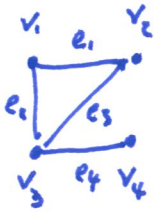
3.6 Datenstrukturen für Graphen

Wie beschreibt man einen Graphen?

(1) Inzidenzmatrix



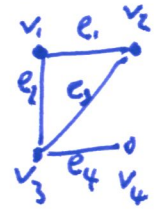
	e_1	e_2	e_3	e_4
v_1	1	1	0	0
v_2	1	0	1	0
v_3	0	1	1	1
v_4	0	0	0	1

(„inzident": zusammentreffend)

Also: $A \in \{0,1\}^{n \times m}$ mit $a_{v,e} := \begin{cases} 1 & \text{für } v \in e \\ 0 & \text{sonst} \end{cases}$

Größe: nm für einen Graphen mit n Knoten, m Kanten. (Viele Nullen!)

(2) Adjazenzmatrix



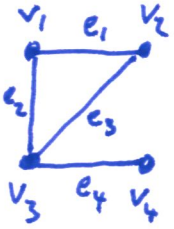
	v_1	v_2	v_3	v_4
v_1	0	1	1	0
v_2	1	0	1	0
v_3	1	1	0	1
v_4	0	0	1	0

(„adjazent": verbunden)

Also: $A \in \{0,1\}^{n \times n}$ mit $a_{v,w} := \begin{cases} 1 & \text{für } \{v,w\} \in E \\ 0 & \text{sonst} \end{cases}$

Größe: n^2 für einen Graphen mit n Knoten.

(3) Kantenliste



$\{v_1,v_2\}, \{v_1,v_3\}, \{v_2,v_3\}, \{v_3,v_4\}$

Benötigt wird eine Kantennummerierung!

→ Jeder Index ist eine Binärzahl mit $(\lfloor \log_2 n \rfloor + 1)$ Bits, bzw. eine Dezimalzahl mit $(\lfloor \log_{10} n \rfloor + 1)$ Stellen.

Unterschied in Codierungslänge: Ein Faktor $\log_2 10 = 3{,}3219\ldots$, denn $\log_2 n = \log_2 10 \cdot \log_{10} n$

Einschub: Wie viele Ziffern benötigt man für eine Zahl in Binär- oder Dezimaldarstellung?

Beispiele:

Binär	Dezimal	Stellen (binär)	(dezimal)
1	1	1	1
10	2	2	1
101	5	3	1
1111111111	1.023	10	2
10000000000000000	65.536	17	5

Eine Zahl n mit b Binärstellen liegt zwischen 2^{b-1} und 2^b,
mit d Dezimalstellen liegt zwischen 10^{d-1} und 10^d:

$$2^{b-1} \leq n < 2^b$$
bzw. $$10^{d-1} \leq n < 10^d$$

Also gilt
$$b-1 \leq \log_2 n < b$$
bzw.
$$d-1 \leq \log_{10} n < d,$$

d.h.
$$b = \lfloor \log_2 n \rfloor + 1$$
bzw.
$$d = \lfloor \log_{10} n \rfloor + 1$$

Außerdem gilt mit
$$2^b = \left(2^{\log_2 10}\right)^{\frac{b}{\log_2 10}} = 10^{\frac{b}{\log_2 10}},$$

also $d \approx \frac{1}{\log_2 10} b$, d.h. zwischen der binären und dezimalen Stellenzahl gibt es eine Umrechnungskonstante von $\log_2 10 \approx 3{,}3219\ldots$ (Das gilt nur ungefähr, weil wir ja ganze Zahlen haben und runden!)

Für einen Graphen mit m Kanten und n Knoten ergibt sich in obiger (ausführlicher) Codierung ein Platzbedarf von

$$(6m-1) + 2m\left(\lfloor \log_{10} n \rfloor + 1\right).$$

Dabei kann man an ein paar Stellen sparen (z.B. „v" oder „{", „}" weglassen) aber auch mehr Platz investieren (z.B. in ASCII codieren bzw. binär statt dezimal codieren). So wäre auch

$$(2m-1) + 2m\left(\lfloor \log_2 n \rfloor + 1\right) \quad \text{denkbar.}$$

Was ist wirklich wichtig dabei?!

(i) Die Kantenliste ist sparsamer als die Inzidenzmatrix, denn wenn n nicht zu klein ist, dann ist

$$n \geq 2 + 2\left(\lfloor \log_2 n \rfloor + 1\right).$$

(was heißt „nicht zu klein"? $n \geq 16$ reicht!)

Also ist auch

$$mn > (2m-1) + 2m\left(\lfloor \log_2 n \rfloor + 1\right).$$

(ii) Unabhängig von der Codierung ist für die Größe des Speicherplatzes der zweite Ausdruck wichtig, denn

$$2m\left(\log_2 n + 1\right) \leq (2m-1) + 2m\left(\lfloor \log_2 n \rfloor + 1\right)$$
$$\leq 4m\left(\log_2 n + 1\right) \quad \text{für } n \geq 2.$$

(iii) Letztlich kommt es also gar nicht so sehr auf die Vorfaktoren an (die sind codierungsabhängig!), sondern auf den Ausdruck

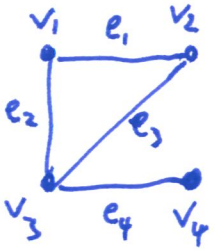
$$m \log n.$$

(iv) In $m \log n$ steckt das Wesen der Kantenliste:
Zähle für alle m Kanten die Nummern der beteiligten Knoten auf!

Wie wir im nächsten Abschnitt sehen werden, lohnt sich dafür eine eigene Notation:

Die Kantenliste benötigt $\Theta(m \log n)$ Speicherplatz.

(4) Adjazenzliste



V_1 : V_2, V_3 ;
V_2 : V_1, V_3 ;
V_3 : V_1, V_2, V_4 ;
V_4 : V_3 ;

Das ist etwas praktischer als die Kantenliste, wenn man für Graphenalgorithmen direkten Zugriff auf die Nachbarn eines Knotens benötigt! Man muss nicht die Nachbarn erst mühsam aus einer Liste herraussuchen.

Länge:

Jede Kante taucht doppelt auf, einmal für jeden Knoten.

So also:

$$2n + 4m + n(\lfloor \log_{10} n \rfloor + 1) + 2m(\lfloor \log_{10} n \rfloor + 1)$$

- d.h. $\Theta(n \log n + m \log n)$.

Im allgemeinen sind Graphen mit vielen isolierten Knoten (ohne Kanten!) uninteressant, d.h. z.B. $m \geq \frac{n}{2}$

$m \geq n$ o.ä.

Also wieder $\Theta(m \log n)$.

Jetzt wollen wir noch etwas mehr: den direkten Zugriff auf die Nachbarn der Knoten, wenn wir sie brauchen!

Also:

Liste (!) $\quad \underset{\downarrow}{V_2}, V_3 ; \underset{\downarrow}{V_1}, V_3 ; \underset{\downarrow}{V_1}, V_2, V_4 ; \underset{\downarrow}{V_3}$

Zugriff auf Nachbarn jeweils ab Semikolon.

Algorithmisch: Gehe Liste durch, zähle Semikolons → das dauert

Datenstruktur: Speichere die Stelle ab, an der die Nachbarn von V_3 zu finden sind.

→ Zeiger (engl. "Pointer")

Bekannt bei Webseiten:

Speicherinhalt: z.B. YouTube-Video (etliche MB)

Zeiger: URL (einige Byte)

„Man muss nicht alles wissen, man muss nur wissen wo's steht!"

Für den direkten Zugriff brauchen wir n Zeiger; jeder codiert eine Speicherzelle, d.h. die Nummer eines Bits in der Liste.

So ein Zeiger braucht also selber

$$\log_2 \left\lfloor \left(2n + 4m + n\left(\lfloor \log_2 n \rfloor + 1\right) + 2m\left(\lfloor \log_2 n \rfloor + 1\right)\right)\right\rfloor + 1$$

für $n \geq 10$, $m \geq n$

$$\leq \log_2 \left(9m\left(\log_2 n + 1\right) + 1\right)$$

$$\leq \log_2 9 + \log_2 m + \log_2\left(\log_2 n + 1\right) + 1 \leq 2\log_2 m$$

Bits

Insgesamt benötigen wir also

$$\Theta(n \log_2 m) \quad \text{Bits.}$$

Jetzt ist $m \leq n^2$,

also $\log_2 m \leq \log_2 n^2 = 2 \log_2 n$,

d.h. $n \log m \leq 2n \log n$,

und der insgesamt verbrauchte Speicherplatz ist

$$\Theta(n \log m + m \log n) = \Theta(m \log n).$$

3.7 Wachstum von Funktionen

Im letzten Abschnitt haben wir Funktionen abgeschätzt und auf die „wesentlichen" Bestandteile reduziert, um Größenordnungen und Wachstumsverhalten zu beschreiben. Ein bisschen formaler:

Definition 3.9 (Θ-Notation)

Seien $f, g : \mathbb{N} \to \mathbb{R}$ Funktionen.

Dann gilt $f \in \Theta(g) :\Leftrightarrow$ Es gibt positive Konstanten c_1, c_2, n_0 mit

$$0 \leq c_1 g(n) \leq f(n) \leq c_2 g(n) \text{ für alle } n \geq n_0.$$

Man sagt: f wächst asymptotisch in derselben Größenordnung wie g.

Beispiele: $2n^2 - 1 \in \Theta(n^2)$

$$\frac{n^3}{1000} + n^2 + n \log n \in \Theta(n^3)$$

Manchmal hat man nur untere oder obere Abschätzungen:

DEFINITION 3.10 (O-Notation)

Seien $f, g : \mathbb{N} \to \mathbb{R}$ Funktionen.

Dann gilt $f \in \Theta(g) :\Leftrightarrow$ Es gibt positive Konstanten c, n_0 mit

$$0 \leq f(n) \leq c\,g(n) \quad \text{für alle } n \geq n_0.$$

Man sagt: f wächst ~~exponentiell~~ höchstens in derselben Größenordnung wie g.

Beispiele:
$2n^2 - 1 \in O(n^2)$
$2n^2 - 1 \in O(n^3)$
$n \log n \in O(n^2)$

DEFINITION 3.11 (Ω-Notation)

Seien $f, g : \mathbb{N} \to \mathbb{R}$ Funktionen.

Dann gilt $f \in \Omega(g) :\Leftrightarrow$ Es gibt positive Konstanten c, n_0 mit

$$0 \leq c\,g(n) \leq f(n) \quad \text{für alle } n \geq n_0.$$

Beispiele: $2n^2 - 1 \in \Omega(n^2)$
$2n^2 - 1 \in \Omega(n^3)$

Einige einfache Eigenschaften:

SATZ 3.12

Seien $f, g : \mathbb{N} \to \mathbb{R}$ Funktionen.

Dann gilt
(1) $f \in \Theta(g) \Leftrightarrow g \in \Theta(f)$
(2) $f \in \Theta(g) \Leftrightarrow f \in O(g)$ und $f \in \Omega(g)$.
(3) $f \in O(g) \Leftrightarrow g \in \Omega(f)$

Beweis: Übung!