

Seminarausarbeitungskonzept
On the Power of Color Refinement

Florian Lüdiger
17. Dezember 2017

im Rahmen des Seminars
Algorithm Engineering
von Prof. Dr. Petra Mutzel
Wintersemester 2017/18

Betreuer:
Christopher Morris

Basierend auf:
V. Arvind, Johannes Köbler, Gaurav Rattan und Oleg Verbitsky
On the Power of Color Refinement
https://link.springer.com/chapter/10.1007/978-3-319-22177-9_26

Fakultät für Informatik
Algorithm Engineering (Ls11)
Technische Universität Dortmund

Inhaltsverzeichnis

1	Einführung	4
2	Graph-Isomorphie und Color Refinement	5
2.1	Graph-Isomorphie	5
2.2	Color Refinement	5
2.3	Formale Definition von CR-Graphen	7
3	Lokale Struktur von CR-Graphen	8
4	Globale Struktur von CR-Graphen	10
4.1	Mögliche Eigenschaften von Zellgraphen	10
4.2	Allgemeine globale Eigenschaften von CR-Graphen	11
4.3	Baumartige Struktur von CR-Graphen	15
5	Ergebnis und Laufzeit	16
5.1	Hinreichende Bedingungen für das Erkennen von CR-Graphen	16
5.2	Laufzeit	16
6	Fazit	17
	Literaturverzeichnis	18

Glossar

bipartiter Graph

Ein Graph heißt bipartit, wenn sich seine Knoten in zwei Teilmengen aufteilen lassen, sodass Kanten nur zwischen den beiden Mengen aber nicht innerhalb existieren. $G[X, Y]$ ist der bipartite Graph, welcher durch die beiden disjunkten Teilmengen $X, Y \subseteq V(G)$ und allen Kanten, die Knoten aus X und Y verbinden, gebildet wird. 3

bipartites Komplement

Das bipartite Komplement eines Graphen G mit Knotenklassen X und Y stellt der bipartite Graph G' dar, welcher die selben Knotenklassen enthält, allerdings das Komplement der Kanten zwischen den beiden Knotenklassen. 3, 8, 11

biregulärer Graph

Ein bipartiter Graph G mit Knotenklassen X und Y ist biregulär, wenn alle Knoten in X und Y den gleichen Grad besitzen. 3

disjunkte Vereinigung von Sternen

Bei einer *disjunkten Vereinigung* von *Sternen* $sK_{1,t}$ bezeichnet s die Anzahl der Sterne und t die Anzahl der äußeren Knoten jedes Sterns. 3, 8

disjunkte Vereinigung

Die knotendisjunkte Vereinigung von G und H wird $G + H$ genannt. Die disjunkte Vereinigung von m Kopien des Graphen G wird als mG geschrieben. 3, 7

Hypergraph

Ein Hypergraph ist ein Graph, in dem eine Kante, auch Hyperkante genannt, mehr als zwei Knoten verbinden kann. 3, 13

matching Graph

Ein matching Graph ist ein Graph, bei dem kein Knoten mehr als eine inzidente Kante besitzt. Somit gibt es nur Zusammenhangskomponenten mit maximal einer Kante und zwei Knoten. 3, 8

Multimenge

Eine Multimenge unterscheidet eine Menge dadurch, dass Elemente mehrfach vorkommen können. Multimengen werden hier durch doppelte geschweifte Klammern dargestellt. 3, 5, 7

Nachbarschaft

Die Nachbarschaft $N(u)$ bildet die Menge der Knoten, die adjazent zu $u \in V(G)$ sind. 3, 5

regulärer Graph

Ein Graph ist regulär wenn alle seine Knoten den gleichen Grad besitzen. 3

Stern

In einem Sterngraphen $K_{1,t}$ gibt es einen zentralen Knoten, welcher mit allen anderen Knoten des Graphen, t an der Zahl, durch eine Kante verbunden ist. Die anderen Knoten sind untereinander nicht verbunden. 3

Subgraph

Der Subgraph $G[X]$ ist ein Teilgraph von G , der durch die Knotenmenge $X \subseteq V(G)$ und deren inzidenten Kanten gebildet wird. 3

Unigraph

Die Isomorphieeigenschaften eines Unigraphen sind durch die Sequenz der Knotengrade genau definiert. Dies bedeutet, dass allein anhand der Knotengrade zweier Unigraphen bestimmt werden kann, ob diese isomorph sind. 3, 8

vollständiger bipartiter Graph

In einem vollständigen bipartiten Graphen mit den Knotenmengen X und Y sind alle Knoten aus X mit allen Knoten aus Y verbunden. Somit haben alle Knoten aus X den Grad $|Y|$ und alle Knoten aus Y den Grad $|X|$. 3, 8

vollständiger Graph

In einem vollständigen Graphen K_n mit n Knoten, ist jeder Knoten mit jedem anderen Knoten verbunden und besitzt somit den Grad $n - 1$. 3, 8

Zyklus

Ein geschlossener Pfad eines Graphen über n Knoten wird Zyklus C_n genannt. 3, 8

1 Einführung

Dieses Dokument stellt eine Aufarbeitung der Veröffentlichung „On the Power of Color Refinement“ von Arvind et al. (2015) dar. In dieser wird ein Verfahren zur Überprüfung von Isomorphieeigenschaften von Graphen vorgestellt, welches die bekannte Color Refinement Heuristik erweitert. Graph-Isomorphie ist ein viel gefragtes Thema, da diese in vielen unterschiedlichen Bereichen von Wissenschaft und Technik benötigt wird. Ein bekanntes Anwendungsgebiet stellt die Chemieinformatik dar, welche die Gleichheit von Molekülen untersucht und zur Verarbeitung großer Datenmengen die Unterstützung eines performanten Algorithmus benötigt.

Eine weit verbreitete Heuristik zur Lösung des Isomorphie-Problems stellt der Color Refinement Algorithmus dar, welcher in vielen Fällen zwei nicht isomorphe Graphen unterscheiden kann, nicht aber in allen Fällen. Somit wird die Frage aufgeworfen, ob sich eine Klasse von Graphen, nachfolgend CR-Graphen genannt, definieren lässt, für die der Algorithmus sicher sein kann, dass sowohl nicht isomorphe, als auch isomorphe Paare von Graphen korrekt erkannt werden können. Diese Frage wird im Folgenden beantwortet und eine Definition der Menge der CR-Graphen aufgestellt werden.

Zunächst werden Graph-Isomorphie und Color Refinement genauer vorgestellt und eine formale Definition für die Anforderungen an CR-Graphen aufgestellt. Wie sich herausstellen wird, lassen sich die erforderlichen Eigenschaften an CR-Graphen in eine lokale Struktur und eine globale Struktur einordnen, welche in den folgenden Kapiteln detailliert definiert werden. Die Ergebnisse bezüglich hinreichender Bedingungen zur Erkennung von CR-Graphen sowie Laufzeit und Fazit werden im abschließenden Kapitel behandelt.

2 Graph-Isomorphie und Color Refinement

2.1 Graph-Isomorphie

Zwei Graphen G und H sind isomorph, kurz $G \simeq H$, wenn es eine bijektive Abbildung ϕ der Knoten von G auf H gibt, sodass die Adjazenz aller Knoten erhalten bleibt. Es gilt also: $(u, v) \in E_G \Leftrightarrow (\phi(u), \phi(v)) \in E_H$ für alle $u, v \in V_G$. Eine so definierte Abbildung ϕ wird **Isomorphismus** genannt.

Es wurde gezeigt, dass das Graph-Isomorphie-Problem in NP liegt, die NP -Vollständigkeit ist allerdings noch unklar. Es konnte außerdem noch kein Algorithmus gefunden werden, welcher das Problem in polynomieller Zeit löst, jedoch wird davon ausgegangen, dass es nicht NP -vollständig ist. Siehe Goldreich et al. (1991).

2.2 Color Refinement

Der Color Refinement Algorithmus stellt eine Heuristik dar, mit der in polynomieller Zeit festgestellt werden kann, dass zwei Graphen nicht isomorph sind. Der Algorithmus geht wie folgt vor.

1. Sämtliche Knoten des Graphen werden in der selben Farbe gefärbt.
2. Die Knotenfärbungen werden verfeinert, indem überprüft wird, ob zwei Knoten gleicher Farbe unterschiedliche *Nachbarschaften* mit Berücksichtigung der Farbe besitzen. Ist dies der Fall, werden die Knoten in unterschiedlichen Farben gefärbt und das Verfeinern wird fortgeführt.
3. Ist die Bedingung für kein Knotenpaar mehr erfüllt, terminiert der Algorithmus.
4. Sind die *Multimengen* der Farben beider Graphen unterschiedlich, sind diese nicht isomorph.

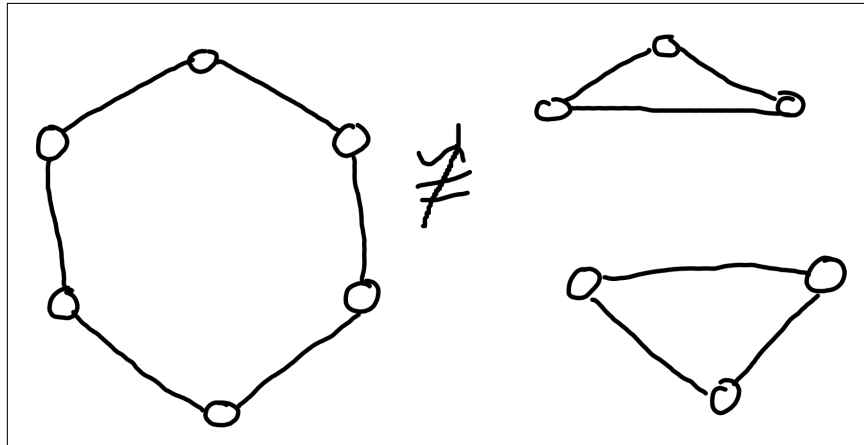


Abbildung 1: Zwei nicht isomorphe Graphen, welche vom Color Refinement nicht unterschieden werden können

Es kann bei diesem Vorgehen vorkommen, dass zwei nicht isomorphe Graphen G und H nicht voneinander unterschieden werden können. Ein Beispiel hierfür wird in Abbildung 1 dargestellt. Zu erkennen ist, dass alle Knoten beider Graphen initial in einer Farbe gefärbt wurden, danach allerdings kein Verfeinerungsschritt mehr nötig ist, da die Nachbarschaften aller Knoten identisch gefärbt sind. Obwohl offensichtlich erkennbar ist, dass die beiden Graphen nicht isomorph sind, gelingt es dem Algorithmus also nicht diese zu unterscheiden. Auf Basis dieser Erkenntnisse wird die folgende Klasse von Graphen definiert.

Definition 1 Ein Graph G wird **CR-Graph** genannt, wenn der Color Refinement Algorithmus diesen von jedem beliebigen, nicht zu G isomorphen Graphen H unterscheiden kann. Die Arbeit, auf der diese Ausarbeitung beruht (Arvind et al. (2015)) nennt diese Kerneigenschaft von CR-Graphen **amenable**

Durch die Eigenschaften des Color Refinement Algorithmus ist folgende Beziehung bekannt: Color Refinement erkennt zwei Graphen als nicht isomorph \Rightarrow die Graphen sind nicht isomorph. Aus Definition 1 folgt außerdem: zwei CR-Graphen sind nicht isomorph \Rightarrow Color Refinement kann diese als nicht isomorph unterscheiden. Somit lässt sich wie in Korollar 1 formuliert die Äquivalenz der Aussagen für die Klasse der CR-Graphen schlussfolgern.

Korollar 1 Für zwei Graphen aus der Klasse der CR-Graphen gilt: Color Refinement erkennt die Graphen als isomorph \Leftrightarrow die Graphen sind isomorph.

Neben dem klassischen Isomorphie-Test ist der Color Refinement Algorithmus auch für andere Anwendungsfelder geeignet. Als Beispiel ist hier die Verkleinerung von linearen Programmen durch Reduktion der Dimensionen der Gleichungssysteme und somit eine Erhöhung der Effizienz zu nennen. Siehe Grohe et al. (2014). Weitere Anwendungsfelder finden sich in Shervashidze et al. (2011) und Kersting et al. (2014).

2.3 Formale Definition von CR-Graphen

Der Color Refinement Algorithmus berechnet iterativ eine Sequenz von Knotenfärbungen C^i für einen Graphen G . Die initiale Färbung C^i weist jedem Knoten die selbe Farbe zu. Nachfolgend wird in jeder Iteration nach folgender Regel eine neue Färbung gebildet.

$$C^{i+1}(u) = (C^i(u), \{\{C^i(a) : a \in N(u)\}\}) \quad (1)$$

Die doppelten geschweiften Klammern $\{\{\}$ markieren hierbei eine *Multimenge*. Die Knoten von G werden mithilfe einer Partitionierung \mathcal{P} in die entsprechenden Farbklassen unterteilt, wodurch also zu jeder Färbung C^i eine Partition \mathcal{P}^i gehört.

Bei der Ausführung des Algorithmus wird irgendwann eine Partitionierung erreicht, die sich bei weiteren Verfeinerungsschritten nicht mehr verändert, da keine neue Färbung mehr generiert wird. Für diese gilt das Folgende.

Definition 2 Wenn sich eine Partitionierung bei weiteren Verfeinerungsschritten nicht mehr verändert, stellt dies die stabile Partitionierung \mathcal{P}^s dar. Für sie gilt $\mathcal{P}^s = \mathcal{P}^i$ für alle $i \geq s$.

Die stabile Partitionierung ist bei jedem Graphen nach maximal $n - 1$ Verfeinerungsschritten erreicht, da spätestens dann jeder Knoten eine unterschiedliche Farbe besitzt und somit keine weitere Verfeinerung mehr möglich ist.

Eine Partitionierung kann außerdem die Eigenschaft **equitable** haben. Dafür müssen die im Folgenden definierten Eigenschaften erfüllt sein. Die Elemente der Partitionierung werden hier Zellen genannt.

Definition 3 Eine Partitionierung wird **equitable** genannt, wenn die folgenden Eigenschaften erfüllt sind:

1. Jede Zelle $X \in \mathcal{P}$ ist einfarbig und enthält somit nur Knoten einer einzigen Farbe.
2. Für jede Zelle $X \in \mathcal{P}$ ist der Graph $G[X]$ ein regulärer Graph.
3. Für beliebige Zellen $X, Y \in \mathcal{P}$ ist der bipartite Graph $G[X, Y]$ ein biregulärer Graph.

Es ist somit leicht zu erkennen, dass die durch das Color Refinement erstellte stabile Partitionierung \mathcal{P}^s equitable ist.

Eine weitere Eigenschaft der Graphfärbungen C^i ist der Erhalt der Färbung über Isomorphismen hinweg.

Lemma 1 Für die Färbungen zweier isomorpher Graphen G und H und ihren Isomorphismus ϕ muss gelten: $C^i(u) = C^i(\phi(u))$ für alle $u \in V_G$.

Aus dieser Aussage leitet sich die folgende Gleichung her.

$$\forall i \geq 0 : \{\{C^i(u) : u \in V_G\}\} = \{\{C^i(v) : v \in V_H\}\} \quad (2)$$

Das Color Refinement akzeptiert somit zwei Graphen G und H bei der Eingabe der *disjunkten Vereinigung* $G + H$ genau dann, wenn die Gleichung (2) erfüllt ist.

Das Überprüfen dieser Bedingung ist in endlicher Zeit berechenbar. Für einen Zeugen i dafür, dass die Gleichung nicht erfüllt ist, gilt $i < 2n$, wobei n die Anzahl der Knoten jedes Graphens ist. Dies beruht darauf, dass \mathcal{P}^{2n-1} in jedem Fall die stabile Partitionierung von $G + H$ ist und weitere Verfeinerungsschritte keine neue Partitionierung erstellen würden. Es genügt sogar die Gleichung für $i = n$ zu verifizieren, da die Existenz einer Partitionierung $\mathcal{P}^{i+1} \neq \mathcal{P}^i$ bedeuten würde, dass es mehr als n Partitionen, da in jedem Partitionierungsschritt mindestens eine Partition hinzukommt. Mehr als n Partitionen sind ein Indikator dafür, dass die Graphen G und H nicht isomorph sind, da die n Knoten jedes Graphen unmöglich in mehr als n Partitionen unterteilt werden können und Gleichheit somit ausgeschlossen ist.

Aus den gewonnenen Erkenntnissen ergibt sich folgende Erweiterung zu Definition 1.

Definition 4 Ein Graph G wird **CR-Graph** genannt, wenn für jeden beliebigen, nicht isomorphen Graphen H die Gleichung (2) für $i = n$ nicht erfüllt ist.

3 Lokale Struktur von CR-Graphen

Um schlussendlich die Frage beantworten zu können von welcher Beschaffenheit CR-Graphen sein müssen, damit sie die in Definition 4 genannte Eigenschaft erfüllen, werden hier zunächst notwendige, lokale Eigenschaften solcher Graphen vorgestellt. Die Basis dieses Kapitels bildet das folgende Lemma, welches für die Zellen X und Y der stabilen Partition \mathcal{P}_G eines CR-Graphen G einige Merkmale definiert.

Lemma 2 Die Zellen der stabilen Partition \mathcal{P}_G eines CR-Graphen erfüllen folgende Eigenschaften:

- (A) Für beliebige Zellen $X \in \mathcal{P}_G$ ist $G[X]$ ein leerer Graph, vollständiger Graph, matching Graph mK_2 , das Komplement eines matching Graphen oder der 5er Zyklus
- (B) Für beliebige Zellen $X, Y \in \mathcal{P}_G$ ist $G[X, Y]$ ein leerer Graph, vollständiger bipartiter Graph, eine disjunkte Vereinigung von Sternen $sK_{1,t}$, bei der X die Menge der s inneren Knoten und Y die Menge der st Blätter ist, oder das bipartite Komplement des zuletzt genannten Graphen

Wichtig ist an dieser Stelle, dass die in **B** angesprochene Vereinigung von Sternen ausschließlich Sterne enthält, die die gleiche Anzahl, nämlich t , von Blättern besitzen. Um das Lemma beweisen zu können, werden zunächst folgende Hilfsaussagen benötigt.

Lemma 3 Ein regulärer Graph mit Grad d und n Knoten ist ein Unigraph, genau dann wenn $d \in \{0, 1, n-2, n-1\}$ oder $d = 2$ und $n = 5$.

Tabelle 1: Abbildung der Aussagen aus Lemma 2 **A** und Lemma 3

Graphentyp aus Lemma 2 A	Definition aus Lemma 3
Leerer Graph	$d = 0$
Vollständiger Graph	$d = n - 1$
Matching Graph	$d = 1$
Komplement eines Matching Graphen	$d = n - 2$
5er Zyklus	$d = 2$ und $n = 5$

Tabelle 2: Abbildung der Aussagen aus Lemma 2 **B** und Lemma 4

Graphentyp aus Lemma 2 B	Definition aus Lemma 4
Leerer Graph	$d = c = 0$
Vollständiger bipartiter Graph	$d = m$ und $c = n$
Disjunkte Vereinigung von Sternen	$d = 1$ oder $c = 1$
Komplement einer disjunkten Vereinigung von Sternen	$d = m - 1$ oder $c = n - 1$

Der Beweis zu dieser Aussage findet sich in Johnson (1975). Bei genauerer Betrachtung fällt auf, dass die in **A** genannten Graphentypen sich in diesem Lemma wieder finden. Tabelle 1 gibt dazu eine Übersicht über die Abbildung der Aussagen.

Lemma 4 *Sei G ein bipartiter Graph mit den beiden Komponenten X und Y . Die Isomorphieeigenschaften von G sind genau dann vollständig dadurch definiert, dass die m Knoten aus X den Grad c und die n Knoten aus Y den Grad d haben, wenn $c \in \{0, 1, n - 1, n\}$ oder $d \in \{0, 1, m - 1, m\}$ gilt.*

Bewiesen wird diese Aussage in Koren (1976). Auch für bipartite Graphen können die Aussagen von Lemma 2 (B) und Lemma 4 aufeinander abgebildet werden, was in Tabelle 2 übersichtlich dargestellt wird.

Wenn ein Graph G nun einen Subgraphen $G[X]$ oder $G[X, Y]$ mit $X, Y \in \mathcal{P}_G$ enthält, dessen Typ allerdings nicht in Lemma 2 aufgelistet wird, so lässt sich dieser durch einen nicht-isomorphen regulären oder biregulären Graphen mit gleichem Grad ersetzen. Diese Aussage folgt daraus, dass dessen Isomorphieeigenschaften wie in Lemma 3 und 4 dargestellt nicht allein durch ihre Parameter definiert sind, was das Ersetzen durch einen nicht-isomorphen Graphen mit eben diesen Parametern unmöglich machen würde.

Diese Ergebnisse lassen sich für den Beweis von Lemma 2 nutzen, wenn gezeigt wird, dass der durch das Ersetzen des Subgraphen entstandene Graph H vom Color Refinement nicht von G unterschieden werden kann. Im Folgenden gilt dazu, dass G und H die gleiche Knotenmenge enthalten und die Farben für einen Knoten u durch $C_G^i(u)$ und $C_H^i(u)$ für die jeweiligen Graphen definiert sind.

Lemma 5 *Seien X und Y Zellen der stabilen Partition eines Graphen G .*

- (a) *Ist der Graph H aus G erstellt worden, indem die Kanten eines Subgraphen $G[X]$ mit denen eines regulären Graphen mit dem gleichen Grad auf der glei-*

chen Knotenmenge X ersetzt wurde, so gilt: $C_G^i(u) = C_H^i(u)$ für alle $u \in V(G)$ und beliebiges i .

- (b) Ist der Graph H aus G erstellt worden, indem die Kanten eines Subgraphen $G[X, Y]$ mit denen eines biregulären Graphen mit dem gleichen Grad und den gleichen Partitionen ersetzt wurden, sodass die Knotengrade erhalten bleiben, gilt somit: $C_G^i(u) = C_H^i(u)$ für alle $u \in V(G)$ und beliebiges i .

Beweis von Lemma 2:

(A) Sei $G[X]$ ein Graph, dessen Struktur nicht in Lemma 2 (A) aufgelistet wird, so ist dieser wie in Lemma 3 dargestellt, kein Unigraph. Somit ist es möglich $G[X]$ innerhalb von G durch einen nicht-isomorphen Graphen mit gleichen Parametern zu ersetzen. Der so entstandene Graph H ist nicht isomorph zum Ursprungsgraphen G , da die Subgraphen $G[X]$ und $H[X]$ nicht isomorph sind. Teil (a) von Lemma 5 hingegen zeigt, dass Gleichung 2 für beliebige i erfüllt ist und das Color Refinement die beiden Graphen somit nicht unterscheiden kann. Aus dieser Aussage folgt, dass der Graph G nicht zu der Klasse der CR-Graphen gehört. Durch Kontraposition folgt nun, dass ein CR-Graph die in Lemma 2 (A) genannten Eigenschaften in jedem Fall erfüllt.

(B) Diese Aussage lässt sich unter Zuhilfenahme von Lemma 4 und Lemma 5 (ii) sehr ähnlich zu (A) begründen.

4 Globale Struktur von CR-Graphen

Zusätzlich zur lokalen Struktur ist außerdem die globale Struktur von CR-Graphen interessant. Der Begriff globale Struktur gibt hier an, dass der Zellgraph von G bestimmte Eigenschaften erfüllen muss.

Definition 5 Der Zellgraph $C(G)$ eines Graphen G wird aus dessen stabilen Partition \mathcal{P}_G gebildet. Es handelt sich dabei um einen vollständigen Graphen, bei dem die Knoten die Zellen von \mathcal{P}_G darstellen.

4.1 Mögliche Eigenschaften von Zellgraphen

Für das Verständnis der folgenden Eigenschaften von CR-Graphen ist es erforderlich einige Begriffe einzuführen und Eigenschaften von Zellgraphen zu erklären und zu benennen, wozu einige Definitionen folgen. Die Knoten eines Zellgraphen, auch als Zellen bezeichnet, können dabei folgende Eigenschaften aufweisen.

Definition 6 Eine Zelle $X \in C(G)$ wird **homogen** genannt, wenn der Graph $G[X]$ vollständig oder leer ist. Anderenfalls wird diese **heterogen** genannt.

Definition 7 Für eine heterogene Zelle $X \in C(G)$ finden sich je nach Beschaffenheit von $G[X]$ die Bezeichnungen **matching**, **co-matching** oder **pentagonal**. Eine homogene Zelle wird dagegen entweder **leer** oder **vollständig** genannt.

Für die Kanten eines Zellgraphen finden sich ebenfalls unterschiedliche Bezeichnungen, welche deren Beschaffenheit beschreiben.

Definition 8 Eine Kante X, Y mit $X, Y \in C(G)$ wird **isotrop** genannt, wenn der bipartite Graph $G[X, Y]$ vollständig oder leer ist. Anderenfalls wird diese **anisotrop** genannt.

Definition 9 Eine anisotrope Kante X, Y wird **Konstellation** genannt, wenn $G[X, Y]$ eine disjunkte Vereinigung von Sternen ist. Anderenfalls wird diese **Co-Konstellation** genannt. Bei Co-Konstellationen bildet das *bipartite Komplement* von $G[X, Y]$ eine disjunkte Vereinigung von Sternen. Eine isotrope Kante dagegen wird entweder **leer** oder **vollständig** genannt.

Definition 10 Ein Pfad $X_1 X_2 \dots X_l$ in $C(G)$, bei dem jede Kante X_i, X_{i+1} anisotrop ist, wird **anisotroper Pfad** genannt. Wenn dieser Pfad einen Kreis schließt, wird er als **anisotroper Zyklus** bezeichnet. Gilt für einen anisotropen Pfad $|X_1| = |X_2| = \dots = |X_l|$ dann wird er **gleichmäßig** genannt.

4.2 Allgemeine globale Eigenschaften von CR-Graphen

Mit dem so gewonnenen Hintergrundwissen kann nun das folgende Lemma formuliert werden, welches globale Eigenschaften von CR-Graphen definiert.

Lemma 6 Der Zellgraph $C(G)$ eines CR-Graphen G erfüllt folgende Eigenschaften:

- (C) $C(G)$ enthält keinen gleichmäßigen, anisotropen Pfad, der zwei heterogene Zellen verbindet
- (D) $C(G)$ enthält keinen gleichmäßigen, anisotropen Zyklus
- (E) $C(G)$ enthält weder einen anisotropen Pfad $XY_1Y_2\dots Y_lZ$, sodass $|X| < |Y_1| = |Y_2| = \dots = |Y_l| > |Z|$, noch einen anisotropen Zyklus $XY_1Y_2\dots Y_l$, sodass $|X| < |Y_1| = |Y_2| = \dots = |Y_l|$ und die Zelle Y_l heterogen ist
- (F) $C(G)$ enthält keinen anisotropen Pfad $XY_1Y_2\dots Y_l$, sodass $|X| < |Y_1| = |Y_2| = \dots = |Y_l|$ und die Zelle Y_l heterogen ist

Beweis C: Angenommen P wäre, entgegen Bedingung C, ein gleichmäßiger, anisotroper Pfad in $C(G)$, welcher die beiden anisotropen Komponenten X und Y verbindet. Da es sich um einen gleichmäßigen Pfad handelt, besitzen sämtliche Komponenten auf dem Pfad die Kardinalität $k = |X| = |Y|$. Alle Kanten in P , die eine Co-Konstellation bilden, werden nun komplementiert, sodass diese eine Konstellation ergeben. Da nun sämtliche Kanten aus P eine Konstellation darstellen und die Kardinalität aller Komponenten identisch ist, können die Sterne, welche die disjunkte Vereinigung für die Konstellationen bilden nur eine einzige Form annehmen. Diese bestehen aus einem Zentralknoten und einem Blatt, wobei diese in unterschiedlichen Komponenten liegen. Dadurch, dass sämtliche Kanten in P diese Form annehmen

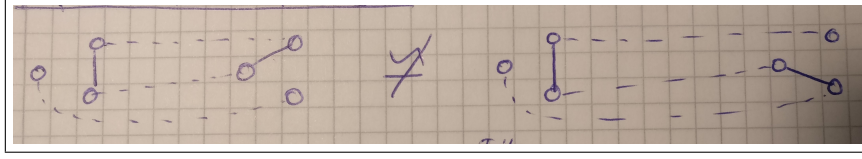


Abbildung 2: Beispiel für das Ersetzen eines Subgraphen in Lemma 6 C

ergeben sich k knotendisjunkte Pfade zwischen X und Y . Es sind also alle Knoten aus X mit genau einem Knoten aus Y durch einen Pfad verbunden und andersherum. Sei $v \in X$ durch einen solchen Pfad mit seinem Gegenstück $v^* \in Y$ verbunden. Der Pfad P wird *conducting* genannt, wenn diese Abbildung von Knoten einen Isomorphismus von $G[X]$ auf $G[Y]$ darstellt. Dazu müssen zwei Knoten $u, v \in X$ genau dann adjazent sein wenn ihre Gegenstücke $u^*, v^* \in Y$ adjazent sind. Angenommen eine der beiden Komponenten ist *matching* und die andere ist *co-matching*, dann wird P ebenfalls *conducting* genannt, wenn die Isomorphie zwischen der *matching* Komponente und dem Komplement der *co-matching* Komponente besteht.

Nachfolgend wird ein Graph H erstellt werden, der nicht zu G isomorph ist, welchen das Color Refinement allerdings nicht von G unterscheiden kann. Da X und Y heterogen sind, können die Kanten des Subgraphen $G[X]$ durch einen isomorphen aber unterschiedlichen Graphen mit der gleichen Knotenmenge X ersetzt werden. Ist P in G *conducting*, so muss $G[X]$ so ersetzt werden, dass der entstehende Pfad in H nicht *conducting* ist, anderenfalls muss der entsprechende Pfad in H *conducting* sein. Ein Beispiel für ein solches Ersetzen findet sich in Abbildung 2.

Lemma 5 (a) besagt nun, dass der Color Refinement Algorithmus für G und H die gleiche Färbung errechnet und somit nicht zwischen den beiden unterscheiden kann. Außerdem besagt Lemma 1, dass jeder Isomorphismus zwischen G und H Zellen auf sich selbst abbilden muss. Da also gilt $\phi(v^*) = \phi(v)^*$, muss ϕ die Eigenschaft von P erhalten *conducting* oder nicht zu sein, was in diesem konstruierten Beispiel nicht erfüllt ist. Es folgt, dass G und H nicht isomorph sind und G daher kein CR-Graph sein kann.

Beweis D: Es wird wieder entgegen der Bedingung **D** angenommen, dass $C(G)$ einen gleichmäßigen, anisotropen Zyklus Q der Länge m enthält. Da dieser gleichmäßig ist, besitzen sämtliche Zellen in Q die gleiche Kardinalität k . Wie im Beweis von **C** kann nun der Graph $G[A, B]$ für jede Co-Konstellation A, B komplementiert werden, sodass ein Zyklus entsteht, welcher nur Konstellationen als Kanten enthält. Ebenfalls wie im Beweis von **C** gibt es eine eins zu eins Beziehung zwischen den Knoten zweier Benachbarter Komponenten, wodurch eine knotendisjunkte Vereinigung von Kreisen in G entsteht. Alle Kreise besitzen dabei ein Vielfaches der Länge m , wobei so im Extremfall ein Kreis der Länge km oder k Kreise der Länge m vorliegen.

Durch die Anzahl der Kreise und deren Länge können die Isomorphieeigenschaften des entstehenden Subgraphen vollständig beschreiben, sodass der Isomorphietyp für den Zyklus Q nachfolgend als $\tau(Q)$ bezeichnet wird. Sei ϕ ein Isomorphismus von G zu einem anderen Graphen H , so ist ϕ' der auf den Zellgraphen übertragene Isomorphismus von $C(G)$ zu $C(H)$. Durch die Definition von τ ergibt sich, dass $\tau(\phi'(Q)) = \tau(Q)$ gelten muss. Intuitiv bedeutet dies, dass ein Kreis einer bestimm-

ten Größe nur auf einen anderen Kreis abgebildet werden kann, welcher die selbe Größe besitzt.

Seien $X, Y \in Q$ zwei aufeinanderfolgende Zellen, so kann der dadurch entstehende Subgraph $G[X, Y]$ durch einen isomorphen aber nicht identischen Graphen ersetzt werden, da die Kante X, Y anisotrop ist. Somit entsteht ein neuer Graph H , welcher beispielsweise so konstruiert sein kann, dass für den betrachteten Zyklus die beiden eingangs erwähnten Extreme vorliegen. Dadurch, dass sich hier zwei Möglichkeiten ergeben wie $\tau(Q)$ nach dem Austauschen aussehen kann, ist es möglich den Subgraphen $G[X, Y]$ so zu ersetzen, dass $\tau(Q)$ sich ändert.

Es wird wieder das Lemma 5 genutzt, um zu zeigen, dass das Color Refinement nicht zwischen G und H unterschieden werden kann. Außerdem ist G nicht isomorph zu H , da sich H so konstruieren lässt, dass $\tau(Q)$ in beiden Graphen unterschiedlich ist, wodurch der Graph G nicht zu der Klasse der CR-Graphen gehören kann.

Beweis E: Sei $P = XY_1Y_2\dots Y_lZ$ ein anisotroper Pfad in G , sodass dessen Kardinalitäten durch $|X| < |Y_1| = |Y_2| = \dots = |Y_l| > |Z|$ beschrieben werden. Dabei sei es erlaubt, dass $X = Z$ ist und somit ein Kreis entsteht. Für alle Kanten aus P , welche eine Co-Konstellation sind, wird im Folgenden das bipartite Komplement betrachtet, sodass sämtliche Kanten eine Konstellation bilden. Somit stellen die Subgraphen $G[X, Y_1] = sK_{1,t}$ und $G[Z, Y_l] = aK_{1,b}$ mit $s, a, t, b \geq 2$ disjunkte Vereinigungen von Sternen dar. Es folgt, dass $|X| = s, |Z| = a$ und $|Y_1| = |Y_l| = st = ab$ sind. In dem Beweis zu Bedingung **C** wurde bereits gezeigt, dass der gleichmäßige, anisotrope Pfad $Y_1Y_2\dots Y_l$ eine Abbildung der Knoten der Zellen Y_1 und Y_l bildet, welche im Folgenden dazu verwendet wird die Identifikation Y des gesamten Pfades zu bilden. Der gesamte Pfad lässt sich zusammenfassen, indem $Y = Y_1 = Y_l = \{1, 2, \dots, st\}$ als Identifikation genutzt wird. Diese lässt sich unterteilen, indem die Menge Y_x für beliebige $x \in X$ die zu x adjazenten Knoten in Y beschreibt. Eine ähnliche Menge Y_z für alle $z \in Z$ lässt sich durch die zu z adjazenten Knoten aus Y bilden. Daraus folgt, dass für beliebige $x \neq x'$ aus X und $z \neq z'$ aus Z folgendes gilt:

$$|Y_x| = t, |Y_z| = b, Y_x \cap Y_{x'} = \emptyset \text{ und } Y_z \cap Y_{z'} = \emptyset$$

Die hier definierten Symbole werden in Abbildung 3 beispielhaft dargestellt.

Es wird nun der *Hypergraph* \mathcal{Y}_G auf den Knoten von Y gebildet. Dabei wird für jedes $x \in X$ und $z \in Z$ eine Hyperkante aus der Knotenmenge von Y_x beziehungsweise Y_z erstellt, wodurch sich die gesamte Kantenmenge durch $\{Y_x\}_{x \in X} \cup \{Y_z\}_{z \in Z}$ zusammensetzt. Zwei dieser Hyperkanten sind in Abbildung 3 beispielhaft eingetragen, der Graph enthält selbstverständlich aber noch weitere. Falls für irgendwelche x und z die Mengen Y_x und Y_z identisch sind, so enthält dieser Graph auch Mehrfachkanten. Der ursprünglich betrachtete Pfad P wird **flach** genannt, sofern kein Paar $(x, z) \in X \times Y$ existiert, sodass keine der entsprechenden Hyperkanten Y_x und Y_z die andere enthält. Es werden nun die Knoten in Y so umsortiert, dass die Hyperkanten $Y_z, z \in Z$ aufeinanderfolgende Intervalle beschreiben, was in dem abgebildeten Beispiel bereits optisch so dargestellt wurde.

Es wird nun wieder ein zu G nicht isomorpher Graph H gebildet, welcher vom Color Refinement nicht von diesem unterschieden werden kann. Ist P in G flach,

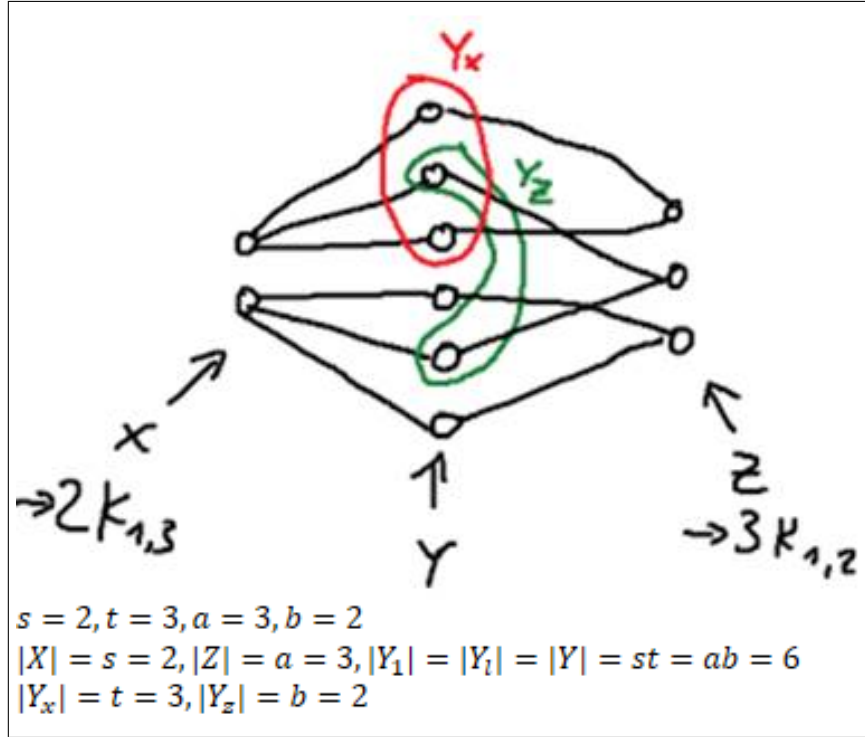


Abbildung 3: Beispiel für die Darstellung der im Beweis für Bedingung C eingeführten Symbole

wird H so konstruiert, dass der Pfad P nicht mehr flach ist. Dies ist möglich, indem die Kanten des Subgraphen $G[X, Y_1]$ durch einen isomorphen aber anderen Subgraphen ersetzt werden, sodass die in \mathcal{Y}_H gebildeten Hyperkanten in aufeinanderfolgenden Intervallen stehen. Formal wird also $\mathcal{Y}_H = \{Y_i\}_{i \in [s]} \cup \{Y_z\}_{z \in Z}$ mit $Y_i = \{(i-1)t+1, 2, \dots, it\}$ definiert. Der Pfad P ist somit nicht mehr flach, da sowohl wie vorher beschrieben die Hyperkanten Y_z , als auch die Hyperkanten $\{Y_x\}$ in aufeinanderfolgender Reihenfolge in Y liegen, wodurch sichergestellt ist, dass es immer ein (x, z) , sodass eine der entsprechenden Hyperkanten die andere einschließt. Für das vorgestellte Beispiel ist in Abbildung 4 diese Erweiterung mit einem ersetzten Subgraphen $G[X, Y_1]$, sowie den Hyperkanten $\{Y_x\}$ in rot und $\{Y_z\}$ in grün zu sehen.

Ist P in G allerdings nicht flach, so wird der Subgraph $G[X, Y_1]$ in H so ersetzt, dass der entsprechende Pfad P in H flach wird. Formal ist also $Y_i = \{i, i+s, \dots, i+(t-1)s\}$.

H ist somit in der Tat nicht zu G isomorph, da für diese Eigenschaft der Pfad P in beiden Graphen entweder flach oder in beiden Graphen nicht flach sein müsste. Aus Teil (a) von Lemma 5 folgt außerdem, dass das Color Refinement die beiden Graphen nicht unterscheiden kann, wodurch klar ist, dass der Graph G kein CR-Graph sein kann.

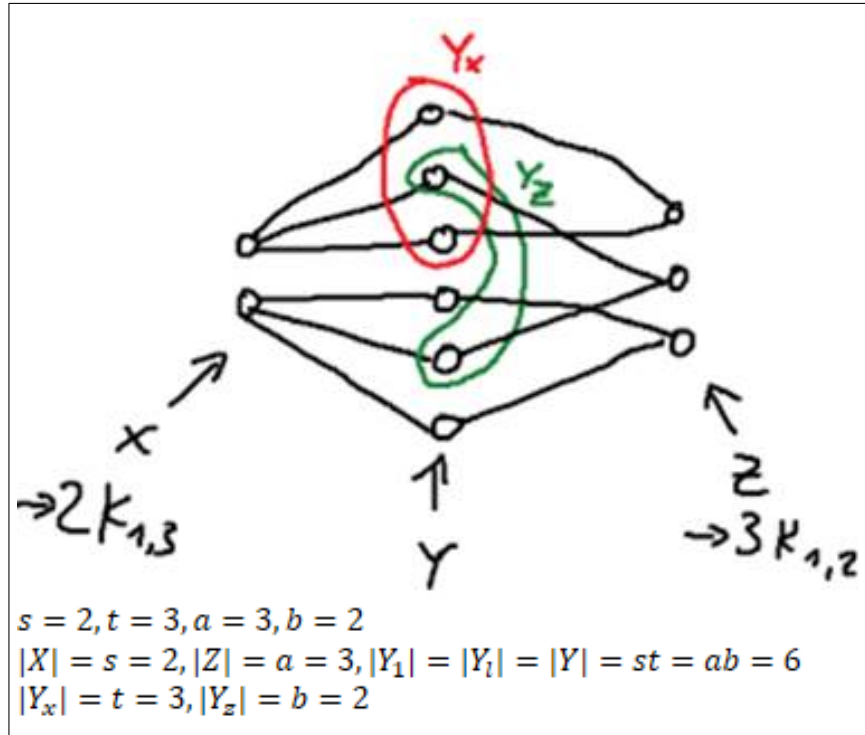


Abbildung 4: Erweiterung des Beispiels, sodass der Pfad nicht mehr flach ist

4.3 Baumartige Struktur von CR-Graphen

Bei genauerer Betrachtung fällt auf, dass der Zellgraph von CR-Graphen eine baumartige Struktur aufweist. Um das folgende Lemma verstehen zu können, wird der Begriff **anisotrope Komponente** benötigt.

Definition 11 In einem Zellgraphen $C(G)$ bezeichnet eine **anisotrope Komponente** einen Subgraphen, dessen Kanten alle isotrop sind. Wenn eine Zelle keine inzidente, anisotrope Kante besitzt, dann ergibt sich daraus eine anisotrope Komponente mit nur einer Zelle.

Lemma 7 Angenommen ein CR-Graph G erfüllt die Bedingungen **A-F** aus den Lemmata 2 und 6. Für jede anisotrope Komponente A von $C(G)$ gelten folgende Eigenschaften:

- (G) A ist ein Baum, der folgende Monotonieeigenschaft erfüllt: Sei R eine Zelle aus A mit minimaler Kardinalität, so ist A_R der gerichtete Baum mit Wurzel R ; Für jede gerichtete Kante (X, Y) aus A_R gilt dann $|X| \leq |Y|$
- (H) A enthält maximal eine heterogene Zelle; Wenn eine solche Zelle existiert, hat diese minimale Kardinalität in A

5 Ergebnis und Laufzeit

Die vorgestellten lokalen und globalen Eigenschaften von CR-Graphen reichen wie im Folgenden gezeigt aus, um hinreichende Bedingungen für CR-Graphen zu formulieren und darauf basierend ein effizientes Verfahren für das Erkennen solcher Graphen zu entwickeln.

5.1 Hinreichende Bedingungen für das Erkennen von CR-Graphen

Theorem 1 *Für einen Graphen G sind folgende Aussagen äquivalent:*

- (a) G ist ein CR-Graph
- (b) G erfüllt Bedingungen **A-F**
- (c) G erfüllt Bedingungen **A, B, G** und **H**

Beweis: Die Äquivalenz der Aussagen wird gezeigt, indem gezeigt wird, dass gilt: $(a) \rightarrow (b) \rightarrow (c) \rightarrow (a)$. Die bisher erlangten Erkenntnisse ermöglichen es bereits einen großen Teil dieser Aussage zu bestätigen. Somit wurde in den Lemmata 2 und 6 gezeigt, dass $(a) \rightarrow (b)$ gilt. Ebenfalls wurde in Lemma 7 gezeigt, dass $(b) \rightarrow (c)$ gilt. Es bleibt also nur noch zu zeigen, dass auch $(c) \rightarrow (a)$ gültig ist.

Sei dazu G ein Graph, welcher die Bedingungen **A, B, G** und **H** erfüllt und H ein durch Color Refinement nicht von G unterscheidbarer Graph, dann ist zu zeigen, dass G und H isomorph sind.

5.2 Laufzeit

Zur Berechnung der Laufzeit wird im Folgenden davon ausgegangen, dass der Graph G in Adjazenzlistendarstellung vorliegt. Nach Cardon und Crochemore (1982) lässt sich die stabile Partition eines Graphen G in Zeit $\mathcal{O}((n + m) \log n)$ berechnen.

Theorem 2 *Die Klasse der CR-Graphen ist in Zeit $\mathcal{O}((n + m) \log n)$ entscheidbar. Dabei bezeichnet n die Anzahl der Knoten und m die Anzahl der Kanten des Eingabegraphen.*

Beweis: Zunächst wird die stabile Partition $\mathcal{P}_G = \{X_1, X_2, \dots, X_k\}$ berechnet, was wie eingangs erwähnt die Laufzeit $\mathcal{O}((n + m) \log n)$ benötigt. Außerdem wird $C^*(G)$ definiert als der Zellgraph, bei dem sämtliche leeren Kanten, also solche bei denen keine Verbindungen zwischen den Elementen der beiden Endpunkte besteht, entfernt wurden.

Für die Elemente $X_i \in C^*(G)$ werden die Adjazenzlisten gebildet, indem die Adjazenzliste eines beliebigen Knoten $u \in X_i$ durchlaufen wird und sämtliche Zellen aufgelistet werden, welche einen zu u adjazenten Knoten enthalten. Die dadurch gewonnenen Informationen sind identisch für alle Knoten aus X_i , da diese alle gleichartige Nachbarschaften besitzen, weshalb es ausreicht die Operation für einen beliebigen Knoten durchzuführen. Durch die Informationen aus der Adjazenzliste lässt

sich der Grad der Knoten innerhalb der Zellen bestimmen und somit leicht Bedingung **A** durch die in Lemma 3 vorgestellten Bedingungen für jeden Subgraphen $G[X_i]$ überprüfen.

Für jede Kante X_i, X_j aus $C^*(G)$ wird der Wert d_{ij} berechnet, welcher die Anzahl der Nachbarn in X_j beschreibt, zu denen jeder Knoten aus X_i adjazent ist. Dieser Wert wird ebenfalls für den Fall $i = j$ betrachtet, wobei die Nachbarn innerhalb der Zelle gezählt werden. Dadurch, dass die Werte $|X_i|$, $|X_j|$ und d_{ij} nun bekannt sind, lässt sich Bedingung **B** durch die in Lemma 4 vorgestellten Bedingungen überprüfen.

Da alle Zellen des Zellgraphen $C^*(G)$ im schlimmsten Falle unterschiedliche Farben haben, gibt es maximal n Zellen. Mit dem beschriebenen Verfahren können die Bedingungen **A** und **B** somit in der Zeit $\mathcal{O}(n)$ überprüft werden.

Zum Überprüfen von Bedingung **H** wird eine Breitensuche auf dem Zellgraphen $C^*(G)$ durchgeführt, welche alle anisotropen Komponenten findet und gleichzeitig überprüft, ob diese eine Baumstruktur aufweisen und nur eine anisotrope Komponente enthält.

Wird diese Breitensuche nun von der Zelle minimaler Kardinalität für jede Komponente wiederholt, so lässt sich die in Bedingung **G** beschriebene Monotonieeigenschaft für jede Kante überprüfen.

Der Breitensuchealgorithmus benötigt eine Laufzeit von $\mathcal{O}(n + m)$, da der Zellgraph maximal n Knoten und maximal m Kanten besitzt.

Es ist somit zu erkennen, dass das Errechnen der stabilen Partition, auf der die ganzen Operationen ausgeführt werden mit $\mathcal{O}((n + m) \log n)$ die dominierende Laufzeit des Algorithmus ist.

6 Fazit

In dem Paper von Arvind et al. (2015) sowie in dieser Ausarbeitung wurde die Klasse der CR-Graphen eingeführt und dessen Eigenschaften bezüglich des Color Refinement Algorithmus aufgezeigt. Durch die Definition von hinreichenden Bedingungen zur Entscheidung der Klasse wurde ein Verfahren skizziert, welches es ermöglicht, zu entscheiden, ob der Color Refinement Algorithmus auf einen Graphen anwendbar ist und sowohl für Wahr- als auch Falschaussagen die korrekte Antwort liefert. Die in Kapitel 1 vorgestellte Frage wurde somit beantwortet und es wurde eine handfeste Definition für alle CR-Graphen aufgestellt.

Wie von Babai et al. (1980) gezeigt wurde, erfüllen zufällig generierte Graphen die Eigenschaften von CR-Graphen mit hoher Wahrscheinlichkeit. Somit sind die hier vorgestellten Ergebnisse für eine Vielzahl von Graphen anwendbar, für welche damit ein Polynomialzeitalgorithmus für das Graph-Isomorphie-Problem existiert.

Literatur

V. Arvind, J. Köbler, G. Rattan, und O. Verbitsky. **On the Power of Color Refinement**, Seiten 339–350. Springer International Publishing, Cham, 2015.

- ISBN 978-3-319-22177-9. doi:10.1007/978-3-319-22177-9_26. URL https://doi.org/10.1007/978-3-319-22177-9_26.
- L. Babai, P. Erdős, und S. M. Selkow. Random graph isomorphism. **SIAM Journal on Computing**, 9(3):628–635, 1980. doi:10.1137/0209047. URL <https://doi.org/10.1137/0209047>.
- A. Cardon und M. Crochemore. Partitioning a graph in $o(|a|\log 2|v|)$. **Theoretical Computer Science**, 19(1):85 – 98, 1982. ISSN 0304-3975. doi:[https://doi.org/10.1016/0304-3975\(82\)90016-0](https://doi.org/10.1016/0304-3975(82)90016-0). URL <http://www.sciencedirect.com/science/article/pii/0304397582900160>.
- O. Goldreich, S. Micali, und A. Wigderson. Proofs that yield nothing but their validity or all languages in np have zero-knowledge proof systems. **J. ACM**, 38(3):690–728, Jull 1991. ISSN 0004-5411. doi:10.1145/116825.116852. URL <http://doi.acm.org/10.1145/116825.116852>.
- M. Grohe, K. Kersting, M. Mladenov, und E. Selman. **Dimension Reduction via Colour Refinement**, Seiten 505–516. Springer Berlin Heidelberg, Berlin, Heidelberg, 2014. ISBN 978-3-662-44777-2. doi:10.1007/978-3-662-44777-2_42. URL https://doi.org/10.1007/978-3-662-44777-2_42.
- R. Johnson. Simple separable graphs. **Pacific Journal of Mathematics**, 56(1):143–158, 1975.
- K. Kersting, M. Mladenov, R. Garnett, und M. Grohe. Power iterated color refinement. In **Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence**, Seiten 1904–1910. AAAI Press, 2014.
- M. Koren. Pairs of sequences with a unique realization by bipartite graphs. **Journal of Combinatorial Theory, Series B**, 21(3):224–234, 1976.
- N. Shervashidze, P. Schweitzer, E. J. v. Leeuwen, K. Mehlhorn, und K. M. Borgwardt. Weisfeiler-lehman graph kernels. **Journal of Machine Learning Research**, 12(Sep):2539–2561, 2011.