

Aufgabe 5-2 HITS-Algorithmus

Das Gesamtgewicht eines gerichteten Graphen $G = (V, E)$ sei $I(G) = \sum_{\substack{i \\ \exists (i,j) \in E}} h_i + \sum_{\substack{j \\ \exists (j,i) \in E}} a_j$

Dabei sei h_i der Hubwert und a_i der Autoritätswert von Knoten i nach dem HITS-Algorithmus.

Der nebenstehende Graph G hat drei bipartite Komponenten G_1, G_2, G_3 .

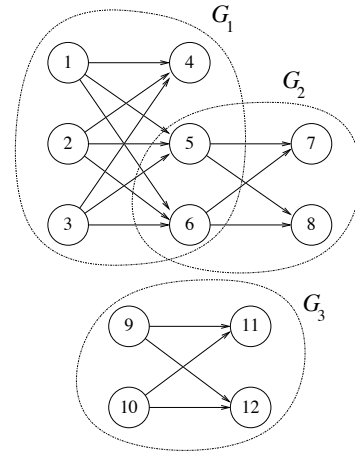
a) Berechnen Sie mit dem HITS-Algorithmus, aber **ohne Normalisierung**, die Vektoren bis \bar{h}^3 und \bar{a}^3 für G . G :

b) Geben Sie zu jedem Schritt k die Gesamtgewichte $I^k(G_1), I^k(G_2), I^k(G_3)$ der drei Komponenten an.

c) Welche der drei bipartiten Komponenten G_1, G_2, G_3 wird asymptotisch das höchste Gesamtgewicht erhalten? Woran liegt das?

d) In G werde die Kante $(3,9)$ eingefügt. Welche *qualitativen* Auswirkungen hat das auf die Gesamtgewichte der drei Komponenten?

e) Zusätzlich werde die Kante $(10,3)$ eingefügt. Welche *qualitativen* Auswirkungen hat das auf die Gesamtgewichte der drei Komponenten?



Zu „bipartit“: Seien V_Q, V_Z zwei endliche Mengen und $V = V_Q \cup V_Z$.

Sei $E = V_Q \times V_Z = \{(i,j) \mid i \in V_Q \wedge j \in V_Z\}$.

(V, E) ist der bipartite Graph mit Quellknotenmenge V_Q und Zielknotenmenge V_Z .

Anschaulich:

von jedem Quellknoten zu jedem Zielknoten gibt es eine Kante; das sind die einzigen Kanten.

Der gegebene Graph $G = (V, E)$ kann damit so dargestellt werden:

$$V = \{1, \dots, 12\} \quad E = (\{1, 2, 3\} \times \{4, 5, 6\}) \cup (\{5, 6\} \times \{7, 8\}) \cup (\{9, 10\} \times \{11, 12\})$$

Die drei Teilmengen von E entsprechen den drei bipartiten Komponenten von G :

$$\begin{aligned} G_1 &= (\{1, \dots, 6\}, \{1, 2, 3\} \times \{4, 5, 6\}) \\ G_2 &= (\{5, \dots, 8\}, \{5, 6\} \times \{7, 8\}) \\ G_3 &= (\{9, \dots, 12\}, \{9, 10\} \times \{11, 12\}) \end{aligned}$$

Man beachte:

die drei Komponenten von G sind nicht disjunkt,
der Graph ist aber trotzdem nicht zusammenhängend
(weder stark noch schwach).

Lösungsvorschlag:

a)

| Hub-Vektor | | | | | | | | | | | | |
|-------------|----|----|----|---|---|---|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| \bar{h}^0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| \bar{h}^1 | 3 | 3 | 3 | 0 | 2 | 2 | 0 | 0 | 2 | 2 | 0 | 0 |
| \bar{h}^2 | 9 | 9 | 9 | 0 | 4 | 4 | 0 | 0 | 4 | 4 | 0 | 0 |
| \bar{h}^3 | 27 | 27 | 27 | 0 | 8 | 8 | 0 | 0 | 8 | 8 | 0 | 0 |

| Authority-Vektor | | | | | | | | | | | | |
|------------------|---|---|---|----|----|----|---|---|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| \bar{a}^0 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| \bar{a}^1 | 0 | 0 | 0 | 3 | 3 | 3 | 2 | 2 | 0 | 0 | 2 | 2 |
| \bar{a}^2 | 0 | 0 | 0 | 9 | 9 | 9 | 4 | 4 | 0 | 0 | 4 | 4 |
| \bar{a}^3 | 0 | 0 | 0 | 27 | 27 | 27 | 8 | 8 | 0 | 0 | 8 | 8 |

In jeder der drei bipartiten Komponenten
haben die Quellknoten hohe Hubwerte und die Zielknoten hohe Autoritätswerte.

Die relativ höchsten Werte haben dabei die Knoten der größten Komponente G_1 .

Je mehr Zielknoten, desto höher die Hubwerte.

Je mehr Quellknoten, desto höher die Autoritätswerte.

b) Gesamtgewichte der drei Komponenten:

$$l(G) = \sum_{\exists(i,j) \in E} h_i + \sum_{\exists(j,i) \in E} a_i$$

| | $l^k(G_1)$ | $l^k(G_2)$ | $l^k(G_3)$ |
|-------|-----------------|----------------|----------------|
| $k=0$ | $3 + 3 = 6$ | $2 + 2 = 4$ | $2 + 2 = 4$ |
| $k=1$ | $9 + 9 = 18$ | $4 + 4 = 8$ | $4 + 4 = 8$ |
| $k=2$ | $27 + 27 = 54$ | $8 + 8 = 16$ | $8 + 8 = 16$ |
| $k=3$ | $81 + 81 = 162$ | $16 + 16 = 32$ | $16 + 16 = 32$ |

c) Asymptotisches Verhalten:

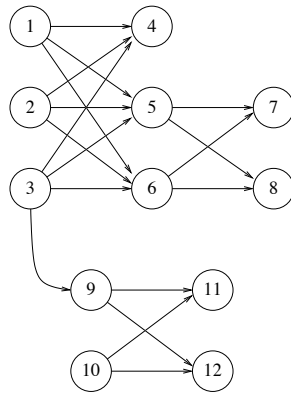
Für jedes k ist das Gesamtgewicht von G_1 am höchsten.

Mit wachsendem k nimmt der Unterschied zu
(auch unter Berücksichtigung der Normalisierung).

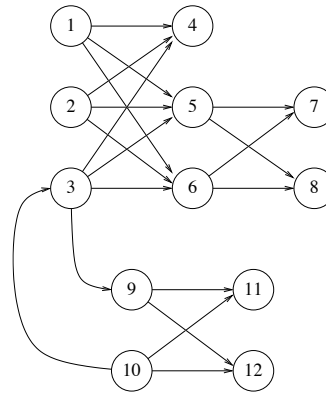
Auch asymptotisch hat G_1 das höchste Gesamtgewicht.

Der HITS-Algorithmus favorisiert die größte bipartite Komponente
aus den Gründen, die wir in Teilaufgabe a) beobachtet haben.

$$G' = G + (3, 9):$$



$$G'' = G' + (10, 3):$$



- d) In G werde Kante $(3, 9)$ eingefügt.

Qualitative Auswirkungen:

Der Autoritätswert von Knoten 9 war bisher 0 und ist jetzt der Hubwert von Knoten 3, also > 0 .

Da der Autoritätswert von Knoten 9 wächst, nimmt der Hubwert von Knoten 3 zu und damit die Autoritätswerte von Knoten 4, 5, 6.

Damit wächst $I(G_1)$.

Aber $I(G_3)$ bleibt unverändert, da Knoten 9 in Komponente G_3 ein Quellknoten ist, so dass sein höherer Autoritätswert keine Rolle für $I(G_3)$ spielt.

- e) In den erweiterten Graphen werde zusätzlich noch Kante $(10, 3)$ eingefügt.

Qualitative Auswirkungen:

Der Autoritätswert von Knoten 3 war bisher 0 und ist jetzt der Hubwert von Knoten 10, also > 0 .

Da der Autoritätswert von Knoten 3 wächst, nimmt der Hubwert von Knoten 10 zu und damit die Autoritätswerte von Knoten 11, 12. Das führt zu einer weiteren Erhöhung der Hubwerte von Knoten 9, 10.

Damit wächst $I(G_3)$.

Aber $I(G_1)$ bleibt unverändert, da Knoten 3 in Komponente G_1 ein Quellknoten ist, so dass sein höherer Autoritätswert keine Rolle für $I(G_3)$ spielt.

Beobachtung:

die eingefügte Kante ging in beiden Fällen von einem Quellknoten einer Komponente zu einem Quellknoten einer anderen Komponente. Das hat jeweils das Gesamtgewicht der ersten Komponente erhöht, aber das der anderen Komponente nicht.