

Markov-láncok további tulajdonságai

1 Periódus

Tekintsünk egy irreducibilis Markov-láncot (reducibilis lánc esetén tekinthetjük külön-külön az egyes rekurrens osztályokat). Az i állapot periódusán a

$$J_i = \{n > 0 : p_n(i, i) > 0\}$$

halmaz elemeinek legnagyobb közös osztóját értjük. Ha ez az érték 1, akkor azt mondjuk, hogy a lánc *aperiodikus* (nem periodikus). A periódus definíciója a következőt jelenti: vegyünk az összes olyan lépésszámot (n), amellyel visszatérhetünk valamekkora nem nulla valószínűséggel i -ből indulva az i állapotba (ezek a lépésszámok alkotják J_i elemeit), majd vegyünk ezen értékek legnagyobb közös osztóját. Pl. ha 3 és 5 lépéssel is visszatérhetünk a kiindulási állapotba, akkor a periódus 1; ha csak 2, 4, ... lépésszámmal térhetünk vissza, akkor a periódus kettő stb.

A periódus osztálytulajdonság, azaz egy rekurrens osztályon belül minden állapot periódusa ugyanaz. Ha a Markov-lánc irreducibilis, akkor minden állapot periódusa ugyanaz, de ha reducibilis, akkor lehetséges, hogy vannak különböző periódusú állapotok is.

Az aperiodikusság segítségével is megfogalmazhatunk feltételt az invariáns eloszlás létezésére:

Legyen P egy irreducibilis, aperiodikus Markov-lánc átmenetvalószínűség mátrixa. Ekkor

- *egyértelműen létezik invariáns eloszlás, mely kielégíti a*

$$\underline{\pi} = \underline{\pi}P$$

egyenletet, továbbá

- tetszőleges φ_0 kezdeti eloszlás esetén

$$\lim_{n \rightarrow \infty} \varphi_0 P^n = \underline{\pi}.$$

Példa

Tekintsük a szimmetrikus véletlen bolyongást visszaverő falakkal, az állapottér pedig legyen $S = \{0, 1, 2, 3, 4\}$. A Markov-lánc nyilvánvalóan irreducibilis, hiszen bármely állapotból elérhetünk bármely más állapotba, másrészt arról is könnyű meggyőződni, hogy a periódus 2. Az átmenetvalószínűség mátrix az alábbi:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0,5 & 0 & 0,5 & 0 & 0 \\ 0 & 0,5 & 0 & 0,5 & 0 \\ 0 & 0 & 0,5 & 0 & 0,5 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

P^n viselkedése függeni fog n paritásától. Ez is nyilvánvaló, hiszen páros számú, illetve páratlan számú lépéssel más helyekre juthatunk el, például az 1-esből csak páros számú lépéssel juthatunk el a 3-asba, míg a 4-esbe csak páratlan számúval.

- Ha n páratlan, akkor

$$P^n \approx \begin{bmatrix} 0 & 0,5 & 0 & 0,5 & 0 \\ 0,25 & 0 & 0,5 & 0 & 0,25 \\ 0 & 0,5 & 0 & 0,5 & 0 \\ 0,25 & 0 & 0,5 & 0 & 0,25 \\ 0 & 0,5 & 0 & 0,5 & 0 \end{bmatrix},$$

- Ha n páros, akkor

$$P^n \approx \begin{bmatrix} 0,25 & 0 & 0,5 & 0 & 0,25 \\ 0 & 0,5 & 0 & 0,5 & 0 \\ 0,25 & 0 & 0,5 & 0 & 0,25 \\ 0 & 0,5 & 0 & 0,5 & 0 \\ 0,25 & 0 & 0,5 & 0 & 0,25 \end{bmatrix}.$$

Tudjuk, hogy aperiodikus esetben a hosszú távú viselkedést az invariáns eloszlás írja le, vagyis tetszőleges kezdeti eloszlás esetén a határeloszlás az invariáns eloszlás lesz. Mi lehet periodikus esetben az invariáns eloszlás szerepe?

A $\underline{\pi} = \underline{\pi}P$ egyenlet megoldása, vagyis az invariáns eloszlás a példában

$$\underline{\pi} = \begin{bmatrix} 0,125 & 0,25 & 0,25 & 0,25 & 0,125 \end{bmatrix}.$$

Láthatólag ez nem egyezik meg P^n soraival egyik esetben sem, de vegyük észre, hogy a fenti két mátrix átlaga a következő:

$$\lim_{n \rightarrow \infty} \frac{1}{2} (P^n + P^{n+1}) = \begin{bmatrix} 0,125 & 0,25 & 0,25 & 0,25 & 0,125 \\ 0,125 & 0,25 & 0,25 & 0,25 & 0,125 \\ 0,125 & 0,25 & 0,25 & 0,25 & 0,125 \\ 0,125 & 0,25 & 0,25 & 0,25 & 0,125 \\ 0,125 & 0,25 & 0,25 & 0,25 & 0,125 \end{bmatrix}.$$

Azt mondhatjuk tehát, hogy ekkor $\pi(i)$ nem annak valószínűsége, hogy hosszú távon a rendszer az i állapotban lesz (hiszen ez a valószínűség függ n paritásától), hanem az i állapotban eltöltött átlagos időt adja meg.

Általánosan, ha P egy irreducibilis Markov-lánc átmenetvalószínűség mátrixa, melynek periódusa d , akkor létezik invariáns eloszlás ($\underline{\pi}$), de tetszőleges φ_0 kezdeti eloszlásnál nagy n esetén $\varphi_0 P^n$ d különböző eloszláson halad végig ciklikusan. Az invariáns eloszlás ezek átlaga lesz, azaz

$$\lim_{n \rightarrow \infty} \frac{1}{d} (\varphi_0 P^n + \varphi_0 P^{n+1} + \dots + \varphi_0 P^{n+d-1}) = \underline{\pi}.$$

2 Visszatérési idő

Legyen X_n egy irreducibilis Markov-lánc, az átmenetvalószínűség mátrixa pedig legyen P . Tekintsük az n idő alatt a j állapotban eltöltött időt:

$$Y(j, n) = \sum_{m=0}^n I(X_m = j),$$

ahol I az indikátor függvényt jelöli, vagyis értéke 1, ha a jelölt esemény bekövetkezik, és 0, ha nem. Az előző szakaszból következik, hogy ha az invariáns eloszlás $\underline{\pi}$, akkor

$$\lim_{n \rightarrow \infty} \frac{1}{n+1} E[Y(j, n) | X_0 = i] = \lim_{n \rightarrow \infty} \frac{1}{n+1} \sum_{m=0}^n P(X_m = j | X_0 = i) = \pi(j),$$

azaz $\pi(j)$ azt adja meg, hogy hosszú távon az idő hányadrészét tölti a rendszer a j állapotban.

Tegyük fel, hogy $X_0 = i$. Legyen T az első visszatérésig eltelt idő, azaz

$$T = \min \{n \geq 1 : X_n = i\} .$$

Mivel a lánc irreducibilis, ezért $T < \infty$ 1 valószínűséggel, és megmutatható, hogy $E(T) < \infty$ is teljesül.

Tekintsük most a k -edik visszatérésig eltelt időt. Ez az idő megadható független, azonos eloszlású valószínűségi változók összegeként: $T_1 + T_2 + \dots + T_k$, ahol T_m az $m - 1$ -edik és az m -edik visszatérés között eltelt időt jelöli. Nagy k esetén a nagy számok törvénye szerint

$$\frac{1}{k}(T_1 + T_2 + \dots + T_k) \approx E(T) ,$$

vagyis $kE(T)$ lépés alatt átlagosan k visszatérés történik. De tudjuk azt is, hogy n lépésből átlagosan $n\pi(i)$ -szer lesz a rendszer az i állapotban. Így n helyére $kE(T)$ -t helyettesítve kapjuk a következőt:

$$kE(T)\pi(i) = k \implies E(T) = \frac{1}{\pi(i)} .$$

Vagyis az i állapotból indulva a visszatéréshez szükséges lépések számának várható értéke az invariáns valószínűség reciproka. A fenti gondolatmenet természetesen nem teljesen precíz, de a matematikailag teljesen korrekt levezetés is ezt az eredményt szolgáltatja.

Példa Tekintsük az időjárásos példát. Az állapotok: napos (N), felhős (F). Az átmenetvalószínűség mátrix pedig

$$P = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}$$

Tegyük fel, hogy ma napos az idő, vagyis a folyamat az N állapotban van. Átlagosan hány nap múlva áll elő újra ez az állapot (vagyis átlagosan mennyit kell várni egy napos nap után, hogy újra ilyen legyen az időjárás)?

Láttuk, hogy az invariáns eloszlás $\pi = [4/7, 3/7]$, tehát $\pi_N = 4/7$. A napos állapotból indulva az ugyanide történő visszatéréshez szükséges lépések számának várható értéke az invariáns valószínűség reciproka, azaz most $7/4 = 1.75$. Tehát *átlagosan* 1.75 nap múlva lesz ugyanez az állapot.

Vizsgáljuk meg most általánosabban, hogy mennyi az i állapotból indulva a j állapotba jutáshoz szükséges lépések számának várható értéke. Legyen T_{ij} az i -ből a j -be való eljutáshoz szükséges idő, azaz

$$T_{ij} = \min \{n \geq 1 : X_n = j | X_0 = i\} .$$

Azt már tudjuk, hogy $E(T_{ii}) = \frac{1}{\pi(i)}$. Nézzük azt az esetet, amikor $i \neq j$. Módosítsuk a Markov-láncot egy újabb állapot (h) bevezetésével, a rá vonatkozó átmenetvalószínűségek pedig legyenek a következők:

$$P(X_{n+1} = i | X_n = h) = 1 \quad \text{és} \quad P(X_{n+1} = h | X_n = j) = 1 .$$

Így a h -ból h -ba jutás útvonala: $h \rightarrow i \rightarrow j \rightarrow h$. Ebből következően a visszajutáshoz szükséges lépések számának várható értéke:

$$E(T_{hh}) = 1 + E(T_{ij}) + 1 = 2 + E(T_{ij}) ,$$

ebből pedig $E(T_{ij})$ meghatározható, hiszen

$$E(T_{ij}) = E(T_{hh}) - 2$$

Az eljárás lépésként:

- Felírjuk a módosított átmenetvalószínűség mátrixot: az új h állapotból csak az i -be lehessen lépni (1 valószínűséggel), a j -ből pedig csak a h -ba lehessen lépni (1 valószínűséggel). Tehát a j -ből máshova (nem h -ba) lépések valószínűségét nullára változtatjuk.
- Meghatározzuk az új mátrixhoz tartozó invariáns eloszlás h -hoz tartozó komponensét (π_h).
- A h -ból h -ba jutáshoz lépések számának várható értéke $E(T_{hh}) = 1/\pi_h$.

- Végül:

$$E(T_{ij}) = \frac{1}{\pi_h} - 2$$

Példa Tekintsük a újra az időjárásos példát. Az állapotok: napos (N), felhős (F).

Az átmenetvalószínűség mátrix

$$P = \begin{bmatrix} 0.4 & 0.6 \\ 0.8 & 0.2 \end{bmatrix}.$$

Tegyük fel, hogy ma napos az idő. Átlagosan hány nap múlva lesz felhős?

A kérdés az, hogy az N állapotból átlagosan hány lépéssel jutunk el F -be. Vezessük be az új h állapotot úgy, hogy h -ból csak N -be lehessen lépni, F -ből pedig csak h -ba. Ekkor a módosított mátrix:

$$P' = \begin{matrix} & \begin{matrix} N & F & h \end{matrix} \\ \begin{matrix} N \\ F \\ h \end{matrix} & \begin{bmatrix} 0.4 & 0.6 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \end{matrix}$$

Megoldandó a $\underline{\pi}' = \underline{\pi}' \cdot P'$ egyenlet, ahol $\underline{\pi}' = [g]$. Ebből a megoldandó egyenletrendszer (plusz még az az információ, hogy $\pi_N + \pi_F + \pi_h = 1$):

$$\pi_N = 0.4\pi_N + \pi_h$$

$$\pi_F = 0.6\pi_N$$

$$\pi_h = \pi_F$$

Ebből $\pi_h = 6/22$ (a többi nem fontos számunkra). Tehát az N állapotból az F állapotba jutáshoz szükséges lépések számának várható értéke:

$$E(T_{NF}) = \frac{1}{\pi_h} - 2 = \frac{10}{6}.$$

Vegyük észre, hogy ebben az egyszerű esetben ezt az eredményt másképp is megkaphattuk volna. Mivel csak két állapot van, ezért tekinthetjük úgy a rendszert, hogy addig várunk, amíg az $N \rightarrow F$ átmenet be nem következik. Ekkor a szükséges lépések száma geometriai eloszlású, a paraméter (a valószínűség) 0.6, így a várható érték ennek a reciproka, azaz $10/6$.

3 Tranziens állapotok

Legyen P az X_n Markov-lánc átmenetvalószínűség mátrixa. Tegyük fel, hogy létezik néhány tranziens állapot is. Ekkor az átmenetvalószínűség mátrix átrendezhető az alábbi alakra, ahol $\hat{P}$ a rekurrens, Q pedig a tranziens állapotoknak felel meg:

$$P = \left[\begin{array}{c|c} \hat{P} & 0 \\ \hline S & Q \end{array} \right] \implies P^n = \left[\begin{array}{c|c} \hat{P}^n & 0 \\ \hline S_n & Q^n \end{array} \right].$$

Példa Tekintsük a szimmetrikus véletlen bolyongást elnyelő falakkal az $S = \{0, 1, 2, 3, 4\}$ állapottéren. Ekkor az átmenetvalószínűség mátrix:

$$P = \left[\begin{array}{ccccc} 1 & 0 & 0 & 0 & 0 \\ 0,5 & 0 & 0,5 & 0 & 0 \\ 0 & 0,5 & 0 & 0,5 & 0 \\ 0 & 0 & 0,5 & 0 & 0,5 \\ 0 & 0 & 0 & 0 & 1 \end{array} \right].$$

Átrendezve az állapotokat $(0, 4, 1, 2, 3)$:

$$P = \left[\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 0,5 & 0 & 0 & 0,5 & 0 \\ 0 & 0 & 0,5 & 0 & 0,5 \\ 0 & 0,5 & 0 & 0,5 & 0 \end{array} \right].$$

A blokkmátrixban Q egy ún. szubsztochasticus mátrix, vagyis elemei nemnegatívak és összegük minden sorban kisebb vagy egyenlő, mint egy. Mivel a Q mátrix által reprezentált állapotok tranziensek, ezért $Q^n \rightarrow 0$, ha $n \rightarrow \infty$. Ebből következően Q minden sajátértéke abszolút értékben szigorúan kisebb egynél, így az $I - Q$ mátrix invertálható, vagyis definiálhatjuk a következőt:

$$M = (I - Q)^{-1}.$$

Legyen i egy tranziens állapot, és legyen Y_i az i -be tett látogatások teljes száma, azaz

$$Y_i = \sum_{n=0}^{\infty} I(X_n = i).$$

Mivel i egy tranziens állapot, ezért ez az összeg 1 valószínűséggel véges. Tegyük fel, hogy a $X_0 = j$, ahol j egy másik tranziens állapot. Ekkor

$$E[Y_i|X_0 = j] = E\left[\sum_{n=0}^{\infty} I(X_n = i)|X_0 = j\right] = \sum_{n=0}^{\infty} P(X_n = i|X_0 = j) = \sum_{n=0}^{\infty} p_n(j, i).$$

Vagyis azt kaptuk, hogy j -ből indulva az i -be tett látogatások számának várható értéke az $I + P + P^2 + \dots$ mátrix (j, i) eleme. Mivel mindkét állapot tranziens, ezért ez ugyanaz, mint az $I + Q + Q^2 + \dots$ mátrix (j, i) eleme. Egyszerű számítással igazolható, hogy

$$(I + Q + Q^2 + \dots)(I - Q) = I,$$

azaz

$$I + Q + Q^2 + \dots = (I - Q)^{-1} = M.$$

Tehát azt mutattuk meg, hogy j -ből indulva az i -be tett látogatások számának várható értéke az M mátrix (j, i) eleme. Az előző példában:

$$Q = \begin{bmatrix} 0 & 0,5 & 0 \\ 0,5 & 0 & 0,5 \\ 0 & 0,5 & 0 \end{bmatrix}, \quad M = (I - Q)^{-1} = \begin{bmatrix} 1,5 & 1 & 0,5 \\ 1 & 2 & 1 \\ 0,5 & 1 & 1,5 \end{bmatrix}.$$

Tehát például az 1-es állapotból indulva a 3-as állapotba tett látogatások számának várható értéke 0,5. Az elnyelődés előtt a lépések számának várható értéke pedig $1,5 + 1 + 0,5 = 3$.

Az itt megismert eljárást alkalmazhatjuk akkor is, amikor egy irreducibilis Markov-lánc esetében a j állapotból az i állapotba jutáshoz szükséges lépések számának várható értékét akarjuk meghatározni. Először rendezzük át az átmenetvalószínűség mátrixot úgy, hogy az i állapot szerepeljen az első helyen:

$$P = \left[\begin{array}{c|c} p(i, i) & R \\ \hline S & Q \end{array} \right].$$

Ezután változtassuk i -t elnyelő állapottá:

$$P = \left[\begin{array}{c|c} 1 & 0 \\ \hline S & Q \end{array} \right].$$

Legyen T_i az i állapot eléréséhez szükséges lépések száma, vagyis a legkisebb n , amelyre $X_n = i$. A többi k állapot esetében $T_{i,k}$ a k állapotba tett látogatások száma az i állapot elérése előtt. Ekkor

$$E[T_i | X_0 = j] = E\left[\sum_{k \neq i} T_{i,k} | X_0 = j\right] = \sum_{k \neq i} M_{jk}.$$

Tehát a j állapotból indulva az i állapot eléréséhez szükséges lépések számának várható értéke az M mátrix j -edik sorában levő elemek összege.

Példa

Tekintsük a szimmetrikus véletlen bolyongást visszaverő falakkal az $S = \{0, 1, 2, 3, 4\}$ állapottéren. Ekkor az átmenetvalószínűség mátrix:

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0,5 & 0 & 0,5 & 0 & 0 \\ 0 & 0,5 & 0 & 0,5 & 0 \\ 0 & 0 & 0,5 & 0 & 0,5 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix}.$$

Meghatározzuk, hogy a 4-es állapotból várhatóan hány lépésben jutunk el a 0-ba.

$$Q = \begin{bmatrix} 0 & 0,5 & 0 & 0 \\ 0,5 & 0 & 0,5 & 0 \\ 0 & 0,5 & 0 & 0,5 \\ 0 & 0 & 1 & 0 \end{bmatrix}, \quad M = (I - Q)^{-1} = \begin{bmatrix} 2 & 2 & 2 & 1 \\ 2 & 4 & 4 & 2 \\ 2 & 4 & 6 & 3 \\ 2 & 4 & 6 & 4 \end{bmatrix}.$$

Tehát a 4 állapotból várhatóan $2 + 4 + 6 + 4 = 16$ lépésből jutunk el a 0 állapotba.

Tegyük fel most, hogy legalább két rekurrens osztály van. Vizsgáljuk azt, hogy egy adott j tranziens állapotból indulva mi a valószínűsége, hogy egy előre kiszemelt rekurrens osztályba kerül a rendszer. Feltehetjük azt, hogy a rekurrens osztályok mindegyike csupán egyetlen elemet tartalmaz (uis. most nem vizsgáljuk, hogy a rekurrens osztályon belül mi történik).

Legyenek a rekurrens állapotok $r_1, r_2, \dots, r_k$, a tranziens állapotok pedig $t_1, t_2, \dots, t_s$. Rendezzük át az átmenetvalószínűség mátrixot oly módon, hogy a rekurrens állapotok megelőzzék a tranziens állapotokat:

$$P = \left[\begin{array}{c|c} I & 0 \\ \hline S & Q \end{array} \right].$$

Jelölje $\alpha(t_i, r_j)$ ($i = 1, \dots, s$, $j = 1, \dots, k$) annak valószínűségét, hogy a t_i tranziens állapotból indulva valamikor az r_j rekurrens állapotba érünk. Legyen továbbá (természetesen) $\alpha(r_{j_1}, r_{j_2}) = 1$, ha $j_1 = j_2$, és $\alpha(r_{j_1}, r_{j_2}) = 0$, ha $j_1 \neq j_2$. Ekkor minden t_i tranziens állapot esetén:

$$\begin{aligned} \alpha(t_i, r_j) &= P(X_n = r_j \text{ valamikor} \mid X_0 = t_i) \\ &= \sum_{x \in S} P(X_1 = x \mid X_0 = t_i) P(X_n = r_j \text{ valamikor} \mid X_1 = x) \\ &= \sum_{x \in S} p(t_i, x) \alpha(x, r_j). \end{aligned}$$

Legyen A az $\alpha(t_i, r_j)$ elemeket tartalmazó $s \times k$ -as mátrix. Ekkor a fenti egyenlőség mátrix formában felírva:

$$A = S + QA.$$

Ebből az A mátrix:

$$A = (I - Q)^{-1}S = MS.$$

A kapott mátrix (i, j) eleme tehát annak valószínűségét adja meg, hogy a t_i tranziens állapotból indulva végül az r_j rekurrens állapotba jutunk.

Példa

Tekintsük a szimmetrikus véletlen bolyongást elnyelő falakkal az $S = \{0, 1, 2, 3, 4\}$ állapottéren. Rendezzük az állapotokat 0, 4, 1, 2, 3 sorrendbe, mert a 0 és a 4 rekurrens állapotok (egyelemű rekurrens osztlyok), az 1, 2 és 3 pedig tranziensek. Így az átmenetvalószínűség mátrix:

$$P = \left[\begin{array}{cc|ccc} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ \hline 0,5 & 0 & 0 & 0,5 & 0 \\ 0 & 0 & 0,5 & 0 & 0,5 \\ 0 & 0,5 & 0 & 0,5 & 0 \end{array} \right].$$

Az ebből meghatározott mátrixok:

$$S = \begin{bmatrix} 0,5 & 0 \\ 0 & 0 \\ 0 & 0,5 \end{bmatrix}, \quad M = (I - Q)^{-1} = \begin{bmatrix} 1,5 & 1 & 0,5 \\ 1 & 2 & 1 \\ 0,5 & 1 & 1,5 \end{bmatrix}, \quad A = MS = \begin{bmatrix} 0,75 & 0,25 \\ 0,5 & 0,5 \\ 0,25 & 0,75 \end{bmatrix}.$$

A kapott A mátrixból kiolvasható, hogy pl. az 1-es állapotból indulva 0,75 a valószínűsége annak, hogy végül a 0 állapotba jutunk.

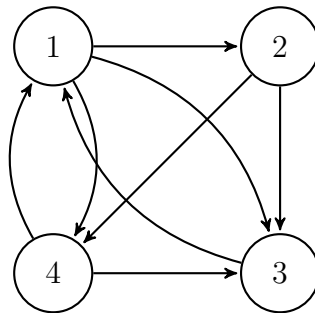
4 Egy fontos alkalmazás: PageRank

Adott n darab weboldal, melyeket *fontosság* szerint szeretnénk sorba rendezni. Az egyes oldalakhoz rendelt fontosságok legyenek $x_1, x_2, \dots, x_n$. A fontosságok értékének meghatározásához az oldalak közötti linkek struktúráját használhatjuk fel, vagyis egy irányított gráfot. Tulajdonképpen tekinthetünk úgy minden egyes bemenő élre, hogy egy szavazat az adott oldalra fontosság szempontjából.

Az alábbiakban a következő példán fogjuk bemutatni az egyes próbálkozásokat, azok hátrányait, majd a megoldást: Legyen $n = 4$, az egyes csúcsok közötti kapcsolatok pedig az alábbiak:

- Az 1-esből mutat link az összes többibe;
- a 2-esből mutat link a 3-asba és a 4-esbe;
- a 3-asból csak az 1-esbe mutat link;
- a 4-esből az 1-esbe és a 3-asba.

Tehát a gráf struktúrája:



Az első próbálkozás az, hogy minden egyes oldal fontossága legyen annyi, mint az oda bemenő linkek száma. A példában ekkor az egyes csúcsok a következő értékeket kapnák: $x_1 = 2$, $x_2 = 1$, $x_3 = 3$ és $x_4 = 2$. Természetes elvárás azonban az, hogy egy fontosabb oldalról bejövő link kapjon nagyobb súlyt, mint egy random blogról jövő, így ezt az ötletet elvethetjük.

Próbálkozzunk esetleg azzal, hogy x_k legyen a k -ba beérkező fontosságok összege. Ekkor a példában a következő egyenletrendszerrel kapjuk:

$$x_1 = x_3 + x_4$$

$$x_2 = x_1$$

$$x_3 = x_1 + x_2 + x_4$$

$$x_4 = x_1 + x_2$$

Itt a jó hír az, hogy ennek az egyenletrendszernek van megoldása. A rossz hír viszont az, hogy csak triviális megoldása van ($x_1 = x_2 = x_3 = x_4 = 0$). További probléma, hogy ha egy oldalról sok kimenő link van, akkor az befolyásosabb lesz (random blogról linkek mindenhova).

A rangsorolási problémára megoldást Sergey Brin és Larry Page adott 1998-ban, az ez alapján kifejlesztett PageRank algoritmus a Google keresőmotor legfontosabb eleme. Legyen az x_k értéke az x_j/n_j értékek összege, ahol az összegzés azon j oldalakra vonatkozik, melyekből mutat link k -ba, n_j értéke pedig a j -ből kimenő linkek száma, azaz

$$x_k = \sum \frac{x_j}{n_j}.$$

Tehát minden egyes oldal tulajdonképpen szétosztja a fontosságát azon oldalak között, ahová belőle link mutat. A példánkon ez a következő egyenletrendszerre vezet:

$$\begin{aligned} x_1 &= x_3 + \frac{1}{2}x_4 \\ x_2 &= \frac{1}{3}x_1 \\ x_3 &= \frac{1}{3}x_1 + \frac{1}{2}x_2 + \frac{1}{2}x_4 \\ x_4 &= \frac{1}{3}x_1 + \frac{1}{2}x_2 \end{aligned}$$

Az egyenletrendszer mátrixos alakban:

$$\begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & x_3 & x_4 \end{bmatrix} \underbrace{\begin{bmatrix} 0 & 1/3 & 1/3 & 1/3 \\ 0 & 0 & 1/2 & 1/2 \\ 1 & 0 & 0 & 0 \\ 1/2 & 0 & 1/2 & 0 \end{bmatrix}}_P$$

Könnyen észrevehetjük, hogy a P mátrix egy átmenetvalószínűség mátrix, a feladat pedig tulajdonképpen az invariáns eloszlás meghatározása. Az egyenletrendszert megoldva a példánkban a következő értékeket kapjuk:

$$\begin{aligned} x_1 &= \frac{12}{31} \approx 0.387 \\ x_2 &= \frac{4}{31} \approx 0.129 \\ x_3 &= \frac{9}{31} \approx 0.290 \\ x_4 &= \frac{6}{31} \approx 0.193 \end{aligned}$$

Összegzés: Adott n darab weboldal, a linkstruktúra alapján készítsük el a P mátrixot. Az i -edik sor j -edik eleme:

$$p_{ij} = \begin{cases} \frac{1}{n_i}, & \text{ha } i\text{-ből mutat link } j\text{-be} \\ 0 & \text{különben.} \end{cases}$$

, ahol n_i az i -ből kiinduló linkek száma. Mivel $\sum_i \frac{1}{n_i} = 1$ minden sorra, ezért P átmenetvalószínűség mátrix lesz.

Problémák:

- létezik-e mindig megoldás? (nem)
- egyértelmű-e a megoldás? (nem)

Láttuk, hogy ha az átmenetvalószínűség mátrix minden eleme pozitív, akkor létezik invariáns eloszlás, továbbá az invariáns eloszlás egyértelmű. Ezért a kapott P mátrixot kicsit módosítjuk, hogy mindig létezzen és egyértelmű legyen az invariáns eloszlás:

$$P' = (1 - d) \cdot P + d \cdot [1/n]$$

ahol $[1/n]$ egy olyan mátrix, aminek minden eleme $1/n$, d pedig egy 0 és 1 közötti szorzótényező (csillapítás), szokásos értéke 0.15. Mivel a kapott P' mátrix minden eleme pozitív, ezért létezik és egyértelmű az invariáns eloszlás.

Megjegyzések:

- Az $[1/n]$ mátrix tekinthető a véletlenszerű netező megjelenítésének, aki nem mindig a linkek közül választ, hanem egy valamekkora (0.15) valószínűséggel új oldalt tölt be.
- A PageRank alkalmazásakor nem szükséges az invariáns eloszlást pontosan meghatározni, ezt nem is teszik meg. Addig folytatják az iterációt ($\varphi_1, \varphi_2, \dots$ sorozatos kiszámítását), amíg az eloszlásban az egyes oldalak valószínűségek szerinti sorrendje nem állandósul.