 Begriffe und Notation

Definition 1 (Graph) Ein *Graph* ist ein Paar (V, E), wobei V eine endliche Menge und $E \subseteq \binom{V}{2}$ eine Teilmenge der zwei-elementigen Teilmengen von V ist. Die Elemente von V heißen Knoten und die Elemente von E Kanten.

Definition 2 (Subgraph, induzierter Subgraph) Ein Graph $H = (V', E')$ heißt *Subgraph* von $G = (V, E)$ wenn $V' \subseteq V$ und $E' \subseteq E \cap \binom{V'}{2}$ gilt. Wenn sogar $E' = E \cap \binom{V'}{2}$ gilt, so wird H auch als *induzierter Subgraph* bezeichnet.

Definition 3 (Nachbarschaft, Grad, Minimalgrad, Maximalgrad) Sei $G = (V, E)$ ein Graph und sei $v \in V$ ein Knoten in G. Die *Nachbarschaft* von v ist die Menge aller Knoten $N(v) \subseteq V$, mit denen v durch eine Kante verbunden ist. Die Anzahl dieser Knoten heißt *Grad* von v (notiert als $d(v)$). Der *Minimalgrad* von G $\delta(G)$ ist definiert als $\delta(G) = \min_{v \in V} d(v)$. Analog ist der *Maximalgrad* von G $\Delta(G)$ definiert durch $\Delta(G) = \max_{v \in V} d(v)$.

Definition 4 (Clique, vollständig) Sei $G = (V, E)$ ein Graph. Eine Teilmenge $C \subseteq V$ seiner Knoten heißt *Clique*, wenn jeder Knoten in C mit jedem anderen Knoten aus C durch eine Kante verbunden ist, d.h. wenn gilt: $\binom{C}{2} \subseteq E$. G heißt *vollständig*, wenn die Knotenmenge V eine Clique ist.

Definition 5 (stabile Menge, Stabilitätszahl) Sei $G = (V, E)$ ein Graph und $S \subseteq V$ eine Teilmenge seiner Knoten. Die Knotenmenge S heißt *stabil*, wenn keinerlei Kanten zwischen den enthaltenen Knoten verlaufen, d.h. wenn gilt: $E \cap \binom{S}{2} = \emptyset$. Die Größe der größten stabilen Menge in G wird als *Stabilitätszahl* $\beta(G)$ bezeichnet.

Definition 6 (Färbung, chromatische Zahl, bipartit) Sei $G = (V, E)$ ein Graph, dann heißt eine Abbildung $f : V(G) \to \mathbb{N}$ von den Knoten des Graphen in die natürlichen Zahlen *Färbung*, wenn gilt: $\forall_{x,y \in V} : \{x, y\} \in E \Rightarrow f(x) \neq f(y)$. Die Elemente von $f(V)$ werden in diesem Zusammenhang auch als Farben bezeichnet. Für $k = |f(V)|$ (d.h. k entspricht der Anzahl der verwendeten Farben) wird f als eine k-Färbung bezeichnet. Die *chromatische Zahl* von G ist das kleinste k, für dass G eine k-Färbung besitzt. Sie wird als $\chi(G)$ notiert. Ein Graph mit einer chromatischen Zahl $\chi(G) \leq 2$ heißt *bipartit*.

Definition 7 (Approximationsgüte) Die *Approximationsgüte* eines Algorithmus zum Färben eines Graphen bzw. zum Finden der größten stabilen Menge ist definiert als Quotient aus der Anzahl benötigter Farben und chromatischer Zahl, bzw. aus der Größe der gefundenen stabilen Menge und der Stabilitätszahl.

Im Rahmen dieser Arbeit werden zufällige Graphen betrachtet. Dabei entstehen die Graphen als Ergebnis eines Zufallsexperiments. Typische Fragestellungen zu solchen Graphen betreffen die erwartete Laufzeit von Algorithmen sowie im Fall von Approximationsalgorithmen die erwartete Approximationsgüte. Das in dieser Arbeit verwendete Modell zufälliger Graphen heißt zufällige Schnittgraphen (im Englischen auch Random Intersection Graphs).

Definition 8 (Projektion, bipartiter Generator, Feature, Featureclique) Sei $B = (V \cup W, E_B)$ ein bipartiter Graph mit $V \cap W = \emptyset$ und $\binom{V}{2} \cap E_B = \binom{W}{2} \cap E_B = \emptyset$. Die *Projektion* von B ist nun ein Graph $G = (V, E)$, dessen Kantenmenge genau die Kanten $\{v_1, v_2\}$ enthält, für die v_1 und v_2 einen gemeinsamen Nachbarn in W haben. Der Graph B wird auch als *bipartiter Generator* von G bezeichnet. Die Elemente von W heißen *Features* von G, und jedes von ihnen induziert eine Clique in G. Daher werden diese Cliquen auch *Featurecliquen* genannt.

Definition 9 (zufälliger Schnittgraph, Featurewahrscheinlichkeit) Das Modell der *zufälligen Schnittgraphen* $\mathcal{G}_{n,m,p}$ beschreibt ein Zufallsexperiment zur Erzeugung von Graphen. Seien dazu V und W zwei disjunkte endliche Mengen mit $|V| = n$ und $|W| = m$ und sei $p \in [0, 1]$ eine reelle Zahl zwischen 0 und 1. Das Experiment besteht aus zwei Schritten:

1. Erzeugen eines bipartiten Graphen $B = (V \cup W, E_B)$, indem für jedes Paar $(v, w) \in (V, W)$ die Kante $\{v, w\}$ mit Wahrscheinlichkeit p zu E_B hinzufügt wird.
2. Erzeugen des Graphen $G = (V, E)$ durch Projektion von B.

G ist das Ergebnis des Zufallsexperiments. Die so erzeugten Graphen bilden die Instanzen von $\mathcal{G}_{n,m,p}$. Die Größe p wird im Rahmen dieser Arbeit als *Featurewahrscheinlichkeit* bezeichnet, da sie die Wahrscheinlichkeit der Zuordnung eines beliebigen Features zu einem beliebigen Knoten beschreibt.

Die Beweise der vorliegenden Arbeit benutzen häufig folgende Formulierung:

> *Sei $G = (V, E)$ ein zufälliger Schnittgraph und sei $B = (V \cup W, E_B)$ sein bipartiter Generator.*

Damit ist gemeint, dass mit dem hier definierten Zufallsexperiment für gegebene n, m, p ein bipartiter Generator B erzeugt wird, dessen Projektion den Graphen G liefert. Allgemein kann einem beliebigen Graphen nicht eindeutig ein bipartiter Generator zugeordnet werden.

Für eine beliebige Grapheneigenschaft A bezeichne '$\mathbb{P}\left[\mathcal{G}_{n,m,p} \text{ besitzt } A\right]$' die Wahrscheinlichkeit, dass eine Instanz von $\mathcal{G}_{n,m,p}$ diese Eigenschaft besitzt. In der Theorie der zufälligen Graphen ist es üblich zu untersuchen, wie sich diese Wahrscheinlichkeit für Knotenzahlen n gegen unendlich entwickelt. Man sagt dass $\mathcal{G}_{n,m,p}$ *fast sicher* eine Eigenschaft A hat, bzw. synonym *fast alle* zufälligen Schnittgraphen die Eigenschaft A besitzen, wenn $\mathbb{P}\left[\mathcal{G}_{n,m,p} \text{ besitzt } A\right]$ für n gegen unendlich gegen 1 strebt. Geht diese Wahrscheinlichkeit gegen 0, so besitzt *fast kein* zufälliger Schnittgraph die Eigenschaft A.

Um auch die Anzahl der Features m in diese Untersuchung einzubeziehen wird in dieser Arbeit m immer von der Gestalt $m = n^\alpha$ sein, für eine Konstante $\alpha > 0$. Weiterhin wird auch p als eine Funktion von n betrachtet, da die zufälligen Schnittgraphen mit zunehmender Knotenzahl für konstante p immer dichter und schließlich fast sicher vollständig werden. Der Autor orientiert sich bei dieser Herangehensweise an [7] und [8].

In diesem Zusammenhang ist die folgende Notation nützlich.

Für zwei Funktion $p, f : \mathbb{N} \to \mathbb{R}$ gilt:

$$p(n) \ll f(n) \quad \Leftrightarrow \quad \lim_{n\to\infty} \frac{p(n)}{f(n)} = 0$$
$$p(n) \gg f(n) \quad \Leftrightarrow \quad \lim_{n\to\infty} \frac{p(n)}{f(n)} = \infty.$$

An dieser Stelle setzen auch die Symbole $o(g(n))$ und $O(g(n))$ an. Sei $A\Big(f(n)\Big)$ eine Aussage über eine Funktion f. Dann gilt:

$$\begin{aligned} A\Big(o(g(n))\Big) \quad &\Leftrightarrow \quad \text{es existiert eine Funktion } f(n) \ll g(n), \text{ so dass } A\Big(f(n)\Big) \text{ gilt} \\ A\Big(O(g(n))\Big) \quad &\Leftrightarrow \quad \text{es existiert eine Konstante } c \text{ und eine Funktion } f(n) \le cg(n), \\ &\qquad \text{so dass } A\Big(f(n)\Big) \text{ gilt.} \end{aligned}$$

Definition 10 (Schwellenwertfunktion) Eine Funktion $f(n)$ heißt *Schwellenwertfunktion* für eine Grapheneigenschaft A, wenn die folgenden beiden Aussagen gelten.

$$\begin{array}{lll} (1) & p(n) \ll f(n) \quad \Rightarrow & \lim_{n\to\infty} \mathbb{P}\left[\mathcal{G}_{n,m,p} \text{ hat } A\right] = 0 \\ (2) & p(n) \gg f(n) \quad \Rightarrow & \lim_{n\to\infty} \mathbb{P}\left[\mathcal{G}_{n,m,p} \text{ hat } A\right] = 1 \end{array}$$

Die folgende Tabelle enthält eine kurze Aufstellung der verwendeten Notation.

$\mathbb{P}[A]$	die Wahrscheinlichkeit, mit der das Ereignis A eintritt
$\mathbb{P}[A\|B]$	die Wahrscheinlichkeit, mit der das Ereignis A eintritt
	unter der Bedingung/Voraussetzung das B eintritt
$\overline{A}$	das Komplementereignis zum Ereignis A
$\|V\|$	die Anzahl der Elemente der Menge V
$\mathbb{E}[X]$	der Erwartungswert der Zufallsvariable X
$\mathrm{Var}[X]$	die Varianz der Zufallsvariable X
$N(v)$	die Menge aller Nachbarknoten des Knotens v
$\Delta(G)$	der maximale Knotengrad des Graphen G
$\delta(G)$	der minimale Knotengrad des Graphen G
$\chi(G)$	die chromatische Zahl des Graphen G
$\beta(G)$	die Stabilitätszahl des Graphen G
$G_{n,p}$	das Modell der klassischen zufälligen Graphen
$\mathcal{G}_{n,m,p}$	das Modell der zufälligen Schnittgraphen
$\binom{V}{k}$	Die Menge aller k-elementigen Teilmengen von V
$(n)_k$	das Produkt der natürlichen Zahlen n bis $n-k+1$
	d.h. $(n)_k = n(n-1)(n-2)\dots(n-k+1)$
$p \ll f$	für zwei Funktionen $p, f : \mathbb{N} \to \mathbb{R}$ gilt: $\lim_{n\to\infty} \frac{p(n)}{f(n)} = 0$
$A\Big(o(g(n))\Big)$	es gibt eine Funktion $f(n) \ll g(n)$, so dass Aussage $A\Big(f(n)\Big)$ gilt
$A\Big(O(g(n))\Big)$	es gibt eine Konstante c und eine Funktion $f(n) \le cg(n)$,
	so dass Aussage $A\Big(f(n)\Big)$ gilt

2.2 Beweisverfahren

2.2.1 1. und 2. Moment Methode

In der Theorie der zufälligen Graphen werden häufig Beweise geführt um zu zeigen, dass fast kein Graph bzw. fast alle Graphen eine bestimmte Eigenschaft besitzen. Dazu wird in vielen Fällen die Wahrscheinlichkeit betrachtet, mit der eine Zufallsvariable den Wert 0 annimmt.

Die 1. Moment Methode benutzt dazu eine Aussage über den Erwartungswert der entsprechenden Zufallsvariable, die aus der Markoffschen Ungleichung

$$\mathbb{P}[X \geq t] \leq \frac{1}{t}\mathbb{E}[X] \quad \text{für } t > 0 \text{ und eine nichtnegative Zufallsvariable } X \tag{1}$$

abgeleitet wird. Denn für den Spezialfall einer ganzzahligen nichtnegativen Zufallsvariable X erhält man aus (1) mit $t = 1$:

$$\mathbb{P}[X > 0] \;\leq\; \mathbb{E}[X]. \tag{2}$$

Das führt nach (3.1) in [9] zu folgender Aussage für eine Folge ganzzahliger nichtnegativer Zufallsvariablen X_n:

$$\lim_{n\to\infty} \mathbb{E}[X_n] = 0 \quad \Rightarrow \quad \lim_{n\to\infty} \mathbb{P}[X_n = 0] = 1.$$

In dieser Arbeit wird die Abhängigkeit der Zufallsvariablen von der Knotenzahl n der Graphen nicht explizit notiert. Mitunter wird daher anstatt von einer Folge von Zufallsvariablen von einer einzelnen Zufallsvariable gesprochen.

Die 2. Moment Methode liefert eine hinreichende Bedingung dafür, dass eine Zufallsvariable fast sicher ungleich 0 ist. Sie basiert auf der Tschebyscheffschen Ungleichung:

$$\mathbb{P}[|X - \mathbb{E}[X]| \geq t] \;\leq\; \frac{\mathrm{Var}[X]}{t^2}, \qquad \text{für } t > 0.$$

Aus dieser Ungleichung kann man ableiten (siehe (3.2) in [9]), dass für eine beliebige Folge von Zufallsvariablen X_n gilt:

$$\lim_{n\to\infty} \mathbb{E}[X_n] = \infty \text{ und } \mathrm{Var}[X_n] \ll \mathbb{E}[X_n]^2 \quad \Rightarrow \quad \lim_{n\to\infty} \mathbb{P}[X_n = 0] = 0.$$

Verglichen mit der 1. Moment Methode ist die 2. Moment Methode oft deutlich aufwendiger, da neben dem meist recht einfach zu bestimmenden Erwartungswert die Varianz der Zufallsvariable abgeschätzt werden muss.

Übertragen auf die Untersuchung der Anzahl isolierter Knoten in zufälligen Graphen liefern die beiden Methoden:

> Sei X eine Zufallsvariable, welche die Anzahl der isolierten Knoten in zufälligen Graphen zählt. Wenn gezeigt werden kann, dass der Erwartungswert von X für n gegen unendlich gegen 0 strebt, so enthält fast kein Graph einen isolierten Knoten (1. Moment Methode).
> Wenn dagegen der Erwartungswert von X gegen unendlich strebt und die Varianz von X klein gegen das Quadrat des Erwartungswerts ist, so enthält fast jeder Graph einen isolierten Knoten (2. Moment Methode).

2.2.2 Chernoff Ungleichungen

In den Beweisen der Arbeit wird häufig eine Aussage darüber benötigt, wie stark sich die Werte einer Zufallsvariable um ihren Erwartungswert konzentrieren. Für binomialverteilte Zufallsvariablen liefern die Chernoff-Ungleichungen sehr gute Schranken.

Definition 11 (binomialverteilt) Eine Zufallsvariable heißt *binomialverteilt*, wenn sie als Summe unabhängiger 0-1-Zufallsvariablen (d.h. Zufallsvariablen, die nur die Werte 0 und 1 annehmen) geschrieben werden kann.

Nach Chernoff gilt für eine binomialverteilte Zufallsgröße X und $t \geq 0$:

$$\mathbb{P}\left[X \geq \mathbb{E}\left[X\right] + t\right] \leq \exp\left(-\frac{t^2}{2\mathbb{E}\left[X\right] + \frac{2}{3}t}\right) \qquad \text{nach 2.5 in [9]} \tag{3}$$

$$\mathbb{P}\left[X \leq \mathbb{E}\left[X\right] - t\right] \leq \exp\left(-\frac{t^2}{2\mathbb{E}\left[X\right]}\right) \qquad \text{nach 2.6 in [9].} \tag{4}$$

X wird also nur selten stark von seinem Erwartungswert abweichen.

2.2.3 Satz über die totale Wahrscheinlichkeit

Im Laufe der Arbeit werden Aussagen bewiesen, die für fast alle zufälligen Schnittgraphen in bestimmten Bereichen für p und α gelten. Sei A eine Grapheneigenschaft, die fast alle Graphen einer Teilmenge der zufälligen Schnittgraphen haben. D.h. die Wahrscheinlichkeit dafür, dass ein beliebig gewählter Graph aus der Teilmenge die Eigenschaft A hat, konvergiert für n gegen unendlich gegen 1. Sei $\mathbb{P}\left[A\right]$ diese Wahrscheinlichkeit. Diese Aussage kann für den Beweis einer anderen Eigenschaft B von fast allen Graphen derselben Teilmenge herangezogen werden. Denn nach dem Satz über die totale Wahrscheinlichkeit gilt:

$$\mathbb{P}\left[B\right] = 1-\mathbb{P}\left[\overline{B}\right] = 1-\mathbb{P}\left[\overline{B}|A\right]\mathbb{P}\left[A\right]-\mathbb{P}\left[\overline{B}|\overline{A}\right]\mathbb{P}\left[\overline{A}\right] \geq 1-\mathbb{P}\left[\overline{B}|A\right]-\mathbb{P}\left[\overline{A}\right].$$

Da $\mathbb{P}\left[A\right]$ gegen 1 strebt, muss $\mathbb{P}\left[\overline{A}\right]$ gegen 0 gehen. Es genügt daher $\mathbb{P}\left[\overline{B}|A\right] \to 0$ zu zeigen, weil dann sofort $\mathbb{P}\left[B\right] \to 1$ folgt. Es ist somit zulässig nur Graphen mit der Eigenschaft A zu betrachten um zu zeigen, dass fast alle zufälligen Schnittgraphen der oben erwähnten Teilmenge B erfüllen.

2.3 Verwendete Abschätzungen und Formeln

Die folgenden Identitäten und Abschätzungen werden in dieser Arbeit ohne Beweis verwendet.

- Für $q \neq 1$ gilt:

$$\sum_{i=0}^{n} q^i = \frac{q^{n+1}-1}{q-1}. \qquad \text{(abgebrochene geometrische Reihe)} \tag{5}$$

- Für $|q| < 1$ gilt:

$$\sum_{i=0}^{\infty} q^i = \frac{1}{q-1}. \qquad \text{(Grenzwert der geometrische Reihe)} \tag{6}$$

- Für $a > -1$ und $b \in I\!N$ gilt:

$$1 + ab \;\leq\; (1+a)^b. \qquad \text{(Bernoullische Ungleichung)} \tag{7}$$

- Für alle $x, n \in I\!R$ gilt:

$$\exp(-2xn) \;\leq\; (1-x)^n \;\leq\; \exp(-xn). \tag{8}$$

- Für alle $n, k \in I\!N$ gilt:

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \left(\frac{ne}{k}\right)^k. \tag{9}$$

- Für jedes beliebige $\epsilon > 0$ gilt:

$$\ln n \ll n^\epsilon. \tag{10}$$

2.4 Bekannte Ergebnisse zu zufälligen Schnittgraphen

Die für die vorliegende Arbeit wichtigsten Resultate zu zufälligen Schnittgraphen stammen aus [7]. Zunächst soll der für die Featurewahrscheinlichkeit p interessante Bereich abgesteckt werden. Das dazu von Singer aufgestellte Theorem lautet in leicht vereinfachter Form:

Singer (1995). *Für $p \ll \frac{1}{n\sqrt{m}}$ haben zufällige Schnittgraphen fast sicher keine Kanten und für $p \gg \sqrt{\frac{\ln n}{m}}$ sind zufällige Schnittgraphen fast sicher vollständig.*

Das Färben von kantenlosen oder vollständigen Graphen ist natürlich keine Herausforderung. Die vorliegende Arbeit wird sich daher ausschließlich mit Featurewahrscheinlichkeiten innerhalb dieses Intervalls beschäftigen. Singer konnte außerdem zeigen, dass die oben angegebenen Grenzen für p auch Schwellenwerte für die jeweilige Eigenschaft sind. D.h. für $\frac{1}{n\sqrt{m}} \ll p \ll \sqrt{\frac{\ln n}{m}}$ enthält fast jeder zufällige Schnittgraph Kanten sowie Knoten, die nicht direkt durch eine Kante verbunden sind.

Andere Ergebnisse von Singer betreffen die Äquivalenz von zufälligen Schnittgraphen zu $G_{n,p}$. Hier konnte die Autorin zeigen, dass zufällige Schnittgraphen und klassische zufällige Graphen für $\alpha > 6$ bei gleicher Knoten- und Kantenzahl sehr ähnlich und nach dem Verständnis der Graphentheorie äquivalent sind.

Singer (1995). *Sei $m = n^\alpha$ für ein festes $\alpha > 6$ und $\frac{1}{n\sqrt{m}} \ll p \ll \sqrt{\frac{\ln n}{m}}$, dann gilt für jede Grapheneigenschaft A und jedes $a \in [0,1]$:*

$$\lim_{n\to\infty} I\!P[\mathcal{G}_{n,m,p} \textit{ hat } A] = a \;\Leftrightarrow\; \lim_{n\to\infty} I\!P[G_{n,\hat{p}} \textit{ hat } A] = a \qquad \textit{für } \hat{p} = 1-(1-p^2)^m.$$

Zufällige Schnittgraphen und klassische Zufallsgraphen haben also genau dieselben Grapheneigenschaften für das relevante Intervall der Featurewahrscheinlichkeit p.

Singer konnte ferner sogar für kleinere α eine partielle Äquivalenz nachweisen.

Singer (1995). *Sei $m = n^\alpha$ für ein festes α. Wenn nun gilt*

$$\begin{array}{llll} & 3 < \alpha \leq 6 & \textit{und} & \frac{1}{n\sqrt{m}} \ll p \ll \sqrt{\frac{\ln n}{nm}} \\ \textit{oder} & \alpha > 1.5 & \textit{und} & \frac{1}{n\sqrt{m}} \ll p \ll \frac{1}{m}, \end{array}$$

dann gilt für jede Grapheneigenschaft A und jedes $a \in [0,1]$:

$$\lim_{n\to\infty} \mathbb{P}[\mathcal{G}_{n,m,p} \text{ hat } A] = a \;\Leftrightarrow\; \lim_{n\to\infty} \mathbb{P}[G_{n,\hat{p}} \text{ hat } A] = a \qquad \text{für } \hat{p} = 1-(1-p^2)^m.$$

Aufgrund dieser Ergebnisse wird sich der Autor bei der Untersuchung von zufälligen Schnittgraphen mit kleinen p vor allem auf kleine α konzentrieren.

3 Zufällige Schnittgraphen optimal färben

Das Finden einer optimalen Färbung für Graphen ist im Allgemeinen ein NP-vollständiges Problem. D.h. wenn $P \neq NP$ gilt, so gibt es keinen polynomiellen Algorithmus für diese Aufgabe. Auch die Approximation ist schwierig. So konnte gezeigt werden, dass es für $P \neq NP$ ein $\epsilon > 0$ gibt, so dass kein polynomieller Färbealgorithmus alle Graphen mit weniger als $n^\epsilon\ \chi$ Farben färbt. [3]

Daher erscheint es wenig aussichtsreich sich mit optimalen Färbungen im Allgemeinen zu beschäftigen. Häufig werden zusätzliche Bedingungen formuliert, wenn dieses Thema angegangen wird.

Eine erste Möglichkeit besteht in der Einschränkung der Betrachtung auf bestimmte Klassen von Graphen. So ist es beispielsweise sehr einfach Bäume (d.h. zusammenhängende, kreisfreie Graphen) in linearer Zeit optimal zu färben. Eine zweite Möglichkeit ist der Übergang zu zufälligen Graphen. Zufällige Graphen ermöglichen die Betrachtung der erwarteten Laufzeit von Algorithmen. Auf diese Weise kann das Problem der wohl unvermeidlichen Worst-Case-Laufzeit umgangen werden, da eine exponentielle Laufzeit auf nur wenigen Instanzen durch eine gute Laufzeit auf dem Großteil der Instanzen aufgewogen werden kann. Ebenso ist die Betrachtung von Approximationsalgorithmen auf zufälligen Graphen um einiges vielversprechender, da auch hier eine gute Approximationsgüte auf der Mehrzahl der Instanzen vereinzelte schlechte Approximationen aufwiegen kann.

Der Vollständigkeit halber sei noch ein dritter Ansatz erwähnt: der Übergang zu randomisierten Algorithmen. Diese Algorithmen nutzen Zufallsentscheidungen zur Laufzeit, was zum Teil erstaunliche Vereinfachungen gegenüber deterministischen Algorithmen zum selben Problem ermöglicht. Bei randomisierten Algorithmen kann ebenfalls die erwartete Laufzeit anstelle der Worst-Case-Laufzeit untersucht werden. Dieser Ansatz wird jedoch in der vorliegenden Arbeit nicht weiter betrachtet.

In diesem Kapitel werden Möglichkeiten untersucht, fast alle zufälligen Schnittgraphen optimal zu färben. D.h. es geht um Färbealgorithmen deren Erfolgswahrscheinlichkeit mit wachsender Knotenzahl der Eingabe gegen 1 konvergiert, wobei nur das Finden einer optimalen Färbung als Erfolg betrachtet wird.

3.1 Chordale zufällige Schnittgraphen

Definition 12 (chordal) Ein Graph G heißt *chordal*, wenn er keinen induzierten Kreis der Länge $k \geq 4$ enthält.

Chordale Graphen bilden eine Klasse von Graphen, in der viele normalerweise NP-schwere Probleme in Polynomialzeit, bzw. sogar in linearer Zeit gelöst werden können. So sind in [3] unter anderem lineare Algorithmen zum Bestimmen der größten Clique und zum Finden einer optimalen Färbung für chordale Graphen angegeben. Darüberhinaus kann ebenfalls in linearer Zeit erkannt werden, ob ein Graph chordal ist oder nicht.

Der folgende Satz beweist nun, dass für hinreichend kleine p fast alle zufälligen Schnittgraphen chordal sind.

Satz 1. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$ und $p \ll \sqrt{\frac{1}{nm}}$. Dann ist fast jede Instanz von $\mathcal{G}_{n,m,p}$ chordal.*

Beweis. Sei $G = (V, E) \in \mathcal{G}_{n,m,p}$ ein zufälliger Schnittgraph und $B = (V \cup W, E_B)$ der bipartite Generator von G. G ist genau dann chordal, wenn er keinen induzierten Kreis der Länge mindestens vier enthält. Da ein solcher Kreis kein Dreieck enthält muss also jede seiner Kanten durch eine andere Featureclique generiert werden. In Abbildung 1 ist eine für einen induzierten Kreis der Länge k notwendige (jedoch nicht hinreichende) Konfiguration im bipartiten Generator abgebildet. Im Folgenden wird ein solcher Kreis als C_k bezeichnet.

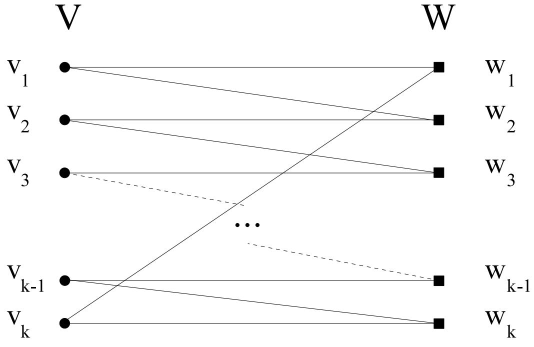

Abbildung 1: eine für einen C_k notwendige Konfiguration in B

Seien $v_1, \ldots, v_k \in V$ eine Anordnung von k Knoten und $w_1, \ldots, w_k \in W$ eine Anordnung von k Featureknoten für ein festes k. Damit die Knoten $v_1, \ldots, v_k$ einen Kreis der Länge k in G induzieren, müssen sie sukzessiv durch Kanten $\{v_i, v_{i+1}\}$ und $\{v_k, v_1\}$ (für $1 \leq i < k$) verbunden sein. Die Wahrscheinlichkeit dafür, dass diese Kanten durch die Featureknoten $w_1, \ldots, w_k$ entsprechend der in Abbildung 1 dargestellten Konfiguration erzeugt werden, lässt sich wie folgt abschätzen:

$$\mathbb{P}\left[v_1, \ldots, v_k \text{ und } w_1, \ldots, w_k \text{ induzieren einen } C_k \text{ in } G\right] \;\leq\; p^{2k}(1-p)^{k^2-2k} \;\leq\; p^{2k}.$$

Außerdem dürfen die nicht durch $w_1, \ldots, w_k$ verbundenen Knoten auch sonst keine gemeinsamen Featureknoten kennen. D.h. die reale Wahrscheinlichkeit dafür, dass $v_1, \ldots, v_k$ entsprechend der Abbildung 1 einen C_k bilden, ist sogar noch kleiner.

Durch Multiplikation mit der Anzahl der Möglichkeiten die Knoten $v_1, \ldots, v_k$ und die Features $w_1, \ldots, w_k$ auszuwählen erhält man die Wahrscheinlichkeit, dass G einen induzierten C_k enthält:

$$\mathbb{P}\left[G \text{ enthält einen induzierten } C_k\right] \;\leq\; n^k m^k p^{2k} \;=\; (nmp^2)^k.$$

Damit kann die Wahrscheinlichkeit dafür, dass G nicht chordal ist, abgeschätzt werden:

$$\begin{aligned}
\mathbb{P}\left[G \text{ ist nicht chordal}\right] &\leq \sum_{k=4}^{min(n,m)} \mathbb{P}\left[G \text{ enthält einen induzierten } C_k\right] \\
&\leq \sum_{k=4}^{min(n,m)} (nmp^2)^k \\
&= \sum_{k=0}^{min(n,m)} (nmp^2)^k - \sum_{k=0}^{3}(nmp^2)^k \\
&\overset{(5)}{=} \frac{(nmp^2)^{min(n,m)+1} - 1}{nmp^2 - 1} - \sum_{k=0}^{3}(nmp^2)^k.
\end{aligned}$$

Aus der Voraussetzung $p \ll \sqrt{\frac{1}{nm}}$ folgt, dass nmp^2 für n gegen unendlich gegen 0 strebt. Somit ergibt sich als Grenzwert für die Wahrscheinlichkeit, dass G nicht chordal ist:

$$\begin{aligned}
\lim_{n\to\infty} P(G \text{ ist nicht chordal}) &\leq \lim_{n\to\infty} \left(\frac{(nmp^2)^{min(n,m)+1} - 1}{nmp^2 - 1} - \sum_{k=0}^{3}(nmp^2)^k \right) \\
&= \frac{-1}{-1} - (1 + 0 + 0 + 0) \\
&= 0.
\end{aligned}$$

Fast jede Instanz von $\mathcal{G}_{n,m,p}$ ist folglich chordal. □

In [7] wird gezeigt, dass $p = \sqrt{\frac{1}{nm}}$ die Schwellenwertfunktion für das Auftauchen eines induzierten Kreises fester Länge k ($k \geq 4$) in zufälligen Schnittgraphen ist. D.h. für $p \gg \sqrt{\frac{1}{nm}}$ enthalten fast alle zufälligen Schnittgraphen einen Kreis der Länge vier und sind somit nicht mehr chordal. Das obige Resultat lässt sich daher nicht auf wesentlich größere p erweitern.

3.2 Greedy-Färbungen mit smallest-last Heuristik

Dieser Abschnitt soll beleuchten, wie optimale Färbungen für zufällige Schnittgraphen mit Hilfe des Greedy-Algorithmus gefunden werden können. Die nachfolgenden Resultate zum Greedy-Algorithmus und verschiedenen Heuristiken inklusive aller Beweise zu den aufgeführten Aussagen können in [3] nachgelesen werden.

Der Greedy-Algorithmus zum Färben von Graphen ist ein sehr einfacher Approximationsalgorithmus mit linearer Laufzeit.

Greedy-Färbung des Graphen $G = (V, E)$ mit $V = \{v_1, \ldots, v_n\}$

```
1  färbe v_1 mit Farbe 1
2  for i := 2 to n do
3     färbe Knoten v_i mit der kleinsten Farbe,
      die in seiner Nachbarschaft nicht vorkommt.
```

Algorithmus 1: Der Greedy-Algorithmus zum Färben von Graphen

Wie unschwer zu erkennen ist, liefert dieser Algorithmus immer eine Färbung, die jedoch nicht optimal sein muss (da sie ggf. mehr als $\chi(G)$ Farben verwendet). Die Anzahl der Farben, die dieser Algorithmus benötigt, hängt stark von der Reihenfolge ab, in der die Knoten gefärbt werden. So gibt es auch immer eine Reihenfolge, die zu einer optimalen Färbung führt. An dieser Stelle setzen Heuristiken an, welche die Knoten vor der Anwendung des Greedy-Algorithmus in eine möglichst vielversprechende Reihenfolge bringen sollen.

Definition 13 (Rückwärtsgrad) Sei $G = (V, E)$ ein Graph und $\sigma = v_1, \ldots, v_n$ eine Reihenfolge seiner Knoten. Dann ist der *Rückwärtsgrad* $d_\sigma(v_i)$ des Knotens v_i definiert als die Anzahl der Kanten zwischen v_i und den Knoten $v_1, \ldots, v_{i-1}$. Der *maximale Rückwärtsgrad* der Anordnung σ wird als $b(\sigma)$ notiert und entspricht dem Maximum über die Rückwärtsgrade der einzelnen Knoten in σ. D.h.:

$$b(\sigma) \;=\; \max_{i=1,\ldots,n} d_\sigma(v_i).$$

Der Greedy-Algorithmus benötigt auf einer Anordnung der Knoten σ höchstens $b(\sigma)+1$ Farben. Daher liegt es nahe nach Anordnungen zu suchen, die den maximalen Rückwärtsgrad minimieren. Eine solche Anordnung wird smallest-last Anordnung genannt und lässt sich recht einfach berechnen.

Bestimmen einer smallest-last Knotenanordnung für $G = (V, E)$

`for` $i := n$ `downto 1 do`
    `wähle Knoten` $v_i \in V(G)$ `mit minimalem Grad in` G
    $G := G - v_i$

Algorithmus 2: Erstellen einer smallest-last Knotenanordnung nach [3]

Sei nun σ_{SL} eine smallest-last Anordnung der Knoten des Graphen G. Dann lässt sich zeigen, dass der maximale Rückwärtsgrad von σ_{SL} gleich dem Maximum über die Minimalgrade aller Subgraphen von G ist [3]. Zusammengefasst gilt:

$$b(\sigma_{SL}) \;=\; \min_{\sigma} b(\sigma) \;=\; \max_{H \subseteq G} \delta(H) \;\geq\; \chi(G) - 1.$$

In zufälligen Schnittgraphen ist die chromatische Zahl nach unten durch die Anzahl der Knoten in der größten Featureclique beschränkt. Es gilt daher immer:

$$\max_{H \subseteq G} \delta(H) + 1 \;\geq\; \chi(G) \;\geq\; \hat{\omega}_W,$$

wobei $\hat{\omega}_W$ die Knotenzahl der größten Featureclique bezeichnet.

Nachfolgend wird nun gezeigt, dass unter bestimmten Bedingungen fast kein zufälliger Schnittgraph einen Subgraphen enthält, dessen Minimalgrad größer oder gleich $\hat{\omega}_W$ ist. Denn wenn kein solcher Subgraph existiert, gilt:

$$b(\sigma_{SL}) \;=\; \max_{H \subseteq G} \delta(H) + 1 \;=\; \chi(G) \;=\; \hat{\omega}_W.$$

Der Greedy-Algorithmus findet in diesem Fall eine optimale Färbung.

Für diese Untersuchung benötigt man zunächst eine Aussage zur Größe der Featurecliquen in zufälligen Schnittgraphen.

Lemma 2. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$ und $p \gg \frac{\ln n}{n}$. Dann hat fast jede Instanz von $\mathcal{G}_{n,m,p}$ ausschließlich Featurecliquen mit mehr als $(1-\epsilon)np$ und weniger als $(1+\epsilon)np$ Knoten für ein beliebiges aber festes $\epsilon > 0$.*

Beweis. Sei $G = (V, E) \in \mathcal{G}_{n,m,p}$ ein zufälliger Schnittgraph und $B = (V \cup W, E_B)$ der bipartite Generator von G. Sei außerdem $p = \frac{\ln n}{n}\gamma(n)$, wobei $\gamma(n)$ eine gegen unendlich gehende Funktion ist.

Für $v \in V$ und $w \in W$ sei $X_{v,w}$ eine Zufallsvariable mit

$$X_{v,w} = \begin{cases} 1 \text{ falls } (v,w) \in E_B \text{ (d.h. } v \text{ gehört zu der von } w \text{ erzeugten Featureclique)} \\ 0 \text{ sonst} \end{cases}$$

und $X_w = \sum_{v \in V} X_{v,w}$ die Zufallsvariable, welche die Größe der von w erzeugten Featureclique misst. Dann gilt für den Erwartungswert von X_w:

$$\mathbb{E}[X_w] = \sum_{v \in V} \mathbb{P}[X_{v,w} = 1] = np.$$

Da X_w binomialverteilt ist, kann die Ungleichung von Chernoff angewandt werden (Abschnitt 2.2.2). Sei dazu Y_w^- eine Zufallsvariable, welche den Wert 1 hat, wenn $X_w \leq (1-\epsilon)np$ ist und welche ansonsten den Wert 0 annimmt, und sei Y_w^+ eine Zufallsvariable, die den Wert 1 annimmt, wenn $X_w \geq (1+\epsilon)np$ ist, und welche sonst den Wert 0 hat.

$$\begin{aligned}
\mathbb{P}\left[Y_w^- = 1\right] &= \mathbb{P}[X_w \leq (1-\epsilon)np] \\
&= \mathbb{P}[X_w \leq \mathbb{E}[X_w] - \epsilon np] \\
&\overset{(4)}{\leq} \exp\left(-\frac{(\epsilon np)^2}{2np}\right) \\
&= \exp\left(-\frac{1}{2}\epsilon^2 np\right) \\
&= \exp\left(-\frac{1}{2}\epsilon^2 \gamma(n) \ln n\right) \\
&= n^{-\frac{1}{2}\epsilon^2\gamma(n)}
\end{aligned}$$

$$\begin{aligned}
\mathbb{P}\left[Y_w^+ = 1\right] &= \mathbb{P}[X_w \geq (1+\epsilon)np] \\
&= \mathbb{P}[X_w \geq \mathbb{E}[X_w] + \epsilon np] \\
&\overset{(3)}{\leq} \exp\left(-\frac{(\epsilon np)^2}{2np + \frac{2}{3}\epsilon np}\right) \\
&< \exp\left(-\frac{\epsilon^2}{2+\epsilon} np\right) \\
&= \exp\left(-\frac{\epsilon^2}{2+\epsilon}\gamma(n) \ln n\right) \\
&= n^{-\frac{\epsilon^2}{2+\epsilon}\gamma(n)}.
\end{aligned}$$

Seien nun $Y^- = \sum_{w \in W} Y_w^-$ und $Y^+ = \sum_{w \in W} Y_w^+$ die Zufallsvariablen, welche die Anzahl der Featurecliquen mit nicht mehr als $(1-\epsilon)np$ bzw. mit nicht weniger als $(1+\epsilon)np$

Knoten messen. Nach der Markoffschen Ungleichung ergibt sich für die Wahrscheinlichkeiten, dass es jeweils mindestens eine solche Featureclique gibt:

$$\mathbb{P}\left[Y^- > 0\right] \overset{(2)}{\leq} \mathbb{E}\left[Y^-\right] = \sum_{w \in W} \mathbb{P}\left[Y_w^- = 1\right] \leq mn^{-\frac{1}{2}\epsilon^2\gamma(n)} = n^{\alpha-\frac{1}{2}\epsilon^2\gamma(n)}.$$

$$\mathbb{P}\left[Y^+ > 0\right] \overset{(2)}{\leq} \mathbb{E}\left[Y^+\right] = \sum_{w \in W} \mathbb{P}\left[Y_w^+ = 1\right] \leq mn^{-\frac{\epsilon^2}{2+\epsilon}\gamma(n)} = n^{\alpha-\frac{\epsilon^2}{2+\epsilon}\gamma(n)}.$$

Im Grenzwert für n gegen unendlich ergibt sich für die Summe der beiden Wahrscheinlichkeiten:

$$\lim_{n\to\infty}\Big(\mathbb{P}\left[Y^- > 0\right] + \mathbb{P}\left[Y^+ > 0\right]\Big) \leq \lim_{n\to\infty}\left(n^{\alpha-\frac{1}{2}\epsilon^2\gamma(n)} + n^{\alpha-\frac{\epsilon^2}{2+\epsilon}\gamma(n)}\right) = 0.$$

Fast jede Instanz von $\mathcal{G}_{n,m,p}$ enthält ausschließlich Featurecliquen mit mehr als $(1-\epsilon)np$ aber weniger als $(1+\epsilon)np$ Knoten. □

Wie muss ein Subgraph eines zufälligen Schnittgraphen zusammengesetzt sein, damit sein Minimalgrad größer oder gleich der Anzahl der Knoten in der größten Featureclique ist? Zunächst ist festzustellen, dass ein solcher Graph ausschließlich Knoten enthalten darf, die zu mehr als einer Featureclique gehören. Denn der Grad eines Knotens mit nur einem oder gar keinem Feature ist offensichtlich kleiner als $\hat{\omega}_W$.

Definition 14 (dick) Sei $G = (V, E)$ ein zufälliger Schnittgraph und $B = (V \cup W, E_B)$ sein bipartiter Generator. Ein Knoten $v \in V$ heißt *dick*, wenn er mindestens zwei Features hat (d.h. v hat mindestens zwei Nachbarn in W). Eine Kante $\{v_1, v_2\} \in E$ heißt *dick*, wenn sie zwischen zwei dicken Knoten verläuft.

Der folgende Satz beruht darauf, dass ein färbetechnisch problematischer Subgraph ausschließlich dicke Kanten enthalten darf.

Satz 3. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$ und $\frac{\ln n}{n} \ll p \ll \sqrt{\frac{1}{m^3}}$. Dann färbt der Greedy-Algorithmus auf einer smallest-last Anordnung fast jede Instanz von $\mathcal{G}_{n,m,p}$ optimal.*

Beweis. Sei $G = (V, E) \in \mathcal{G}_{n,m,p}$ ein zufälliger Schnittgraph. Nach dem Satz über die totale Wahrscheinlichkeit (siehe Abschnitt 2.2.3) genügt es die Schnittgraphen zu betrachten, in denen alle Featurecliquen mindestens $(1-\epsilon)np$ Knoten haben. Sei G ein solcher Graph.

Ein Subgraph von G mit einem Minimalgrad $\delta \geq \chi(G)$ muss mindestens $(1-\epsilon)np$ dicke Knoten enthalten, von denen jeder einzelne mit mindestens ebenso vielen anderen dicken Knoten verbunden ist. Im Folgenden soll die Wahrscheinlichkeit dafür bestimmt werden.

Für zwei beliebige Knoten $v, v' \in V$ sei $X_{v,v'}$ eine 0-1-Zufallsvariable, welche genau dann den Wert 1 annimmt, wenn v und v' dick und durch eine Kante verbunden sind. $X_{v,v'}$ ist also genau dann 1, wenn eine der beiden in Abbildung 2 dargestellten Konfigurationen im bipartiten Generator von G auftritt.

Die Wahrscheinlichkeit dafür lässt sich abschätzen durch

$$\mathbb{P}\left[X_{v,v'} = 1\right] \leq m^3p^4 + m^2p^4 \leq 2m^3p^4.$$

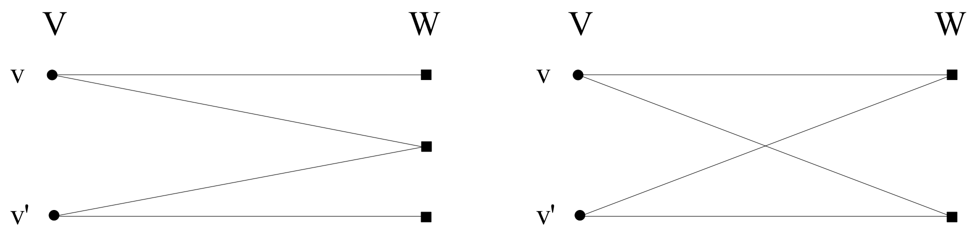

Abbildung 2: die beiden Konfigurationen, welche implizieren, dass v und v' dick und miteinander verbunden sind

Die Zufallsvariable $X = \frac{1}{2}\sum_{v\in V}\sum_{v'\in V\setminus\{v\}} X_{v,v'}$ misst die Anzahl der Kanten, die zwischen dicken Knoten verlaufen. Für den Erwartungswert von X gilt:

$$\mathbb{E}\left[X\right] \;=\; \frac{1}{2}\sum_{v\in V}\sum_{v'\in V\setminus\{v\}} \mathbb{P}\left[X_{v,v'}=1\right] \;\leq\; n^2m^3p^4.$$

Ein Subgraph mit mindestens $(1-\epsilon)np$ dicken Knoten, von denen jeder einzelne mit mindestens $(1-\epsilon)np$ anderen dicken Knoten verbunden ist, erfordert mindestens $\frac{1}{2}(1-\epsilon)^2n^2p^2$ dicke Kanten. Nach der Markoffschen Ungleichung kann die Wahrscheinlichkeit dafür abgeschätzt werden:

$$\begin{aligned}\mathbb{P}\left[X \geq \frac{1}{2}(1-\epsilon)^2n^2p^2\right] &\overset{(1)}{\leq} \frac{2}{(1-\epsilon)^2n^2p^2}\mathbb{E}\left[X\right]\\ &\leq \frac{2}{(1-\epsilon)^2}\; m^3p^2.\end{aligned}$$

Diese Wahrscheinlichkeit geht für n gegen unendlich gegen 0 (da nach Voraussetzung $p \ll \sqrt{\frac{1}{m^3}}$ gilt). Also enthält fast keine Instanz von $\mathcal{G}_{n,m,p}$ einen Subgraphen mit Minimalgrad $\delta \geq \chi(G)$. Der Greedy-Algorithmus färbt somit fast jede Instanz von $\mathcal{G}_{n,m,p}$ optimal auf einer smallest-last Anordnung der Knoten. $\square$

Dieses Resultat lässt sich noch verbessern, wenn man anstelle der Anzahl der dicken Kanten des zufälligen Schnittgraphen die Anzahl der dicken Nachbarn eines Knotens abschätzt. Dazu werden die folgenden beiden Lemmata benötigt.

Lemma 4. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$. Wenn $p \ll \frac{1}{m}$, dann gibt es eine Konstante $c > 0$, so dass fast keine Instanz von $\mathcal{G}_{n,m,p}$ einen Knoten mit mehr als $mp + c\ln n$ Features enthält.*

Beweis. Sei $G = (V,E) \in \mathcal{G}_{n,m,p}$ ein zufälliger Schnittgraph und sei X_v $(v \in V)$ eine Zufallsvariable, welche die Anzahl der Features eines Knotens zählt. Dann gilt:

$$\mathbb{E}\left[X_v\right] \;=\; mp.$$

Nach der Chernoffschen Ungleichung (siehe Abschnitt 2.2.2) gilt daher:

$$\begin{aligned}\mathbb{P}\left[X_v \geq mp + c\ln n\right] &\overset{(3)}{\leq} \exp\left(-\frac{(c\ln n)^2}{2mp+\frac{2}{3}c\ln n}\right)\\ &\leq \exp\left(-\frac{(c\ln n)^2}{c\ln n}\right)\\ &= \exp\left(-c\ln n\right).\end{aligned}$$

Für die zweite Abschätzung wurde benutzt, dass mp gegen 0 geht und damit ab einem gewissen n durch $\frac{1}{3}c\ln n$ abgeschätzt werden kann. Sei nun Y_v eine Zufallsvariable, welche den Wert 1 für $X_v \geq mp+c\ln n$ annimmt, und welche sonst 0 ist. Die Markoffsche Ungleichung liefert dann für $Y = \sum_{v\in V} Y_v$:

$$\mathbb{P}[Y > 0] \overset{(2)}{\leq} \mathbb{E}[Y] \leq n\exp(-c\ln n) = n^{1-c}.$$

Diese Wahrscheinlichkeit strebt für n gegen unendlich gegen 0, wenn $c > 1$. Demzufolge gibt es ein c, so dass fast keine Instanz von $\mathcal{G}_{n,m,p}$ Knoten mit mehr als $mp + c\ln n$ Features hat. □

Lemma 5. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$. Falls $\frac{\ln n}{n} \ll p \ll \sqrt{\frac{\ln n}{nm}}$, gibt es eine Konstante $c > 0$, so dass fast keine Instanz von $\mathcal{G}_{n,m,p}$ Featurecliquen mit mehr als $(1+\epsilon)nmp^2 + c\ln n$ dicken Knoten enthält.*

Beweis. Nach Lemma 2 (Größe der Featurecliquen) und dem Satz über die totale Wahrscheinlichkeit (siehe Abschnitt 2.2.3) genügt es Graphen zu betrachten, bei denen die Größe der Featurecliquen durch $(1-\epsilon)np$ und $(1+\epsilon)np$ beschränkt ist. Sei $G = (V,E) \in \mathcal{G}_{n,m,p}$ ein solcher zufälliger Schnittgraph. Sei nun $C \subseteq V$ eine beliebige Featureclique in G und bezeichne $|C|$ die Anzahl der Knoten in C. Sei außerdem $X_{C,v}$ für $v \in C$ eine Zufallsvariable, die den Wert 1 annimmt, wenn v dick ist, und welche ansonsten 0 ist. Dann gilt:

$$\mathbb{P}[X_{C,v} = 1] = 1-(1-p)^{m-1} \overset{(7)}{\leq} 1-\Big(1-(m-1)p\Big) \leq mp.$$

Sei $X_C = \sum_{v\in C} X_{C,v}$ die Zufallsvariable, welche die Anzahl der dicken Knoten in C zählt. Für den Erwartungswert von X_C gilt dann:

$$\mathbb{E}[X_C] = \sum_{v\in C} \mathbb{P}[X_{C,v} = 1] \leq |C|mp \leq (1+\epsilon)nmp^2.$$

Die Chernoff-Ungleichung (siehe Abschnitt 2.2.2) liefert nun folgende Abschätzung:

$$\begin{aligned}\mathbb{P}\left[X_C \geq (1+\epsilon)nmp^2 + c\ln n\right] &\leq \mathbb{P}[X_C \geq \mathbb{E}[X_C] + c\ln n] \\ &\overset{(3)}{\leq} \exp\left(-\frac{(c\ln n)^2}{2\mathbb{E}[X_C] + \frac{2}{3}c\ln n}\right) \\ &\leq \exp\left(-\frac{(c\ln n)^2}{2(1+\epsilon)nmp^2 + \frac{2}{3}c\ln n}\right).\end{aligned}$$

Sei Y_C eine Zufallsvariable, die für $X_C \geq (1+\epsilon)nmp^2 + c\ln n$ den Wert 1 annimmt und sonst 0 ist. Sei außerdem $Y = \sum_{C\subseteq V} Y_C$ die Anzahl der Featurecliquen mit mindestens $(1+\epsilon)nmp^2 + c\ln n$ dicken Knoten. Nach der Markoffschen Ungleichung gilt:

$$\mathbb{P}[Y > 0] \overset{(2)}{\leq} \mathbb{E}[Y] = \sum_{C\subseteq V} \mathbb{P}[Y_C = 1] \leq m\exp\left(-\frac{(c\ln n)^2}{2(1+\epsilon)nmp^2 + \frac{2}{3}c\ln n}\right).$$

Da $nmp^2 \ll \ln n$ ist, kann $2(1+\epsilon)nmp^2$ durch $\frac{1}{3}c\ln n$ abgeschätzt werden.

$$\mathbb{P}[Y > 0] \leq m\exp\left(-\frac{(c\ln n)^2}{c\ln n}\right) = m\exp(-c\ln n) = n^{\alpha-c}.$$

Für n gegen unendlich strebt diese Wahrscheinlichkeit gegen 0, wenn $c > \alpha$. Also gibt es eine Konstante c, so dass fast keine Instanz von $\mathcal{G}_{n,m,p}$ Featurecliquen mit mehr als $(1+\epsilon)nmp^2 + c\ln n$ dicken Knoten enthält. □

Aus diesen beiden Lemmata kann der folgende Satz abgeleitet werden.

Satz 6. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$, dann färbt der Greedy-Algorithmus für $\frac{\ln^2 n}{n} \ll p \ll \sqrt{\frac{\ln n}{nm}}$ fast jede Instanz von $\mathcal{G}_{n,m,p}$ optimal auf einer smallest-last Anordnung der Knoten.*

Beweis. Auch dieser Beweis benutzt, dass Greedy auf einer smallest-last Anordnung der Knoten optimal färbt, wenn alle Subgraphen einen Minimalgrad kleiner der größten Featureclique haben. In Lemma 4 wird eine obere Schranke für die Anzahl der Features eines Knotens hergeleitet und Lemma 5 beschränkt die Anzahl der dicken Knoten je Featureclique. Das Produkt dieser beiden oberen Schranken liefert eine Abschätzung für die Anzahl der dicken Nachbarn eines dicken Knotens, und somit auch eine Abschätzung für den Minimalgrad von Subgraphen, die ausschließlich aus dicken Knoten bestehen.

Zunächst ist anzumerken, dass aus der Voraussetzung $\alpha < 1$ folgt. Denn für $\alpha \geq 1$ gilt

$$\sqrt{\frac{\ln n}{nm}} \leq \frac{\sqrt{\ln n}}{n} \leq \frac{\ln^2 n}{n}.$$

D.h. es gibt kein p, welches die Voraussetzung unter diesen Bedingungen erfüllen könnte.

Es gilt:

$$p \ll \sqrt{\frac{\ln n}{nm}} \ll \sqrt{\frac{\ln n}{nm}\,\frac{n}{m\ln n}} = \frac{1}{m}.$$

Damit ist die Voraussetzung für Lemma 4 erfüllt und es kann zusammen mit dem Satz über die totale Wahrscheinlichkeit nach Abschnitt 2.2.3 angenommen werden, dass G keine Knoten enthält, die zu mehr als $mp + O(\ln n)$ Featurecliquen gehören. Analog darf nach Lemma 5 angenommen werden, dass G keine Featurecliquen mit mehr als $(1+\epsilon)nmp^2 + O(\ln n)$ dicken Knoten enthält. Der Ausdruck

$$D = \Big(mp + O(\ln n)\Big)\Big((1+\epsilon)nmp^2 + O(\ln n)\Big)$$

liefert daher eine obere Schranke für die Anzahl dicker Nachbarn eines dicken Knotens in G.

Da mp für n gegen unendlich gegen 0 geht, kann es durch $\ln n$ nach oben abgeschätzt werden. Außerdem ist nach Voraussetzung $(1+\epsilon)nmp^2 \ll \ln n$ und daher ebenfalls durch $\ln n$ beschränkt. Es ergibt sich:

$$D \leq \Big(O(\ln n)\Big)\Big(O(\ln n)\Big) = O(\ln^2 n).$$

D ist eine obere Schranke für den Minimalgrad von Subgraphen von G, die ausschließlich aus dicken Knoten bestehen. Sollte D nun kleiner der Anzahl der Knoten in der größten Featureclique von G sein, so färbt Greedy auf smallest-last Anordnungen optimal. Da nach Voraussetzung $p \gg \frac{\ln n}{n}$ gilt, genügt es nach dem Satz über die totale Wahrscheinlichkeit Graphen zu betrachten, in denen jede Featureclique mindestens $(1-\epsilon)np$

Knoten hat, für ein beliebiges ϵ mit $0 < \epsilon < 1$. Also genügt es zu zeigen, dass D kleiner als $(1-\epsilon)np$ ist. Nach Voraussetzung gilt nun:

$$\begin{aligned} p &\gg \frac{\ln^2 n}{n} \\ \Rightarrow (1-\epsilon)np &\gg \ln^2 n \\ \Rightarrow (1-\epsilon)np &\geq D. \end{aligned}$$

□

Die Voraussetzung $p \ll \sqrt{\frac{\ln n}{nm}}$ wird dabei vor allem benötigt um Lemma 5 anwenden zu können. In diesem Lemma wird dann benutzt, dass der Erwartungswert für die Anzahl dicker Knoten einer Featureclique kleiner oder gleich $(1+\epsilon)nmp^2$ ist und somit langsamer als $\ln n$ wächst. Das ist jedoch keineswegs eine zwingende Voraussetzung. Eine ähnliche Rechnung kann auch für deutlich größere Erwartungswerte durchgeführt werden. Für diesen Fall nimmt das Lemma die folgende Form an.

Lemma 7. *Sei $m = n^\alpha$ für ein festes α mit $0 < \alpha < 1$. Falls $p \gg \sqrt{\frac{1}{nm}}$, so enthält fast keine Instanz von $\mathcal{G}_{n,m,p}$ Featurecliquen mit mehr als $(1+\epsilon)nmp^2 + \sqrt{nmp^2}\ln n$ dicken Knoten.*

Beweis. Sei $G = (V, E) \in \mathcal{G}_{n,m,p}$ ein zufälliger Schnittgraph. Wegen

$$p \gg \sqrt{\frac{1}{nm}} \gg \sqrt{\frac{1}{n^{1+\alpha}} \frac{\ln^2 n}{n^{1-\alpha}}} = \frac{\ln n}{n}$$

ist die Voraussetzung für Lemma 2 erfüllt. Zusammen mit dem Satz über die totale Wahrscheinlichkeit (siehe Abschnitt 2.2.3) kann daher angenommen werden, dass die Größe der Featurecliquen von G durch $(1-\epsilon)np$ und $(1+\epsilon)np$ beschränkt ist (für ein beliebiges ϵ mit $0 < \epsilon < 1$). Sei nun $C \subseteq V$ eine beliebige Featureclique in G und bezeichne $|C|$ die Anzahl ihrer Knoten. Sei außerdem $X_{C,v}$ ($v \in C$) eine Zufallsvariable, die den Wert 1 annimmt, wenn v dick ist, und welche ansonsten 0 ist. Dann gilt:

$$\mathbb{P}[X_{C,v} = 1] = 1 - (1-p)^{m-1} \overset{(7)}{\leq} 1 - \Big(1 - (m-1)p\Big) \leq mp.$$

Sei $X_C = \sum_{v \in C} X_{C,v}$ die Zufallsvariable, welche die Anzahl der dicken Knoten in C zählt. Für den Erwartungswert von X_C gilt dann:

$$\mathbb{E}[X_C] = \sum_{v \in C} \mathbb{P}[X_{C,v} = 1] \leq |C|mp \leq (1+\epsilon)nmp^2.$$

Die Chernoff-Ungleichung (siehe Abschnitt 2.2.2) liefert folgende Abschätzung:

$$\begin{aligned} \mathbb{P}\left[X_C \geq (1+\epsilon)nmp^2 + \sqrt{nmp^2}\ln n\right] &\leq \mathbb{P}\left[X_C \geq \mathbb{E}[X_C] + \sqrt{nmp^2}\ln n\right] \\ &\overset{(3)}{\leq} exp\left(-\frac{nmp^2 \ln^2 n}{2\mathbb{E}[X_C] + \frac{2}{3}\sqrt{nmp^2}\ln n}\right) \\ &\leq exp\left(-\frac{nmp^2 \ln^2 n}{4nmp^2 + \sqrt{nmp^2}\ln n}\right). \end{aligned}$$

Sei Y_C eine Zufallsvariable, die für $X_C \geq (1+\epsilon)nmp^2 + \sqrt{nmp^2}\ln n$ den Wert 1 annimmt und sonst 0 ist. Sei außerdem $Y = \sum_{C \subseteq V} Y_C$ die Anzahl der Featurecliquen mit mindestens $(1+\epsilon)nmp^2 + \sqrt{nmp^2}\ln n$ dicken Knoten. Die Markoffsche Ungleichung liefert für $\mathbb{P}[Y > 0]$:

$$\mathbb{P}[Y > 0] \overset{(2)}{\leq} \mathbb{E}[Y] = \sum_{C \subseteq V} \mathbb{P}[Y_C = 1] \leq m \exp\left(-\frac{nmp^2 \ln^2 n}{4nmp^2 + \sqrt{nmp^2}\ln n}\right).$$

Hier setzt eine Fallunterscheidung an.

1. Fall: $4nmp^2 \geq \sqrt{nmp^2}\ln n$

$$\begin{aligned} \mathbb{P}[Y > 0] &\leq m \exp\left(-\frac{nmp^2 \ln^2 n}{8nmp^2}\right) \\ &= m \exp\left(-\frac{1}{8}\ln^2 n\right) \\ &= n^{\alpha - \frac{1}{8}\ln n} \\ &\to 0 \end{aligned}$$

2. Fall: $4nmp^2 < \sqrt{nmp^2}\ln n$

$$\begin{aligned} \mathbb{P}[Y > 0] &\leq m \exp\left(-\frac{nmp^2 \ln^2 n}{2\sqrt{nmp^2}\ln n}\right) \\ &= m \exp\left(-\frac{1}{2}\sqrt{nmp^2}\ln n\right) \\ &= n^{\alpha - \frac{1}{2}\sqrt{nmp^2}} \\ &\to 0. \end{aligned}$$

Für den letzten Schritt wurde verwendet, dass nach Voraussetzung $\sqrt{nmp^2}$ für n gegen unendlich gegen ∞ läuft.

$\mathbb{P}[Y > 0]$ ist für jedes n durch das Maximum dieser beiden Terme beschränkt und geht für n gegen unendlich gegen 0. Somit enthält fast keine Instanz von $\mathcal{G}_{n,m,p}$ Featurecliquen mit mehr als $(1+\epsilon)nmp^2 + \sqrt{nmp^2}\ln n$ dicken Knoten. □

Lemma 4 und Lemma 7 erlauben nun eine zweite Version von Satz 6 für einen anderen Bereich von p.

Satz 8. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$. Wenn nun gilt:*

$$\sqrt{\frac{1}{nm}} \ll p \ll \frac{1}{m \ln n},$$

so färbt der Greedy-Algorithmus auf einer smallest-last Anordnung fast jede Instanz von $\mathcal{G}_{n,m,p}$ optimal.

Beweis. Der Beweis läuft nach demselben Schema wie bei Satz 6. Wie in Satz 6 folgt aus der Voraussetzung dass $\alpha < 1$ ist. Denn für $\alpha \geq 1$ gilt:

$$\frac{1}{m \ln n} \ll \frac{1}{m} \leq \sqrt{\frac{1}{nm}},$$

so dass kein p die Voraussetzung erfüllen kann.

Sei G ein zufälliger Schnittgraph. Ein Subgraph von G, dessen Minimalgrad größer oder gleich der die Größe der größten Featureclique von G ist, darf nur aus dicken Knoten bestehen. Der Minimalgrad eines solchen Subgraphen ist durch die maximale Anzahl von dicken Nachbarn eines dicken Knotens in G begrenzt. Nach Voraussetzung ist $p \ll \frac{1}{m}$ und somit die Voraussetzung für Lemma 4 erfüllt. Zusammen mit dem Satz über die totale Wahrscheinlichkeit folgt, dass es genügt Graphen zu betrachten, welche keine Knoten enthalten, die zu mehr als $mp + O(\ln n)$ Featurecliquen gehören. Analog darf nach Lemma 7 angenommen werden, dass keine Featurecliquen mit mehr als $(1+\epsilon)nmp^2 + \sqrt{nmp^2}\ln n$ dicken Knoten vorliegen. Folglich liefert

$$D = \Big(mp + O(\ln n)\Big)\Big((1+\epsilon)nmp^2 + \sqrt{nmp^2}\ln n\Big)$$

eine obere Schranke für die Anzahl dicker Nachbarn eines dicken Knoten in G.

Da mp für n gegen unendlich gegen 0 strebt, kann es ab einem gewissen n durch $\ln n$ abgeschätzt werden. Es ergibt sich:

$$\begin{aligned} D &\leq \Big(O(\ln n)\Big)\Big(2nmp^2 + \sqrt{nmp^2}\ln n\Big) \\ &= nmp^2 O(\ln n) + \sqrt{nmp^2}\,O(\ln^2 n) \\ &= \left(mpO(\ln n) + \sqrt{\frac{m}{n}}\,O(\ln^2 n)\right) np. \end{aligned}$$

Nach Voraussetzung geht $mp\ln n$ für n gegen unendlich gegen 0. Ebenso strebt $\sqrt{\frac{m}{n}}\ln^2 n$ gegen 0, da $\alpha < 1$ ist. Somit folgt für ein $\epsilon < 1$:

$$D \leq (1-\epsilon)np.$$

D ist eine obere Schranke für den Minimalgrad von Subgraphen in G, die ausschließlich aus dicken Knoten bestehen. Wie in Lemma 7 darf angenommen werden, dass die Größe der Featurecliquen nach unten durch $(1-\epsilon)np$ beschränkt ist. Damit folgt dann nach der selben Argumentation wie in Satz 6, dass der Greedy-Algorithmus den Graphen G auf einer smallest-last Anordnung der Knoten fast sicher optimal färbt. □

3.3 Zusammenfassung

In diesem Kapitel wurde gezeigt unter welchen Bedingungen zufällige Schnittgraphen optimal gefärbt werden können. Die folgende Tabelle liefert eine Übersicht zu den erzielten Resultaten.

Satz 1		p	$\ll$	$\sqrt{\frac{1}{nm}}$	fast alle Instanzen sind chordal
Satz 3	$\frac{\ln n}{n} \ll$	p	$\ll$	$\sqrt{\frac{1}{m^3}}$	Greedy färbt fast alle Instanzen
Satz 6	$\frac{\ln^2 n}{n} \ll$	p	$\ll$	$\sqrt{\frac{\ln n}{nm}}$	auf smallest-last Knotenanordnungen
Satz 8	$\sqrt{\frac{1}{nm}} \ll$	p	$\ll$	$\frac{1}{m\ln n}$	optimal

Damit die Voraussetzungen von Satz 3 erfüllbar sind, muss $\alpha < \frac{2}{3}$ sein. Die Sätze 6 und 8 erzwingen $\alpha < 1$. Die von den Sätzen 6 und 8 abgedeckten Bereiche für p können zusammengefasst werden und ersetzen damit Satz 3 fast vollständig. Insgesamt ergibt sich damit folgendes Bild.

	p	$\ll \sqrt{\frac{1}{nm}}$		fast alle Instanzen sind chordal
$\frac{\ln n}{n} \ll$	p	$\ll n^{\epsilon-1}$	$\left(\alpha < \frac{2}{3}\right)$	Greedy färbt fast alle Instanzen auf
$\frac{\ln^2 n}{n} \ll$	p	$\ll \frac{1}{m \ln n}$	$(\alpha < 1)$	smallest-last Anordnungen optimal

Diese Resultate liefern ein Verfahren, welches fast alle zufälligen Schnittgraphen in der Vereinigung der Parameterbereiche (d.h. im Wesentlichen $p \ll \sqrt{\frac{1}{nm}}$ und $p \ll \frac{1}{m \ln n}$ bei $\alpha < 1$) optimal färbt. Dazu überprüft man zunächst, ob der vorliegende Graph chordal ist (in linearer Zeit möglich). Wenn ja, so kann der Graph mit einem Färbealgorithmus für chordale Graphen in linearer Zeit optimal gefärbt werden. Wenn nicht, so wird eine smallest-Anordnung der Knoten bestimmt und der Greedy Algorithmus darauf angewandt.

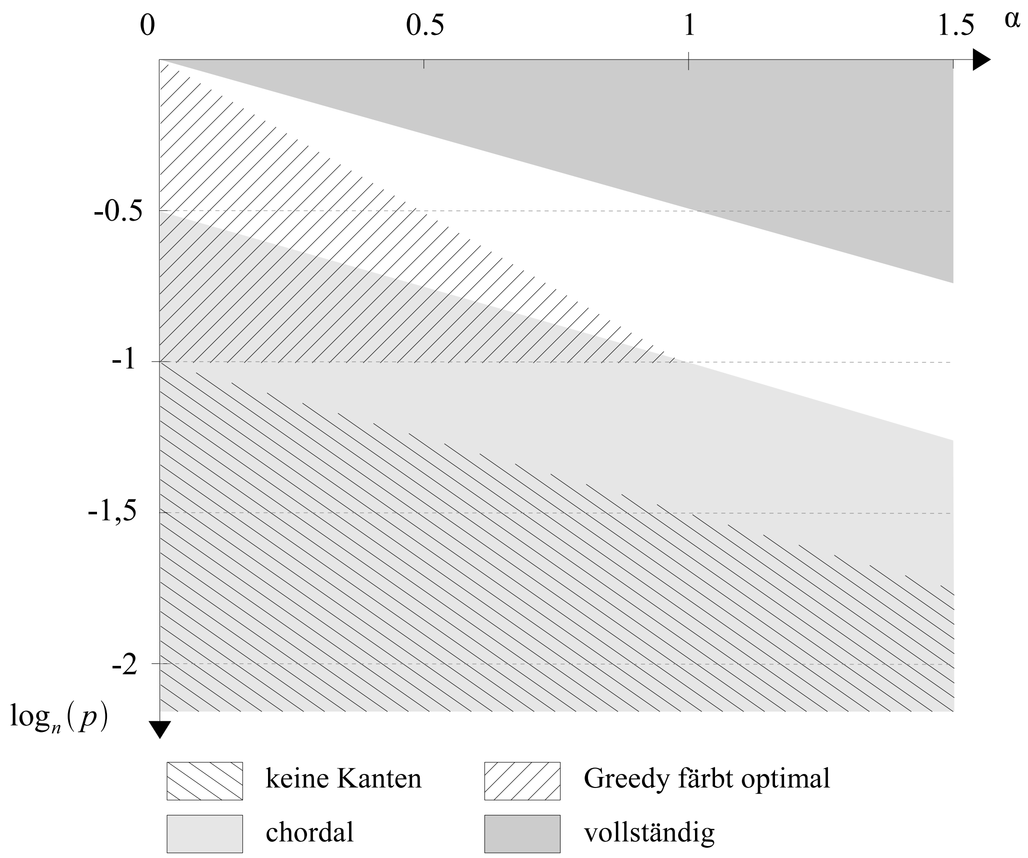

Abbildung 3: Bereiche, in denen zufällige Schnittgraphen optimal gefärbt werden können

Abbildung 3 fasst die Aussagen dieses Kapitels zusammen und enthält zusätzlich die Resultate von Singer bezüglich des Auftretens von Kanten sowie Vollständigkeit von $\mathcal{G}_{n,m,p}$ (siehe Abschnitt 2.4).

Das weiße nicht-schraffierte Gebiet beschreibt den Bereich, zu dem bisher keine Aussagen gemacht werden können. Wie lassen sich also Graphen aus diesem Bereich färben? Zunächst stellt sich die Frage, inwieweit sich die Ergebnisse dieses Kapitels verbessern lassen. Wie bereits erwähnt wurde in [7] gezeigt, dass $\sqrt{\frac{1}{nm}}$ eine Schwellenwertfunktion für das Auftauchen von induzierten Kreisen fester Länge (≥ 4) darstellt. Die erste Aussage zu chordalen Graphen kann daher nicht weiter verbessert werden. Bei den Aussagen zum Greedy-Algorithmus ist eine Überprüfung der Grenzen wesentlich schwieriger. Die unteren Grenzen für p scheinen vor allem ein technisches Problem zu sein. Der Autor hat keinen Anlass zu vermuten, dass der Greedy-Algorithmus für kleinere p (d.h. Graphen mit noch weniger Kanten) plötzlich schlechtere Resultate liefert. Weit interessanter ist jedoch die Frage, wie scharf die obere Grenze für p ist.

Dazu ist zunächst anzumerken, dass das bisher beste Resultat zum Greedy Algorithmus (die Sätze 6 und 8) darauf basiert, dass der Knotengrad in Subgraphen aus dicken Knoten kleiner als die chromatischen Zahl ist. Sobald mp gegen unendlich strebt, sind jedoch fast alle Knoten dick. D.h. das Verfahren der genannten Sätze liefe im Wesentlichen auf eine Untersuchung des Maximalgrads der Graphen hinaus, der sogar unabhängig von der Knotenanordnung eine Schranke für die Anzahl der vom Greedy-Algorithmus verwendeten Farben liefert. Außerdem ist zu erwarten, dass die bisherige untere Schranke für die chromatische Zahl (die Größe der Featurecliquen) mit steigendem p zunehmend schlechter wird. Aus diesem Grund beschäftigt sich das nächste Kapitel mit der Stabilitätszahl von zufälligen Schnittgraphen. Dabei wird besonderes Augenmerk auf den bisher nicht abgedeckten Parameterraum gelegt. Mit den Resultaten zur Stabilitätszahl soll dann in Kapitel 5 eine neue untere Schranke für die chromatische Zahl abgeleitet werden, mit deren Hilfe die Approximationsgüte des Greedy-Färbealgorithmus abgeschätzt werden kann.

4 Die Stabilitätszahl von zufälligen Schnittgraphen

Ebenso wie das Färben von Graphen stellt auch das Finden der größten stabilen Knotenmenge bzw. das Ermitteln der Stabilitätszahl ein NP-vollständiges Problem dar. Auch die Approximation dieser Größe ist schwierig. So zeigten Arora, Lund, Motwani, Sudan und Szegedy 1992, dass unter der Voraussetzung $P \neq NP$ ein $\epsilon > 0$ existiert, so dass es keinen polynomiellen Approximationsalgorithmus für die Stabilitätszahl mit einer Güte kleiner oder gleich n^ϵ gibt [3]. Erstaunlicherweise konnte dennoch gezeigt werden, dass bei fast allen $G_{n,p}$-Instanzen die Stabilitätszahl in einem vergleichsweise kleinen Bereich liegt. So wurde nachgewiesen, dass für die Stabilitätszahl fast aller $G_{n,p}$-Instanzen mit $n^{-c} \leq p \leq c$ (für $c < 1$ beliebig) gilt:

$$k_{-\epsilon} \;\leq\; \beta \;\leq\; k_\epsilon$$

$$\text{für } k_{\pm\epsilon} = \left\lfloor \frac{2}{p}(2\log np - \log\log np + 1 - \log 2 \pm \epsilon) \right\rfloor \text{ und } \epsilon > 0 \text{ beliebig [9].}$$

In diesem Kapitel soll nun untersucht werden, wie groß die größte stabile Menge in zufälligen Schnittgraphen ist. In [7] wurde bereits gezeigt dass:

1. für $p \ll \frac{1}{n\sqrt{m}}$ fast kein Instanz von $\mathcal{G}_{n,m,p}$ Kanten enthält und
2. für $p \gg \sqrt{\frac{\ln n}{m}}$ fast alle Instanzen von $\mathcal{G}_{n,m,p}$ vollständig sind.

Die Untersuchung wird sich daher auf den Bereich zwischen diesen beiden Schranken konzentrieren.

Außerdem wurde in [7] auch gezeigt, dass für fast alle Instanzen von $\mathcal{G}_{n,m,p}$ bei $mp^2 \to c$ gilt:

$$\lfloor d - \epsilon_1 \rfloor \;\leq\; \beta \;\leq\; \lfloor d + \epsilon_2 \rfloor,$$

wobei $d = 2\log_b n - 2\log_b \log_b n + 1 + 2\log_b\left(\frac{e}{2}\right)$ mit $b = e^c$, $\epsilon_1 > \frac{c^2}{4}$ und $\epsilon_2 > 0$ beliebig sind.

Wenn man versucht die Stabilitätszahl für einen möglichst großen Bereich für p abzuschätzen, so hilft dieses Resultat sicherlich nur begrenzt weiter, da es nur ein sehr kleines p-Intervall betrifft. Die Herleitung der unteren Schranke war jedoch bei der Ausarbeitung der Beweise in 4.2 eine wertvolle Hilfe.

4.1 Schranken für die Stabilitätszahl

Zunächst soll die Stabilitätszahl für kleine Featurewahrscheinlichkeiten p untersucht werden. Der folgende Satz benutzt, dass die Anzahl der Knoten ohne Features eine untere Schranke für die Stabilitätszahl liefert, da diese Knoten isoliert sind.

Satz 9. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$ und sei $p \ll \frac{1}{m}$, dann gibt es in fast allen Instanzen von $\mathcal{G}_{n,m,p}$ eine stabile Menge der Größe $\Big(1 - o(1)\Big)n$.*

Beweis. Sei $G = (V, E) \in \mathcal{G}_{n,m,p}$ ein zufälliger Schnittgraph. Es genügt zu zeigen, dass eine Funktion $\gamma(n) \ll 1$ existiert, so dass G fast sicher $\Big(1 - \gamma(n)\Big)n$ Knoten ohne

Features hat. Sei dazu $\gamma(n) = \sqrt{\frac{\ln n}{n} + mp}$, dann ist $\gamma(n) \ll 1$, da $\sqrt{\ln n} \ll \sqrt{n}$ ist und mp für n gegen unendlich gegen 0 strebt.

Sei außerdem X_v für einen Knoten $v \in V$ eine 0-1-Zufallsvariable, welche genau dann 1 ist, wenn v keiner Featureclique angehört. Es gilt:

$$\mathbb{P}[X_v = 1] \;=\; (1-p)^m.$$

Sei nun $X = \sum_{v \in V} X_v$ die Anzahl der Knoten ohne Features. Für den Erwartungswert von X gilt:

$$\mathbb{E}[X] \;=\; \sum_{v \in V} \mathbb{P}[X_v = 1] \;=\; n(1-p)^m \;\stackrel{(7)}{\geq}\; (1-mp)n.$$

Da X binomialverteilt ist, kann die Ungleichung von Chernoff angewandt werden (siehe Abschnitt 2.2.2).

$$\begin{aligned}
\mathbb{P}\left[X \leq \Big(1-\gamma(n)\Big)n\right] &= \mathbb{P}\left[X \leq \mathbb{E}[X] - \gamma(n)n - \mathbb{E}[X] + n\right] \\
&\leq \mathbb{P}\left[X \leq \mathbb{E}[X] - \Big(\gamma(n) + 1 - mp - 1\Big)n\right] \\
&= \mathbb{P}\left[X \leq \mathbb{E}[X] - \Big(\gamma(n) - mp\Big)n\right] \\
&\stackrel{(4)}{\leq} \exp\left(-\frac{\Big(\gamma(n)-mp\Big)^2 n^2}{2\mathbb{E}[X]}\right) \\
&\leq \exp\left(-\frac{1}{2}\Big(\gamma(n)-mp\Big)^2 n\right) \qquad \text{da } \mathbb{E}[X] \leq n \\
&\leq \exp\left(-\frac{1}{2}\ln n\right) \\
&\to 0.
\end{aligned}$$

In fast allen Graphen aus $\mathcal{G}_{n,m,p}$ gibt es folglich mehr als $\Big(1-o(1)\Big)n$ isolierte Knoten und somit eine stabile Menge dieser Größe. ☐

Die Stabilitätszahl von fast allen zufälligen Schnittgraphen liegt damit für $p \ll \frac{1}{m}$ zwischen $(1-o(1))n$ und n. Es bleibt zu untersuchen, wie sie sich für größere p verhält. Ein erstes Resultat ergibt sich durch Betrachtung des Maximalgrads.

Satz 10. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$ und sei $p \gg \max\left(\frac{\ln n}{n}, \frac{\ln n}{m}\right)$, dann gibt es für ein beliebiges $\epsilon > 0$ in fast allen Instanzen von $\mathcal{G}_{n,m,p}$ eine stabile Menge der Größe $(1-\epsilon)\frac{1}{mp^2}$.*

Beweis. Sei $G = (V,E) \in \mathcal{G}_{n,m,p}$ ein zufälliger Schnittgraph. Der Beweis beruht darauf, dass der Maximalgrad von G nach oben beschränkt ist durch die maximale Anzahl der Features eines Knotens multipliziert mit der maximalen Anzahl von Knoten eines Features. Diese Beschränkung für den Maximalgrad liefert eine obere Schranke für die chromatische Zahl und damit eine untere Schranke für die Stabilitätszahl, denn es gilt:

$$\begin{aligned}
\Delta(G) + 1 \;\geq\; \chi(G) \;&\geq\; \frac{n}{\beta(G)} \\
\Rightarrow \quad \beta(G) \;&\geq\; \frac{n}{\Delta(G)+1}.
\end{aligned}$$

Für einen Knoten $v \in V$ sei X_v eine Zufallsvariable, welche die Anzahl der Features dieses Knotens zählt. Dann gilt:

$$\mathbb{E}[X_v] \ = \ mp.$$

Da X_v binomialverteilt ist, kann die Ungleichung von Chernoff angewandt werden (siehe Abschnitt 2.2.2). Für ein genügend kleines $\epsilon' > 0$ gilt:

$$\begin{aligned} \mathbb{P}\left[X_v \geq (1+\epsilon')mp\right] &= \mathbb{P}\left[X_v \geq \mathbb{E}[X_v] + \epsilon' mp\right] \\ &\overset{(3)}{\leq} \exp\left(-\frac{(\epsilon' mp)^2}{2\mathbb{E}[X_v] + \frac{2}{3}\epsilon' mp}\right) \\ &\leq \exp\left(-\frac{\epsilon'^2}{3}mp\right). \end{aligned}$$

Da $p \gg \frac{\ln n}{m}$ ist, kann es als $\gamma(n)\frac{\ln n}{m}$ geschrieben werden für eine gegen unendlich strebende Funktion γ. Die Wahrscheinlichkeit für die Existenz eines Knotens in G mit mehr als $(1+\epsilon')mp$ Features lässt sich abschätzen durch:

$$\begin{aligned} \mathbb{P}\left[\exists_{v\in V} : X_v > (1+\epsilon')mp\right] &\leq n\mathbb{P}\left[X_v \geq (1+\epsilon')mp\right] \\ &\leq n \exp\left(-\frac{\epsilon'^2}{3}\gamma(n)\ln n\right) \\ &\leq n^{1-\frac{1}{3}\epsilon'^2\gamma(n)} \\ &\to 0. \end{aligned}$$

Nach dem Satz über die totale Wahrscheinlichkeit genügt es den Fall zu betrachten, dass G keinen Knoten mit mehr $(1+\epsilon')mp$ Features hat. Außerdem gilt $p \gg \frac{\ln n}{n}$, und es kann nach Lemma 2 angenommen werden, dass G ausschließlich Featurecliquen mit weniger als $(1+\epsilon')np$ Knoten enthält. Damit gilt für den Maximalgrad von G:

$$\Delta(G) \ \leq \ (1+\epsilon')^2 nmp^2.$$

Für die Stabilitätszahl von G bedeutet das bei hinreichend kleinem ϵ':

$$\beta(G) \ \geq \ \frac{n}{\Delta(G)+1} \ \geq \ \frac{1}{(1+\epsilon')^2 mp^2 + \frac{1}{n}} \ \geq \ (1-\epsilon)\frac{1}{mp^2}.$$

□

Auf den ersten Blick mag es erstaunlich erscheinen, dass dieser Beweis nicht für beliebig kleine p gilt, da ja kleinere Werte für p zu dünneren Graphen mit höherer Stabilitätszahl führen. Allerdings ist die hier bewiesene Schranke für die Stabilitätszahl abhängig von p und steigt, wenn p sinkt.

Nach dieser unteren Schranke folgt die Bestimmung einer oberen Schranke, welche mit Hilfe der 1. Moment Methode abgeleitet wird.

Satz 11. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$ und sei $\frac{\ln n}{m} \ll p \ll \sqrt{\frac{\ln n}{m}}$, dann ist in fast allen Instanzen von $\mathcal{G}_{n,m,p}$ die Stabilitätszahl kleiner als $(2+\epsilon)\frac{\ln n}{mp^2}$ für ein beliebiges aber festes $\epsilon > 0$.*

Beweis. Es wird gezeigt, dass die erwartete Anzahl stabiler Mengen der Größe $(2+\epsilon)\frac{\ln n}{mp^2}$ für n gegen unendlich gegen 0 strebt. Daraus folgt nach der 1. Moment Methode, dass die Wahrscheinlichkeit für die Existenz einer solchen stabilen Menge ebenfalls gegen 0 läuft, und somit fast keine Instanz von $\mathcal{G}_{n,m,p}$ eine stabile Menge dieser Größe enthält.

Sei $G = (V, E) \in \mathcal{G}_{n,m,p}$ ein zufälliger Schnittgraph. Eine Menge von k Knoten ist genau dann stabil, wenn keine Kanten zwischen ihnen verlaufen. Also muss für jede Featureclique von G gelten:

1. sie enthält keinen der k Knoten
2. oder sie enthält genau einen der k Knoten.

Die Wahrscheinlichkeit für eine gegebene Menge $S_k \subseteq V$ mit k Knoten stabil zu sein beträgt daher:

$$\mathbb{P}\left[S_k \text{ ist stabil}\right] = \Big((1-p)^k + kp(1-p)^{k-1}\Big)^m.$$

Sei X_k die Zufallsvariable, welche die Anzahl der stabilen Mengen mit k Knoten in G zählt. Dann gilt wegen der Linearität des Erwartungswerts:

$$\begin{aligned}
\mathbb{E}\left[X_k\right] &= \binom{n}{k}\Big((1-p)^k + kp(1-p)^{k-1}\Big)^m \\
&\overset{(9)}{\leq} \left(\frac{ne}{k}\right)^k (1-p)^{(k-1)m}\Big(1-p+kp\Big)^m \\
&= \left(\frac{n}{k}\right)^k e^k (1-p)^{(k-1)m}\Big(1+(k-1)p\Big)^m.
\end{aligned}$$

Sei $k = (2+\epsilon)\frac{\ln n}{mp^2}$. Um zu zeigen dass $\mathbb{E}\left[X_k\right]$ gegen 0 geht genügt es zu zeigen, dass $\ln(\mathbb{E}\left[X_k\right])$ gegen $-\infty$ läuft.

$$\begin{aligned}
\ln(\mathbb{E}\left[X_k\right]) &\leq \ln\left(\left(\frac{n}{k}\right)^k e^k (1-p)^{(k-1)m}\Big(1+(k-1)p\Big)^m\right) \\
&= k\ln\left(\frac{n}{k}\right) + k + (k-1)m\ln(1-p) + m\ln\Big(1+(k-1)p\Big).
\end{aligned}$$

Über die Taylorentwicklung erhält man die folgenden beiden Abschätzungen:

$$\ln(1-p) = -\sum_{i=1}^{\infty}\frac{1}{i}p^i \leq -p - \frac{1}{2}p^2 - \frac{1}{3}p^3$$

$$\begin{aligned}
\ln\Big(1+(k-1)p\Big) &= \sum_{i=1}^{\infty}(-1)^{(i-1)}\frac{1}{i}(k-1)^i p^i \\
&\leq (k-1)p - \frac{1}{2}(k-1)^2p^2 + \frac{1}{3}(k-1)^3p^3.
\end{aligned}$$

Im letzten Schritt der zweiten Abschätzung wurde verwendet, dass die Beträge der Reihenglieder mit fortschreitendem i immer kleiner werden. Die nach einem positiven Glied abgebrochene Reihe liefert dann eine obere Schranke, weil in dem Rest der Reihe immer zunächst etwas abgezogen und danach etwas betragsmäßig kleineres addiert wird.

Für den Erwartungswert ergibt sich:

$$\begin{aligned}
\ln(\mathbb{E}\left[X_k\right]) &\leq k\ln\left(\frac{n}{k}\right) + k - (k-1)m\left(p + \frac{1}{2}p^2 + \frac{1}{3}p^3\right) \\
&\quad + m\left((k-1)p - \frac{1}{2}(k-1)^2p^2 + \frac{1}{3}(k-1)^3p^3\right) \\
&= k\ln\left(\frac{n}{k}\right) + k + \frac{1}{2}kmp^2 + \frac{1}{3}k^3mp^3 + \frac{2}{3}kmp^3 - \frac{1}{2}k^2mp^2 - k^2mp^3 \\
&\leq k\left(\ln\left(\frac{n}{k}\right) + 1 + \frac{1}{2}mp^2 + \frac{1}{3}k^2mp^3 + \frac{2}{3}mp^3 - \frac{1}{2}kmp^2\right) \\
&\leq k\left(\ln\left(\frac{n}{k}\right) + 1 + \underbrace{mp^2}_{=:A} + \underbrace{2\frac{\ln^2 n}{mp}}_{=:B} + \underbrace{mp^3}_{=:C} - \frac{2+\epsilon}{2}\ln n\right).
\end{aligned}$$

Die 1, sowie die Terme A, B und C können jeweils durch $\frac{1}{16}\epsilon\ln n$ abgeschätzt werden, weil gilt:

$$\begin{array}{lllll}
A & \ll & \ln n, & & \text{da nach Voraussetzung } p \ll \sqrt{\frac{\ln n}{m}} \\
B & = & 2\,\frac{\ln n}{mp}\ln n & \ll \ \ln n, & \text{da nach Voraussetzung } p \gg \frac{\ln n}{m} \\
C & \ll & A & \ll \ \ln n. &
\end{array}$$

Damit ergibt sich:

$$\begin{aligned}
\ln(\mathbb{E}\left[X_k\right]) &\leq k\left(\ln\left(\frac{n}{k}\right) - \left(1 + \frac{\epsilon}{4}\right)\ln n\right) \\
&\leq k\left(\ln n - \left(1 + \frac{\epsilon}{4}\right)\ln n\right) \\
&= -\frac{\epsilon}{4}k\ln n.
\end{aligned}$$

Dieser Term strebt für n gegen unendlich gegen $-\infty$, weil $k = (2+\epsilon)\frac{\ln n}{mp^2}$ gegen unendlich strebt. Der Erwartungswert von X_k strebt daher gegen 0. Nach der 1. Moment Methode folgt, dass fast keine Instanz von $\mathcal{G}_{n,m,p}$ eine stabile Menge der Größe k besitzt. □

Dieses Resultat vervollständigt das Bild zur Stabilitätszahl. Zusammen mit Satz 10 ergibt sich ein relativ kleiner Bereich, in den die Stabilitätszahl von fast allen zufälligen Schnittgraphen fällt. Für p oberhalb von $\sqrt{\frac{\ln n}{m}}$ sind fast alle zufälligen Schnittgraphen vollständig, und die Stabilitätszahl folglich fast sicher 1.

Noch offen ist die Frage, wie scharf diese Schranken sind. Für $\alpha \leq 1$ liefern die Sätze 10 und 11 die folgende Aussage für fast alle zufälligen Schnittgraphen:

$$\text{Wenn } \frac{\ln n}{m} \ll p \ll \sqrt{\frac{\ln n}{m}}\text{, so gilt:} \qquad (1-\epsilon)\frac{1}{mp^2} \ \leq \ \beta \ \leq \ (2+\epsilon)\frac{\ln n}{mp^2}.$$

Obere und untere Schranke liegen um einen Faktor von ca. $2\ln n$ auseinander.

Im Vergleich dazu ist die Stabilitätszahl von $G_{n,p}$ auf einen sehr kleinen Bereich beschränkt. Die dort hergeleiteten oberen und unteren Schranken unterscheiden sich lediglich um eine Konstante $4\epsilon\frac{1}{p}$ (7.1 und 7.4 in [9]). Zwar kann auch dieser Wert für hinreichend kleine p sehr groß werden, dafür beschreibt er jedoch bereits die echte Breite des Intervalls und nicht den Quotient zwischen oberer und unterer Schranke.

Verglichen mit dem obigen Resultat zur Stabilitätszahl in zufälligen Schnittgraphen sind die bekannten Resultate auf einfachen zufälligen Graphen wesentlich schärfer. Damit stellt sich die Frage, ob das $\mathcal{G}_{n,m,p}$-Resultat noch weiter verbessert werden kann. In der Ableitung der oberen Schranke wurde bereits sehr vorsichtig abgeschätzt, und es ist dem Autor nicht gelungen dieses Ergebnis zu verbessern. Es liegt daher nahe die untere Schranke einer genaueren Analyse zu unterwerfen.

4.2 Untere Schranken für die Stabilitätszahl nach der 2. Moment Methode

Die 2. Moment Methode stellt ein gängiges Mittel zum Beweisen solcher unteren Schranken dar. Um diese Methode anzuwenden wird eine Abschätzung des Erwartungswerts sowie eine Abschätzung der Varianz der zu untersuchenden Zufallsvariable benötigt. Diese Abschätzungen werden in den folgenden beiden Lemmata bereitgestellt.

Lemma 12. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$. Sei außerdem X_k eine Zufallsvariable, welche die Anzahl der stabilen Mengen der Größe k in Instanzen von $\mathcal{G}_{n,m,p}$ misst. Wenn für n gegen unendlich $kp \to 0$ geht, dann gilt für den Erwartungswert von X_k:*

$$\ln \mathbb{E}[X_k] \;\geq\; k\left(\ln\binom{n}{k} + \frac{1-\epsilon}{2}mp^2 - \frac{1}{2}kmp^2\right).$$

Beweis. Sei $S_k \in V(G)$ eine Menge mit k Knoten. Nach dem Beweis von Satz 11 gilt für die Wahrscheinlichkeit dass S_k stabil ist:

$$\mathbb{P}[S_k \text{ ist stabil}] \;=\; \Big((1-p)^k + kp(1-p)^{k-1}\Big)^m.$$

Daraus ergibt sich für den Erwartungswert:

$$\begin{aligned}\mathbb{E}[X_k] \;&=\; \binom{n}{k}\Big((1-p)^k + kp(1-p)^{k-1}\Big)^m \\ &\overset{(9)}{\geq}\; \left(\frac{n}{k}\right)^k (1-p)^{(k-1)m}\Big(1+(k-1)p\Big)^m.\end{aligned}$$

Für den Logarithmus des Erwartungswerts bedeutet das:

$$\ln \mathbb{E}[X_k] \;\geq\; k\ln\left(\frac{n}{k}\right) + (k-1)m\ln(1-p) + m\ln\Big(1+(k-1)p\Big).$$

Über die Taylorentwicklung erhält man die folgenden beiden Abschätzungen:

$$\ln(1-p) \;=\; -\sum_{i=1}^{\infty}\frac{1}{i}p^i \;\geq\; -\sum_{i=1}^{\infty}p^i + \frac{p^2}{2} \;\overset{(6)}{=}\; -\frac{1}{1-p} + 1 + \frac{p^2}{2} \;=\; -\frac{p}{1-p} + \frac{p^2}{2},$$

$$\ln\Big(1+(k-1)p\Big) \;=\; \sum_{i=1}^{\infty}(-1)^{(i-1)}\frac{1}{i}(k-1)^i p^i \;\geq\; (k-1)p - \frac{1}{2}(k-1)^2p^2.$$

Bei der zweiten Abschätzung wurde benutzt, dass $kp \to 0$ geht und daher die nach einem negativen Reihenglied abgebrochene Reihe eine untere Schranke liefert.

Es ergibt sich:

$$\begin{aligned}
\ln \mathbb{E}\left[X_k\right] &\geq k \ln \binom{n}{k} + (k-1)m\left(\frac{p^2}{2} - \frac{p}{1-p}\right) + m\left((k-1)p - \frac{1}{2}(k-1)^2 p^2\right) \\
&= k \ln \binom{n}{k} + \frac{1}{2}kmp^2 + \frac{mp}{1-p} + kmp + kmp^2 \\
&\quad -\frac{1}{2}mp^2 - \frac{kmp}{1-p} - mp - \frac{1}{2}k^2mp^2 - \frac{1}{2}mp^2 \\
&= k \ln \binom{n}{k} + \frac{3}{2}kmp^2 + mp\left(\frac{1}{1-p} - 1\right) \\
&\quad -mp^2 - \frac{1}{2}k^2mp^2 - kmp\left(\frac{1}{1-p} - 1\right) \\
&= k \ln \binom{n}{k} + kmp^2\left(\frac{3}{2} - \frac{1}{1-p}\right) + mp^2\left(\frac{1}{1-p} - 1\right) - \frac{1}{2}k^2mp^2 \\
&= k \ln \binom{n}{k} + kmp^2\frac{1-3p}{2(1-p)} + \frac{mp^3}{1-p} - \frac{1}{2}k^2mp^2 \\
&\geq k\left(\ln \binom{n}{k} + \frac{1-\epsilon}{2}mp^2 - \frac{1}{2}kmp^2\right).
\end{aligned}$$

□

Die Abschätzung für die Varianz der Zufallsvariable, welche die Anzahl stabiler Mengen einer bestimmten Größe misst, ist recht aufwendig. Um die 2. Moment Methode anzuwenden, muss die Varianz der entsprechenden Zufallsvariable klein gegen das Quadrat des Erwartungswerts sein. Die Varianz ist ein Maß für die Abweichung einer Zufallsvariable von ihrem Erwartungswert. Für eine beliebige Zufallsvariable X gilt:

$$\mathrm{Var}\left[X\right] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2$$

Der folgende Hilfssatz liefert eine Abschätzung für das Verhältnis von $\mathbb{E}\left[X_k^2\right]$ zu $\mathbb{E}\left[X_k\right]^2$, wobei die Zufallsvariable X_k die Anzahl stabiler Mengen der Größe k misst. Aus diesem Verhältnis kann dann die Aussage zu Varianz und dem Quadrat des Erwartungswerts abgeleitet werden. Der Ansatz dazu, sowie Teile der Rechnung orientieren sich dabei an Satz 5.8 in [7].

Lemma 13. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$. Sei außerdem X_k eine Zufallsvariable, welche die Anzahl der stabilen Mengen der Größe $k \geq 1$ in Instanzen von $\mathcal{G}_{n,m,p}$ misst. Wenn für n gegen unendlich $kp \to 0$ geht, dann gibt es eine Konstante M so dass gilt:*

$$\frac{\mathbb{E}\left[X_k^2\right]}{\mathbb{E}\left[X_k\right]^2} \leq 1 + M \sum_{s=1}^{k} \left(\frac{k^2}{n} \exp\left(1 + \frac{1+o(1)}{2}kmp^2 + 3k^3mp^3\right)\right)^s.$$

Beweis. Sei $G = (V, E) \in \mathcal{G}_{n,m,p}$ ein zufälliger Schnittgraph. Für eine k-elementige Knotenmenge $A \subseteq V$ sei X_A die Zufallsvariable, welche genau dann den Wert 1 annimmt, wenn A stabil ist und welche sonst 0 ist. Dann ist $X_k = \sum_{A \in \binom{V}{k}} X_A$. Wegen der Linearität des Erwartungswerts erhält man für $\mathbb{E}\left[X_k^2\right]$:

$$\mathbb{E}\left[X_k^2\right] = \mathbb{E}\left[\left(\sum_{A \in \binom{V}{k}} X_A\right)^2\right] = \sum_{A,B \in \binom{V}{k}} \mathbb{E}\left[X_A X_B\right] = \sum_{A,B \in \binom{V}{k}} P_{A,B},$$

wobei $P_{A,B}$ die Wahrscheinlichkeit dafür notiert, dass sowohl A als auch B stabil sind. Seien nun $A, B \in \binom{V}{k}$ zwei k-elementige Knotenmengen mit $|A \cap B| = s$ und sei P_k die Wahrscheinlichkeit dafür, dass eine gegebene Menge von k Knoten stabil ist. Für $s = 0$ ist dann $P_{A,B} = P_k^2$. Für $s \geq 1$ muss jede Featureclique eine der folgenden Bedingungen erfüllen damit A und B stabil sind:

1. die Featureclique enthält weder einen Knoten aus A noch einen aus B
2. die Featureclique enthält genau einen Knoten aus $A \cup B$
3. die Featureclique enthält genau einen Knoten aus $A \backslash B$ und genau einen aus $B \backslash A$.

Für $s \geq 1$ ergibt sich:

$$P_{A,B} \;=\; \left((1-p)^{2k-s} + (2k-s)p(1-p)^{2k-s-1} + (k-s)^2 p^2 (1-p)^{2k-s-2}\right)^m .$$

Weiterhin gilt nach dem Beweis von Satz 11:

$$\mathbb{E}\left[X_k\right] \;=\; \binom{n}{k} P_k \;=\; \binom{n}{k} \left((1-p)^k + kp(1-p)^{k-1}\right)^m .$$

Damit ergbit sich für den zu untersuchenden Term:

$$\begin{aligned}
\frac{\mathbb{E}\left[X_k^2\right]}{\mathbb{E}\left[X_k\right]^2} \;&=\; \frac{\sum_{A,B \in \binom{V}{k}} P_{A,B}}{\mathbb{E}\left[X_k\right]^2} \;=\; \sum_{s=0}^{k} \frac{\sum_{A,B \in \binom{V}{k} \text{ mit } |A \cap B| = s} P_{A,B}}{\mathbb{E}\left[X_k\right]^2} \\
&=\; \frac{\binom{n}{k}\binom{n-k}{k}}{\binom{n}{k}^2} \frac{P_k^2}{P_k^2} \;+\; \sum_{s=1}^{k} \frac{\binom{n}{k}\binom{k}{s}\binom{n-k}{k-s}}{\binom{n}{k}^2} \frac{P_{A,B}}{P_k^2} \\
&=\; \frac{\binom{n-k}{k}}{\binom{n}{k}} \;+\; \sum_{s=1}^{k} \frac{\binom{k}{s}\binom{n-k}{k-s}}{\binom{n}{k}} \frac{P_{A,B}}{P_k^2}.
\end{aligned}$$

Zunächst soll $\frac{P_{A,B}}{P_k^2}$ (für $|A \cap B| = s \geq 1$) betrachtet werden:

$$\begin{aligned}
\frac{P_{A,B}}{P_k^2} &= \frac{\Big((1-p)^{2k-s} + (2k-s)p(1-p)^{2k-s-1} + (k-s)^2p^2(1-p)^{2k-s-2}\Big)^m}{\Big((1-p)^k + kp(1-p)^{k-1}\Big)^{2m}} \\
&= \frac{\Big((1-p)^{2k-s} + (2k-s)p(1-p)^{2k-s-1} + (k-s)^2p^2(1-p)^{2k-s-2}\Big)^m}{\Big((1-p)^{2k} + 2kp(1-p)^{2k-1} + k^2p^2(1-p)^{2k-2}\Big)^m} \\
&= \frac{1}{(1-p)^{sm}} \left(\frac{1 + \frac{(2k-s)p}{1-p} + \frac{(k-s)^2p^2}{(1-p)^2}}{1 + \frac{2kp}{1-p} + \frac{k^2p^2}{(1-p)^2}}\right)^m \\
&= \frac{1}{(1-p)^{sm}} \exp\Bigg(m \ln\Bigg(1 + \frac{(2k-s)p}{1-p} + \frac{(k-s)^2p^2}{(1-p)^2}\Bigg) \\
&\quad -m \ln\Bigg(1 + \frac{2kp}{1-p} + \frac{k^2p^2}{(1-p)^2}\Bigg)\Bigg) \\
&\overset{(Taylor)}{=} \frac{1}{(1-p)^{sm}} \exp\Bigg(m\Bigg(\frac{(2k-s)p}{1-p} + \frac{(k-s)^2p^2}{(1-p)^2} - \frac{1}{2}\frac{(2k-s)^2p^2}{(1-p)^2} \\
&\quad -\frac{(2k-s)(k-s)^2p^3}{(1-p)^3} - \frac{1}{2}\frac{(k-s)^4p^4}{(1-p)^4} + \frac{1}{3}\frac{(2k-s)^3p^3}{(1-p)^3} \\
&\quad +\frac{(2k-s)^2(k-s)^2p^4}{(1-p)^4} + \frac{(2k-s)(k-s)^4p^5}{(1-p)^5} + \frac{1}{3}\frac{(k-s)^6p^6}{(1-p)^6)}\Bigg) \\
&\quad -m\Bigg(\frac{2kp}{1-p} + \frac{k^2p^2}{(1-p)^2} - \frac{1}{2}\frac{4k^2p^2}{(1-p)^2} - \frac{2k^3p^3}{(1-p)^3} - \frac{1}{2}\frac{k^4p^4}{(1-p)^4}\Bigg)\Bigg) \\
&\leq \frac{1}{(1-p)^{sm}} \exp\Bigg(m\Bigg(\frac{(2k-s)p - 2kp}{1-p} + \frac{(k-s)^2p^2 - \frac{1}{2}(2k-s)^2p^2 + k^2p^2}{(1-p)^2} \\
&\quad +\frac{\frac{1}{3}(2k-s)^3p^3 + 2k^3p^3}{(1-p)^3}\Bigg)\Bigg) \\
&\leq \frac{1}{(1-p)^{sm}} \exp\left(-\frac{smp}{1-p} + \frac{s^2mp^2}{2(1-p)^2} + 4k^3mp^3\right).
\end{aligned}$$

$(1-p)^{-sm}$ kann über Taylorentwicklung wie folgt abgeschätzt werden:

$$\begin{aligned}
\frac{1}{(1-p)^{sm}} &= \exp\Big(-sm\ln(1-p)\Big) = \exp\left(sm\sum_{i=1}^{\infty}\frac{p^i}{i}\right) \leq \exp\left(smp\sum_{i=0}^{\infty}p^i - \frac{1}{2}smp^2\right) \\
&\overset{(6)}{=} \exp\left(\frac{smp}{1-p} - \frac{1}{2}smp^2\right).
\end{aligned}$$

Für $\frac{P_{A,B}}{P_k^2}$ ergibt sich:

$$\begin{aligned}
\frac{P_{A,B}}{P_k^2} &\leq \exp\left(\frac{smp}{1-p} - \frac{1}{2}smp^2 - \frac{smp}{1-p} + \frac{s^2mp^2}{2(1-p)^2} + 4k^3mp^3\right) \\
&= \exp\left(-\frac{1}{2}smp^2 + \frac{s^2mp^2}{2(1-p)^2} + 4k^3mp^3\right).
\end{aligned}$$

Um den Quotienten der Erwartungswerte abzuschätzen müssen noch die Binomialkoeffizienten betrachtet werden. Mit Hilfe der Formel 3.6 aus [10]

$$\frac{(n)_s}{n^s} = O(1)\exp\left(-\frac{s^2}{2n}-\frac{s^3}{6n^2}\right)$$

ergibt sich für die Binomialkoeffizienten:

$$\begin{aligned}
\frac{\binom{k}{s}\binom{n-k}{k-s}}{\binom{n}{k}} &= \frac{(k)_s}{s!}\frac{(n-k)_{k-s}}{(k-s)!}\frac{k!}{(n)_k} \le (k)_s\frac{1}{(k-s)!}\frac{k!}{(n)_s} = \frac{(k)_s^2}{(n)_s}\\
&= \frac{k^{2s}}{O(1)\exp\left(-\frac{s^2}{2n}-\frac{s^3}{6n^2}\right)n^s}\\
&\le M\exp\left(\frac{s^2}{2n}+\frac{s^3}{6n^2}\right)\frac{k^{2s}}{n^s}
\end{aligned}$$

für eine Konstante M.

Durch Zusammenfügen der Teilergebnisse ergibt sich die Behauptung:

$$\begin{aligned}
\frac{\mathbb{E}\left[X_k^2\right]}{\mathbb{E}\left[X_k\right]^2} &= \frac{\binom{n-k}{k}}{\binom{n}{k}} + \sum_{s=1}^{k}\frac{\binom{k}{s}\binom{n-k}{k-s}}{\binom{n}{k}}\frac{P_{A,B}}{P_k^2}\\
&\le 1 + M\sum_{s=1}^{k}\exp\left(\frac{s^2}{2n}+\frac{s^3}{6n^2}-\frac{1}{2}smp^2+\frac{s^2mp^2}{2(1-p)^2}+4k^3mp^3\right)\frac{k^{2s}}{n^s}\\
&\le 1+M\sum_{s=1}^{k}\left(\frac{k^2}{n}\exp\left(\frac{1}{2}+\frac{1}{6}+\frac{1+o(1)}{2}smp^2+4k^3mp^3\right)\right)^s\\
&\le 1+M\sum_{s=1}^{k}\left(\frac{k^2}{n}\exp\left(1+\frac{1+o(1)}{2}kmp^2+4k^3mp^3\right)\right)^s.
\end{aligned}$$

□

Mit diesen beiden Hilfssätzen können nun Aussagen zur unteren Schranke der Stabilitätszahl bewiesen werden. Die bisherige untere Schranke aus Satz 10 ist $\beta \ge (1-\epsilon)\frac{1}{mp^2}$ für $p \gg \max\left(\frac{\ln n}{n}, \frac{\ln n}{m}\right)$, während $\beta \le (2+\epsilon)\frac{\ln n}{mp^2}$ in Satz 11 für $\frac{\ln n}{m} \ll p \ll \sqrt{\frac{\ln n}{m}}$ gezeigt wurde. Eine erste Verbesserung der unteren Schranke liefert der folgende Satz.

Satz 14. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$. Sei außerdem für eine beliebiges $\epsilon > 0$:*

$$\epsilon\left(\frac{1}{m}\right)^{\frac{2}{3}} \le p \qquad \text{und} \qquad \left(\frac{1}{m}\right)^{\frac{1}{2}}\left(\frac{1}{n}\right)^{\frac{1}{4}} \ll p,$$

dann ist in fast allen Instanzen von $\mathcal{G}_{n,m,p}$ die Stabilitätszahl $\beta \ge \frac{c}{mp^2}$ für eine beliebige Konstante c.

Beweis. Für $mp^2 \ge c$, ist die Aussage erfüllt, da jeder nichtleere Graph eine stabile Menge der Größe 1 enthält. Sei daher im Folgenden $mp^2 < c$.

Die Argumentation basiert auf der 2. Moment Methode. Sei dazu X_k eine Zufallsvariable, welche die Anzahl stabiler Mengen der Größe k in Instanzen von $\mathcal{G}_{n,m,p}$ zählt. Wenn nun gezeigt werden kann, dass für n gegen unendlich

1. $\mathbb{E}[X_k] \to \infty$ und
2. $\mathrm{Var}[X_k] \ll \mathbb{E}[X_k]^2$

gilt, folgt nach der 2. Moment Methode, dass $\mathbb{P}[X_k = 0]$ für n gegen unendlich gegen 0 geht.

Zuerst soll die erste Bedingung überprüft werden. Sei $k = \frac{c}{mp^2}$. Mit der Voraussetzung $\epsilon \left(\frac{1}{m}\right)^{\frac{2}{3}} \leq p$ folgt:

$$kp \;=\; \frac{c}{mp} \;\leq\; \frac{c}{\epsilon}\left(\frac{1}{m}\right)^{\frac{1}{3}} \;\to\; 0.$$

Folglich kann Lemma 12 angewandt werden und es gilt:

$$\ln \mathbb{E}[X_k] \;\geq\; k\left(\ln\left(\frac{n}{k}\right) + \frac{1-\epsilon}{2}mp^2 - \frac{1}{2}kmp^2\right).$$

Durch Einsetzen für k ergibt sich:

$$\ln \mathbb{E}[X_k] \;\geq\; \frac{c}{mp^2}\left(\ln\left(\frac{1}{c}nmp^2\right) - \frac{c}{2}\right).$$

Wegen $mp^2 < c$ ist $\frac{c}{mp^2}$ nach unten beschränkt. Außerdem folgt aus $\left(\frac{1}{m}\right)^{\frac{1}{2}}\left(\frac{1}{n}\right)^{\frac{1}{4}} \ll p$:

$$\frac{1}{c}nmp^2 \;\gg\; \sqrt{n} \;\to\; \infty.$$

Für n gegen unendlich gehen also $\ln \mathbb{E}[X_k]$ und $\mathbb{E}[X_k]$ gegen unendlich.

Nun muss noch die zweite Bedingung überprüft werden. Dafür genügt es zu zeigen, dass:

$$\frac{\mathrm{Var}[X_k]}{\mathbb{E}[X_k]^2} \;=\; \frac{\mathbb{E}\left[X_k^2\right] - \mathbb{E}[X_k]^2}{\mathbb{E}[X_k]^2} \;=\; \frac{\mathbb{E}\left[X_k^2\right]}{\mathbb{E}[X_k]^2} - 1 \;\to\; 0.$$

Da kp gegen 0 strebt, sind die Voraussetzungen von Lemma 13 erfüllt und es gibt eine Konstante M mit:

$$\frac{\mathbb{E}\left[X_k^2\right]}{\mathbb{E}[X_k]^2} \;\leq\; 1 + M\sum_{s=1}^{k}\left(\frac{k^2}{n}\exp\left(1 + \frac{1+o(1)}{2}kmp^2 + 4k^3mp^3\right)\right)^s.$$

Es ergibt sich:

$$\begin{aligned}
\frac{\mathbb{E}\left[X_k^2\right]}{\mathbb{E}[X_k]^2} - 1 \;&\leq\; M\sum_{s=1}^{k}\left(\frac{k^2}{n}\exp\left(1 + kmp^2 + 4k^3mp^3\right)\right)^s \\
&\overset{(5)}{=}\; M\left(\frac{\left(\frac{k^2}{n}\exp\left(1 + kmp^2 + 4k^3mp^3\right)\right)^{k+1} - 1}{\frac{k^2}{n}\exp\left(1 + kmp^2 + 4k^3mp^3\right) - 1} - 1\right).
\end{aligned}$$

Damit genügt es zu zeigen, dass für n gegen unendlich gilt:

$$\frac{k^2}{n} \exp\left(1 + kmp^2 + 4k^3mp^3\right) \;\to\; 0.$$

Nach Voraussetzung ist $p \gg \left(\frac{1}{m}\right)^{\frac{1}{2}} \left(\frac{1}{n}\right)^{\frac{1}{4}}$ und es gilt:

$$\frac{k^2}{n} \;=\; c^2 \frac{n}{m^2p^4} \;\to\; 0.$$

Außerdem gilt für das Argument der Exponentialfunktion wegen $k = \frac{c}{mp^2}$:

$$1 + kmp^2 + 4k^3mp^3 \;=\; 1 + c + \frac{4c^3}{m^2p^3}.$$

Dieser Term ist wegen der Voraussetzung $p \geq \epsilon \left(\frac{1}{m}\right)^{\frac{2}{3}}$ nach oben beschränkt. Damit wurde gezeigt:

$$\frac{\operatorname{Var}[X_k]}{\mathbb{E}[X_k]^2} \to \; 0.$$

Die beiden Voraussetzungen um die 2. Moment Methode anwenden zu können sind erfüllt. Es folgt, dass $\mathbb{P}[X_k = 0]$ für $k = \frac{c}{mp^2}$ und n gegen unendlich gegen 0 konvergiert, d.h. es gibt in fast allen Instanzen von $\mathcal{G}_{n,m,p}$ stabile Mengen der Größe k. □

Die untere Schranke konnte leicht verbessert werden, allerdings nicht für den kompletten Bereich für p. Damit stellt sich die Frage, ob durch sorgfältigere Abschätzungen ein größerer p-Bereich hätte abgedeckt werden können. Der Autor hat daher alle Abschätzungen der beiden Lemmata 12 und 13 sowie des Satzes 14 überprüft und versucht die Genauigkeit der Rechnung an diesen Stellen zu verbessern. Die dabei erzielten Verbesserungen in den Zwischenergebnissen schlugen sich leider nicht in einem besseren Endergebnis nieder.

Im nächsten Satz wird untersucht unter welchen Bedingungen $\frac{c\ln n}{mp^2}$ eine untere Schranke für die Stabilitätszahl ist. Es ist natürlich nicht zu erwarten, dass diese Schranke für ein beliebiges c gilt, da in Satz 11 bereits die obere Schranke mit $(2+\epsilon)\frac{\ln n}{mp^2}$ angegeben wurde.

Satz 15. *Sei $m = n^\alpha$ für eine Konstante $\alpha > 0$ und sei c eine Konstante mit $c > 2 - \frac{4}{3}\alpha$. Wenn nun gilt:*

$$p \geq \sqrt{\frac{1}{m^{\frac{2}{3}} n^{1-\frac{c}{2}}}},$$

dann ist in fast allen Instanzen von $\mathcal{G}_{n,m,p}$ die Stabilitätszahl $\beta \geq \frac{c\ln n}{mp^2}$.

Beweis. Zunächst ist anzumerken, dass c für $\alpha > \frac{3}{2}$ in der Voraussetzung durch eine negative Zahl nach unten beschränkt wird. Sollte c kleiner oder gleich 0 sein, so gilt die Behauptung, da jeder nichtleere Graph eine stabile Menge der Größe 1 enthält. Sei im Folgenden daher $c > 0$.

Wie in Satz 14 nutzt der Beweis die 2. Moment Methode. Sei dazu X_k eine Zufallsvariable, welche die Anzahl stabiler Mengen der Größe k in Instanzen von $\mathcal{G}_{n,m,p}$ zählt.

Zuerst soll überprüft werden, ob der Erwartungswert $\mathbb{E}[X_k]$ für n gegen unendlich gegen unendlich strebt. Sei $k = \frac{c \ln n}{mp^2}$. Wegen $p \geq \sqrt{\frac{1}{m^{\frac{2}{3}} n^{1-\frac{c}{2}}}}$ und $c > 2 - \frac{4}{3}\alpha$ folgt für kp:

$$\begin{aligned} kp &= \frac{c \ln n}{mp} \\ &\leq \frac{c \ln n \; m^{\frac{1}{3}} n^{\frac{1}{2}-\frac{c}{4}}}{m} \\ &= c \ln n \; n^{\frac{1}{2}-\frac{2}{3}\alpha-\frac{c}{4}} \\ &< c \ln n \; n^{\frac{1}{2}-\frac{2}{3}\alpha-\frac{1}{4}(2-\frac{4}{3}\alpha)} \\ &= c \ln n \; n^{-\frac{1}{3}\alpha} \\ &\to 0. \end{aligned}$$

Folglich darf Lemma 12 angewandt werden und es gilt:

$$\ln \mathbb{E}[X_k] \;\geq\; k \left(\ln \left(\frac{n}{k} \right) + \frac{1-\epsilon}{2} mp^2 - \frac{1}{2} kmp^2 \right).$$

Durch Einsetzen für k erhält man:

$$\ln \mathbb{E}[X_k] \;\geq\; \frac{c \ln n}{mp^2} \left(\ln \left(\frac{1}{c \ln n} nmp^2 \right) - \frac{c}{2} \ln n \right) + \frac{1-\epsilon}{2} c \ln n.$$

Der Erwartungswert strebt nun gegen unendlich, wenn $\ln\left(\frac{1}{c \ln n} nmp^2\right) - \frac{c}{2} \ln n$ nicht negativ ist. Nach Voraussetzung ist $p \geq \sqrt{\frac{1}{m^{\frac{2}{3}} n^{1-\frac{c}{2}}}}$. Damit folgt:

$$\begin{aligned} \ln \left(\frac{1}{c \ln n} nmp^2 \right) - \frac{c}{2} \ln n &\geq \ln \left(\frac{1}{c \ln n} \; \frac{nm}{m^{\frac{2}{3}} n^{1-\frac{c}{2}}} \right) - \frac{c}{2} \ln n \\ &= \ln \left(\frac{1}{c \ln n} \; n^{\frac{c}{2}} m^{\frac{1}{3}} \right) - \frac{c}{2} \ln n \\ &\geq \ln \left(n^{\frac{c}{2}} \right) - \frac{c}{2} \ln n \\ &= 0. \end{aligned}$$

Der Erwartungswert von X_k läuft gegen unendlich.

Um die 2. Moment Methode anwenden zu können muss außerdem überprüft werden, ob die Varianz von X_k klein gegen das Quadrat des Erwartungswerts ist. Dazu genügt es zu zeigen:

$$\frac{\mathrm{Var}[X_k]}{\mathbb{E}[X_k]^2} \;=\; \frac{\mathbb{E}\left[X_k^2\right] - \mathbb{E}[X_k]^2}{\mathbb{E}[X_k]^2} \;=\; \frac{\mathbb{E}\left[X_k^2\right]}{\mathbb{E}[X_k]^2} - 1 \;\to\; 0.$$

Da kp gegen 0 geht, darf Lemma 13 angewandt werden und es gibt eine Konstante M mit:

$$\frac{\mathbb{E}\left[X_k^2\right]}{\mathbb{E}[X_k]^2} \;\leq\; 1 + M \sum_{s=1}^{k} \left(\frac{k^2}{n} \exp\left(1 + \frac{1+o(1)}{2} kmp^2 + 4k^3 mp^3 \right) \right)^s.$$

Das kann umgeformt werden zu:

$$\begin{aligned}
\frac{\mathbb{E}\left[X_k^2\right]}{\mathbb{E}\left[X_k\right]^2}-1 &\leq M\sum_{s=1}^{k}\left(\frac{k^2}{n}\exp\left(1+\frac{1+o(1)}{2}kmp^2+4k^3mp^3\right)\right)^s \\
&\overset{(5)}{=} M\left(\frac{\left(\frac{k^2}{n}\exp\left(1+\frac{1+o(1)}{2}kmp^2+4k^3mp^3\right)\right)^{k+1}-1}{\frac{k^2}{n}\exp\left(1+\frac{1+o(1)}{2}kmp^2+4k^3mp^3\right)-1}-1\right).
\end{aligned}$$

Dieser Term geht gegen 0 wenn $\frac{k^2}{n}\exp\left(\frac{1+o(1)}{2}kmp^2+4k^3mp^3\right)$ gegen 0 strebt. Durch Einsetzen für k und mittels der Voraussetzung $p \geq \sqrt{\frac{1}{m^{\frac{2}{3}}n^{1-\frac{c}{2}}}}$ erhält man:

$$\begin{aligned}
\frac{k^2}{n}\exp\left(\frac{1+o(1)}{2}kmp^2+4k^3mp^3\right) &= c^2\frac{\ln^2 n}{nm^2p^4}\exp\left((1+o(1))\frac{c}{2}\ln n+4c^3\frac{\ln^3 n}{m^2p^3}\right) \\
&= c^2\frac{\ln^2 n}{nm^2p^4}\; n^{(1+o(1))\frac{c}{2}+4c^3\frac{\ln^2 n}{m^2p^3}} \\
&= c^2\frac{\ln^2 n}{m^2p^4}\; n^{(1+o(1))\frac{c}{2}+4c^3\frac{\ln^2 n}{m^2p^3}-1} \\
&\leq c^2\frac{m^{\frac{4}{3}}n^{2-c}\ln^2 n}{m^2}\; n^{(1+o(1))\frac{c}{2}+4c^3\frac{\ln^2 n}{m^2p^3}-1} \\
&= c^2\ln^2 n\; n^{2-\frac{2}{3}\alpha-c}\; n^{(1+o(1))\frac{c}{2}+4c^3\frac{\ln^2 n}{m^2p^3}-1} \\
&= c^2\ln^2 n\; n^{1-\frac{2}{3}\alpha-(1-o(1))\frac{c}{2}+4c^3\frac{\ln^2 n}{m^2p^3}}.
\end{aligned}$$

Es bleibt zu zeigen, dass der Exponent von n durch eine negative Konstante nach oben beschränkt ist. Nach Voraussetzung gibt es ein $\epsilon > 0$, so dass $c = 2-\frac{4}{3}\alpha+\epsilon$. Durch erneutes Einsetzen der Voraussetzung für p ergibt sich für den Exponenten von n:

$$\begin{aligned}
1-\frac{2}{3}\alpha-(1-o(1))\frac{c}{2}+4c^3\frac{\ln^2 n}{m^2p^3} &\leq 1-\frac{2}{3}\alpha-(1-o(1))\frac{c}{2}+4c^3\frac{mn^{\frac{3}{2}-\frac{3}{4}c}\ln^2 n}{m^2} \\
&= 1-\frac{2}{3}\alpha-(1-o(1))\frac{c}{2}+4c^3\ln^2 n\; n^{\frac{3}{2}-\alpha-\frac{3}{4}c} \\
&= 1-\frac{2}{3}\alpha-(1-o(1))\frac{1}{2}\left(2-\frac{4}{3}\alpha+\epsilon\right) \\
&\quad +4c^3\ln^2 n\; n^{\frac{3}{2}-\alpha-\frac{3}{4}\left(2-\frac{4}{3}\alpha+\epsilon\right)} \\
&= -\frac{1}{2}\epsilon+\frac{o(1)}{2}\left(2-\frac{4}{3}\alpha+\epsilon\right)+4c^3\ln^2 n\; n^{-\frac{3}{4}\epsilon} \\
&= -\frac{1}{2}\epsilon+o(1).
\end{aligned}$$

Die Varianz von X_k ist für $k = \frac{c\ln n}{mp^2}$ klein gegen das Quadrat des Erwartungswerts. Damit sind die Voraussetzungen für die 2. Moment Methode erfüllt und es folgt, dass $P(X_k = 0)$ für n gegen unendlich gegen 0 strebt. Also gibt es in fast allen Instanzen von $\mathcal{G}_{n,m,p}$ eine stabile Menge dieser Größe. $\square$

Die Voraussetzung von Satz 15 ist nur für $c < 2$ interessant. Denn für $c \geq 2$ gilt die Schranke erst für p mit

$$p \;\geq\; \sqrt{\frac{1}{m^{\frac{2}{3}} n^{1-\frac{c}{2}}}} \;\geq\; \sqrt{\frac{1}{m^{\frac{2}{3}}}} \;\gg\; \sqrt{\frac{\ln n}{m}},$$

so dass die vom Satz aufgestellte Schranke für die Stabilitätszahl $\frac{c \ln n}{mp^2}$ gegen 0 läuft.

Inwiefern sind nun die Bedingungen an p in Satz 15 schärfer als in Satz 14? Um diese Frage zu beantworten muss man die Fälle $\alpha < \frac{3}{2}$ und $\alpha \geq \frac{3}{2}$ unterscheiden. Für $\alpha < \frac{3}{2}$ liefert Satz 14:

$$\text{für } p \;\geq\; \epsilon \left(\frac{1}{m}\right)^{\frac{2}{3}} \quad \text{gilt:} \quad \beta \geq \frac{c}{mp^2}.$$

Satz 15 liefert für dasselbe α und $c = 2 - \frac{4}{3}\alpha + \epsilon$:

$$\text{für } p \;\geq\; \sqrt{\frac{1}{m^{\frac{2}{3}} n^{1-\frac{c}{2}}}} \;=\; n^{\frac{\epsilon}{4}} \left(\frac{1}{m}\right)^{\frac{2}{3}} \quad \text{gilt:} \quad \beta \geq \frac{c \ln n}{mp^2}.$$

Satz 14 deckt also mit der schwächeren Aussage einen nur geringfügig größeren Bereich für p ab. Die Voraussetzungen der beiden Sätze sind in diesem Bereich fast gleichwertig.

Im Fall $\alpha \geq \frac{3}{2}$ ergibt sich dagegen bei Satz 14:

$$\text{für } p \;\gg\; \left(\frac{1}{n}\right)^{\frac{1}{4}} \left(\frac{1}{m}\right)^{\frac{1}{2}} \quad \text{gilt:} \quad \beta \geq \frac{c}{mp^2}$$

und bei Satz 15 für $c = \epsilon$:

$$\text{für } p \;\geq\; \sqrt{\frac{1}{m^{\frac{2}{3}} n^{1-\frac{c}{2}}}} \;=\; \left(\frac{1}{n}\right)^{\frac{1}{2}-\frac{\epsilon}{4}} \left(\frac{1}{m}\right)^{\frac{1}{3}} \quad \text{gilt:} \quad \beta \geq \frac{c \ln n}{mp^2}.$$

Mit steigendem α wächst das von Satz 14 abgedeckte Intervall für p somit wesentlich schneller als bei Satz 15.

4.3 Zusammenfassung

In diesem Kapitel wurde die Stabilitätszahl von zufälligen Schnittgraphen untersucht. Nachfolgend zunächst ein Überblick über die Aussagen der einzelnen Sätze.

[7]		p	$\ll \frac{1}{n\sqrt{m}}$	$\beta = n$	
[7]	$\sqrt{\frac{\ln n}{m}} \ll$	p		$\beta = 1$	
Satz 9		p	$\ll \frac{1}{m}$	$\beta \geq \big(1 - o(1)\big)n$	
Satz 10	$\max\left(\frac{\ln n}{m}, \frac{\ln n}{n}\right) \ll$	p		$\beta \geq (1-\epsilon)\frac{1}{mp^2}$	
Satz 11	$\frac{\ln n}{m} \ll$	p	$\ll \sqrt{\frac{\ln n}{m}}$	$\beta < (2+\epsilon)\frac{\ln n}{mp^2}$	
Satz 14	$\epsilon\left(\frac{1}{m}\right)^{\frac{2}{3}} \leq$	p		$\beta \geq \frac{c}{mp^2}$	(für $\alpha < \frac{3}{2}$, c beliebig)
Satz 14	$\left(\frac{1}{n}\right)^{\frac{1}{4}}\left(\frac{1}{m}\right)^{\frac{1}{2}} \ll$	p		$\beta \geq \frac{c}{mp^2}$	(für $\alpha \geq \frac{3}{2}$, c beliebig)
Satz 15	$n^{\frac{\epsilon}{4}}\left(\frac{1}{m}\right)^{\frac{2}{3}} \leq$	p		$\beta \geq c\frac{\ln n}{mp^2}$	(für $\alpha < \frac{3}{2}$, $c = 2 - \frac{4}{3}\alpha + \epsilon$)
Satz 15	$\left(\frac{1}{n}\right)^{\frac{1}{2}-\frac{\epsilon}{4}}\left(\frac{1}{m}\right)^{\frac{1}{3}} \leq$	p		$\beta \geq c\frac{\ln n}{mp^2}$	(für $\alpha \geq \frac{3}{2}$, c beliebig)

Betrachtet man die Stabilitätszahl für das komplette Intervall von p, ergibt sich das in Abbildung 4 dargestellte Bild. Für $p \ll \frac{1}{n\sqrt{m}}$ enthalten zufällige Schnittgraphen fast sicher keine Kanten. Für etwas größere p mit $p \ll \frac{1}{m}$ gibt es dann in fast allen Instanzen eine stabile Menge, welche $n - o(n)$ Knoten umfasst. Im Bereich $\frac{\ln n}{m} \ll p \ll \sqrt{\frac{\ln n}{m}}$ ist die Stabilitätszahl nach oben durch $(2+\epsilon)\frac{\ln n}{mp^2}$ beschränkt. Für den oberen Teil dieses Gebiets kommt die untere Schranke bis auf einen konstanten Faktor an diesen Wert heran. In einem anderen Teil $(\max\left(\frac{\ln n}{m}, \frac{\ln n}{n}\right) \ll p)$ konnte zumindest $(1-\epsilon)\frac{1}{mp^2}$ als untere Schranke gezeigt werden. Für den Rest (im Wesentlichen $\frac{1}{m} \ll p \ll \frac{\ln n}{n}$ bei $\alpha > 1$) ist keine untere Schranke bekannt. Ab $p \gg \sqrt{\frac{\ln n}{m}}$ schließlich sind die zufälligen Schnittgraphen fast sicher vollständig.

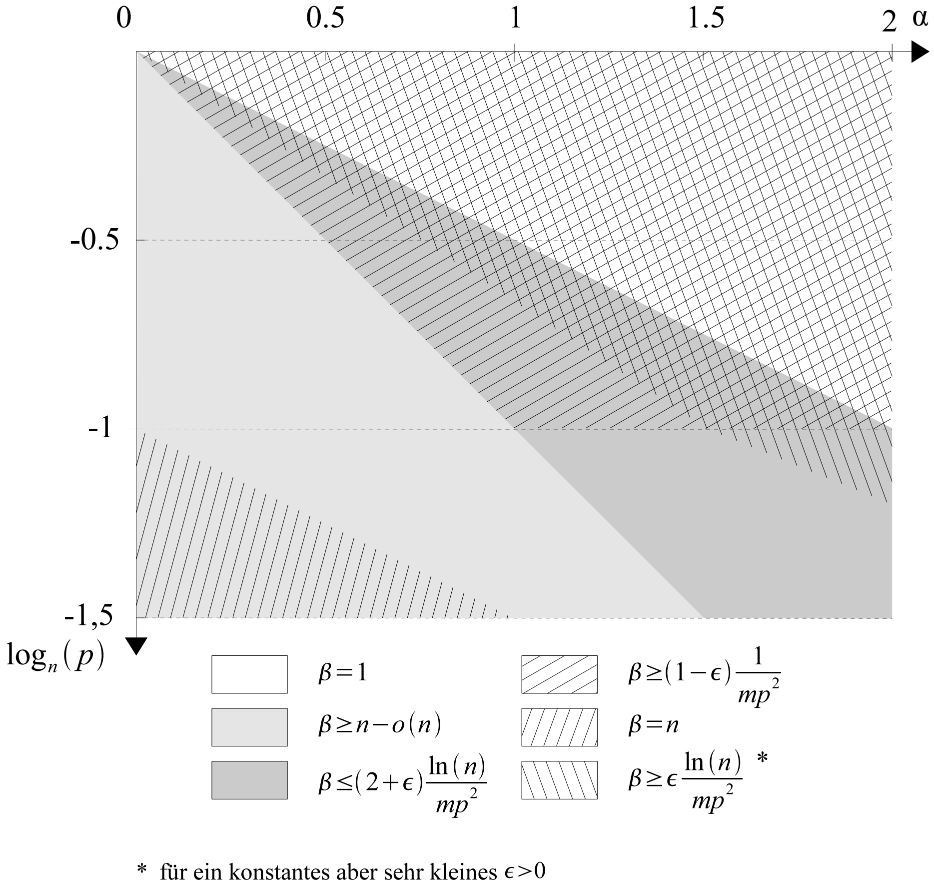

Abbildung 4: Aussagen zur Stabilitätszahl

Damit ist die Stabilitätszahl für die meisten Bereiche von p schon recht gut abgesteckt. Insbesondere konnte eine obere Grenze gefunden werden, welche den von Kapitel 3 offengelassenen Parameterbereich fast vollständig abdeckt. Damit ist die Grundlage für die Abschätzung der Approximationsgüte des Greedy-Algorithmus in diesem Bereich geschaffen.

Es bleiben im Wesentlichen zwei offene Punkte. Erstens konnte für einen Großteil der Featurewahrscheinlichkeiten zwischen $\frac{1}{m}$ und $\frac{\ln n}{n}$ bei $\alpha > 1$ keine untere Schranke für die Stabilitätszahl gefunden werden. Und zweitens unterscheiden sich obere und untere

Schranke im Bereich $\frac{\ln n}{m} \ll p \ll \left(\frac{1}{m}\right)^{\frac{2}{3}}$ um den Faktor $2 \ln n$. Das bedeutet entweder, dass die sich Stabilitätszahlen von $\mathcal{G}_{n,m,p}$ für dieses p-Intervall einfach nicht so stark um einen Wert konzentrieren, oder dass die Schranken noch nicht bestmöglich sind. An dieser Stelle sieht der Autor einen Ansatzpunkt für weitere Untersuchungen.

5 Approximation der chromatischen Zahl von zufälligen Schnittgraphen

Dieses Kapitel beschäftigt sich wieder mit dem Färben von zufälligen Schnittgraphen. Für $p \ll \frac{1}{m \ln n}$ bei $\alpha < 1$ sowie für $p \ll \sqrt{\frac{1}{nm}}$ können fast alle Instanzen optimal gefärbt werden (Kapitel 3) und ab $p \gg \sqrt{\frac{\ln n}{m}}$ sind fast alle zufälligen Schnittgraphen vollständig (Abschnitt 2.4). Dieses Kapitel betrachtet daher Featurewahrscheinlichkeiten p zwischen diesen beiden Grenzen. Wie bereits in der Zusammenfassung von Kapitel 3 versprochen, soll dazu zunächst eine untere Schranke für die chromatische Zahl von zufälligen Schnittgraphen aus den Resultaten von Kapitel 4 abgeleitet werden.

Weil die Farbklassen einer Färbung stabile Mengen sind, gilt:

$$\chi \geq \frac{n}{\beta}.$$

Nach Satz 11 ist die größte stabile Menge in fast allen Instanzen von $\mathcal{G}_{n,m,p}$ kleiner als $(2+\epsilon)\frac{\ln n}{mp^2}$, wenn $\frac{\ln n}{m} \ll p \ll \sqrt{\frac{\ln n}{m}}$. Damit folgt für die chromatische Zahl dieser Instanzen:

$$\text{wenn } \frac{\ln n}{m} \ll p \ll \sqrt{\frac{\ln n}{m}}, \text{ dann gilt:} \quad \chi \geq \frac{n}{\beta} \geq \frac{1}{2+\epsilon}\frac{mp}{\ln n}np \gg np.$$

Nach Lemma 2 ist für $p \gg \frac{\ln n}{n}$ die Größe der Featurecliquen und damit auch die chromatische Zahl in fast allen zufälligen Schnittgraphen durch $(1-\epsilon)np$ nach unten beschränkt. Die aus den Abschätzungen zur Stabilitätszahl abgeleitete Schranke ist also für größere p deutlich besser als die in Kapitel 3 verwendete, welche aus der Größe der Featurecliquen abgeleitet wurde. Im Folgenden wird daher die neue Schranke benutzt um die Approximationsgüte von Färbealgorithmen abzuschätzen.

Wie lassen sich zufällige Schnittgraphen mit größeren p färben? In einer ersten Überlegung kann man untersuchen, inwiefern die Ergebnisse aus Kapitel 3 für Approximationen verwendet werden können. In Abschnitt 3.2 wird der Greedy Algorithmus auf smallest-last Knotenanordnungen untersucht. Die Resultate basieren darauf, dass die Anzahl der dabei benötigten Farben durch das Maximum der Minimalgrade aller Subgraphen der zu färbenden Instanz beschränkt ist. Eine einfache Überlegung zeigte dann, dass es genügt Subgraphen zu betrachten, die ausschließlich aus dicken Knoten (Knoten mit mindestens zwei Features) bestehen. Die besten Resultate konnten mit den Sätzen 6 und 8 erzielt werden, indem eine obere Schranke für die maximale Anzahl dicker Nachbarn eines dicken Knoten und damit auch für den Knotengrad in solchen Subgraphen gefunden wurde.

Für die zufälligen Schnittgraphen im hier betrachteten Bereich von p sind jedoch fast alle Knoten dick. Ein Ansatz wie in Satz 6 bzw. Satz 8 läuft somit im Wesentlichen auf eine Betrachtung des normalen Maximalgrads in zufälligen Schnittgraphen hinaus. Im Beweis von Satz 10 wurde gezeigt, dass für $p \gg \max\left(\frac{\ln n}{n}, \frac{\ln n}{m}\right)$ der Maximalgrad fast sicher durch $(1+\epsilon)nmp^2$ nach oben beschränkt ist (bei beliebigem $\epsilon > 0$). Die Anzahl der vom Greedy-Algorithmus benötigten Farben ist durch $\Delta + 1$ beschränkt, d.h. der einfache Greedy-Algorithmus liefert bei $p \gg \max\left(\frac{\ln n}{n}, \frac{\ln n}{m}\right)$ fast sicher eine Färbung mit $(1+\epsilon)nmp^2$ Farben.

Wie nah ist dieses Ergebnis an einer optimalen Färbung? Mit Hilfe der oben abgeleiteten unteren Schranke für die chromatische Zahl erhält man folgende Abschätzung für die Approximationsgüte des einfachen Greedy-Algorithmus:

$$\frac{\Delta+1}{\chi} \leq \frac{(1+\epsilon)nmp^2}{\frac{1}{2+\epsilon}\frac{mp}{\ln n}np} = (1+\epsilon)(2+\epsilon)\ln n \leq (2+\epsilon')\ln n$$

für ein beliebiges $\epsilon' > 0$. Die gefundene Färbung enthält daher schlimmstenfalls $(2+\epsilon)\ln n\ \chi$ Farben.

Von dem zu betrachtenden Intervall bleibt nun lediglich noch der Teil $p \ll \frac{1}{n}$ bei $\alpha \geq 1$ zu untersuchen. In diesem Bereich gilt die oben benutzte Schranke für den Maximalgrad nicht, da Lemma 2 nicht angewandt werden kann und somit keine Schranke für die Größe der Featurecliquen zur Verfügung steht. Ganz analog zu diesem Lemma lässt sich jedoch zeigen, dass fast kein zufälliger Schnittgraph bei $p \ll \frac{1}{n}$ eine Featureclique mit mehr als $O(\ln n)$ Knoten enthält.

Lemma 16. *Sei $m = n^\alpha$ für ein festes $\alpha > 0$ und $p \ll \frac{1}{n}$, dann hat fast jede Instanz von $\mathcal{G}_{n,m,p}$ ausschließlich Featurecliquen mit weniger als $c\ln n$ Knoten für eine beliebige Konstante $c > \frac{2}{3}\alpha$.*

Beweis. Sei $G = (V,E) \in \mathcal{G}_{n,m,p}$ ein zufälliger Schnittgraph und $B = (V \cup W, E_B)$ der bipartite Generator von G. Für $w \in W$ sei X_w die Zufallsvariable, welche die Größe der von w erzeugten Featureclique misst. Sei $c = \frac{2}{3}\alpha + \epsilon$ und $c' = c - \frac{\epsilon}{2}$. Da X_w binomialverteilt ist, kann die Ungleichung von Chernoff angewandt werden (Abschnitt 2.2.2).

$$\begin{aligned}
\mathbb{P}\left[X_w \geq c\ln n\right] &\leq \mathbb{P}\left[X_w \geq \mathbb{E}\left[X_w\right] + c'\ln n\right], \qquad \text{da } \mathbb{E}\left[X_w\right] = np \to 0 \\
&\overset{(3)}{\leq} \exp\left(-\frac{(c'\ln n)^2}{2np + \frac{2}{3}c'\ln n}\right) \\
&\leq \exp\left(-\frac{3}{2+\epsilon'}c'\ln n\right) \\
&= n^{-\frac{3}{2+\epsilon'}c'}.
\end{aligned}$$

Für die Wahrscheinlichkeit der Existenz einer Featureclique mit mehr als $c\ln n$ Knoten ergibt sich:

$$\mathbb{P}\left[\exists_{w\in W} : X_w \geq c\ln n\right] \leq m\mathbb{P}\left[X_w \geq c\ln n\right] \leq n^{\alpha - \frac{3}{2+\epsilon'}c'} \to 0$$

für ein genügend kleines $\epsilon' > 0$.

Folglich enthält fast jede Instanz von $\mathcal{G}_{n,m,p}$ ausschließlich Featurecliquen mit weniger als $c\ln n$ Knoten. □

Der Maximalgrad ist beschränkt durch das Produkt der maximalen Anzahl von Features eines Knotens (nach dem Beweis von Satz 10 $(1+\epsilon)mp$) und der oberen Schranke für die Größe der Featurecliquen. Damit ergibt sich für die Güte des Greedy-Algorithmus:

$$\frac{\Delta+1}{\chi} \leq \frac{(1+\epsilon)mpO(\ln n)}{\frac{1}{2+\epsilon}\frac{mp}{\ln n}np} = O\left(\frac{\ln^2 n}{np}\right).$$

Je schneller np gegen 0 strebt, desto schlechter ist diese Approximationsgüte. Das Ergebnis lässt sich jedoch verbessern. Der Erwartungswert für die Größe einer Featureclique (np) strebt gegen 0, d.h. die zufälligen Schnittgraphen in diesem Bereich enthalten sehr viele leere oder kleine Featurecliquen. Um eine Kante zu erzeugen muss eine Featureclique mindestens zwei Knoten enthalten. Bei der Betrachtung des Maximalgrads können also alle Featurecliquen ignoriert werden, die gar keinen oder nur einen Knoten enthalten. Analog zu den dicken Knoten (Knoten mit mindestens zwei Features) können dicke Features bzw. dicke Featurecliquen definiert werden.

Definition 15 (dicke Featureclique) Sei G ein zufälliger Schnittgraph. Eine Featureclique in G heißt *dick*, wenn sie mindestens zwei Knoten enthält.

Sei G ein zufälliger Schnittgraph und $B = (V \cup W, E_B)$ sein bipartiter Generator. Der Maximalgrad von G ist beschränkt durch die maximale Anzahl dicker Featurecliquen eines Knotens multipliziert mit der maximalen Größe der Featurecliquen. Letzteres ist nach Lemma 16 für $p \ll \frac{1}{n}$ durch $O(\ln n)$ beschränkt. Betrachtet man den Schnittgraphen G' der entsteht, wenn die Featuremenge W als Knotenmenge und die Knotenmenge V als Featuremenge aufgefasst wird, so gibt es zu jeder dicken Featureclique von G einen dicken Knoten in G'. Die maximale Anzahl dicker Featurecliquen eines Knotens von G entspricht somit der maximalen Anzahl dicker Knoten einer Featureclique in G', welche mit Lemma 7 abgeschätzt werden kann. Auf diese Weise interpretiert liefert Lemma 7:

Lemma 7'. *Sei $m = n^\alpha$ für ein festes $\alpha \geq 1$. Falls $p \gg \sqrt{\frac{1}{nm}}$, so enthält fast keine Instanz von $\mathcal{G}_{n,m,p}$ Knoten, die in mehr als $(1+\epsilon)nmp^2 + \sqrt{nmp^2}\ln m$ dicken Featurecliquen enthalten sind.*

Für die Schranke der Approximationsgüte des Greedy-Algorithmus ergibt sich damit bei $\sqrt{\frac{1}{nm}} \ll p \ll \frac{1}{n}$:

$$\begin{aligned}
\frac{\Delta+1}{\chi} &\leq \frac{\left((1+\epsilon)nmp^2 + \sqrt{nmp^2}\ln m\right)O(\ln n)}{\frac{1}{2+\epsilon}\frac{mp}{\ln n}np} \\
&= \left((1+\epsilon) + \sqrt{\frac{1}{nmp^2}}\ln m\right) O(\ln^2) \\
&= O(\ln^2 n) + O\left(\frac{\ln^3 n}{\sqrt{nmp^2}}\right).
\end{aligned}$$

Abhängig davon, wie schnell nmp^2 gegen unendlich strebt, liegt die Approximationsgüte daher zwischen $O(\ln^2 n)$ und $o(\ln^3 n)$. Sie ist damit potentiell deutlich besser als die oben ermittelte Approximationsgüte von $O\left(\frac{\ln n}{np}\right)$.

Abbildung 5 stellt überblicksartig die bislang erhaltenen Ergebnisse zum Färben von zufälligen Schnittgraphen dar.

Der Parameterraum, für den keine optimale Färbung angenommen werden kann, wurde komplett durch Approximationsaussagen (Approximationsgüte $o(\ln^3 n)$ oder $(2+\epsilon)\ln n$) abgedeckt. Lässt sich diese Aussage noch verbessern? An dieser Stelle scheint wieder ein vergleichender Blick zum $G_{n,p}$-Modell angebracht. Für dieses Modell zufälliger Graphen konnte gezeigt werden, dass ein modifizierter Greedy-Algorithmus zum Färben fast sicher nur etwa doppelt so viele Farben wie unbedingt erforderlich benötigt [9].

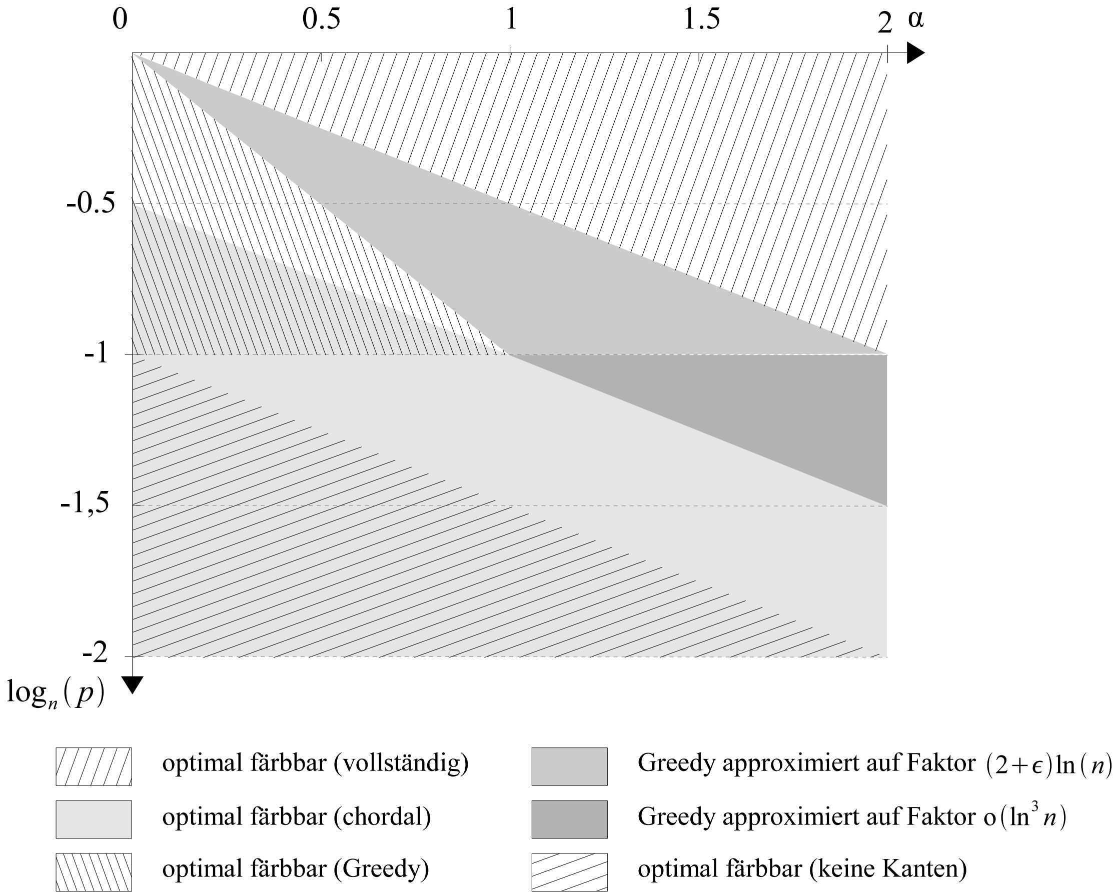

Abbildung 5: Färben von zufälligen Schnittgraphen

Grimmet, McDiarmid (1975). *Es existiert eine Konstante c_0, so dass der unten vorgestellte Färbealgorithmus (Algorithmus 3) für $\frac{c_0}{n} \leq p \leq 1 - \epsilon$ (bei einem beliebigen $\epsilon > 0$) fast sicher eine Approximationsgüte von $2 + o(1)$ auf $G_{n,p}$ erreicht.*

Es werden maximale stabile Mengen gesucht, aus dem Graphen entfernt und mit jeweils einer Farbe gefärbt. Sobald nur noch ein kleiner Rest der Knoten übrig ist, bricht der Algorithmus ab und gibt jedem dieser Knoten eine eigene Farbe. Bei der Analyse des Algorithmus wird davon ausgegangen, dass die Kanten des zu färbenden Graphen erst während des Laufs des Algorithmus ausgewürfelt werden. Erst in Zeile 6 werden die Nachbarn des Knotens v bestimmt. Am Anfang eines jeden Durchlaufs durch die äußere Schleife sind alle Kanten des Restgraphen G unbestimmt. Während des Durchlaufs durch die äußere Schleife werden dann genau die Kanten bestimmt (gemäß des $G_{n,p}$-Verfahrens), die am Ende durch das Entfernen von S mitentfernt werden. Damit liegt dem Algorithmus am Anfang eines jeden Durchlaufs durch die äußere Schleife ein Graph mit ausschließlich unbestimmten Kanten vor, und es kann angenommen werden, dass es sich um eine $G_{n,p}$-Instanz handelt. Details zum Algorithmus und zum Beweis der Approximationsgüte können in [3] nachgelesen werden.

Leider kann diese Analyse nicht ohne Weiteres auf zufällige Schnittgraphen übertragen werden. Das Problem liegt darin in einem Durchlauf nur genau die Kanten zu bestimmen, die dann mit der gefunden maximalen stabilen Menge S entfernt werden. Würde

Greedy-Färbung des Graphen $G = (V, E)$

1 $c := 0,\ \ n := |V|$
2 `while` $|V| \geq \frac{n}{\log^2 n}$ `do`
3 $\quad S := \emptyset,\ \ H := G$
4 $\quad$ `while` $H \neq \emptyset$ `do`
5 $\qquad S := S \cup \{v\}$ `(für` $v \in V(H)$ `beliebig)`
6 $\qquad H := H \setminus \Big(\{v\} \cup N(v)\Big)$
7 $\quad G := G \setminus S,\ \ c := c + 1$
8 $\quad$ `färbe alle Knoten in` S `mit Farbe` c
9 `färbe die restlichen Knoten in V mit jeweils einer Farbe`

Algorithmus 3: modifizierter Greedy-Algorithmus

man zuerst die komplette Kantenmenge bestimmen und dann den Algorithmus laufen lassen, funktioniert das Analyse-Verfahren nicht, denn bereits im ersten Durchlauf würden z.B. alle Knoten ohne Features entfernt werden. Nach dem ersten Durchlauf durch die äußere Schleife enthält der Restgraph daher nur noch Knoten mit mindestens einem Feature. Man darf dann nicht mehr davon ausgehen, dass eine typische Instanz eines zufälligen Schnittgraphen vorliegt. Bei den klassischen zufälligen Graphen wurde dieses Problem umgangen, indem in jedem Durchlauf durch die äußere Schleife nur die Kanten bestimmt werden, die dann beim Entfernen von S aus G mit entfernt werden. In zufälligen Schnittgraphen ist es jedoch nicht möglich so selektiv Kanten zu bestimmen. Denn um hier die Nachbarschaft eines Knotens v zu bestimmen, müssen im entsprechenden bipartiten Generator zunächst die Features von v und dann alle Knoten dieser Features bestimmt werden. Damit würden jedoch gleichzeitig auch Kanten innerhalb der Nachbarschaft von v festgelegt. Auch nach dem Herausnehmen von S aus G in Zeile 7 würde der Restgraph noch Kanten enthalten. Man kann also nicht sicher sein auch nach dem ersten Durchlauf durch die äußere Schleife noch einen zufälligen Schnittgraph vorliegen zu haben.

Zusammenfassend kann man festhalten, dass der einfache Greedy-Färbealgorithmus auf zufälligen Schnittgraphen fast sicher eine Approximationsgüte von $o(\ln^3 n)$ hat, wenn $\sqrt{\frac{1}{nm}} \ll p \ll \frac{1}{n}$. Für $\max\left(\frac{\ln n}{n}, \frac{\ln n}{m}\right) \ll p \ll \sqrt{\frac{\ln n}{m}}$ erzielt er sogar eine Güte von $(2+\epsilon)\ln n$. Bei größeren Featurewahrscheinlichkeiten $p \gg \sqrt{\frac{\ln n}{m}}$ sind die zufälligen Schnittgraphen fast sicher vollständig, und können optimal gefärbt werden. Die große Differenz zur bekannten Approximationsgüte auf klassischen zufälligen Graphen lässt vermuten, dass sich die hier erhaltenen Ergebnisse noch verbessern lassen. Für klassische Zufallsgraphen konnte gezeigt werden, dass Greedy fast sicher mit 2χ Farben auskommt. Jedoch ist weder der Ansatz aus Kapitel 3 noch das bei der Analyse von Greedy auf klassischen zufälligen Graphen angewandte Verfahren geeignet eine bessere Abschätzung zu erhalten. An dieser Stelle sieht der Autor daher einen Ansatzpunkt für zukünftige Untersuchungen.

6 Experimentelle Ergebnisse

Nachdem in den vorangegangenen Kapiteln untersucht wurde, wie gut der Greedy-Algorithmus in Verbindung mit einer smallest-last Heuristik zufällige Schnittgraphen färbt, soll er nun auf einige reale Graphen angewandt werden. Die drei im Folgenden vorgestellten Instanzen stammen aus [11]. Im Schauspielergraphen repräsentieren die Knoten Schauspieler und werden durch eine Kante verbunden, wenn es einen Film gibt, in dem beide mitgespielt haben. Dieser Graph liegt in Form seines bipartiten Generators vor, d.h. die Zuordnung von Schauspielern (Knoten) zu Filmen (Features) ist bekannt. Der WWW-Graph enthält die Struktur der `nd.edu` domain. Dabei bilden die Webseiten die Knoten, die durch eine Kante verbunden werden, wenn es einen Hyperlink von der einen zur anderen Seite gibt. Der Proteingraph stellt das Zusammenspiel der Proteine für Hefe dar, wobei die Proteine die Knoten bilden. Der vierte und letzte gefärbte Graph stammt aus [12] und ist eine Karte der Router für einen Teil des Internets. Die Knoten repräsentieren Router und werden mit anderen Knoten verbunden, wenn die beiden Router benachbart sind.

Die Ergebnisse des Greedy-Färbealgorithmus mit smallest-last Heuristik auf realen Graphen:

	Schauspieler	WWW	Protein	Router
Knoten (n)	392 340	325 730	2116	720 718
Kanten	15 038 083	1 090 108	2203	860 564
Features (m)	127 823	528 640	1937	860 337
p	$29.32 \cdot 10^{-6}$	$7.97 \cdot 10^{-6}$	$984.95 \cdot 10^{-6}$	$2.78 \cdot 10^{-6}$
α	0.9129	1.0381	0.9885	1.0131
benötigte Farben	294	156	6	4
größte gefundene Clique	294	155	6	3

Die ersten beiden Zeilen der Tabelle enthalten die Anzahl der Knoten bzw. Kanten der jeweiligen Graphen. Um die hier erhaltenen Ergebnisse in die Aussagen der anderen Kapitel einordnen zu können, wird die Anzahl der Features m und die Featurewahrscheinlichkeit p benötigt. Während beim Schauspieler-Graphen die Features (und damit auch p) gegeben sind, müssen sie für die anderen Graphen zusätzlich bestimmt werden. Dafür wurde der in [8] vorgestellte Algorithmus zum Finden von Features verwendet, der in gewisser Weise zu einem Graphen einen wahrscheinlichen bipartiten Generator liefert. In der vorletzten Zeile ist schließlich die Anzahl der Farben aufgeführt, die der Greedy-Algorithmus auf einer smallest-last Anordnung der Knoten benötigt. Zum Vergleich wurde dann in der letzten Zeile die Größe der größten gefundenen Featureclique als eine untere Schranke für die chromatische Zahl aufgeführt.

Auf den getesteten Graphen schneidet der Algorithmus hervorragend ab. Mindestens zwei der vier Graphen werden optimal gefärbt. Bei den anderen beiden wurde höchstens eine Farbe mehr als unbedingt nötig benutzt. Eine exakte Einordnung dieser Ergebnisse in die Aussagen der vorangegangenen Kapitel ist nicht möglich, da alle verwendeten Schranken völlig von konstanten Faktoren abstrahieren und außerdem immer nur fast sicher gelten. Eine Einordnung in die durch Abbildung 5 gegebene Übersicht ist in Abbildung 6 dargestellt.

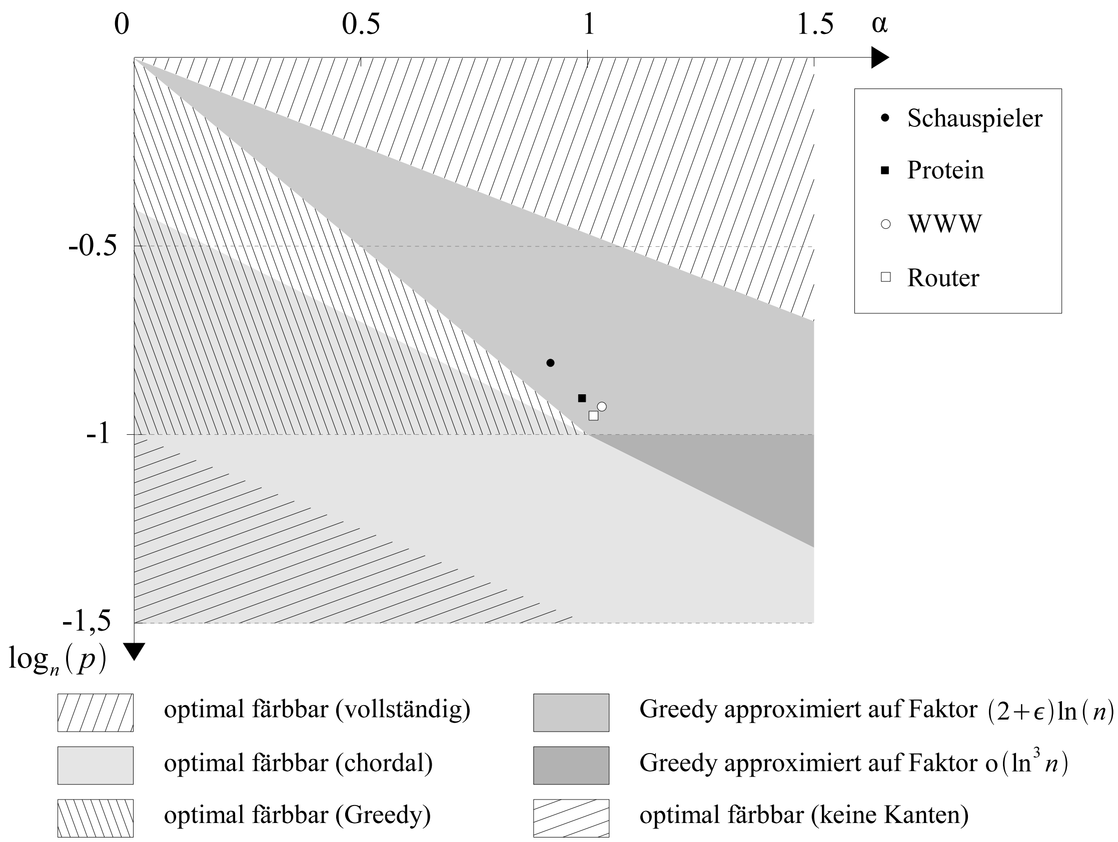

Abbildung 6: reale Graphen in Relation zu den Färbeergebnissen

Die Graphen liegen also alle in einem Bereich, in dem für den Greedy-Algorithmus eine Approximationsgüte von $(2 + \epsilon) \ln n$ nachgewiesen werden konnte. Dennoch werden sie annähernd optimal gefärbt.

7 Zusammenfassung

In den vorangegangen Kapiteln wurde untersucht, wie sich zufällige Schnittgraphen färben lassen. Dabei wurde besonderes Augenmerk auf den Greedy-Algorithmus gelegt, da dieser vergleichsweise einfach aufgebaut und einfach zu untersuchen ist und außerdem auf klassischen zufälligen Graphen sehr gute Ergebnisse liefert.

Es wurde das Modell der zufälligen Schnittgraphen $\mathcal{G}_{n,m,p}$ verwendet, da die Ausgaben dieses Modells oft deutlich näher an realen Graphen liegen als das bei den klassischen zufälligen Graphen der Fall ist. Kapitel 3 beschäftigte sich mit der Frage, wie und in welchen Bereichen für die Anzahl der Features m und die Featurewahrscheinlichkeit p zufällige Schnittgraphen fast sicher optimal gefärbt werden können. Es konnte gezeigt werden, dass für $p \ll \sqrt{\frac{1}{nm}}$ fast alle zufälligen Schnittgraphen chordal sind. Auf der Klasse der chordalen Graphen ist das Färbeproblem nicht NP-vollständig, sondern kann in linearer Zeit gelöst werden. Im Anschluss daran wurde der Greedy-Algorithmus auf einer smallest-last Anordnung der Knoten untersucht. Für $\alpha < 1$ stellte sich heraus, dass dieser Algorithmus für $\frac{\ln^2 n}{n} \ll p \ll \frac{1}{m \ln n}$ fast sicher optimal färbt. Da sich in linearer Zeit auch feststellen lässt ob ein Graph chordal ist, liefert die Kombination beider Ergebnisse ein Verfahren zum optimalen Färben fast aller zufälligen Schnittgraphen in der Vereinigung der beiden Parameterbereiche.

Bei der Analyse in Kapitel 3 wurde die Größe der Featurecliquen als untere Schranke für die chromatische Zahl benutzt. Es war zu erwarten, dass diese Schranke mit zunehmendem α und p schlechter wird. Daher beschäftigte sich das darauffolgende Kapitel mit der Stabilitätszahl von zufälligen Schnittgraphen, mit deren Hilfe eine andere untere Schranke für die chromatische Zahl abgeleitet werden kann. Zunächst konnte gezeigt werden, dass für $p \ll \frac{1}{m}$ fast alle zufälligen Schnittgraphen eine stabile Menge enthalten, die $(1-o(1))n$ Knoten umfasst. Für größere p wurden eine obere und eine untere Schranke der Stabilitätszahl bestimmt, die sich um den Faktor $(2+\epsilon)\ln n$ unterscheiden. Zur weiteren Einschränkung hat der Autor mit Hilfe der 2. Moment Methode und einiger aufwendiger Rechnungen versucht die untere Schranke zu verbessern. Dabei konnte sie für den oberen Teil des p-Intervalls verbessert und der Quotient aus beiden Schranken auf einen konstanten Faktor reduziert werden. Teilweise offen bleibt die Frage, wie die Stabilitätszahl in dem unteren Teil des p-Intervalls aussieht. Die in dieser Arbeit verfolgten Ansätze scheinen nicht geeignet zu sein, die hier erhaltenen Schranken zu verbessern. Es wäre auch denkbar, dass die Stabilitätszahlen der zufälligen Schnittgraphen entsprechend gestreut sind und sich die beiden Schranken trotz des vergleichsweise großen Unterschieds nicht wesentlich verbessern lassen.

Mit Hilfe der gewonnenen Aussagen zur Stabilitätszahl konnte in Kapitel 5 eine neue und deutlich bessere untere Schranke für die chromatische Zahl zufälliger Schnittgraphen im Bereich $\frac{\ln n}{m} \ll p \ll \sqrt{\frac{\ln n}{m}}$ abgeleitet werden. Die Approximationsgüte des Greedy-Algorithmus ließ sich damit unabhängig von der verwendeten Knotenanordnung auf $(2+\epsilon)\ln n$ im Bereich $\max\left(\frac{\ln n}{n}, \frac{\ln n}{m}\right) \ll p \ll \sqrt{\frac{\ln n}{m}}$ und auf $o(\ln^3 n)$ im Bereich $\sqrt{\frac{1}{nm}} \ll p \ll \frac{1}{n}$ abschätzen. Der Vergleich mit der Approximationsgüte von Greedy auf klassischen Zufallsgraphen $(2+o(1))$ sowie die experimentellen Ergebnisse lassen allerdings vermuten, dass die hier bewiesene Approximationsgüte noch nicht bestmöglich ist. Daher wurde der Ansatz zum Beweis der Approximationsgüte auf $G_{n,p}$ auf Anwendbarkeit bei $\mathcal{G}_{n,m,p}$ untersucht, erwies sich jedoch als nicht übertragbar. Das Finden einer besseren Abschätzung für die Approximationsgüte des Greedy-Algorithmus (evtl. in Verbindung mit anderen Heuristiken) sowie die Analyse anderer Färbealgorithmen sind mögliche Anknüpfungspunkte für weitere Untersuchungen.

Die vorliegende Arbeit steckt somit einen Parameterbereich für m und p ab, auf dem zufällige Schnittgraphen fast sicher in polynomieller Zeit optimal gefärbt werden können. In dem Bereich, in dem keine Optimalität nachgewiesen werden konnte, wurde die Approximationsgüte des Greedy-Algorithmus abgeschätzt. Damit liefert diese Arbeit für das komplette Spektrum zufälliger Schnittgraphen Vorschläge für effiziente Färbealgorithmen inklusive einer Abschätzung der Qualität der erzeugten Färbungen. Da das Modell der zufälligen Schnittgraphen aufgestellt wurde um reale Graphen anzunähern, sind die hier erhaltenen Ergebnisse auch für praktische Probleme von Nutzen. Neben der Analyse von Färbealgorithmen stellt die vorliegende Arbeit außerdem eine umfassende Analyse der Stabilitätszahl zufälliger Schnittgraphen zur Verfügung.

Die Stabilitätszahl von $\mathcal{G}_{n,m,p}$ wurde von Singer in [7] im Rahmen einer ersten Untersuchung zufälliger Schnittgraphen punktuell (für $mp^2 = const$) bestimmt. Nach Kenntnis des Autors gibt es bisher keine Veröffentlichung zum Färben solcher Graphen. Die vorliegende Arbeit baut auf den Ergebnissen von Singer auf und führt die von ihr begonnene Analyse fort. Damit liefert sie einen wichtigen Beitrag zum Verständnis zufälliger Schnittgraphen.

Das nach Ansicht des Autors wichtigste offene Problem bei zufälligen Schnittgraphen ist die Bestimmung der durchschnittlichen Pfadlänge sowie des Clustering-Koeffizienten. Die durchschnittliche Pfadlänge entspricht dem Erwartungswert für den Abstand zweier beliebiger Knoten eines Graphen, während der Clustering-Koeffizient die Tendenz der Kanten zur Zusammenballung (clustering) misst (siehe Abschnitt 1.3). Beide Größen sind wichtige Merkmale bei der Bestimmung der Ähnlichkeit zwischen verschiedenen Graphen (siehe [6] und Abschnitt 1.3). Eine genaue Kenntnis dieser Größen für das $\mathcal{G}_{n,m,p}$-Modell würde daher eine Analyse hinsichtlich der Übertragbarkeit von für dieses Modell erhaltenen Ergebnissen auf reale Graphen ermöglichen.

8 Quellen

Literatur

[1] Kenneth Appel, Wolfgang Haken: *Every planar map is four colorable.* American Mathematical Society, Providence, RI, 1989.

[2] Michael R. Garey, David S. Johnson, H. C. So: *An application of graph coloring to printed circuit testing.* IEEE Transactions on Circuits and Systems, 1976.

[3] Thomas Emden-Weinert, Stefan Hougardy, Bernd Kreuter Hans Jürgen Prömel Angelika Steger: *Einführung in Graphen und Algorithmen*, September 1996.

[4] Halldórsson, Magnús M.: *A Still Better Performance Guarantee for Approximate Graph Coloring.* Information Processing Letters, 45, 1993.

[5] Jean-Loup Guillaume, Matthieu Latapy: *Bipartite structure of all complex networks.* Information Processing Letters, 90, 2004.

[6] Réka Albert, Albert-László Barabási: *Statistical Mechanics of Complex Networks.* Reviews of Modern Physics, 74, 2002.

[7] Singer, Karen B.: *Random Intersection Graphs*, John Hopkins University, Baltimore, Maryland, 1995.

[8] Michael Behrisch, Anusch Taraz: *Efficiently covering complex networks with cliques of similar vertices.* Special issue of the international journal Theoretical Computer Science (TCS) on Complex Networks, 2004.

[9] Svante Janson, Tomasz Łuczak, Andrzej Ruciński: *Random Graphs.* John Wiley & Sons, 2000.

[10] Palmer, Edgar M.: *Graphical Evolution - An Introduction to the Theory of Random Graphs.* Wiley-Interscience, 1985.

[11] Réka Albert, Hawoong Jeong, Albert-László Barabási Zoltán N. Oltvai Sean Mason: *Network Databases of the Center for Complex Network Research*, 2005. http://www.nd.edu/~networks/resources.htm.

[12] Ramesh Govindan, Hongsuda Tangmunarunkit: *SCAN+Lucent internet map from the ISI*, 1999. http://www.isi.edu/scan/mercator/maps.html.