# Drei Vorlesungen über Bäume

Jens Vygen

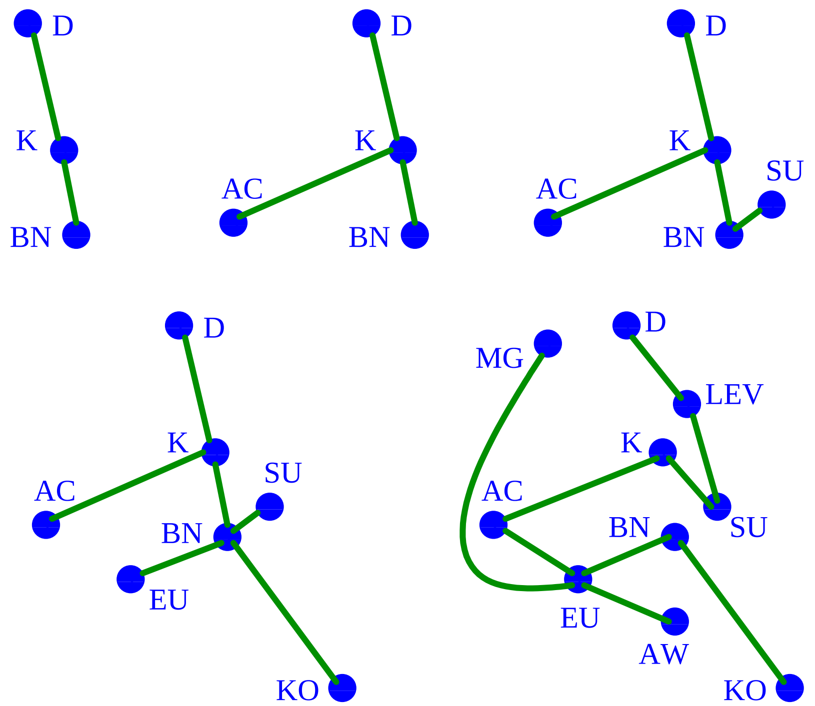
## 1 Bäume

### 1.1 Einleitung

Nehmen wir einmal an, für einen neuen Zug wollen wir ein komplett neues Schienennetz bauen, das einige Städte miteinander verbindet. Oder ein neues Netzwerk aus Kabeln.

Wollen wir zum Beispiel Bonn, Köln und Düsseldorf miteinander verbinden, reicht ein Gleis/Kabel von Bonn nach Köln und ein anderes von Köln nach Düsseldorf. Wer von Bonn nach Düsseldorf möchte, fährt eben über Köln. Um drei Städte miteinander zu verbinden, reichen also zwei Gleise bzw. Kabel.

Will man eine vierte Stadt (z.B. Aachen) anbinden, benötigt man ein weiteres (z.B. Aachen–Köln). Fünf Städte kann man mit vier Gleisen/Kabeln verbinden, sechs Städte mit fünf usw. Wir schauen uns einige weitere Beispiele an.




Ein Gleis oder Kabel verbindet bei uns immer genau zwei Städte miteinander und ist in beiden Richtungen befahrbar. Es kommt uns nicht auf den genauen Verlauf der Gleise oder Kabel an; wichtig ist uns nur das Paar von Städten, das es verbindet; dies nennen wir eine **Kante**.

Ein Netzwerk aus Städten und Kanten, von denen jede genau zwei Städte miteinander verbindet, heißt **Baum**, wenn

- es zusammenhängend ist, d.h. wenn man von jeder Stadt in jede andere gelangen kann, eventuell über Zwischenstationen,
- und man keine Kante weglassen könnte, ohne dass das Netzwerk unzusammenhängend würde.

Mit solchen Bäumen wollen wir uns in diesen drei Vorlesungen beschäftigen. In dieser ersten Vorlesung möchten wir folgende Aussage beweisen:

*Jeder Baum mit $n$ Städten hat genau $n-1$ Kanten.*

Haben wir dies nicht schon oben gesehen? Nun, wir haben uns nur einige Beispiele angesehen, für die das tatsächlich stimmt. Das beweist aber noch keineswegs, dass die Aussage für jedes beliebige $n = 2, 3, 4, 5, 6, \ldots$ und jeden Baum mit $n$ Städten richtig ist. In der Mathematik sind Beispiele hilfreich, um sich zu überlegen, ob eine Aussage stimmen könnte. Bevor man einen Satz aber wirklich akzeptiert, braucht man einen unumstößlichen Beweis. Da wir – wie meistens in der Mathematik – etwas über unendlich viele Objekte beweisen wollen, können wir nicht einfach alle ausprobieren. Wir müssen erst einmal etwas mehr über Bäume lernen.

## 1.2 Wege und Kreise

Sei $B$ ein beliebiger Baum. Eine Folge verschiedener Städte $s_1, s_2, s_3, \ldots, s_k$, so dass je zwei aufeinanderfolgende Städte der Folge durch eine Kante verbunden sind, nennen wir einen **Weg** in $B$ (von $s_1$ nach $s_k$). Den Fall $k = 1$ schließen wir nicht aus.

Wenn es einen Weg von einer Stadt $r$ nach einer Stadt $t$ gibt, dann gibt es auch einen Weg von $t$ nach $r$ (man kann die Folge ja einfach von hinten nach vorne lesen). Den oben benutzten Begriff **zusammenhängend** kann man nun so präzisieren, dass es für jedes Paar von Städten einen Weg geben muss, der sie verbindet.

Haben wir einen Weg $s_1, s_2, s_3, \ldots, s_k$ mit $k \geq 3$, so dass auch $s_1$ und $s_k$ durch eine Kante verbunden sind, so heißt die Folge $s_1, s_2, s_3, \ldots, s_k, s_1$ ein **Kreis** in $B$. Wir beweisen zunächst:

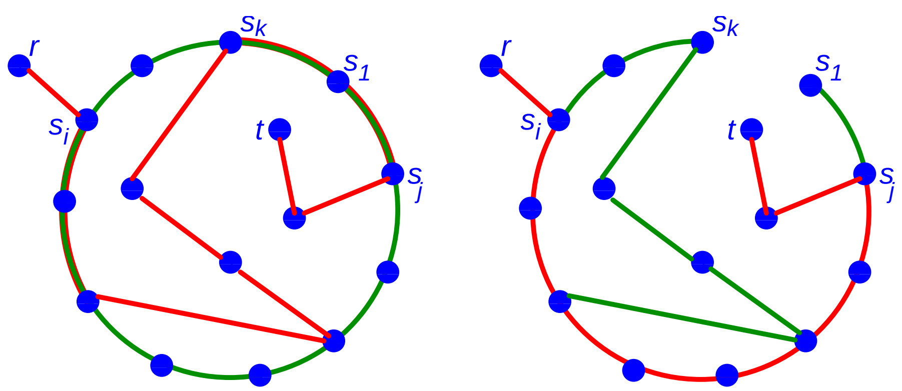
**Satz 1:** Kein Baum enthält einen Kreis.

**Beweis:** Nehmen wir an, es gäbe einen Baum $B$ mit einem Kreis $s_1, s_2, \ldots, s_k, s_1$. Wir entfernen nun einfach die Kante, die $s_1$ und $s_k$ verbindet. Nun behaupten wir, dass das so entstandene Netzwerk $B'$ immer noch zusammenhängend ist. Dies steht aber im Widerspruch zu der Annahme, dass $B$ ein Baum ist.

Den Zusammenhang von $B'$ beweisen wir wie folgt. Wir nehmen zwei beliebige Städte $r$ und $t$ und zeigen, dass es einen Weg von $r$ nach $t$ in $B'$ gibt.

Da $B$ zusammenhängend war (es war ja schließlich ein Baum), muss es in $B$ einen Weg $W$ von $r$ nach $t$ geben. Wenn dieser Weg die entfernte Kante nicht benutzt, so ist $W$ auch ein Weg in $B'$.

Sonst gibt es auf dem Weg mindestens zwei Städte aus unserem Kreis (nämlich insbesondere $s_1$ und $s_k$). Sei $s_i$ der Kreisknoten, der in der Folge $W$ zuerst auftritt, und $s_j$ der letzte. Nun entfernen wir einfach das Mittelstück von $W$, genauer: alle Städte von $s_i$ bis $s_j$. Stattdessen setzen wir die Städtefolge $s_i, s_{i+1}, \ldots, s_j$ (wenn $i < j$ ist) oder $s_i, s_{i-1}, \ldots, s_j$ (wenn $i > j$ ist) ein. Wir erhalten einen Weg von $r$ nach $t$ in $B'$. □




Umgekehrt gilt auch:

**Satz 2:** Ein zusammenhängendes Netzwerk ohne Kreise ist ein Baum.

**Beweis:** Wir müssen nur zeigen, dass man keine Kante entfernen kann, ohne den Zusammenhang zu zerstören. Entfernen wir eine beliebige Kante $e$, etwa zwischen den Städten $s$ und $t$, so kann es nach Entfernen von $e$ keinen Weg von $s$ nach $t$ geben, denn ein solcher Weg $W$ bildete zusammen mit $e$ einen Kreis, den es nach Voraussetzung ja nicht gibt. □

## 1.3 Blätter

Wir benötigen noch einen weiteren Begriff. Ein **Blatt** in einem Baum ist eine Stadt, die mit genau einer Kante verbunden ist.

**Satz 3:** Jeder Baum mit mindestens zwei Städten hat mindestens ein Blatt.

**Beweis:** Wir betrachten einen beliebigen Baum $B$ mit mindestens zwei Städten und wählen einen möglichst langen Weg $s_1, s_2, \ldots, s_k$ in $B$; d.h. $k$ soll möglichst groß sein. Offenbar ist $k$ mindestens 2 (denn irgendeine Kante muss es ja geben); außerdem kann $k$ nicht größer als die Anzahl der Städte von $B$ sein. Wir behaupten, dass dann $s_1$ ein Blatt ist.

Wäre $s_1$ kein Blatt, so wäre diese Stadt außer mit $s_2$ noch mit mindestens einer anderen Stadt, die wir $t$ nennen wollen, durch eine Kante verbunden. Nun gibt es zwei Möglichkeiten.

Entweder kommt $t$ nicht in dem Weg vor; dann könnte man den Weg aber verlängern, indem man $t$ voranstellt. Dies kann also nicht sein, denn wir hatten ja schon einen längstmöglichen Weg gewählt. Bleibt die zweite Möglichkeit: $t$ ist eine der Städte auf dem Weg, also eine von $s_3, \ldots, s_k$, sagen wir $s_i$. Dann ist aber $s_1, s_2, s_3, \ldots, s_i, s_1$ ein Kreis. Gemäß Satz 1 kann ein Baum aber keinen Kreis enthalten. Auch die zweite Möglichkeit scheidet damit aus. Also ist $s_1$ ein Blatt. $\square$

## 1.4 Anzahl der Kanten

Nun können wir den versprochenen Satz beweisen:

**Satz 4:** Jeder Baum mit $n$ Städten hat genau $n - 1$ Kanten.

**Beweis:** Wäre die Aussage falsch, dann gäbe es mindestens einen Baum, für den die Aussage nicht zutrifft; wir wollen solche Bäume Gegenbeispiele nennen. Wir wollen sehen, was aus der Annahme folgt, dass mindestens ein Gegenbeispiel existiert.

Wir schauen uns ein Gegenbeispiel $G$ an, das so wenige Städte wie möglich hat (unter allen, möglicherweise unendlich vielen Gegenbeispielen). Wir nennen $n$ die Anzahl der Städte von $G$ und $m$ die Anzahl der Kanten. Wir wissen zweierlei:

- $m \neq n - 1$
- Wenn es noch andere Gegenbeispiele gibt, dann haben sie alle jeweils mindestens $n$ Städte.

Außerdem können wir uns direkt überzeugen, dass $n \geq 2$ sein muss, denn ein Baum mit nur einer Stadt kann keine Kante haben, ist also kein Gegenbeispiel.

Nach Satz 3 hat jeder Baum, also auch $G$, mindestens ein Blatt. Nun entfernen wir ein Blatt $b$ sowie die einzige Kante $e$, mit der $b$ verbunden ist. Das Ergebnis wollen wir $G'$ nennen.

Wir werden gleich zeigen, dass $G'$ wieder ein Baum ist. $G'$ hat offenbar $n - 1$ Städte und $m - 1$ Kanten. Da $G'$ also kein Gegenbeispiel sein kann, gilt $m - 1 = (n - 1) - 1$, woraus dann eben doch $m = n - 1$ folgt. Ein Widerspruch, der nur so zu erklären ist, dass es überhaupt kein Gegenbeispiel gibt.

Wir müssen also noch zeigen, dass $G'$ tatsächlich ein Baum ist.

Um zu sehen, dass $G'$ zusammenhängend ist, betrachten wir zwei Städte $s$ und $t$ in $G'$. Da $G$ ein Baum ist, gibt es in $G$ einen Weg $W$ von $s$ nach $t$. Dieser Weg kann das Blatt $b$ nicht enthalten, da weder $s$ noch $t$ gleich $b$ sind, und Zwischenstationen eines Weges keine Blätter sein können. Also ist $W$ auch ein Weg in $G'$.

Außerdem kann $G'$ keinen Kreis enthalten, denn ein solcher Kreis wäre auch ein Kreis in $G$; nach Satz 1 enthält $G$ aber keinen Kreis. Also ist $G'$ ein zusammenhängendes Netzwerk ohne Kreise, und damit nach Satz 2 ein Baum. $\square$

## 1.5 Anwendungen

Bäume, wie wir sie studieren, kommen viel häufiger vor, als man vielleicht denkt. Hier einige wenige Beispiele:

- Ein Kabelnetzwerk, z.B. für Strom, Telefon oder Internet, soll Städte miteinander verbinden.
- Eine Gruppe von Computern soll durch ein Netzwerk miteinander verbunden werden.
- In der Nordsee soll ein Windpark installiert werden. Die einzelnen Windräder müssen alle untereinander verbunden werden.
- Auf einem Computerchip muss ein Rechenergebnis an anderen Stellen weiterverarbeitet werden. Dazu müssen die entsprechenden Punkte auf dem Chip miteinander verbunden werden.

Meist möchte man mit so wenig Kanten (Kabeln, Gleisen) auskommen wie möglich. Zusammenhängende Netzwerke ohne überflüssige Verbindungen sind Bäume. Wir wir gesehen haben, haben diese alle gleich viele Kanten.

## 1.6 Anmerkungen

Die Anwendungen zeigen, dass man nicht immer nur Städte verbinden will. Statt von Städten spricht man meist allgemein von Knoten. Bäume sind spezielle Graphen (nämlich solche, die zusammenhängend sind und keine Kreise enthalten). Bäume, und allgemein Graphen, kann man als Paar $(V, E)$ beschreiben, wobei $V$ eine endliche, nicht leere Menge (die der Knoten bzw. Städte) und $E$ eine Menge von zweielementigen Teilmengen von $V$ ist. Ein Element von $E$ (eine Kante) ist also eine Menge $\{v, w\}$ zweier verschiedener Knoten (die durch diese Kante verbunden werden).

## 1.7 Übungen

1. Auf dem Übungsblatt sind drei Mengen von Städten eingezeichnet. Verbinde jeweils die Städte durch einen Baum.
2. Zeichne einen Baum mit den Städten Aachen, Bonn, Köln, Euskirchen, Siegburg und Koblenz. Nun zeichne einen weiteren Baum mit diesen Städten, aber so, dass Köln, Bonn und Aachen Blätter sind, aber keine andere Stadt.
3. Zeichne einen Baum mit 9 Städten und genau 4 Blättern.

4. Zeichne einen Baum mit 10 Städten und folgenden Eigenschaften: (a) Keine Stadt ist mit mehr als drei Kanten verbunden. (b) Zwischen je zwei Städten gibt es einen Weg mit höchstens drei Zwischenstationen.

5. Zeichne einen Baum mit möglichst vielen Städten und folgenden Eigenschaften: (a) Keine Stadt ist mit mehr als drei Kanten verbunden. (b) Zwischen je zwei Städten gibt es einen Weg mit höchstens vier Zwischenstationen.

6. Nimm vier Städte und zeichne einen roten und einen blauen Baum mit diesen Städten ein, so dass kein Städtepaar sowohl durch eine rote als auch durch eine blaue Kante verbunden sind. Geht das auch für fünf Städte? Für drei?

7. Beweise, dass jeder Baum mit mindestens zwei Städten mindestens zwei Blätter besitzt.

8. Beweise, dass für jedes $n \geq 3$ gilt: Es gibt einen Baum mit $n$ Städten und $n-1$ Blättern, aber kein Baum mit $n$ Städten besitzt mehr als $n-1$ Blätter.

9. Starte mit einem Baum mit mindestens drei Städten und füge eine weitere Kante hinzu, die zwei Städte miteinander verbindet, die noch nicht durch eine Kante verbunden waren. Beweise, dass dadurch immer ein Kreis entsteht.

10. Betrachte ein Netzwerk, in dem es eine Stadt $s$ gibt und Wege von $s$ zu allen anderen Städten. Folgere daraus, dass das Netzwerk zusammenhängend ist.

11. Beweise, dass es für alle natürliche Zahlen $n$ und $k$, für die $2 \leq k \leq n-1$ gilt, einen Baum mit genau $n$ Städten und genau $k$ Blättern gibt.

12. Beweise: Wenn es in einem Baum eine Stadt gibt, in der $k$ Gleise enden, dann gibt es mindestens $k$ Blätter.

13. Wenn die Gleise jeweils nur in einer Richtung befahren werden können und man trotzdem von jeder Stadt in jede andere kommen möchte, wieviele Gleise braucht man dann mindestens, für drei, vier, $n$ Städte? (Jetzt sind die Netzwerke keine Bäume mehr.)

# 2 Satz von Cayley

## 2.1 Bäume zählen

Da die Verbindungen zwischen den Städten z.B. Gleise, Straßen oder Kabel sein können und uns deren genauer Verlauf nicht interessiert, sprechen wir allgemein von **Kanten**. Es kommt nur darauf an, welche beiden Städte eine Kante miteinander verbindet. Eine Kante, die die Städte $s$ und $t$ miteinander verbindet, bezeichnen wir einfach mit $\{s, t\}$.

Wiederholen wir noch einmal: Ein **Netzwerk** besteht aus einer Menge von Städten und einer Menge von Kanten, wobei eine Kante ein Paar von (verschiedenen) Städten ist, und es für jedes Paar von Städten höchstens eine Kante gibt. Ein Netzwerk heißt **Baum**, wenn es zusammenhängend ist, man aber keine Kante entfernen kann, ohne den Zusammenhang zu zerstören.

Wir wollen uns nun die Frage stellen, wieviel Bäume es gibt, die $n$ Städte miteinander verbinden. Auf die Namen der Städte kommt es offenbar nicht an (nennen wir sie $s_1, s_2, \ldots, s_n$), nur auf deren Anzahl, die wir wieder mit $n$ bezeichnen.

Für $n = 2$ Städte gibt es offenbar nur eine Möglichkeit: Wir müssen die beiden Städte durch eine Kante verbinden.

Für $n = 3$ gibt es drei Möglichkeiten: Von den drei möglichen Kanten $\{s_1, s_2\}$, $\{s_2, s_3\}$ und $\{s_1, s_3\}$ müssen wir genau zwei nehmen und die dritte weglassen.

Für $n = 4$ wird es schon etwas schwieriger; es gibt bereits 16 verschiedene Bäume. Arthur Cayley hat 1889 ganz allgemein bewiesen:

*Die Anzahl der Bäume mit n Städten beträgt* $n^{n-2}$.

Diese Zahl steigt mit wachsendem $n$ sehr rasch an, wie die nebenstehende Tabelle zeigt. Schon für 15 Städte würde selbst ein Computer sehr lange brauchen, um alle Bäume zu berechnen. Spätestens für 20 Städte würden wir das Ende der Berechnung nicht erleben.

| Anzahl Städte ($n$) | Anzahl Bäume ($n^{n-2}$) |
|---:|---:|
| 2 | 1 |
| 3 | 3 |
| 4 | 16 |
| 5 | 125 |
| 6 | 1 296 |
| 7 | 16 807 |
| 8 | 262 144 |
| 9 | 4 782 969 |
| 10 | 100 000 000 |
| 11 | 2 357 947 691 |
| 12 | 61 917 364 224 |
| 13 | 1 792 160 394 037 |
| 14 | 56 693 912 375 296 |
| 15 | 1 946 195 068 359 375 |

## 2.2 Wurzelwälder

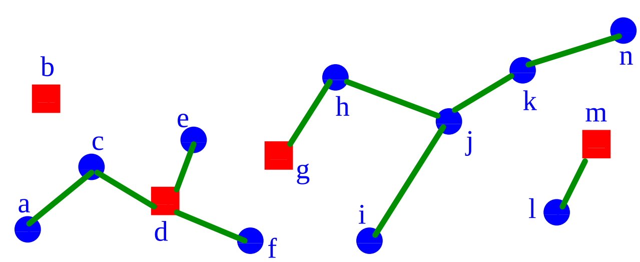
Für unseren Beweis, der vor einigen Jahren von Jim Pitman gefunden wurde, und der erheblich einfacher ist als der ursprüngliche von Cayley, benötigen wir einen neuen Begriff. Ein **Wurzelwald** besteht aus

- einer Menge von Städten,
- einer Menge von Kanten, und
- einer Teilmenge der Städte; diese werden Hauptstädte genannt,

so dass

- kein Kreis existiert, und
- von jeder Stadt aus genau eine Hauptstadt über einen Weg erreichbar ist (dieser Weg kann nur aus der Stadt selbst bestehen, wenn sie nämlich selbst Hauptstadt ist).

Zu jeder Hauptstadt in einem Wurzelwald definieren wir ihren **Bezirk**, bestehend aus den Städten, von denen aus die Hauptstadt erreichbar ist, und den Kanten, die diese Städte miteinander verbinden. Das Bild zeigt einen Wurzelwald mit 14 Städten und vier Bezirken; die Hauptstädte sind als rote Quadrate gezeichnet.




Wir zeigen zunächst:

**Satz 5:** In einem Wurzelwald mit $n$ Städten und $m$ Kanten gibt es genau $n-m$ Hauptstädte.

**Beweis:** Jede Stadt gehört zu genau einem Bezirk; ebenso jede Kante. Wir behaupten, dass jeder Bezirk zusammenhängend ist.

Betrachten wir also einen Bezirk mit Hauptstadt $h$. Nun gibt es ja für je zwei Städte $s$ und $t$ des Bezirks einen Weg $W$ von $s$ nach $h$ und einen Weg $W'$ von $t$ nach $h$. Die erste Stadt von $W$, die auch in $W'$ vorkommt, nennen wir $r$ (mit $h$ gibt es ja mindestens eine Stadt, die in beiden Wegen vorkommt). Dann nehmen wir das Anfangsstück von $W$ bis $r$, und hängen daran das Anfangsstück von $W'$ bis $r$ in umgekehrter Reihenfolge an ($r$ nehmen wir dabei aber nur einmal). Das Ergebnis ist ein Weg von $s$ nach $t$.

Jeder Bezirk ist also zusammenhängend und enthält auch keine Kreise, ist nach Satz 2 also ein Baum.

Also gibt es nach Satz 4 pro Bezirk eine Stadt mehr als Kanten. Insgesamt gibt es aber $n-m$ Städte mehr als Kanten, also ebensoviele Bezirke, und damit ebensoviele Hauptstädte. □

## 2.3 Beweis des Satzes von Cayley

Ein **nummerierter Wurzelwald** ist ein Wurzelwald, in dem zusätzlich die Kanten mit $1, 2, \ldots, m$ durchnummeriert sind (wobei $m$ die Anzahl der Kanten ist).

Damit können wir den Satz von Cayley beweisen:

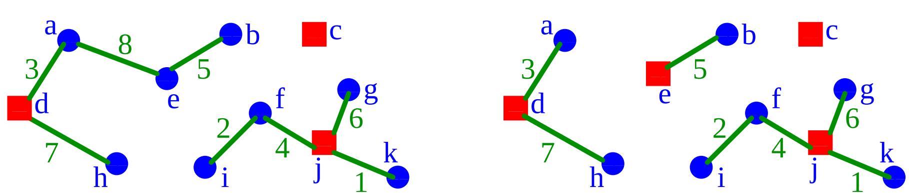
**Satz 6:** Die Anzahl der Bäume mit $n$ Städten beträgt $n^{n-2}$.

**Beweis:** Wir arbeiten mit den Städten $s_1, \ldots, s_n$. Sei $N$ die gesuchte Anzahl der Bäume mit diesen Städten. Mit $\mathcal{B}_m$ bezeichnen wir die Menge der nummerierten Wurzelwälder auf den Städten $s_1, \ldots, s_n$ mit genau $m$ Kanten, und mit $|\mathcal{B}_m|$ ihre Anzahl.

Jeder Baum mit unseren $n$ Städten hat nach Satz 4 genau $n-1$ Kanten, die wir auf genau $(n-1) \cdot (n-2) \cdot \cdots \cdot 3 \cdot 2 \cdot 1 = (n-1)!$ verschiedene Arten nummerieren können. In einem Baum kann es nur eine Hauptstadt geben, und es gibt $n$ Möglichkeiten, diese zu wählen. Ein Baum ist also Bestandteil von genau $(n-1)! \cdot n = n!$ nummerierten Wurzelwäldern (die natürlich alle genau $n-1$ Kanten haben). Damit haben wir $|\mathcal{B}_{n-1}| = n! \cdot N$.

Weiter ist $|\mathcal{B}_0| = 1$, denn wenn es keine Kanten gibt, gibt es auch nichts zu nummerieren, und jede Stadt muss Hauptstadt sein. Wir werden noch $|\mathcal{B}_{i+1}| = n \cdot (n-i-1) \cdot |\mathcal{B}_i|$ beweisen. Daraus folgt dann $|\mathcal{B}_{n-1}| = n^{n-1} \cdot (n-1)! \cdot |\mathcal{B}_0| = n^{n-2} \cdot n \cdot (n-1)! \cdot 1 = n^{n-2} \cdot n!$, und also wie gewünscht $N = n^{n-2}$.

Zu jedem Element $B \in \mathcal{B}_{i+1}$ erhalten wir ein Element $g(B) \in \mathcal{B}_i$, indem wir die Kante mit der höchsten Nummer $i+1$ entfernen und die Stadt zur Hauptstadt küren, die mit dieser Kante verbunden war, von der aus aber nun kein Weg mehr zu einer Hauptstadt führt. Das Bild zeigt ein Beispiel (hier ist $n = 11$; die Städte sind mit den Buchstaben a bis k bezeichnet statt mit $s_1, \ldots, s_{11}$. Links sieht man ein $B \in \mathcal{B}_8$, rechts $g(B) \in \mathcal{B}_7$).




Wenn wir dies für alle Elemente von $\mathcal{B}_{i+1}$ machen, wie oft erhalten wir dann ein bestimmtes $B' \in \mathcal{B}_i$? Wir zeigen, dass dies genau $n \cdot (n-i-1)$ mal passiert, und zwar ganz unabhängig von $B'$.

Hierzu betrachten wir den umgekehrten Vorgang: es gilt $g(B) = B'$ genau dann, wenn $B$ ein nummerierter Wurzelwald ist, der wie folgt aus $B'$ entsteht: wähle zwei Städte $s$ und $t$, wobei $t$ Hauptstadt ist und $s$ in einem anderen Bezirk liegt, füge eine Kante zwischen $s$ und $t$ hinzu, gib ihr die Nummer $i+1$, und entziehe $t$ den Hauptstadtstatus. Hierbei gibt es $n$ Möglichkeiten, $s$ zu wählen (alle Städte), und dann $n-i-1$ Möglichkeiten, $t$ zu wählen (alle von $s$ aus nicht erreichbaren Hauptstädte; wir benutzen Satz 5). Insgesamt also $n \cdot (n-i-1)$, woraus $|\mathcal{B}_{i+1}| = n \cdot (n-i-1) \cdot |\mathcal{B}_i|$ folgt. □

## 2.4 Anmerkungen

Was wir Hauptstadt genannt haben, heißt meist Wurzel. Wurzelwälder werden oft "aus Richtung der Wurzeln" orientiert und heißen dann auch Branchings. Bezirke heißen meist Zusammenhangskomponenten. Für den Satz von Cayley existieren viele, ganz unterschiedliche Beweise; einige sind im schönen Buch von Matoušek und Nešetřil beschrieben (s.u.).

## 2.5 Übungen

1. Zeichne einen Wurzelwald mit 10 Städten und 4 Hauptstädten.
2. Zeichne alle 16 möglichen Bäume für die vier Städte Bonn, Köln, Siegburg, Euskirchen. (Wie gesagt: es kommt nicht auf den Verlauf einer Kante an, sondern nur darauf, welches Städtepaar sie verbindet.)
3. Wieviele Wurzelwälder gibt es mit den drei Städten Bonn, Köln und Düsseldorf?
4. Wieviele Wurzelwälder auf den vier Städten Bonn, Köln, Siegburg und Euskirchen gibt es, in denen Bonn und Köln die einzigen Hauptstädte sind?
5. Zeichne einen Wurzelwald mit 5 Hauptstädten, deren Bezirke alle verschieden viele Städte enthalten.
6. Wieviele Bäume mit $n$ Städten gibt es, die genau zwei Blätter haben?
7. Beweise: Entfernen wir aus einem Wurzelwald eine Kante und machen eine der beiden Städte, die diese Kante verbunden hat, zur Hauptstadt, so entsteht ein neuer Wurzelwald.
8. Benutze die vorige Aufgabe, um einen anderen Beweis von Satz 5 zu führen.
9. Beweise: Jeder Wurzelwald enthält mindestens doppelt so viele Blätter wie Hauptstädte, deren Bezirk noch mindestens eine andere Stadt enthält.

# 3 Kürzeste Bäume

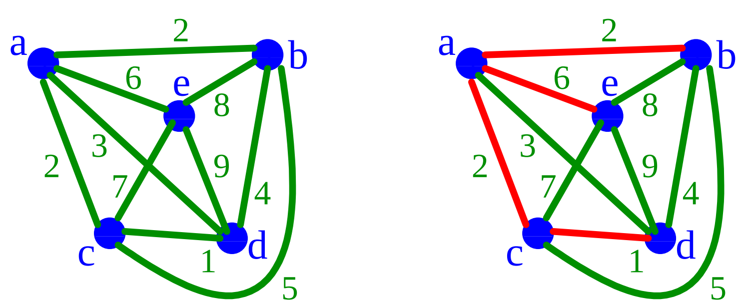
## 3.1 Problemstellung

Nehmen wir wieder an, wir möchten $n$ Städte miteinander verbinden. Der Bau von Kanten (Gleisen, Straßen, Kabeln) kostet Geld, deshalb wollen wir unnötige Kanten vermeiden. Das Netzwerk soll also ein Baum sein. In Satz 4 haben wir gesehen, dass jeder Baum genau $n-1$ Kanten enthält. Allerdings kann es sein, dass der Bau mancher Kanten teurer ist, z.B. weil die Entfernung größer ist.

Für jedes Paar von verschiedenen Städten $s$ und $t$ bezeichnen wir mit $c(\{s,t\})$ die Kosten für den Bau der Kante $\{s,t\}$. Gesucht ist dann ein Baum, dessen Kanten insgesamt möglichst wenig kosten. Einen solchen Baum wollen wir **optimal** nennen.

Eine naheliegende Möglichkeit, einen solchen Baum zu finden, besteht darin, alle möglichen Bäume auf den $n$ Städten auszuprobieren. Allerdings zeigt Satz 6, dass das schon für kleine $n$ sehr viele sind; schon für $n = 20$ kann kein Computer der Welt alle Bäume ausprobieren. Es geht aber viel besser!

Ein viel schnelleres Verfahren hat Otakar Borůvka 1926 vorgeschlagen. 1956 fand Joseph Kruskal ein noch einfacheres Verfahren und bewies, dass es immer einen optimalen Baum liefert. Diesen Algorithmus wollen wir nun vorstellen. Als **Algorithmus** bezeichnen wir ein Rechenverfahren, das man auf einem Computer implementieren kann, und das zu jedem gewünschten Input einen korrekten Output berechnet. In unserem Fall besteht der Input aus den Städten und den Kantenkosten; korrekter Output ist ein optimaler Baum. Das Bild zeigt ein Beispiel: links der Input und rechts in rot der optimale Baum (mit Gesamtkosten $1+2+2+6=11$).




## 3.2 Kruskals Algorithmus

Kruskals Algorithmus ist sehr einfach. Wir beginnen mit unseren Städten ohne jede Kante. Insgesamt gibt es $\frac{n(n-1)}{2}$ Paare von zwei verschiedenen Städten, also ebensoviele Möglichkeiten, eine Kante zu bauen. Wir sortieren zunächst alle diese Städtepaare entsprechend der Kantenkosten, das billigste zuerst und das teuerste zuletzt. Wenn mehrere Städtepaare die gleichen Kosten haben, ist deren Reihenfolge egal.

Dann gehen wir diese Paare der Reihe nach durch. Wenn ein Paar $\{s, t\}$ an der Reihe ist, so prüfen wir, ob die Hinzunahme einer Kante zwischen $s$ und $t$ einen Kreis bilden würde. Falls nein, nehmen wir sie hinzu, sonst natürlich nicht, denn Kreise können wir in einem Baum nicht gebrauchen (Satz 1). Egal wie die Entscheidung ausfällt, machen wir anschließend mit dem nächsten Paar unserer sortierten Liste weiter.

Das ist schon alles. Wir behaupten:

*Kruskals Algorithmus berechnet immer einen optimalen Baum.*

## 3.3 Beweis

Wir benötigen folgende äquivalente Beschreibung von Bäumen:

**Satz 7:** Ein Netzwerk, das keine Kreise enthält, und in dem man keine neue Kante einfügen kann, ohne einen Kreis zu erzeugen, ist ein Baum.

**Beweis:** Wir überlegen uns zuerst, dass das Netzwerk zusammenhängend ist. Für je zwei Städte $s$ und $t$ würde das Einfügen einer Kante zwischen $s$ und $t$ einen Kreis erzeugen. Dieser Kreis ohne die neue Kante ist dann aber ein Weg von $s$ nach $t$. Also ist das Netzwerk zusammenhängend.

Nun folgt die Behauptung aus Satz 2. □

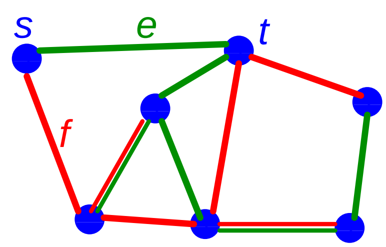
Von diesem Satz gilt übrigens auch die Umkehrung (Übungsaufgabe). Nun können wir beweisen:

**Satz 8:** Kruskals Algorithmus berechnet immer einen optimalen Baum.

**Beweis:** Kruskals Algorithmus berechnet eine Menge von Kanten, die garantiert keinen Kreis enthält (denn das wird stets explizit geprüft). Man kann auch keine Kante hinzufügen ohne einen Kreis zu erzeugen, denn sonst hätte Kruskals Algorithmus eine solche Kante hinzugefügt, als das Städtepaar an der Reihe war. Nach Satz 7 berechnet Kruskals Algorithmus also einen Baum, nennen wir ihn $B$.

Da es nur endlich viele Bäume gibt (in Satz 6 haben wir gesehen, wieviele genau), muss mindestens einer davon optimal sein. Unter den optimalen Bäumen (vielleicht gibt es ja mehr als einen) wählen wir nun einen, der möglichst viele Kanten mit $B$ gemeinsam hat. Diesen nennen wir $B^*$.

Wir nehmen an, $B^*$ sei von $B$ verschieden; sonst sind wir ja fertig. Da $B^*$ und $B$ gleich viele Kanten haben (nach Satz 4 nämlich $n-1$, es sind ja Bäume), gibt es dann mindestens eine Kante, die zu $B^*$ gehört, aber nicht zu $B$. Die erste solche Kante in der sortierten Liste nennen wir $e$ und seine Endpunkte $s$ und $t$. (Das Bild auf der nächsten Seite zeigt ein Beispiel; hier ist $B$ rot und $B^*$ grün gezeichnet.)




Kruskals Algorithmus hat $e$ nicht ausgewählt, weil sonst ein Kreis entstanden wäre. Das heißt: Ergänzt man $B$ um $e$, so entsteht ein Kreis, in dem keine Kante teurer ist als $e$ (teurere Kanten kamen ja erst später als $e$ an die Reihe). Anders ausgedrückt: $B$ enthält einen Weg $W$ von $s$ nach $t$, in dem keine Kante teurer ist als $e$.

Nun entfernen wir aus $B^*$ die Kante $e$ und ergänzen eine Kante $f$ von $W$, so dass wieder ein Baum $B^{**}$ entsteht. Warum gibt es immer eine solche Kante $f$? Nun, wir können z.B. die Endpunkte der entfernten Kante $e$ zu Hauptstädten erklären und erhalten einen Wurzelwald mit zwei Bezirken. Da der Weg $W$ von einem Bezirk in den anderen führt, muss mindestens eine Kante von $W$ von einem Bezirk in den anderen führen; eine solche nehmen wir als $f$ und fügen sie hinzu.

Da $f$ zu $W$ gehört, ist $f$ nicht teurer als $e$, und somit $B^{**}$ insgesamt nicht teurer als $B^*$. Auch $B^{**}$ ist daher ein optimaler Baum. Nun hat aber $B^{**}$ eine Kante mehr mit $B$ gemeinsam als $B^*$ (denn $f$ gehört zu $W$ und damit zu $B$, $e$ gehört aber nicht zu $B$). Dies ist ein Widerspruch zur Wahl von $B^*$. $\square$

## 3.4 Anmerkungen

In der Literatur ist dieses Problem als Minimum-Spanning-Tree-Problem bekannt, oft auch nur unter dem Kürzel MST. Der Algorithmus von Kruskal wird auch als Greedy-Algorithmus bezeichnet, weil er greedy=gierig immer die billigste noch mögliche Kante wählt, ohne darauf zu achten, was später noch kommt. Ein solcher Greedy-Algorithmus liefert nur für wenige Probleme stets eine optimale Lösung (z.B. ist er für das Traveling-Salesman-Problem im allgemeinen sehr schlecht). Es gibt noch schnellere Algorithmen für das MST-Problem (z.B. den von Prim), aber das würde hier zu weit führen.

## 3.5 Übungen

1. Verbinde die Städte auf dem Übungsblatt durch einen optimalen Baum. Die Kantenkosten sind in den Entfernungstabellen angegeben.

2. Zeige die Umkehrung von Satz 7.

3. Finde eine Menge von Städten und Kantenkosten, so dass es mehr als einen optimalen Baum gibt.

4. Die zehn Städtepaare, die man aus den fünf Städten a,b,c,d,e bilden kann, sollen Kantenkosten 1,2,3,4,5,5,6,7,8,9 haben. (je nachdem wie die Kantenkosten auf die Städtepaare zugeordnet sind). Ordne diese Kantenkosten den Städtepaaren so zu, dass es (a) nur einen optimalen Baum gibt; (b) mehr als einen optimalen Baum gibt.

5. Angenommen, wir suchen einen Baum, dessen teuerste Kante möglichst billig ist. Beweise, dass Kruskals Algorithmus auch dieses Problem löst.

6. Wie kann man einen Wurzelwald zu gegebenen Städten und Hauptstädten finden, dessen Kanten insgesamt möglichst wenig kosten?

7. Wir schauen uns einen anderen Algorithmus an: Wir nehmen zunächst für jedes Städtepaar eine Kante und entfernen dann sukzessive die teuerste Kante, deren Entfernung das Netzwerk nicht unzusammenhängend macht. Führe diesen Algorithmus auf dem Übungsblatt (Rückseite) durch. Liefert er auch immer einen optimalen Baum?

8. Auf derselben Menge von Städten seien zwei Wurzelwälder mit unterschiedlich vielen Hauptstädten gegeben. (Die beiden Wurzelwälder können Kanten oder Hauptstädte gemeinsam haben oder auch nicht.) Beweise, dass man eine Kante aus einem der Wurzelwälder zu dem anderen hinzufügen kann, ohne einen Kreis zu erzeugen.

9. Beweise: Falls keine zwei Städtepaare die gleichen Kosten haben, so gibt es nur *einen* optimalen Baum.

# 4 Literatur

Das erste der folgenden Bücher ist ein mathematischer Roman, der relativ leicht verständlich ist. Die beiden anderen Bücher sind anspruchsvoller.


- Peter Gritzmann, René Brandenberg: Das Geheimnis des kürzesten Weges
- Jiří Matoušek, Jaroslav Nešetřil: Diskrete Mathematik
- Martin Aigner, Günter Ziegler: Das Buch der Beweise