Das *Schubfachprinzip* (engl. *pigeonhole principle*) besagt Folgendes: Seien A und B endliche Mengen mit $|A| > |B|$, dann existiert keine injektive Funktion $f : A \to B$. Es kann mit einem einfachen Induktionsargument bewiesen werden. Ein Beispiel für seine Anwendung liefert der folgende Satz.

Satz 2.2. *Für jedes ungerade $q \in \mathbb{N}$ existiert ein $i \geq 1$ so dass $q \mid 2^i - 1$.*

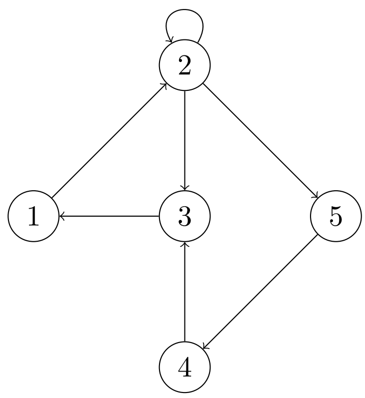
Beweis. Wir kürzen $2^i - 1$ als a_i ab. Wir betrachten $a_1, \ldots, a_q \pmod{q}$. Falls es darunter ein a_i gibt mit $a_i \equiv 0 \pmod{q}$ dann sind wir fertig. Falls nicht, dann gibt es nach dem Schubfachprinzip $i < j$ mit $a_i \equiv a_j \pmod{q}$, d.h. $q \mid a_j - a_i$. Nun ist aber $a_j - a_i = 2^j - 1 - 2^i + 1 = 2^j - 2^i = 2^i(2^{j-i} - 1)$ und da q ungerade muss $q \mid a_{j-i}$, Widerspruch. $\square$

2.2 Graphen

Definition 2.1. Ein *gerichteter Graph* ist ein Paar $G = (V, E)$ wobei V eine beliebige Menge ist und $E \subseteq V \times V$.

Die Elemente von V heißen *Knoten*, die Elemente von E *Kanten*. Graphen werden oft aufgezeichnet indem die Kanten als Pfeile zwischen den Knoten dargestellt werden.

Beispiel 2.7. Der gerichtete Graph $G = (V, E)$ mit $V = \{1, 2, 3, 4, 5\}$ und $E = \{(1, 2), (2, 2), (2, 3), (2, 5), (3, 1), (4, 3), (5, 4)\}$ kann z.B. folgendermaßen gezeichnet werden.



Zwei Knoten $x, y \in V$ heißen *adjazent* falls $(x, y) \in E$ oder $(y, x) \in E$. Der *Ausgangsgrad* von $v \in V$ ist $\mathrm{d}^+(v) = |\{w \in V \mid (v, w) \in E\}|$. Der *Eingangsgrad* von $v \in V$ ist $\mathrm{d}^-(v) = |\{u \in V \mid (u, v) \in E\}|$. Ein Pfad ist eine endliche Folge $v_1, \ldots, v_n \in V$ mit $(v_i, v_{i+1}) \in E$ für $i = 1, \ldots, n-1$ und $i \neq j$ impliziert $v_i \neq v_j$.

Oft werden wir auch ungerichtete Graphen betrachten.

Definition 2.2. Ein *ungerichteter Graph* ist ein Paar (V, E) wobei V eine beliebige Menge ist und $E \subseteq \{\{x, y\} \mid x, y \in X, x \neq y\}$.

Damit ist $|E|$ in einem ungerichteten Graphen die Anzahl ungerichteter Kanten. Weiters definieren wir für $v \in V$ den Grad von v als $\mathrm{d}(v) = |\{w \in V \mid \{v, w\} \in E\}|$, es gibt also keinen getrennten Eingangs- und Ausgangsgrad mehr. Ein ungerichteter Graph kann als gerichteter Graph aufgefasst werden indem eine ungerichtete Kante $\{x, y\}$ durch die gerichteten Kanten (x, y) und (y, x) ersetzt wird. In diesem Sinn stellen die ungerichteten Graphen den allgemeineren Begriff dar. Von nun an werden wir mit "Graph" immer einen ungerichteten Graphen meinen. Wenn wir gerichtete Graphen betrachten wollen, werden wir das explizit erwähnen.

Ein *Pfad* in einem Graphen $G = (V, E)$ ist eine endliche Folge $v_1, \ldots, v_k \in V$ mit $\{v_i, v_{i+1}\} \in E$ für $i = 1, \ldots, k-1$ und $i \neq j$ impliziert $v_i \neq v_j$. G heißt *zusammenhängend* wenn es für alle $v, w \in V$ einen Pfad von v nach w gibt. G heißt *vollständig* falls $E = \{\{x, y\} \mid x, y \in V, x \neq y\}$.

Graphen treten in einer Unzahl von Anwendungskontexten und Berechnungsproblemen auf, wobei man es in der Informatik naturgemäß üblicherweise mit endlichen Graphen zu tun hat. Beispiele für Situationen die durch Graphen modelliert werden können sind: Verkehrsnetze wobei die Knoten z.B. Städten entsprechen und die Kanten Zugverbindungen, das WWW wobei die Knoten Webseiten und die Kanten Hyperlinks entsprechen oder auch Landkarten wobei jedem Land ein Knoten entspricht und zwei Konten durch eine ungerichtete Kante verbunden sind falls sie aneinander grenzen. Beispiele für Berechnungsprobleme aus der Graphentheorie sind:

Kürzester Pfad

Eingabe: endlicher zusammenhängender Graph $G = (V, E)$, Kostenfunktion $c : E \to \mathbb{R}_{>0}$, $s, t \in V$

Ausgabe: Pfad $v_1, \ldots, v_n$ mit $v_1 = s$, $v_n = t$ so dass $\sum_{i=1}^{n-1} c(\{v_i, v_{i+1}\})$ minimal ist

Problem des Handelsreisenden
(engl. *Travelling Salesman Problem (TSP)*)

Eingabe: endlicher vollständiger Graph $G = (V, E)$, Kostenfunktion $c : E \to \mathbb{R}_{>0}$

Ausgabe: Pfad $v_1, \ldots, v_n, v_{n+1} = v_1$ mit $\{v_1, \ldots, v_n\} = V$ so dass $\sum_{i=1}^{n} c(\{v_i, v_{i+1}\})$ minimal ist

Knotenfärbung

Eingabe: endlicher Graph $G = (V, E)$

Ausgabe: Abbildung $f : V \to \{1, \ldots, k\}$ so dass $(x, y) \in E \Rightarrow f(x) \neq f(y)$ und k minimal ist

Wir werden später effiziente Algorithmen zur Bestimmung eines kürzesten Pfades sehen, insb. wird deren Laufzeit polynomial in der Größe der Eingabe sein. Für das Problem des Handelsreisenden oder jenes der Knotenfärbung sind keine Algorithmen mit polynomialer Laufzeit bekannt. Aus der Existenz eines solchen Algorithmus würde $\mathbf{P} = \mathbf{NP}$ folgen und damit die Lösung eines der bedeutendsten offenen Probleme der Mathematik. Das **P** vs. **NP** Problem wird später noch etwas detaillierter besprochen werden.

Zur algorithmischen Behandlung von Graphen müssen diese (etwa im Speicher eines Computers) repräsentiert werden. Es gibt verschiedene Datenstrukturen zur Repräsentation eines Graphen.

Definition 2.3. Die *Adjazenzmatrix* eines gerichteten Graphen $G = (\{1, \ldots, n\}, E)$ ist die Matrix $M = (m_{i,j})_{1 \leq i,j \leq n}$ wobei

$$m_{i,j} = \begin{cases} 1 & \text{falls } (i, j) \in E \\ 0 & \text{falls } (i, j) \notin E \end{cases}$$

Beispiel 2.8. Die Adjazenzmatrix des in Beispiel 2.7 angegebenen Graphen ist

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

Eine Adjazenzmatrix wird im Speicher als ein Datenfeld von Datenfeldern abgelegt. Der für eine Adjazenzmatrix benötigte Speicherplatz ist $\Theta(|V|^2)$. Auf Basis einer Adjazenzmatrix kann von gegebenen Knoten v und w in Zeit $\Theta(1)$ festgestellt werden ob der Graph die Kante (v, w) enthält. Das Durchlaufen aller von einem gegebenen Knoten v aus wegführenden Kanten benötigt $\Theta(|V|)$ Zeit.

Definition 2.4. Die *Adjazenzliste* eines gerichteten Graphen $G = (\{1, \ldots, n\}, E)$ ist ein Datenfeld L der Länge n wobei $L[i]$ die Liste jener $j \in \{1, \ldots, n\}$ ist für die $(i, j) \in E$ gilt.

Beispiel 2.9. Die Adjazenzliste des in Beispiel 2.7 angegebenen Graphen ist:

1: 2

2: 2,3,5

3: 1

4: 3

5: 4

Der für eine Adjazenzliste benötigte Speicherplatz ist $\Theta(|E|)$ was im schlechtesten Fall auch $\Theta(|V|^2)$ ist. Für dünn besetzte Graphen, d.h. also solche die wesentlich weniger als $|V|^2$ viele Kanten haben, stellt die Verwendung einer Adjazenzliste aber eine Speicherersparnis dar. Auf Basis einer Adjazenzliste kann, gegeben Knoten v und w, in Zeit $O(\mathrm{d}^+(v))$ festgestellt werden, ob der Graph die Kante (v, w) enthält. Das Durchlaufen aller von einem gegebenen Knoten v wegführenden Kanten benötigt $\Theta(\mathrm{d}^+(v))$ Zeit.

Beispiel 2.10. Sei $G = (V, E)$ ein Graph wobei V die Menge der Kreuzungen in einer Großstadt ist und $(v, w) \in E$ falls eine Straße von v nach w führt ohne eine andere Kreuzung zu passieren. Der Graph G ist dünn besetzt, die Verwendung einer Adjazenzliste ist also empfehlenswert.

2.3 Bäume

Eine, insbesondere für Algorithmen, besonders wichtige Klasse von Graphen sind Bäume. Um Bäume näher zu untersuchen benötigen wir noch einige Begriffe: Sei $G = (V, E)$ ein Graph. Ein Graph $G' = (V', E')$ heißt *Teilgraph* von G falls $V' \subseteq V$ und $E' \subseteq E$. Sei $V' \subseteq V$, dann ist $G' = (V', E')$ der *von V' in G induzierte Graph* wobei $E' = \{\{v, w\} \in E \mid v, w \in V'\}$.

Definition 2.5. Sei $G = (V, E)$ ein Graph. $G' = (V', E')$ heißt *Zusammenhangskomponente* von G falls

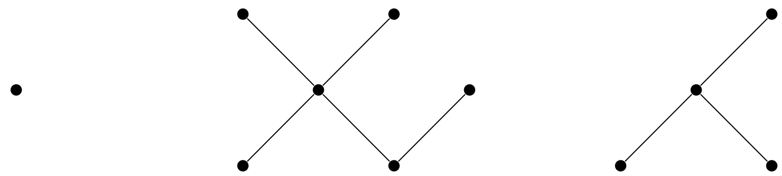
1. $V' \subseteq V$,
2. G' wird durch V' in G induziert,
3. G' ist zusammenhängend und

4. Für alle V'' mit $V' \subset V'' \subseteq V$ gilt: der durch V'' in G induzierte Graph ist nicht zusammenhängend.

Aus dieser Definition folgt unmittelbar dass jeder Graph als disjunkte Vereinigung seiner Zusammenhangskomponenten dargestellt werden kann. Ein *Zyklus* ist ein Pfad $v_1, \dots, v_k$ mit $k \geq 3$ und $\{v_k, v_i\} \in E$.

Definition 2.6. Ein Graph $G = (V, E)$ heißt *Baum* falls er zusammenhängend und zyklenfrei ist. Ein Graph $G = (V, E)$ heißt *Wald* falls er zyklenfrei ist.

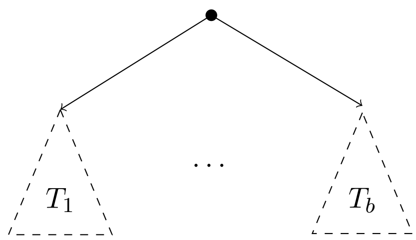
Beispiel 2.11. Ein Wald der aus 3 Bäumen besteht:



In dieser Definition eines Baums ist keine Wurzel ausgezeichnet, jeder Knoten kann als Wurzel designiert werden (je nachdem ändert sich dann die Form des Baums). Ein Wurzelbaum ist ein Tripel (V, E, r) wobei (V, E) ein Baum ist und $r \in V$. Ein Wurzelbaum kann als gerichteter Graph aufgefasst werden, indem, ausgehend von r, alle Kanten von r weg orientiert werden. Wurzelbäume werden uns, wie im folgenden Beispiel, oft auch in Form einer rekursiven Definition begegnen.

Beispiel 2.12. Sei $b \geq 2$. Ein *vollständiger Baum mit Arität* b und Tiefe 0 ist ein einzelner Knoten

Ein solcher Knoten heißt *Blatt*. Ein vollständiger Baum mit Arität b und Tiefe $d + 1$ ist ein gerichteter Graph der Form



wobei $T_1, \dots, T_b$ vollständige Bäume mit Arität b und Tiefe d sind. Jeder in der Wurzel beginnende Pfad in einem vollständigen Baum kann identifiziert werden mit einem Tupel $(p_1, \dots, p_d) \in \{1, \dots, b\}^k$ wobei $k \leq d$. Also hat ein vollständiger Baum b^d *Blätter* sowie

$$b^0 + b^1 + \cdots + b^d = \sum_{i=0}^{d} b^i = \frac{b^{d+1} - 1}{b - 1}$$

Knoten. Ein vollständiger Binärbaum der Tiefe 3 hat also $2^3 = 8$ Blätter und $2^4 - 1 = 15$ Knoten.

Zum Zweck einer Charakterisierung der Bäume unter den Graphen machen wir zunächst noch einige vorbereitende Beobachtungen.

Lemma 2.2. *Sei $G = (V, E)$ ein endlicher, nicht-leerer, zyklenfreier Graph. Dann ist $|E| \leq |V| - 1$.*

Beweis. Wir gehen mit Induktion nach $|V|$ vor. Falls $|V| = 1$, dann muss $|E| = 0$ sein. Sei nun $|V| \geq 2$. Falls $E = \emptyset$ sind wir fertig. Falls $E \neq \emptyset$, dann existiert ein $v \in V$ mit $\mathrm{d}(v) \geq 1$. Seien $\{v, w_1\}, \ldots, \{v, w_k\}$ alle zu v inzidenten Kanten, dann gilt für alle $i, j \in \{1, \ldots, k\}$ mit $i \neq j$: jeder Pfad von w_i nach w_j enthält v, sonst würde nämlich G einen Zyklus enthalten. Für $i = 1, \ldots, k$ sei nun $G_i = (V_i, E_i)$ die Zusammenhangskomponente von w_i im Graphen der aus G entsteht wenn wir v sowie die Kanten $\{v, w_1\}, \ldots, \{v, w_k\}$ entfernen. Dann sind alle G_i zyklenfrei und mit der Induktionshypothese ist also

$$|E| = k + \sum_{i=1}^{k} |E_i| \leq^{\mathrm{IH}} k + \sum_{i=1}^{k} (|V_i| - 1) = \sum_{i=1}^{k} |V_i| = |V| - 1.$$

□