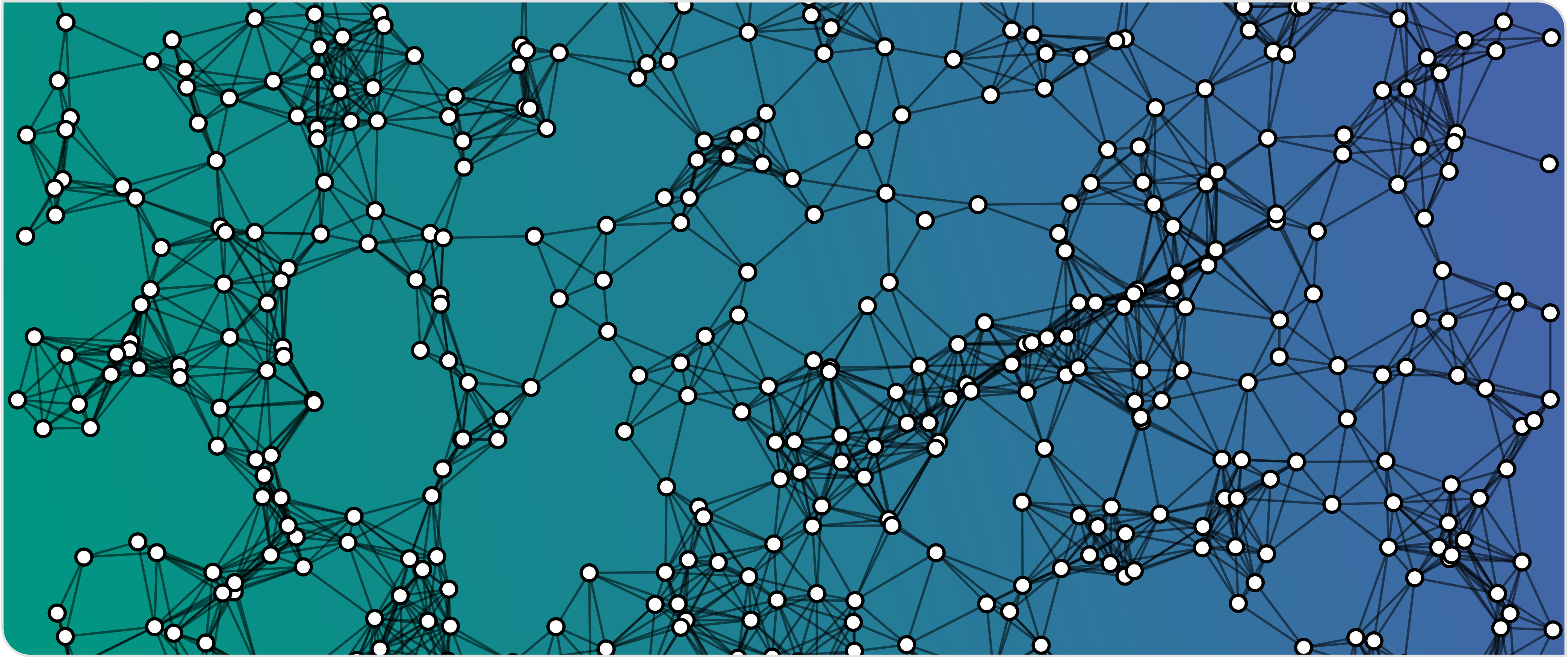
Karlsruher Institut für Technologie

Parametrisierte Algorithmen

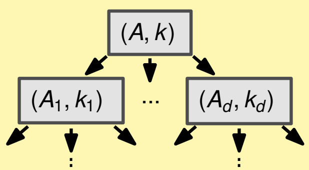
Kernbildung: Ähnliche Bäume

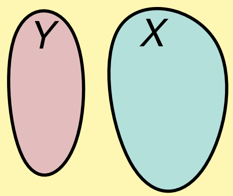


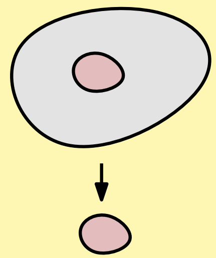
Inhalt

Basic Toolbox

- beschränkte Suchbäume
- iterative Kompression
- Kernbildung

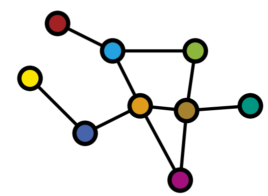
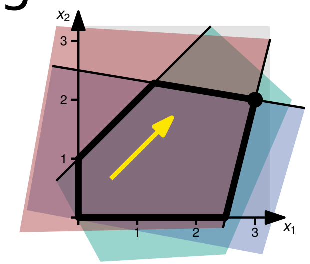
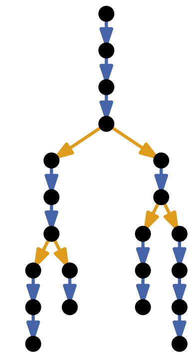






Erweiterte Toolbox

- lineare Programme
- Branch-and-Reduce
- Color Coding

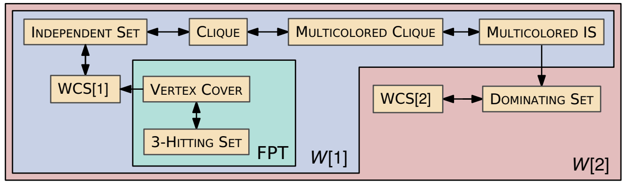




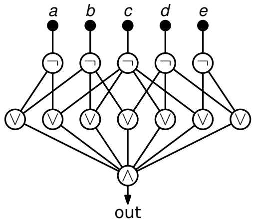
Baumweite

- dynamische Programme
- chordale & planare Graphen
- Courcelles Theorem

Untere Schranken

- parametrisierte Reduktionen
- boolesche Schaltkreise und die W-Hierarchie
- ETH und SETH





Phylogenetische Bäume

Definition
Ein **phylogenetischer Baum** ist ein ungewurzelter und vollständiger Binärbaum, sodass jedes Blatt ein eindeutiges Label ($\equiv$ Spezies) hat.

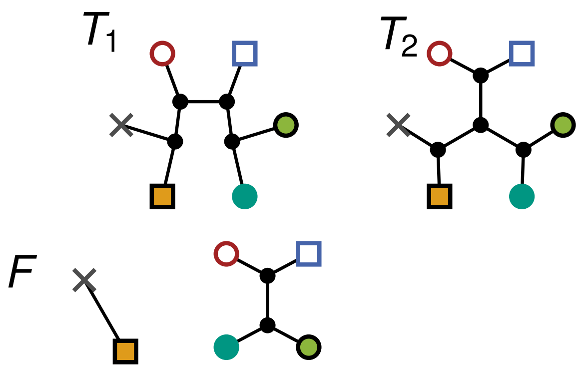
- werden oft automatisiert erstellt
- beispielsweise basierend auf DNA-Sequenzierung

Problem

- unterschiedliche Algorithmen liefern unterschiedliche Bäume
- untersch. Daten (z.B. durch Messfehler) liefern untersch. Bäume
- Wie kann man unterschiedliche Bäume T_1 und T_2 auf der gleichen Blattmenge L miteinander vergleichen?

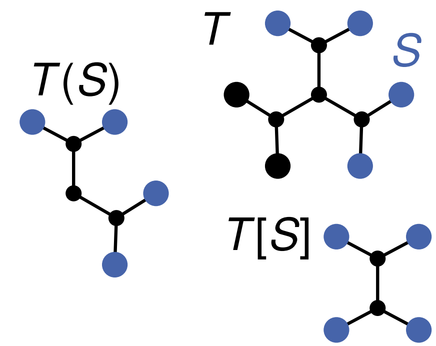
Maximum Agreement Forest

- Wald F aus Binärbäumen mit Blattmenge L
- Bäume in F enthalten in T_1 und in T_2 (als Minor)
- minimiere #Bäume in F



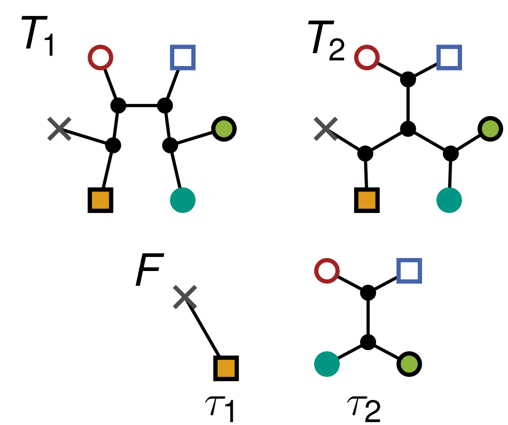
Notation

Blattinduzierte Teilbäume
- sei T ein Baum mit Blättern $L(T)$ und sei $S \subseteq L(T)$
- $T(S)$ = minimaler Teilbaum von T, der S enthält
- kontrahiere alle Knoten mit Grad 2 in $T(S) \rightarrow T[S]$



Agreement Forest
- sei F ein Wald bestehend aus den Bäumen $\tau_1, \dots, \tau_k$
- betrachte zwei Bäume T_1 und T_2 mit $L(T_1) = L(T_2) = L(F)$
- F ist *Agreement Forest* für T_1 und T_2, wenn $T_1[L(\tau_i)] = T_2[L(\tau_i)] = \tau_i$ und die $T_1(L(\tau_i))$ (für $i \in \{1, \dots, k\}$), sowie die $T_2(L(\tau_i))$ sind disjunkt

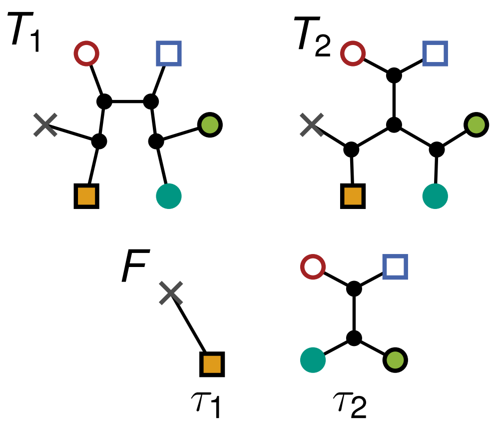
> **Problem: Maximum Agreement Forest**
> Gegeben sind T_1, T_2 und ein Parameter k. Gibt es einen Agreement Forest mit maximal k Bäumen?



Ziel
- finde einen Kernbildungsalgorithmus
- also polynomielle Reduktionsregeln
- Größe der resultierenden Bäume hängt nur von k ab

Grober Fahrplan

Problem: MAXIMUM AGREEMENT FOREST
Gegeben sind T_1, T_2 und ein Parameter k. Gibt es einen Agreement Forest mit maximal k Bäumen?



Ziel

- finde einen Kernbildungsalgorithmus
- also polynomielle Reduktionsregeln
- Größe der resultierenden Bäume hängt nur von k ab

Vorschläge?

Bevor wir losreduzieren: Was ist das Ziel?

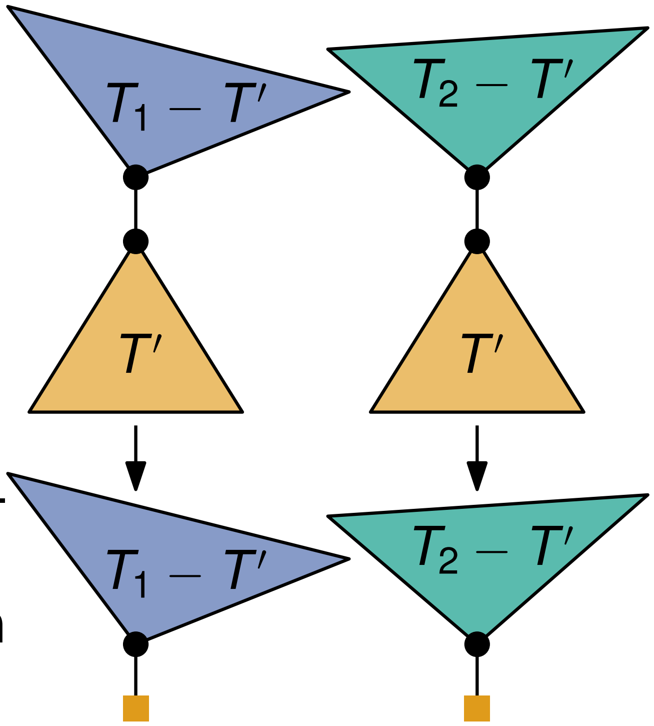
- Größe der resultierenden Bäume hängt nur von k ab
- etwas stärkere, aber konkretere Behauptung:
  - jeder Baum in einem Agreement Forest enthält nur wenige Blätter
    $\Rightarrow$ insgesamt wenige Blätter $\Rightarrow$ kleine Instanz
- Welche Baumstrukturen widersprechen der Behauptung?
- Können wir diese mittels Reduktionsregeln loswerden?

Ungünstige Baumstruktur 1

Behauptung
Jeder Baum in einem Agreement Forest enthält nur wenige Blätter.

Gegenbeispiel

- T_1 und T_2 können einen Teilbaum T' mit vielen Blättern gemeinsam haben
- im Agreement Forest kann man T' als einen der Bäume $\tau_1, \dots, \tau_k$ wählen



Reduktionsregel 1

- finde Kanten in T_1 und T_2, sodass diese den selben Baum T' abtrennen
- ersetze T' durch ein einzelnes Blatt mit neuem (gleichem) Label

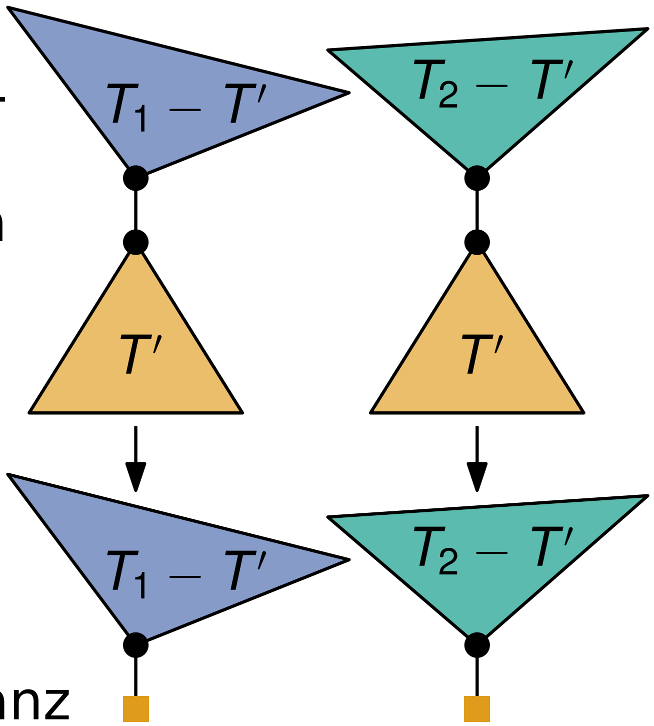
Lemma
Reduktionsregel 1 ist sicher und in polynomieller Zeit ausführbar.

Beweis: nächste Folie

Reductionsregel 1

Reduktionsregel 1

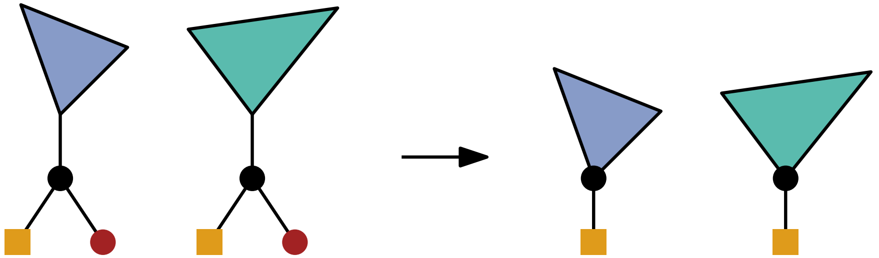
- finde Kanten in T_1 und T_2, sodass diese den selben Baum T' abtrennen
- ersetze T' durch ein einzelnes Blatt mit neuem (gleichem) Label



Lemma
Reduktionsregel 1 ist sicher und in polynomieller Zeit ausführbar.

Beweis

- sei F' ein Agreement Forest der neuen Instanz und sei τ_1' der Baum mit dem neuen Blatt in F'
- Ersetzung dieses Blattes in τ_1' durch T' liefert Agreement Forest der gleichen Größe für T_1 und T_2
- andere Richtung: ähnlich
- Laufzeit: z.B. durch iterative Ersetzung von „Kirschen“



Ungünstige Baumstruktur 2

Behauptung
Jeder Baum in einem Agreement Forest enthält nur wenige Blätter.

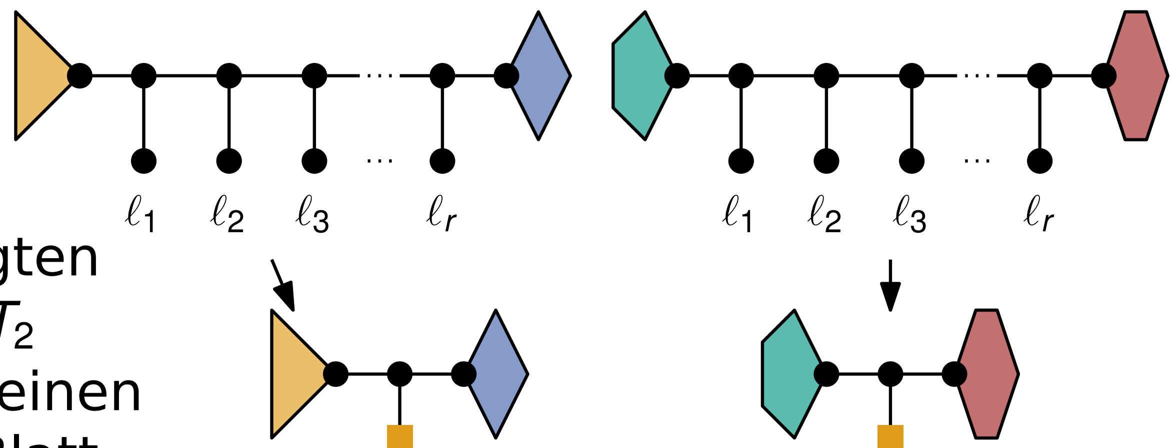
Gegenbeispiel

(trotz Anwendung der Reduktionsregel 1)

- viele Blätter in geteiltem Teilbaum

Reduktionsregel 2

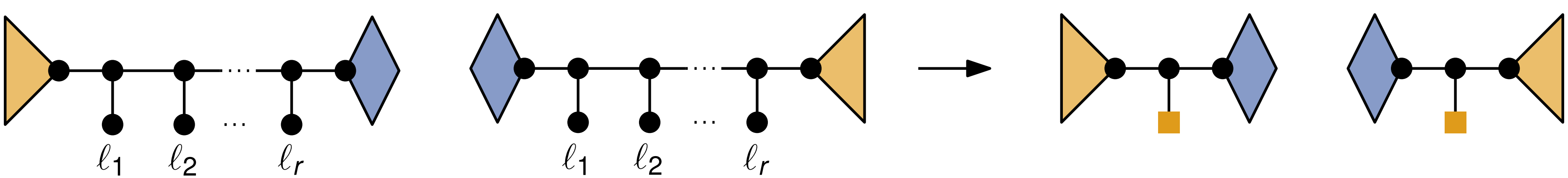
- finde Pfad mit angehängten Blättern $\ell_1, \dots, \ell_r$ in T_1 und T_2
- ersetze den Pfad durch einen Knoten mit einem neuen Blatt



Ist diese Regel sicher?

Problem

- Anzahl notwendiger Bäume wird ggf. verringert



drei Bäume nötig ein Baum genügt

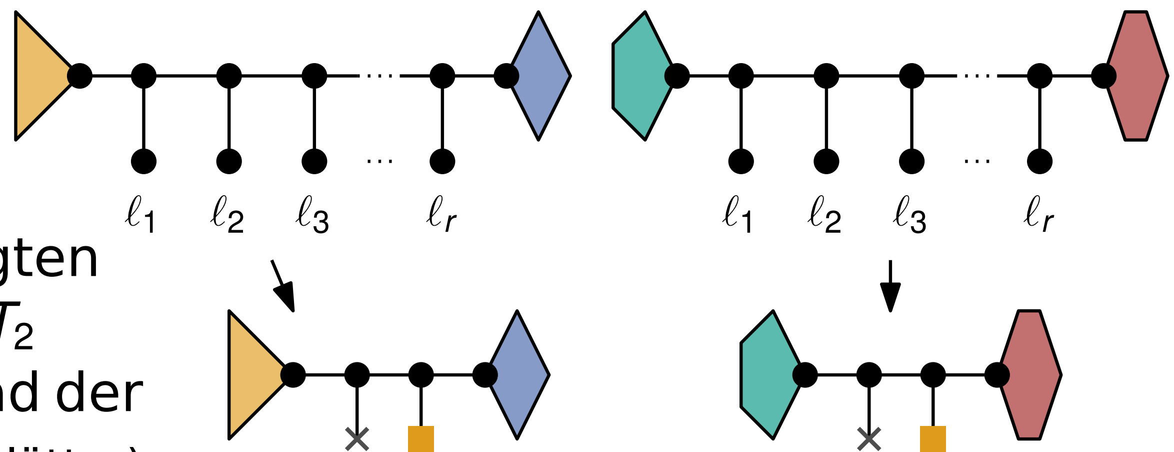
Ungünstige Baumstruktur 2

Behauptung
Jeder Baum in einem Agreement Forest enthält nur wenige Blätter.

Gegenbeispiel

(trotz Anwendung der Reduktionsregel 1)

- viele Blätter in geteiltem Teilbaum



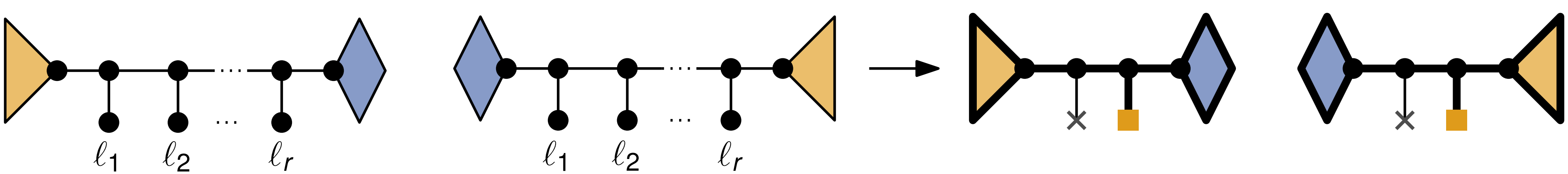
Reduktionsregel 2

- finde Pfad mit angehängten Blättern $\ell_1, \dots, \ell_r$ in T_1 und T_2
- ersetze ihn durch einen Pfad der Länge 2 (zwei neue Blätter)

Ist diese Regel sicher?

Problem

- Anzahl notwendiger Bäume wird ggf. verringert



drei Bäume nötig

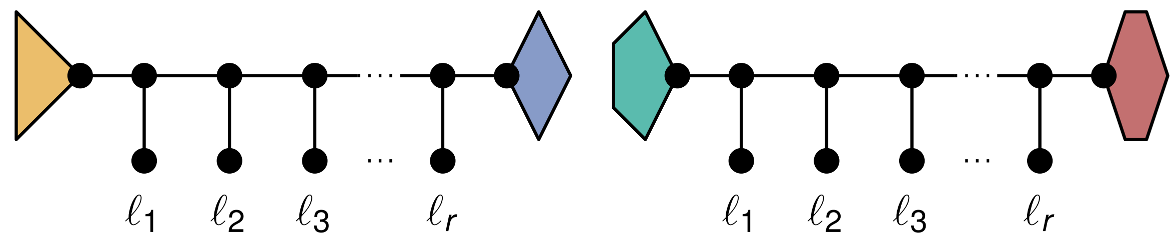
zwei Bäume genügen

Ungünstige Baumstruktur 2

Behauptung
Jeder Baum in einem Agreement Forest enthält nur wenige Blätter.

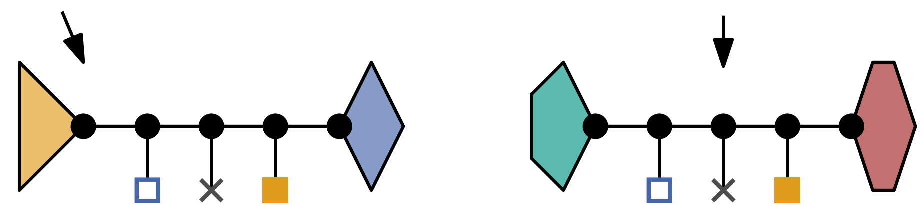
Gegenbeispiel (trotz Anwendung der Reduktionsregel 1)

- viele Blätter in geteiltem Teilbaum



Reduktionsregel 2

- finde Pfad mit angehängten Blättern $\ell_1, \dots, \ell_r$ in T_1 und T_2
- ersetze ihn durch einen Pfad der Länge 3 (drei neue Blätter)



Lemma
Reduktionsregel 2 ist sicher und in polynomieller Zeit ausführbar.

Beweis

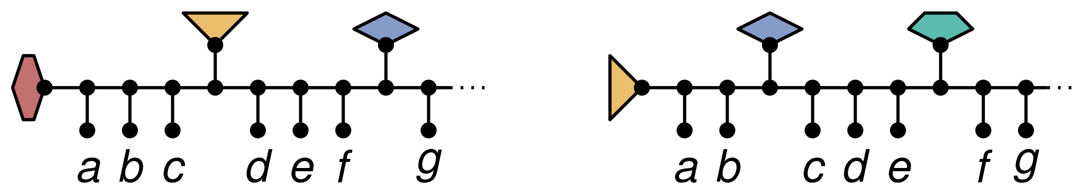
- Fallunterscheidung (nicht hier, aber nicht schwer)
- polynomielle Laufzeit: z.B. durch iteratives Verkürzen von Pfaden der Länge 4

Ungünstige Baumstruktur 3

Behauptung
Jeder Baum in einem Agreement Forest enthält $\overbrace{\text{nur wenige}}^{\text{amortisiert}}$ Blätter.

Gegenbeispiel (trotz Anwendung der Reduktionsregeln 1 und 2)

- viele Blätter ($a, b, c, \ldots$) in geteiltem Teilbaum



Ist das wirklich ein Problem?

Amortisierte Sichtweise

- nach je drei Blättern kommen sich unterscheidende Teilbäume
- jeder dieser Teilbäume sorgt für (mindestens) einen eigenen Baum im Agreement Forest (wenn $a, b, c, d, \ldots$ alle im gleichen Baum liegen)
- wenige Bäume im Agreement Forest $\Rightarrow$ insgesamt wenige Blätter

Sicher, dass es keine anderen ungünstigen Strukturen geben kann?

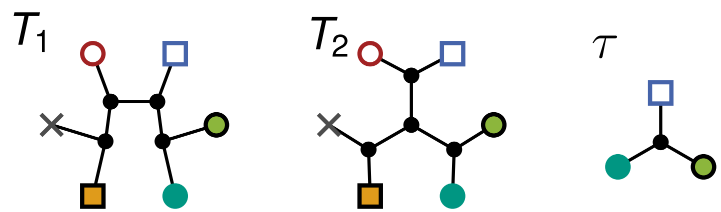
Lemma
In einer lösbaren Instanz liefern die Reduktionsregeln 1 und 2 eine äquivalente Instanz mit maximal ck Blättern (für eine kleine Konstante c).

Grober Plan

Notation

- betrachte einen der Bäume τ in einem Agreement Forest F
- kontrahiere $T_1(L(\tau))$ in T_1 zu einem einzelnen Knoten
- bezeichne den Grad dieses Knotens mit $\deg_1(\tau)$ (analog: $\deg_2(\tau)$)

Beispiel



- $\deg_1(\tau) = 1$
- $\deg_2(\tau) = 2$

Amortisierte Sichtweise

(wie eben, nur etwas formaler)

- τ zerlegt T_i in $\deg_i(\tau)$ viele Teilbäume (für $i \in \{1, 2\}$)
- also: viele Bäume τ mit großem Grad $\deg_i(\tau) \Rightarrow$ viele Bäume in F

Lemma (große Grade $\Rightarrow$ viele Bäume)
Wenn F aus k Bäumen besteht, dann gilt $\sum_{\tau \in F} \deg_i(\tau) \leq 2k - 2$.

Lemma (kleine Grade $\Rightarrow$ wenig Blätter)
Die Reduktionsregeln 1 und 2 sorgen dafür, dass jeder Baum $\tau \in F$ maximal $c(\deg_1(\tau) + \deg_2(\tau))$ Blätter enthält (für kleine Konstante c).

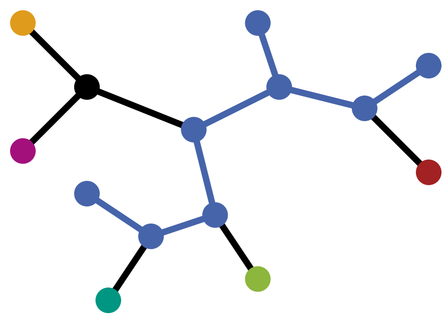
Summe der Grade

Lemma (große Grade $\Rightarrow$ viele Bäume)
Wenn F aus k Bäumen besteht, dann gilt $\sum_{\tau \in F} \deg_i(\tau) \leq 2k - 2$.

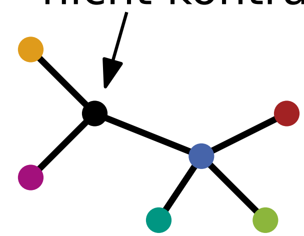
Beweis

- kontrahiere jedes $\tau \in F$ zu einzelnem Knoten in T_i $\rightarrow$ neuer Baum T_i'
- jeder Knoten in T_i' gehört zu einem der k Bäume/ist *nicht kontrahiert*

T_i (Bäume aus F bunt)



T_i' nicht kontrahiert



- nicht kontrahierte Knoten haben Grad 3
- n_3 = Anzahl dieser Knoten
- m' = Anzahl Kanten in T_i'
- n' = Anzahl Knoten in T_i'

- es gilt (handshaking): $\sum_{\tau \in F} \deg_i(\tau) + 3n_3 = 2m'$
- außerdem (T_i' ist Baum): $n' = k + n_3 = m' + 1$

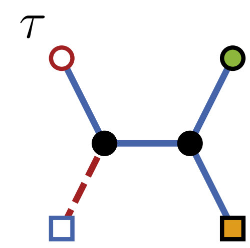
$$\Rightarrow \sum_{\tau \in F} \deg_i(\tau) = 2(k + n_3 - 1) - 3n_3 \leq 2k - 2$$

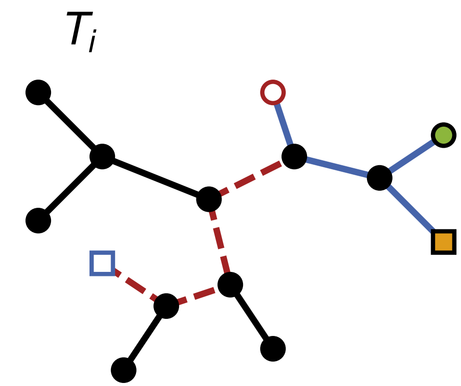
Wenige Blätter pro Baum

Lemma (kleine Grade $\Rightarrow$ wenig Blätter)
Die Reduktionsregeln 1 und 2 sorgen dafür, dass jeder Baum $\tau \in F$ maximal $c(\deg_1(\tau) + \deg_2(\tau))$ Blätter enthält (für kleine Konstante c).

Intuition

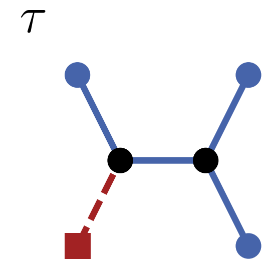
- betrachte wie τ in T_1 bzw. T_2 liegt
- Kanten von τ repräsentieren Kanten/Pfade
- großer Teilbaum ohne Pfade
  $\Rightarrow$ Reduktionsregel anwendbar
- viele Pfade $\Rightarrow$ $\deg_i(\tau)$ ist groß





Beweisplan

- betrachte nur τ mit gefärbten Kanten und Blättern
  - blaue Kante: Kante in T_1 und T_2
  - rote Kante: Pfad in T_1 oder T_2
  - Blattfarbe: entsprechend inzidenter Kante
- Beobachte: #rote Kanten $\leq \deg_1(\tau) + \deg_2(\tau)$
- Zeige: #Blätter $\leq c \cdot$ #rote Kanten

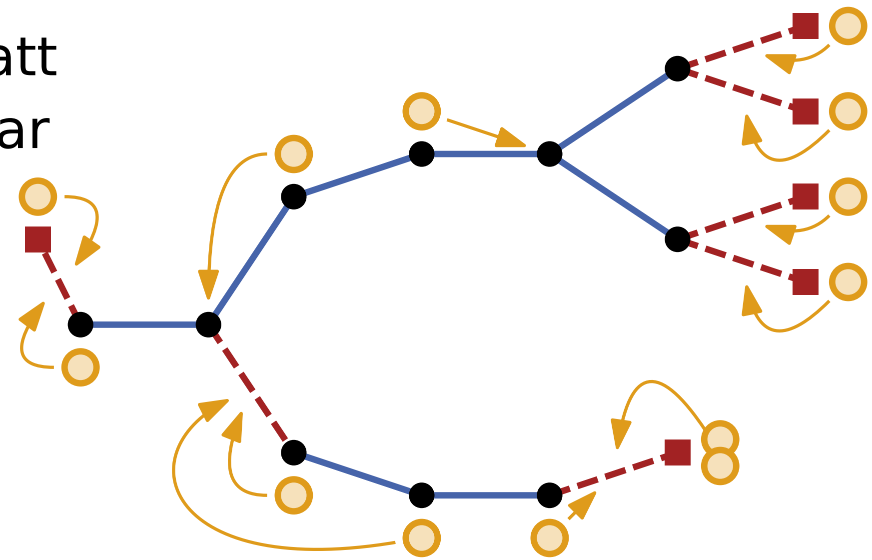


Wenige Blätter pro Baum

Lemma (kleine Grade $\Rightarrow$ wenig Blätter)
Die Reduktionsregeln 1 und 2 sorgen dafür, dass jeder Baum $\tau \in F$ maximal $c(\deg_1(\tau) + \deg_2(\tau))$ Blätter enthält (für kleine Konstante c).

Beweis

- starte mit einem Token bei jedem Blatt
- lösche blaue Blätter; Token $\rightarrow$ Nachbar
  - keine neuen blauen Blätter
  - ~~≤ 2 Token pro rotes Blatt~~
  - ~~≤ 1 Token pro Grad-2 Konten~~
- schiebe alle Token zu: nächstem [Grad-3 Knoten oder rote Kante]
  - ≤ 6 Token pro Grad-3 Knoten
  - ≤ 4 Token pro rote Kante
- #Blätter $= $ #Token $\leq 6n_3 + 4m_r \leq 6n_1 + 4m_r \leq 10m_r \leq 10(\deg_1(\tau) + \deg_2(\tau))$



$n_3 = n_1 - 2$
(siehe Übungsblatt)

nur rote Blätter

n_3: #Grad-3 Knoten n_1: #Blätter m_r: #rote Kanten

Zusammenfassung

Problem: MAXIMUM AGREEMENT FOREST
Gegeben sind T_1, T_2 und ein Parameter k. Gibt es einen Agreement Forest mit maximal k Bäumen?

Theorem
MAXIMUM AGREEMENT FOREST hat einen Kern mit $O(k)$ vielen Blättern, der in polynomieller Zeit berechnet werden kann.

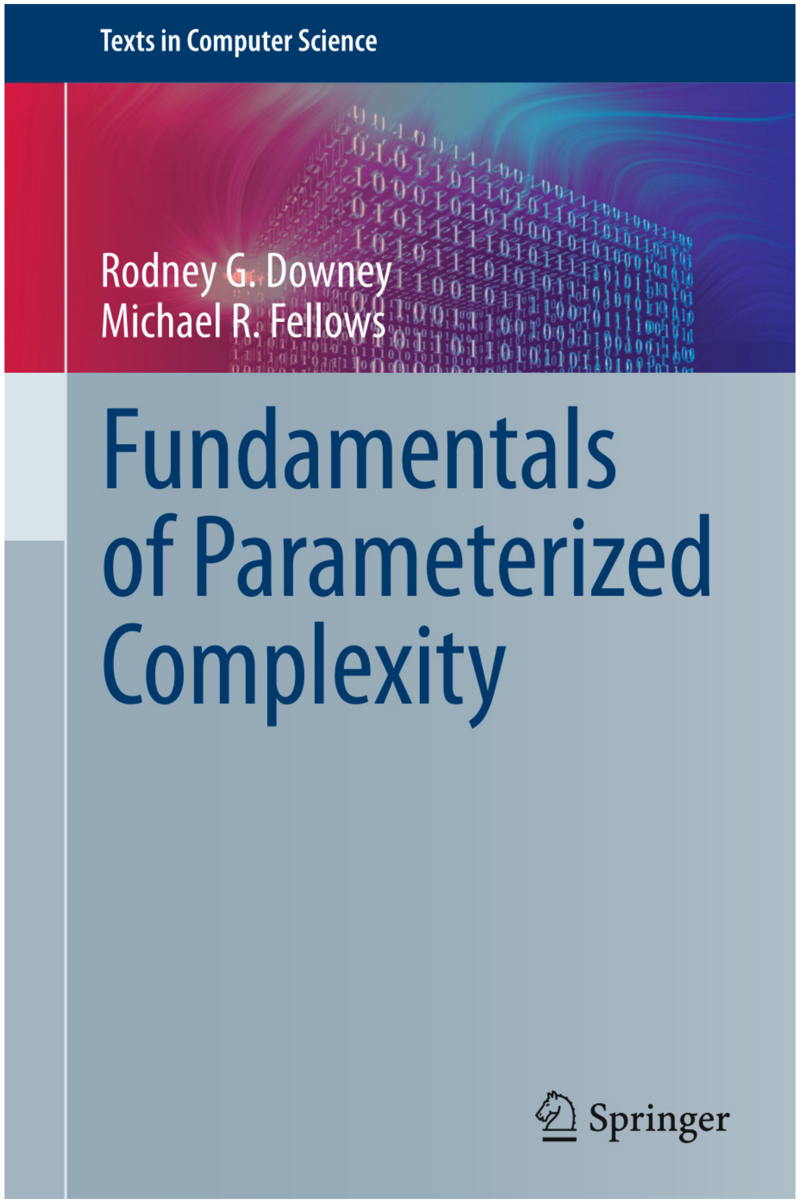
Methodik

- ein konkreteres Ziel als „ich will einen kleinen Kern“ kann helfen
  (z.B.: ich will, dass jeder Baum in F nur wenige Blätter enthält)
- Gegenbeispiele verraten, was wegreduziert werden muss
- ggf. lohnt es, das Ziel später etwas aufzuweichen
  (z.B.: amortisiert über alle Bäume in F statt für jeden einzelnen)

Nicht gesehen heute

- konkrete Laufzeit für Kernbildung
- konkrete Laufzeit für anschließendes Brute-Force im Kern
  $O(4^k \cdot k^5)$ ist möglich

Literaturhinweis



Anmerkungen

- Kapitel 4.10 handelt von dem eben betrachteten Thema
- enthält Links zur Originalliteratur
- aus dem Uninetz kostenlos abrufbar

`link.springer.com/book/10.1007/978-1-4471-5559-1`