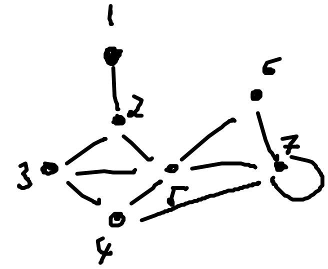
2. Übung zu complex networks



Repräsentationen eines Netzwerks:

(i) Adjazenzmatrix

$$a_{ij} = \begin{cases} 1 & \text{wenn } j \to i \text{ ex.} \\ 0 & \text{sonst} \end{cases} \qquad \underline{\underline{A}} = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 \end{pmatrix}$$

(ii) Linkliste: Liste von Knoten-Tupeln

(1,2), (2,3), (2,5), (3,4), (3,5), (4,5), (5,6), (6,7), (5,7), (7,7), (4,7)

ggf. Richtung codiert durch Reihenfolge

#Links = L (hier L = 11) als Länge der Liste

Auffinden eines bestimmten Links durch Durchgehen der Liste

(iii) Adjazenz-Liste: Speichern der Nachbarn

Knoten	Nachbarn	Grad
1	2	1
2	1, 3, 5	3
3	2, 4, 5	3
4	3, 5, 7	3
5	2, 3, 4, 6, 7	5
6	5, 7	2
7	4, 5, 6, 7	4

Grad = Länge der Nachbarliste

Start und Endpunkte = # Nachbarn summiert

Zentralitätsmaße:

Frage: Wie zentral/wichtig ist ein Knoten?

(i) wichtig ≙ hoher Grad, nicht alle Nachbarn gleich zu werten sind

→ Bestimmte Nachbarn wichtiger als andere (

(ii) Eigenvektor-Zentralität: Wichtigkeit eines Knotens höher, wenn seine Nachbarn wichtig sind.

Wichtigkeit von Knoten j

Wichtigkeit der Nachbarn von Knoten i: $x_i' = \sum_{j=1}^{N} a_{ij}\, x_j \;\Rightarrow\; \underline{x}' = \underline{\underline{A}}\,\underline{x}$

Betrachte Iterationen: $\underline{x}(t) = \underline{\underline{A}}^{t}\, \underline{x}(0)$

Eigenvektoren von $\underline{\underline{A}}$ zum Eigenwert λ_i

Darstellung von $\underline{x}$ in der Basis der Eigenvektoren von $\underline{\underline{A}}$: $\underline{x}(0) = \sum_{i=1}^{N} c_i \underline{v}_i$

$$\underline{x}(t) = \underline{\underline{A}}^{t} \sum_i c_i \underline{v}_i \underset{\underline{\underline{A}}\underline{v}_i = \lambda_i \underline{v}_i}{=} \sum_i c_i \lambda_i^{t} \underline{v}_i \underset{\max_i \lambda_i = \lambda_1}{=} \lambda_1^{t} \sum_i c_i \left(\frac{\lambda_i}{\lambda_1}\right)^{t} \underline{v}_i$$

mit $\frac{\lambda_i}{\lambda_1} < 1$ für alle $i \neq 1$ folgt im Grenzfall $t \to \infty$: $\lim_{t\to\infty} \underline{x}(t) = c_1 \lambda_1^{t} \underline{v}_1$

Bestimmungsgleichung für die Wichtigkeit eines Knotens: $\underline{\underline{A}}\,\underline{x} = \lambda_1 \underline{x}$

implizite Gleichung für x_i

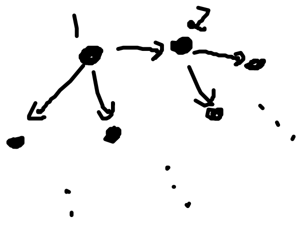
$$x_i = \lambda_1^{-1} \sum_{j=1}^{N} a_{ij} x_j$$

λ_1 wirkt als Skalierungsfaktor zwischen der Summe der Wichtigkeiten der Nachbarn und dem Knoten selbst ($x_i' \longleftrightarrow x_i$)

Fazit: Wichtigkeiten $x_1, x_2, \dots, x_N$ als Lösungen des linearen Gleichungssystems $\underline{x} = \frac{1}{\lambda_1} \underline{\underline{A}}\,\underline{x}$

Vorteil: Wichtigkeit hoch bei vielen Nachbarn (hoher Grad) oder bei wichtigen Nachbarn

Problem: gerichtetes Netzwerk



$x_1 = 0$ weil $k_1^{(in)} = 0$

$\Rightarrow$ Knoten 2 profitiert von Link $1 \to 2$ nicht

Lösung: Katz-Zentralität: $\underbrace{\text{kostenlose Grundwichtigkeit für alle}}$

$\Rightarrow$ Erweitern der Bestimmungsgleichung: $x_i = \alpha \sum_{j=1}^{N} a_{ij} x_j + \beta$

$$\Leftrightarrow \underline{x} = \alpha \underline{\underline{A}}\, \underline{x} + \beta \underline{1} \qquad \begin{pmatrix}1\\ \vdots \\ 1\end{pmatrix}$$

Einheitsmatrix

oder auch $\underline{x} = \beta (\underline{\underline{I}} - \alpha \underline{\underline{A}})^{-1} \underline{1} \quad (\underline{\underline{I}} - \alpha \underline{\underline{A}})\underline{x} = \beta \underline{1}$

Da nur relative Wichtigkeit ausschlaggebend ist, wähle $\beta = 1$.

Achtung: $\det(\underline{\underline{A}} - \alpha^{-1} \underline{\underline{I}}) \neq 0$ gewünscht. Sonst gibt es Divergenzen.
etwa bei $\alpha = \frac{1}{\lambda_1}$

freier Parameter

Wähle Parameter α entsprechend.

Problem: Knoten mit hohem out-degree vererben Wichtigkeit zu allen Nachbarn, aber häufig sind Absender mit vielen Adressaten eher weniger wichtig als bei einem einzelnen Empfänger.

Lösung: Page Rank: gewichte Nachbar wichtig mit dem Ausgangsgrad der Nachbarn

$$x_i = \alpha \sum_{j=1}^{N} a_{ij} \frac{x_j}{k_j^{out}} + \beta$$

$k_i^{out} \overset{!}{=} 1$ wenn kein Ausgangsnachbar vorhanden ist

$\Rightarrow \underline{x} = \alpha \underline{\underline{A}}\, \underline{\underline{D}}^{-1} \underline{x} + \beta \underline{1}$ mit $\underline{\underline{D}}$ ist Diagonalmatrix mit $d_{ii} = \max(k_i^{out}, 1)$

neu im Vgl. zu Katz

$\Rightarrow$ Auflösen nach $\underline{x}$ liefert: $\underline{x} = \beta (\underline{\underline{I}} - \alpha \underline{\underline{A}}\, \underline{\underline{D}}^{-1})^{-1} \underline{1}$

$$= \beta \underline{\underline{D}} (\underline{\underline{D}} - \alpha \underline{\underline{A}})^{-1} \underline{1}$$

Ähnliche Wahl wie bei Katz-Zentralität: $\alpha <$ max. Eigenwert von $\underline{\underline{A}}\,\underline{\underline{D}}^{-1}$

Bsp.: Google verwendet $\alpha = 0.85$, $\beta = 1$

Page Rank wurde 1996 vorgestellt. (Name in Anlehnung an Larry Page)

ohne β-Term: $x_i = \alpha \sum_{j=1}^{N} a_{ij} \frac{x_j}{k_j^{out}}$ $\Rightarrow$ Lösung des Gleichungssystems?

$x_i \sim k_i$