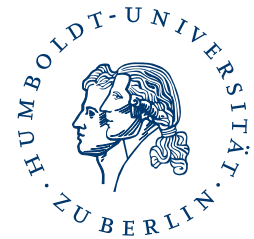
Einführung in die Theoretische Informatik

Johannes Köbler



Institut für Informatik
Humboldt-Universität zu Berlin

WS 2016/17

Zeitkomplexität von Turingmaschinen

Die Laufzeit einer NTM M bei Eingabe x ist die maximale Anzahl an Rechenschritten, die $M(x)$ ausführt.

Definition

- Die Laufzeit einer NTM M bei Eingabe x ist definiert als

$$time_M(x) = \sup\{t \geq 0 \mid \exists K : K_x \vdash^t K\},$$

  wobei $\sup \mathbb{N} = \infty$ ist.
- Sei $t : \mathbb{N} \to \mathbb{N}$ eine monoton wachsende Funktion.
- Dann ist M $t(n)$-zeitbeschränkt, falls für alle Eingaben x gilt:

$$time_M(x) \leq t(|x|).$$

Die Zeitschranke $t(n)$ beschränkt also die Laufzeit bei allen Eingaben der Länge n (worst-case Komplexität).

Zeitkomplexitätsklassen

Wir fassen alle Sprachen und Funktionen, die in einer vorgegebenen Zeitschranke $t(n)$ entscheidbar bzw. berechenbar sind, in folgenden Komplexitätsklassen zusammen.

Definition

- Die in deterministischer Zeit $t(n)$ entscheidbaren Sprachen bilden die Sprachklasse

$$\mathrm{DTIME}(t(n)) = \{L(M) \mid M \text{ ist eine } t(n)\text{-zeitbeschränkte DTM}\}.$$

- Die in nichtdeterministischer Zeit $t(n)$ entscheidbaren Sprachen bilden die Sprachklasse

$$\mathrm{NTIME}(t(n)) = \{L(M) \mid M \text{ ist eine } t(n)\text{-zeitbeschränkte NTM}\}.$$

- Die in deterministischer Zeit $t(n)$ berechenbaren Funktionen bilden die Funktionenklasse

$$\mathrm{FTIME}(t(n)) = \left\{ f \;\middle|\; \begin{array}{l} \text{es gibt eine } t(n)\text{-zeitbeschränkte} \\ \text{DTM } M, \text{ die } f \text{ berechnet} \end{array} \right\}.$$

Die wichtigsten Zeitkomplexitätsklassen

- Die wichtigsten deterministischen Zeitkomplexitätsklassen sind

$$\mathrm{LINTIME} = \bigcup_{c\geq 1} \mathrm{DTIME}(cn+c) \qquad \text{„Linearzeit“}$$

$$\mathrm{P} = \bigcup_{c\geq 1} \mathrm{DTIME}(n^c+c) \qquad \text{„Polynomialzeit“}$$

$$\mathrm{E} = \bigcup_{c\geq 1} \mathrm{DTIME}(2^{cn+c}) \qquad \text{„Lineare Exponentialzeit“}$$

$$\mathrm{EXP} = \bigcup_{c\geq 1} \mathrm{DTIME}(2^{n^c+c}) \qquad \text{„Exponentialzeit“}$$

- Die nichtdeterministischen Klassen NLINTIME, NP, NE, NEXP und die Funktionenklassen FLINTIME, FP, FE, FEXP sind analog definiert.
- Für eine Klasse $\mathcal{F}$ von Funktionen sei $\mathrm{DTIME}(\mathcal{F}) = \bigcup_{t\in\mathcal{F}} \mathrm{DTIME}(t(n))$ (die Klassen $\mathrm{NTIME}(\mathcal{F})$ und $\mathrm{FTIME}(\mathcal{F})$ sind analog definiert).

Asymptotische Laufzeit und Landau-Notation

Definition

Seien f und g Funktionen von $\mathbb{N}$ nach $\mathbb{R}^+ \cup \{0\} = [0, \infty)$.

- Wir schreiben $f(n) = \mathcal{O}(g(n))$, falls es Zahlen n_0 und c gibt mit

  $$\forall n \geq n_0 : f(n) \leq c \cdot g(n).$$

  Bedeutung: „f wächst nicht wesentlich schneller als g."

- Formal bezeichnet der Term $\mathcal{O}(g(n))$ die Klasse aller Funktionen f, die obige Bedingung erfüllen, d.h.

  $$\mathcal{O}(g(n)) = \{f : \mathbb{N} \to [0, \infty) \mid \exists n_0, c \in \mathbb{N}\, \forall n \geq n_0 : f(n) \leq c \cdot g(n)\}.$$

- Die Gleichung $f(n) = \mathcal{O}(g(n))$ drückt also in Wahrheit eine Element-Beziehung $f \in \mathcal{O}(g(n))$ aus.
- O-Terme können auch auf der linken Seite vorkommen. In diesem Fall wird eine Inklusionsbeziehung ausgedrückt.
- So steht $n^2 + \mathcal{O}(n) = \mathcal{O}(n^2)$ für die Aussage

  $$\{n^2 + f \mid f \in \mathcal{O}(n)\} \subseteq \mathcal{O}(n^2).$$

Asymptotische Laufzeit und Landau-Notation

Beispiel

- $7\log(n) + n^3 = \mathcal{O}(n^3)$ ist richtig.
- $7\log(n)n^3 = \mathcal{O}(n^3)$ ist falsch.
- $2^{n+\mathcal{O}(1)} = \mathcal{O}(2^n)$ ist richtig.
- $2^{\mathcal{O}(n)} = \mathcal{O}(2^n)$ ist falsch (siehe Übungen).

Mit der O-Notation lassen sich die wichtigsten deterministischen Zeitkomplexitätsklassen wie folgt charakterisieren:

$$\begin{aligned} \text{LINTIME} &= \text{DTIME}(\mathcal{O}(n)) && \text{„Linearzeit“} \\ \text{P} &= \text{DTIME}(n^{\mathcal{O}(1)}) && \text{„Polynomialzeit“} \\ \text{E} &= \text{DTIME}(2^{\mathcal{O}(n)}) && \text{„Lineare Exponentialzeit“} \\ \text{EXP} &= \text{DTIME}(2^{n^{\mathcal{O}(1)}}) && \text{„Exponentialzeit“} \end{aligned}$$

Das P-NP-Problem

- Wie wir gesehen haben, sind NTMs nicht mächtiger als DTMs, d.h. jede NTM kann von einer DTM simuliert werden.
- Die Frage, wieviel Zeit eine DTM zur Simulation einer NTM benötigt, ist eines der wichtigsten offenen Probleme der Informatik.
- Wegen $\mathrm{NTIME}(t) \subseteq \mathrm{DTIME}(2^{\mathcal{O}(t)})$ erhöht sich die Laufzeit im schlimmsten Fall exponentiell.
- Insbesondere die Klasse NP enthält viele für die Praxis überaus wichtige Probleme, für die kein Polynomialzeitalgorithmus bekannt ist.
- Da jedoch nur Probleme in P als effizient lösbar angesehen werden, hat das so genannte P-NP-Problem, also die Frage, ob alle NP-Probleme effizient lösbar sind, eine immense praktische Bedeutung.

Die Polynomialzeitreduktion

Definition

- Eine Sprache $A \subseteq \Sigma^*$ ist auf $B \subseteq \Gamma^*$ in Polynomialzeit reduzierbar ($A \leq^p B$), falls eine Funktion $f : \Sigma^* \to \Gamma^*$ in FP existiert mit

$$\forall x \in \Sigma^* : x \in A \Leftrightarrow f(x) \in B.$$

- Eine Sprache A heißt $\leq^p$-hart für eine Sprachklasse $\mathcal{C}$ (kurz: $\mathcal{C}$-hart oder $\mathcal{C}$-schwer), falls gilt:

$$\forall L \in \mathcal{C} : L \leq^p A.$$

- Eine $\mathcal{C}$-harte Sprache A, die zu $\mathcal{C}$ gehört, heißt $\mathcal{C}$-vollständig (bzgl. $\leq^p$).
- NPC bezeichnet die Klasse aller NP-vollständigen Sprachen.

Lemma

- Aus $A \leq^p B$ folgt $A \leq B$.
- Die Reduktionsrelation $\leq^p$ ist reflexiv und transitiv (s. Übungen).

Die Polynomialzeitreduktion

Satz

Die Klassen P und NP sind unter $\leq^p$ abgeschlossen.

Beweis

- Sei $B \in \mathrm{P}$ und gelte $A \leq^p B$ mittels einer Funktion $f \in \mathrm{FP}$.
- Seien M und T DTMs mit $L(M) = B$ und $T(x) = f(x)$.
- Weiter seien p und q polynomielle Zeitschranken für M und T.
- Betrachte die DTM M', die bei Eingabe x zuerst T simuliert, um $f(x)$ zu berechnen, und danach M bei Eingabe $f(x)$ simuliert.
- Dann gilt

$$x \in A \Leftrightarrow f(x) \in B \Leftrightarrow f(x) \in L(M) \Leftrightarrow x \in L(M').$$

- Also ist $L(M') = A$ und wegen

$$time_{M'}(x) \leq time_T(x) + time_M(f(x)) \leq q(|x|) + p(q(|x|))$$

ist M' polynomiell zeitbeschränkt und somit A in P. □

NP-Vollständigkeit

Satz

1. $A \leq^p B$ und A ist NP-hart $\Rightarrow$ B ist NP-hart.
2. $A \leq^p B$, A ist NP-hart und $B \in \mathrm{NP}$ $\Rightarrow$ $B \in \mathrm{NPC}$.
3. $\mathrm{NPC} \cap \mathrm{P} \neq \varnothing$ $\Rightarrow$ $\mathrm{P} = \mathrm{NP}$.

Beweis

1. Da A NP-hart ist, ist jede NP-Sprache L auf A reduzierbar.
   Da zudem $A \leq^p B$ gilt und $\leq^p$ transitiv ist, folgt $L \leq^p B$.
2. Klar, da B mit (1) NP-hart und nach Voraussetzung in NP ist.
3. Sei B eine NP-vollständige Sprache in P. Dann ist jede NP-Sprache A auf B reduzierbar und da P unter $\leq^p$ abgeschlossen ist, folgt $A \in \mathrm{P}$. □

Platzkomplexität von Turingmaschinen

- Als nächstes definieren wir den Platzverbrauch von NTMs.
- Intuitiv ist dies die Anzahl aller besuchten Bandfelder.
- Wollen wir auch sublinearen Platz sinnvoll definieren, so dürfen wir hierbei das erste Band offensichtlich nicht berücksichtigen.
- Um sicherzustellen, dass eine NTM M das erste Band nur zum Lesen der Eingabe und nicht auch zum Speichern von weiteren Informationen benutzt, verlangen wir, dass M
  - die Felder auf dem Eingabeband nicht verändert und
  - sich höchstens ein Feld von der Eingabe entfernt.

Definition

Eine NTM M heißt offline-NTM (oder NTM mit Eingabeband), falls für jede von M bei Eingabe x erreichbare Konfiguration

$$K = (q, u_1, a_1, v_1, \ldots, u_k, a_k, v_k)$$

gilt, dass $u_1 a_1 v_1$ ein Teilwort von $\sqcup x \sqcup$ ist.

Platzkomplexität von Turingmaschinen

Definition

- Der Platzverbrauch einer offline-NTM M bei Eingabe x ist definiert als

$$space_M(x) = \sup \left\{ s \geq 1 \;\middle|\; \begin{array}{l} \exists K = (q, u_1, a_1, v_1, \ldots, u_k, a_k, v_k) \\ \text{mit } K_x \vdash^* K \text{ und } s = \sum_{i=2}^{k} |u_i a_i v_i| \end{array} \right\}.$$

- Sei $s : \mathbb{N} \to \mathbb{N}$ eine monoton wachsende Funktion.
- M heißt $s(n)$-platzbeschränkt, falls für alle Eingaben x gilt:

$$space_M(x) \leq s(|x|) \text{ und } time_M(x) < \infty.$$

Platzkomplexitätsklassen

Wir fassen alle Sprachen, die in einer vorgegebenen Platzschranke $s(n)$ entscheidbar sind, in folgenden Platzkomplexitätsklassen zusammen.

Definition

- Die auf deterministischem Platz $s(n)$ entscheidbaren Sprachen bilden die Klasse

$$\mathrm{DSPACE}(s(n)) = \{L(M) \mid M \text{ ist eine } s(n)\text{-platzb. offline-DTM}\}.$$

- Die auf nichtdeterministischem Platz $s(n)$ entscheidbaren Sprachen bilden die Klasse

$$\mathrm{NSPACE}(s(n)) = \{L(M) \mid M \text{ ist eine } s(n)\text{-platzb. offline-NTM}\}.$$

Die wichtigsten Platzkomplexitätsklassen

- Die wichtigsten deterministischen Platzkomplexitätsklassen sind

$$\begin{aligned} \mathsf{L} &= \mathrm{DSPACE}(\mathcal{O}(\log n)) && \text{„Logarithmischer Platz“} \\ \mathsf{LINSPACE} &= \mathrm{DSPACE}(\mathcal{O}(n)) && \text{„Linearer Platz“} \\ \mathsf{PSPACE} &= \mathrm{DSPACE}(n^{\mathcal{O}(1)}) && \text{„Polynomieller Platz“} \end{aligned}$$

- Die nichtdeterministischen Klassen NL, NLINSPACE und NPSPACE sind analog definiert.

Elementare Beziehungen zwischen Komplexitätsklassen

Frage

Welche elementaren Beziehungen gelten zwischen den verschiedenen Zeit- und Platzklassen?

Satz

- Für jede Funktion $t(n) \geq n+2$ gilt

  $$\mathrm{DTIME}(t) \subseteq \mathrm{NTIME}(t) \subseteq \mathrm{DSPACE}(\mathcal{O}(t)).$$

- Für jede Funktion $s(n) \geq \log n$ gilt

  $$\mathrm{DSPACE}(s) \subseteq \mathrm{NSPACE}(s) \subseteq \mathrm{DTIME}(2^{\mathcal{O}(s)}) \text{ und}$$

  $$\mathrm{NSPACE}(s) \subseteq \mathrm{DSPACE}(s^2).$$ (Satz von Savitch)

Korollar

$$\mathrm{L} \subseteq \mathrm{NL} \subseteq \mathrm{P} \subseteq \mathrm{NP} \subseteq \mathrm{PSPACE} = \mathrm{NPSPACE} \subseteq \mathrm{EXP} \subseteq \mathrm{NEXP} \subseteq \mathrm{EXPSPACE}.$$

Komplexität der Stufen der Chomsky-Hierarchie

$$\mathrm{REG} = \mathrm{DSPACE}(\mathcal{O}(1)) = \mathrm{NSPACE}(\mathcal{O}(1)) \subsetneq \mathrm{L},$$

$$\mathrm{DCFL} \subsetneq \mathrm{LINTIME},$$

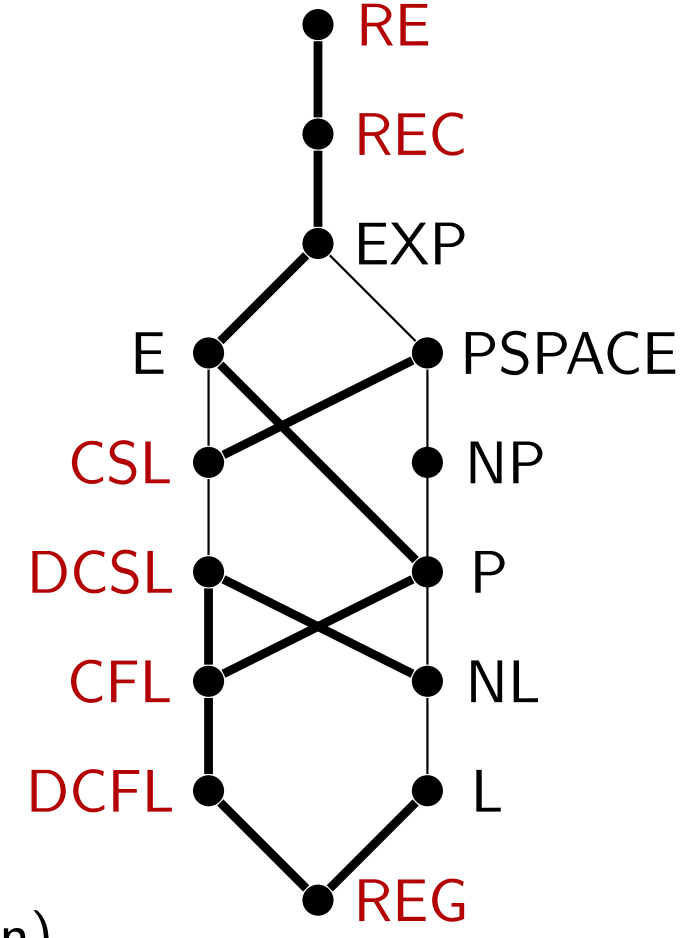
$$\mathrm{CFL} \subsetneq \mathrm{NLINTIME} \cap \mathrm{DTIME}(\mathcal{O}(n^3)) \subsetneq \mathrm{P},$$

$$\mathrm{DCSL} = \mathrm{LINSPACE} \subseteq \mathrm{CSL},$$

$$\mathrm{CSL} = \mathrm{NLINSPACE} \subseteq \mathrm{PSPACE} \cap \mathrm{E},$$

$$\begin{aligned}\mathrm{REC} &= \bigcup_f \mathrm{DSPACE}(f(n)) \\ &= \bigcup_f \mathrm{NSPACE}(f(n)) \\ &= \bigcup_f \mathrm{DTIME}(f(n)) \\ &= \bigcup_f \mathrm{NTIME}(f(n)),\end{aligned}$$

wobei f alle (oder äquivalent: alle berechenbaren) Funktionen $f : \mathbb{N} \to \mathbb{N}$ durchläuft.



Aussagenlogische Formeln

- Die Menge der booleschen (oder aussagenlogischen) Formeln über den Variablen $x_1, \ldots, x_n$, $n \geq 0$, ist induktiv wie folgt definiert:
  - Die Konstanten 0 und 1 sind boolesche Formeln.
  - Jede Variable x_i ist eine boolesche Formel.
  - Mit G und H sind auch die Konjunktion $(G \wedge H)$ und die Disjunktion $(G \vee H)$ von G und H sowie die Negation $\neg G$ von G Formeln.
- Eine Belegung von $x_1, \ldots, x_n$ ist ein Wort $a = a_1 \ldots a_n \in \{0,1\}^n$.
- Der Wert $F(a)$ von F unter a ist induktiv wie folgt definiert:

F	0	1	x_i	$\neg G$	$(G \wedge H)$	$(G \vee H)$
$F(a)$	0	1	a_i	$1 - G(a)$	$G(a)H(a)$	$G(a) + H(a) - G(a)H(a)$

- Durch die Formel F wird also eine n-stellige boolesche Funktion $F : \{0,1\}^n \to \{0,1\}$ definiert, die wir ebenfalls mit F bezeichnen.

Notation

Wir benutzen die Implikation $G \to H$ als Abkürzung für die Formel $\neg G \vee H$ und die Äquivalenz $G \leftrightarrow H$ als Abkürzung für $(G \to H) \wedge (H \to G)$.

Beispiel (Wahrheitswerttabelle)

Die Formel $F = (G \to H)$ mit $G = (\neg x_1 \vee \neg x_2)$ und $H = (x_2 \wedge x_3)$ berechnet folgende boolesche Funktion $F : \{0,1\}^3 \to \{0,1\}$:

a	$G(a)$	$H(a)$	$F(a)$
000	1	0	0
001	1	0	0
010	1	0	0
011	1	1	1
100	1	0	0
101	1	0	0
110	0	0	1
111	0	1	1

◁

Definition

- Zwei Formeln F und G heißen (logisch) äquivalent (kurz $F \equiv G$), wenn sie dieselbe boolesche Funktion berechnen.
- Eine Formel F heißt erfüllbar, falls es eine Belegung a mit $F(a) = 1$ gibt.
- Gilt sogar für alle Belegungen a, dass $F(a) = 1$ ist, so heißt F Tautologie.

Beispiel

- Die Formel $F = (G \to H)$ mit $G = (\neg x_1 \vee \neg x_2)$ und $H = (x_2 \wedge x_3)$ ist erfüllbar, da $F(111) = 1$ ist.
- F ist aber keine Tautologie, da $F(000) = 0$ ist. ◁

Präzedenzregeln zur Klammerersparnis

- Der Junktor $\wedge$ bindet stärker als der Junktor $\vee$ und dieser wiederum stärker als die Junktoren $\rightarrow$ und $\leftrightarrow$.
- Formeln der Form $(x_1 \circ (x_2 \circ (x_3 \circ \cdots \circ x_n)\cdots))$, $\circ \in \{\wedge, \vee\}$, kürzen wir durch $(x_1 \circ \cdots \circ x_n)$ ab.

Beispiel (Formel für die mehrstellige Entweder-Oder Funktion)

- Folgende Formel nimmt unter einer Belegung $a = a_1 \ldots a_n$ genau dann den Wert 1 an, wenn $\sum_{i=1}^{n} a_i = 1$ ist:

$$G(x_1, \ldots, x_n) = (x_1 \vee \cdots \vee x_n) \wedge \bigwedge_{1 \leq i < j \leq n} \neg(x_i \wedge x_j)$$

- D.h. es gilt genau dann $G(a) = 1$, wenn genau eine Variable x_i mit dem Wert $a_i = 1$ belegt ist.
- Diese Formel wird im Beweis des nächsten Satzes benötigt. $\triangleleft$

Das aussagenlogische Erfüllbarkeitsproblem

Erfüllbarkeitsproblem für boolesche Formeln (*satisfiability*, SAT):

Gegeben: Eine boolesche Formel F.
Gefragt: Ist F erfüllbar?

Dabei kodieren wir boolesche Formeln F durch Binärstrings w_F und ordnen umgekehrt jedem Binärstring w eine Formel F_w zu.

Um die Notation zu vereinfachen, werden wir zukünftig jedoch F anstelle von w_F schreiben.

Satz (Cook, Karp, Levin)

SAT ist NP-vollständig.

SAT ist NP-vollständig

SAT ∈ NP

Eine NTM kann bei Eingabe einer booleschen Formel F zunächst eine Belegung a nichtdeterministisch raten und dann in Polynomialzeit testen, ob $F(a) = 1$ ist (*guess and verify* Strategie).

SAT ist NP-hart

- Sei L eine beliebige NP-Sprache und sei $M = (Z, \Sigma, \Gamma, \delta, q_0)$ eine durch ein Polynom p zeitbeschränkte k-NTM mit $L(M) = L$.
- Da sich eine $t(n)$-zeitbeschränkte k-NTM in Zeit $t^2(n)$ durch eine 1-NTM simulieren lässt, können wir $k = 1$ annehmen.
- Unsere Aufgabe besteht nun darin, zu einer beliebigen Eingabe $w = w_1 \dots w_n \in \Sigma^*$ eine Formel F_w zu konstruieren mit
  - $w \in L(M) \iff F_w \in \text{SAT}$,
  - die Reduktionsfunktion $w \mapsto F_w$ ist in FP berechenbar.
- O.B.d.A. sei $Z = \{0, \dots, m\}$, $E = \{m\}$ und $\Gamma = \{a_1, \dots, a_l\}$.
- Zudem gelte $\delta(m, a) = \{(m, a, N)\}$ für alle $a \in \Gamma$.

SAT ist NP-hart

Idee:

Konstruiere F_w so, dass F_w unter einer Belegung a genau dann wahr wird, wenn a eine akzeptierende Rechnung von $M(w)$ beschreibt.

- Wir bilden F_w über den Variablen

$$\begin{aligned} x_{t,q}, &\quad \text{für } 0 \leq t \leq p(n), q \in Z, \\ y_{t,i}, &\quad \text{für } 0 \leq t \leq p(n), -p(n) \leq i \leq p(n), \\ z_{t,i,a}, &\quad \text{für } 0 \leq t \leq p(n), -p(n) \leq i \leq p(n), a \in \Gamma. \end{aligned}$$

- Diese Variablen stehen für folgende Aussagen:

$$\begin{aligned} x_{t,q}: &\quad \text{zum Zeitpunkt } t \text{ befindet sich } M \text{ im Zustand } q, \\ y_{t,i}: &\quad \text{zur Zeit } t \text{ besucht } M \text{ das Feld mit der Nummer } i, \\ z_{t,i,a}: &\quad \text{zur Zeit } t \text{ steht das Zeichen } a \text{ auf dem } i\text{-ten Feld.} \end{aligned}$$

- Konkret sei $F_w = R \wedge S_w \wedge \ddot{U}_1 \wedge \ddot{U}_2 \wedge E$.

SAT ist NP-hart

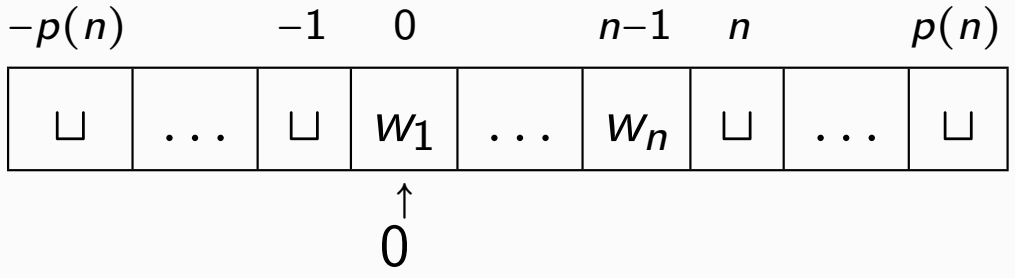
- Konkret sei $F_w = R \wedge S_w \wedge \ddot{U}_1 \wedge \ddot{U}_2 \wedge E$.
- Dabei stellt die Formel $R = \bigwedge_{t=0}^{p(n)} R_t$ (Randbedingungen) sicher, dass wir jeder erfüllenden Belegung von F_w eindeutig eine Folge von Konfigurationen $K_0, \ldots, K_{p(n)}$ zuordnen können:

$$R_t = \; G(x_{t,0}, \ldots, x_{t,m}) \wedge G(y_{t,-p(n)}, \ldots, y_{t,p(n)}) \wedge \bigwedge_{i=-p(n)}^{p(n)} G(z_{t,i,a_1}, \ldots, z_{t,i,a_l}).$$

- Die Teilformel R_t sorgt also dafür, dass zum Zeitpunkt t
  - genau ein Zustand $q \in \{0, \ldots, m\}$ eingenommen wird,
  - genau ein Bandfeld $i \in \{-p(n), \ldots, p(n)\}$ besucht wird und
  - auf jedem Feld i genau ein Zeichen $a_j \in \Gamma$ steht.

SAT ist NP-hart

- Die Formel S_w (wie Startbedingung) stellt sicher, dass zum Zeitpunkt 0 tatsächlich die Startkonfiguration vorliegt:



$$S_w = x_{0,0} \wedge y_{0,0} \wedge \bigwedge_{i=-p(n)}^{-1} z_{0,i,\sqcup} \wedge \bigwedge_{i=0}^{n-1} z_{0,i,w_{i+1}} \wedge \bigwedge_{i=n}^{p(n)} z_{0,i,\sqcup}$$

- Die Formel $\ddot{U}_1$ sorgt dafür, dass der Inhalt von nicht besuchten Feldern beim Übergang von K_t zu K_{t+1} unverändert bleibt:

$$\ddot{U}_1 = \bigwedge_{t=0}^{p(n)-1} \bigwedge_{i=-p(n)}^{p(n)} \bigwedge_{a \in \Gamma} \left(\neg y_{t,i} \wedge z_{t,i,a} \rightarrow z_{t+1,i,a} \right)$$

SAT ist NP-hart

- $\ddot{U}_2$ achtet darauf, dass sich bei jedem Rechenschritt der Zustand, die Kopfposition und das gerade gelesene Zeichen gemäß einer Anweisung in δ verändern:

$$\ddot{U}_2 = \bigwedge_{t=0}^{p(n)-1} \bigwedge_{i=-p(n)}^{p(n)} \bigwedge_{a \in \Gamma} \bigwedge_{q \in Z} \Big(x_{t,q} \wedge y_{t,i} \wedge z_{t,i,a} \rightarrow \bigvee_{(q',b,D) \in \delta(q,a)} x_{t+1,q'} \wedge y_{t+1,i+D} \wedge z_{t+1,i,b} \Big),$$

wobei

$$i + D = \begin{cases} i-1, & D = L \\ i, & D = N \\ i+1, & D = R \end{cases}$$

- Schließlich überprüft E, ob M zur Zeit $p(n)$ den Endzustand m erreicht hat:

$$E = x_{p(n),m}$$

SAT ist NP-hart

- Da der Aufbau der Formel $f(w) = F_w$ einem einfachen Bildungsgesetz folgt und ihre Länge polynomiell in n ist, folgt $f \in \mathrm{FP}$.
- Es ist klar, dass F_w im Fall $w \in L(M)$ erfüllbar ist, indem wir die Variablen von F_w gemäß einer akz. Rechnung von $M(w)$ belegen.
- Umgekehrt führt eine Belegung a mit $F_w(a) = 1$ wegen $R(a) = 1$ eindeutig auf eine Konfigurationenfolge $K_0, \ldots, K_{p(n)}$.
- Für diese gilt:
  - K_0 ist Startkonfiguration von $M(w)$ (wegen $S_w(a) = 1$)
  - $K_i \vdash K_{i+1}$ für $i = 0, \ldots, p(n) - 1$ (wegen $\ddot{U}_1(a) = \ddot{U}_2(a) = 1$)
  - der Zustand von $K_{p(n)}$ ist m (wegen $E(a) = 1$)
- Also gilt für alle $w \in \Sigma^*$ die Äquivalenz

  $$w \in L(M) \Leftrightarrow F_w \in \mathrm{SAT},$$

  d.h. die FP-Funktion $f : w \mapsto F_w$ reduziert $L(M)$ auf SAT. □

Boolesche Schaltkreise

Als nächstes betrachten wir das Erfüllbarkeitsproblem für Schaltkreise.

Definition

- Ein boolescher Schaltkreis über den Variablen $x_1, \ldots, x_n$ ist eine Folge $S = (g_1, \ldots, g_m)$ von Gattern

  $$g_l \in \{0, 1, x_1, \ldots, x_n, (\neg, j), (\wedge, j, k), (\vee, j, k)\} \text{ mit } 1 \leq j, k < l.$$

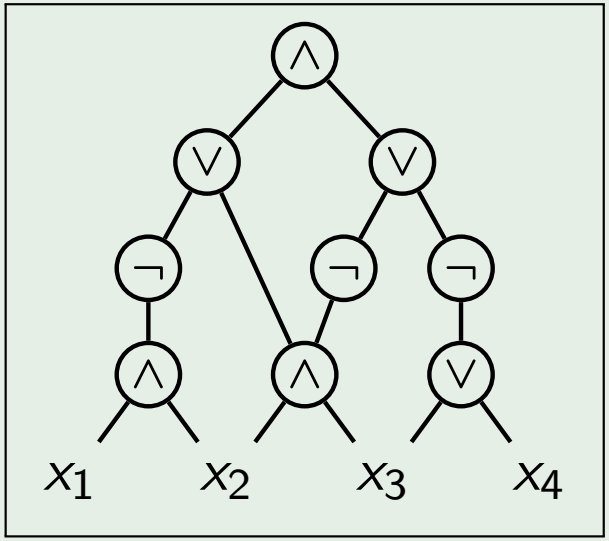
- Jedes Gatter g_l berechnet eine n-stellige boolesche Funktion $g_l : \{0,1\}^n \to \{0,1\}$.
- Für $a = a_1 \ldots a_n \in \{0,1\}^n$ ist $g_l(a)$ induktiv wie folgt definiert:

g_l	0	1	x_i	$(\neg, j)$	$(\wedge, j, k)$	$(\vee, j, k)$
$g_l(a)$	0	1	a_i	$1 - g_j(a)$	$g_j(a)g_k(a)$	$g_j(a) + g_k(a) - g_j(a)g_k(a)$

- S berechnet die boolesche Funktion $S(a) = g_m(a)$.
- S heißt erfüllbar, wenn eine Eingabe $a \in \{0,1\}^n$ ex. mit $S(a) = 1$.

Boolesche Schaltkreise

Beispiel



Graphische Darstellung des Schaltkreises

$S = (x_1, x_2, x_3, x_4, (\wedge, 1, 2), (\wedge, 2, 3), (\vee, 3, 4), (\neg, 5), (\neg, 6), (\neg, 7), (\vee, 6, 8), (\vee, 9, 10), (\wedge, 11, 12))$.

Bemerkung

- Die Anzahl der Eingänge eines Gatters g wird als Fanin von g bezeichnet,
- die Anzahl der Ausgänge von g (also die Anzahl der Gatter, die g als Eingabe benutzen) als Fanout.
- Boolesche Formeln entsprechen also genau den booleschen Schaltkreisen $S = (g_1, \ldots, g_m)$, bei denen jedes Gatter g_i, $1 \leq i \leq m-1$, Fanout 1 hat.
- Eine boolesche Formel F kann somit leicht in einen äquivalenten Schaltkreis S mit $S(a) = F(a)$ für alle Belegungen a transformiert werden.

Das Erfüllbarkeitsproblem für boolesche Schaltkreise

Erfüllbarkeitsproblem für boolesche Schaltkreise (CirSat):

Gegeben: Ein boolescher Schaltkreis S.
Gefragt: Ist S erfüllbar?

Satz

CirSat ist NP-vollständig.

Beweis

Klar, da Sat $\leq^p$ CirSat und CirSat $\in$ NP gilt. □

Bemerkung

- Da Sat NP-vollständig ist, ist CirSat auf Sat reduzierbar.
- Dies bedeutet, dass sich jeder Schaltkreis S in Polynomialzeit in eine erfüllbarkeitsäquivalente Formel F_S überführen lässt.
  F_S und S müssen aber nicht logisch äquivalent sein.
- CirSat ist sogar auf eine spezielle Sat-Variante reduzierbar.

Formeln in konjunktiver Normalform (KNF)

- Ein Literal ist eine Variable x_i oder eine negierte Variable $\neg x_i$, die wir auch kurz mit $\bar{x}_i$ bezeichnen.
- Eine Klausel ist eine Disjunktion $C = \bigvee_{j=1}^{k} l_j$ von Literalen $l_1, \ldots, l_k$.
- Hierbei ist auch $k = 0$ zulässig, d.h. die leere Klausel repräsentiert die Konstante 0 und wird üblicherweise mit $\square$ bezeichnet.
- Eine boolesche Formel F ist in konjunktiver Normalform (kurz KNF), falls $F = \bigwedge_{j=1}^{m} C_j$ eine Konjunktion von Klauseln $C_1, \ldots, C_m$ ist.
- Auch hier ist $m = 0$ zulässig, wobei die leere Konjunktion die Konstante 1 repräsentiert.
- Enthält jede Klausel höchstens k Literale, so heißt F in k-KNF.
- Klauseln werden oft als Menge $C = \{l_1, \ldots, l_k\}$ ihrer Literale und KNF-Formeln als Menge $F = \{C_1, \ldots, C_m\}$ ihrer Klauseln dargestellt.
- Enthält F die leere Klausel, so ist F unerfüllbar.
- Dagegen ist die leere Formel eine Tautologie.

Erfüllbarkeitsprobleme für Formeln in KNF

Beispiel

Der 3-KNF Formel $F = (x_1 \vee \bar{x}_2) \wedge (\bar{x}_1 \vee x_3) \wedge (x_2 \vee \bar{x}_3 \vee x_4)$ entspricht die Klauselmenge $F = \{\{x_1, \bar{x}_2\}, \{\bar{x}_1, x_3\}, \{x_2, \bar{x}_3, x_4\}\}$.

Erfüllbarkeitsproblem für k-KNF Formeln (k-SAT):

Gegeben: Eine boolesche Formel F in k-KNF.

Gefragt: Ist F erfüllbar?

Not-All-Equal-SAT (NAESAT):

Gegeben: Eine Formel F in 3-KNF.

Gefragt: Hat F eine (erfüllende) Belegung, unter der in keiner Klausel alle Literale denselben Wahrheitswert haben?

Beispiel (Fortsetzung)

Offenbar ist $F(0000) = 1$, d.h. $F \in$ 3-SAT. Zudem gilt $F \in$ NAESAT.

Komplexität von Erfüllbarkeitsproblemen

Satz

- 1-SAT und 2-SAT sind in P entscheidbar.
- 3-SAT und NAESAT sind NP-vollständig.

3-SAT ist NP-vollständig

Reduktion von CIRSAT auf 3-SAT

- Wir transformieren einen Schaltkreis $S = (g_1, \ldots, g_m)$ mit n Eingängen in eine Formel F_S über den Variablen $x_1, \ldots, x_n, y_1, \ldots, y_m$.
- F_S enthält die Klausel $\{y_m\}$ und für jedes Gatter g_i die Klauseln einer 3-KNF F_i, die zu folgender Formel G_i äquivalent ist:

Gatter g_i	G_i	Klauseln von F_i
0	$y_i \leftrightarrow 0$	$\{\bar{y}_i\}$
1	$y_i \leftrightarrow 1$	$\{y_i\}$
x_j	$y_i \leftrightarrow x_j$	$\{\bar{y}_i, x_j\}, \{\bar{x}_j, y_i\}$
$(\neg, j)$	$y_i \leftrightarrow \bar{y}_j$	$\{\bar{y}_i, \bar{y}_j\}, \{y_j, y_i\}$
$(\wedge, j, k)$	$y_i \leftrightarrow y_j \wedge y_k$	$\{\bar{y}_i, y_j\}, \{\bar{y}_i, y_k\}, \{\bar{y}_j, \bar{y}_k, y_i\}$
$(\vee, j, k)$	$y_i \leftrightarrow y_j \vee y_k$	$\{\bar{y}_j, y_i\}, \{\bar{y}_k, y_i\}, \{\bar{y}_i, y_j, y_k\}$

3-SAT ist NP-vollständig

Reduktion von CIRSAT auf 3-SAT

- Wir zeigen, dass für alle $a \in \{0,1\}^n$ folgende Äquivalenz gilt:

  $$S(a) = 1 \Leftrightarrow \exists b \in \{0,1\}^m : F_S(ab) = 1.$$

- Ist nämlich $a \in \{0,1\}^n$ eine Eingabe mit $S(a) = 1$. Dann erhalten wir mit

  $$b_i = g_i(a) \text{ für } i = 1, \ldots, m$$

  eine erfüllende Belegung $ab_1 \ldots b_m$ für F_S.
- Ist umgekehrt $ab_1 \ldots b_m$ eine erfüllende Belegung für F_S, so muss
  - $b_m = 1$ sein, da $\{y_m\}$ eine Klausel in F_S ist, und
  - durch Induktion über $i = 1, \ldots, m$ folgt

    $$g_i(a) = b_i,$$

    d.h. insbesondere folgt $S(a) = g_m(a) = b_m = 1$.

3-SAT ist NP-vollständig

Reduktion von CIRSAT auf 3-SAT

- Wir wissen bereits, dass für alle $a \in \{0,1\}^n$ die Äquivalenz

  $$S(a) = 1 \Leftrightarrow \exists b \in \{0,1\}^m : F_S(ab) = 1.$$

  gilt.
- Dies bedeutet, dass der Schaltkreis S und die 3-KNF-Formel F_S erfüllbarkeitsäquivalent sind, d.h.

  $$S \in \text{CIRSAT} \Leftrightarrow F_S \in \text{3-SAT}.$$

- Da zudem die Reduktionsfunktion $S \mapsto F_S$ in FP berechenbar ist, folgt $\text{CIRSAT} \leq^p \text{3-SAT}$. □

Notation – ungerichtete Graphen

- Ein (ungerichteter) Graph ist ein Paar $G = (V, E)$, wobei
  V - eine endliche Menge von Knoten/Ecken und
  E - die Menge der Kanten ist.

  Hierbei gilt

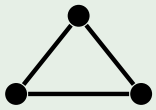
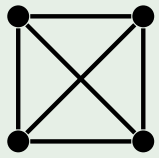
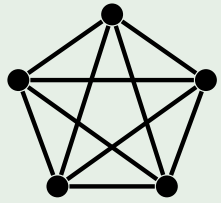
  $$E \subseteq \binom{V}{2} := \big\{\{u, v\} \subseteq V \mid u \neq v\big\}.$$

- Die Knotenzahl von G ist $n(G) = \|V\|$.
- Die Kantenzahl von G ist $m(G) = \|E\|$.
- Die Nachbarschaft von $v \in V$ ist $N_G(v) = \{u \in V \mid \{u, v\} \in E\}$ und die Nachbarschaft von $U \subseteq V$ ist $N_G(U) = \bigcup_{u \in U} N_G(u)$.
- Der Grad von $v \in V$ ist $\deg_G(v) = \|N_G(v)\|$.
- Der Minimalgrad von G ist $\delta(G) := \min_{v \in V} \deg_G(v)$ und
- der Maximalgrad von G ist $\Delta(G) := \max_{v \in V} \deg_G(v)$.

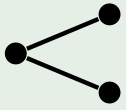
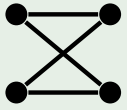
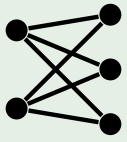
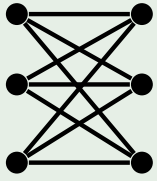
Falls G aus dem Kontext ersichtlich ist, schreiben wir auch einfach n, m, $N(v)$, $\deg(v)$, δ usw.

Beispiel

- Der vollständige Graph (V, E) mit $\|V\| = n$ und $E = \binom{V}{2}$ wird mit K_n und der leere Graph $(V, \varnothing)$ wird mit E_n bezeichnet.

K_1:  K_2:  K_3:  K_4:  K_5: 

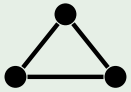
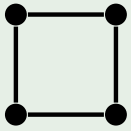
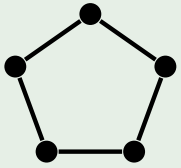
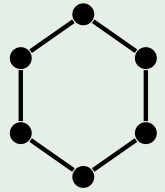
- Der vollständige bipartite Graph (A, B, E) auf $a + b$ Knoten, d.h. $A \cap B = \varnothing$, $\|A\| = a$, $\|B\| = b$ und $E = \{\{u, v\} \mid u \in A, v \in B\}$ wird mit $K_{a,b}$ bezeichnet.

$K_{1,1}$:  $K_{1,2}$:  $K_{2,2}$:  $K_{2,3}$:  $K_{3,3}$: 

- Der Pfadgraph (oder lineare Graph) mit n Knoten heißt P_n:

P_2:  P_3:  P_4:  P_5: 

- Der Kreisgraph (kurz Kreis) mit n Knoten heißt C_n:

C_3:  C_4:  C_5:  C_6: 

Notation – ungerichtete Graphen (Fortsetzung)

- Ein Graph $H = (V', E')$ heißt Sub-/Teil-/Untergraph von $G = (V, E)$, falls $V' \subseteq V$ und $E' \subseteq E$ ist.
- Ein Weg ist eine Folge von Knoten $v_0, \ldots, v_j$ mit $\{v_i, v_{i+1}\} \in E$ für $i = 0, \ldots, j-1$.

  Ein Weg heißt einfach oder Pfad, falls alle durchlaufenen Knoten verschieden sind.

  Die Länge des Weges ist die Anzahl der Kanten, also j.

  Ein Weg $v_0, \ldots, v_j$ heißt auch v_0-v_j-Weg.
- Ein Zyklus ist ein u-v-Weg der Länge $j \geq 2$ mit $u = v$.
- Ein Kreis ist ein Zyklus $v_0, v_1 \ldots, v_{j-1}, v_0$ der Länge $j \geq 3$, für den $v_0, v_1, \ldots, v_{j-1}$ paarweise verschieden sind.

Cliquen, Stabilität und Matchings

- Eine Knotenmenge $U \subseteq V$ heißt Clique, wenn jede Kante mit beiden Endpunkten in U in E ist, d.h. es gilt $\binom{U}{2} \subseteq E$.

  Die Cliquenzahl ist

  $$\omega(G) := \max\{\|U\| \mid U \text{ ist Clique in } G\}.$$

- Eine Knotenmenge $U \subseteq V$ heißt stabil oder unabhängig, wenn keine Kante in G beide Endpunkte in U hat, d.h. es gilt $E \cap \binom{U}{2} = \varnothing$.

  Die Stabilitätszahl ist

  $$\alpha(G) := \max\{\|U\| \mid U \text{ ist stabile Menge in } G\}.$$

- Zwei Kanten $e, e' \in E$ heißen unabhängig, falls $e \cap e' = \varnothing$ ist.

  Eine Kantenmenge $M \subseteq E$ heißt Matching in G, falls alle Kanten in M paarweise unabhängig sind.

  Die Matchingzahl von G ist

  $$\mu(G) = \max\{\|M\| \mid M \text{ ist ein Matching in } G\}.$$

Knotenüberdeckungen und Färbungen

- Eine Knotenmenge $U \subseteq V$ heißt Knotenüberdeckung (engl. *vertex cover*), wenn jede Kante $e \in E$ mindestens einen Endpunkt in U hat, d.h. es gilt $e \cap U \neq \varnothing$ für alle Kanten $e \in E$.

  Die Überdeckungszahl ist

  $$\beta(G) = \min\{\|U\| \mid U \text{ ist eine Knotenüberdeckung in } G\}.$$

- Eine Abbildung $f\colon V \to \mathbb{N}$ heißt Färbung von G, wenn $f(u) \neq f(v)$ für alle $\{u, v\} \in E$ gilt.

  G heißt k-färbbar, falls eine Färbung $f\colon V \to \{1, \ldots, k\}$ existiert.

  Die chromatische Zahl ist

  $$\chi(G) = \min\{k \in \mathbb{N} \mid G \text{ ist } k\text{-färbbar}\}.$$

Eulerlinien und -touren

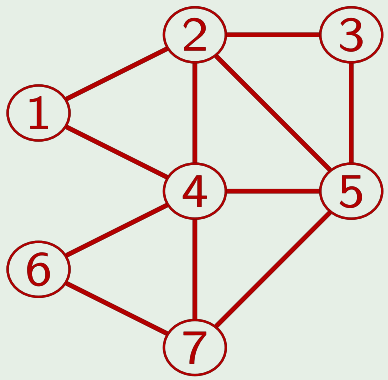
Definition

Sei $s = (v_0, v_1, \ldots, v_l)$ ein Weg in einem Graphen $G = (V, E)$, d.h. $\{v_i, v_{i+1}\} \in E$ für $i = 0, \ldots, l-1$.

- Dann heißt s Eulerlinie (auch Eulerzug oder Eulerweg) in G, falls s jede Kante in E genau einmal durchläuft, d.h. es gilt $l = \|E\|$ und
$$\left\{ \{v_i, v_{i+1}\} \,\middle|\, i = 0, \ldots, l-1 \right\} = E.$$
- Ist s zudem ein Zyklus, d.h. es gilt $v_l = v_0$, so heißt s Eulerkreis (auch Eulerzyklus oder Eulertour).

Beispiel (Eulerlinie)

$$s = (4, 1, 2, 3, 5, 7, 6, 4, 5, 2, 4, 7)$$



Hamiltonpfade und -kreise

Definition

Sei $s = (v_0, v_1, \ldots, v_l)$ ein Pfad in einem Graphen $G = (V, E)$, d.h. $\{v_i, v_{i+1}\} \in E$ für $i = 0, \ldots, l-1$ und $v_0, \ldots, v_l$ sind paarweise verschieden.

- Dann heißt s Hamiltonpfad in G, falls s jeden Knoten in V genau einmal durchläuft, d.h. es gilt

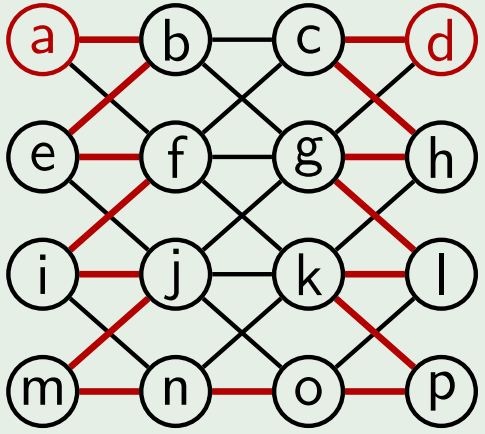
  $$V = \{v_0, \ldots, v_l\} \text{ und } l = \|V\| - 1.$$

- Ist zudem $\{v_0, v_l\} \in E$, d.h. $s' = (v_0, v_1, \ldots, v_l, v_0)$ ist ein Kreis, so heißt s' Hamiltonkreis.

Hamiltonpfade und -kreise

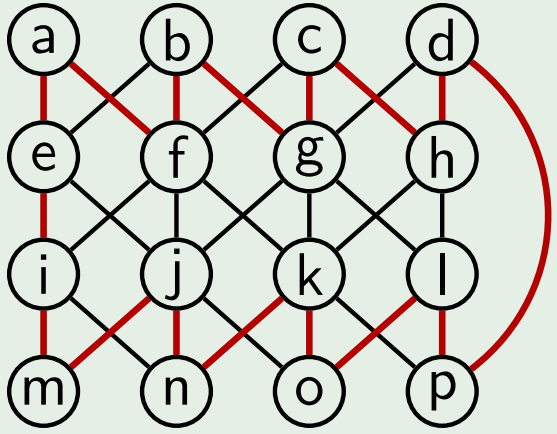
Beispiel (Hamiltonpfad)

$s = (a, b, e, f, i, j, m, n, o, p, k, l, g, h, c, d)$



Beispiel (Hamiltonkreis)

$s = (a, f, b, g, c, h, d, p, l, o, k, n, j, m, i, e, a)$



Seien ein Graph G und eine Zahl $k \geq 1$ gegeben.

Clique:

Gefragt: Hat G eine Clique der Größe k?

Matching:

Gefragt: Hat G ein Matching der Größe k?

Independent Set (IS):

Gefragt: Hat G eine stabile Menge der Größe k?

Vertex Cover (VC):

Gefragt: Hat G eine Knotenüberdeckung der Größe k?

Färbbarkeit (Coloring):

Gefragt: Ist G k-färbbar?

Zudem betrachten wir für einen gegebenen Graphen G folgende Probleme:

k-Färbbarkeit (k-Coloring):

Gefragt: Ist G k-färbbar?

Das Eulerkreisproblem (EulerCycle):

Gefragt: Hat G einen Eulerkreis?

Das Hamiltonkreisproblem (HamCycle):

Gefragt: Hat G einen Hamiltonkreis?

und für einen Graphen G und zwei Knoten s und t folgende Probleme:

Das Eulerlinienproblem (EulerPath):

Gefragt: Hat G eine Eulerlinie von s nach t?

Das Hamiltonpfadproblem (HamPath):

Gefragt: Hat G einen Hamiltonpfad von s nach t?