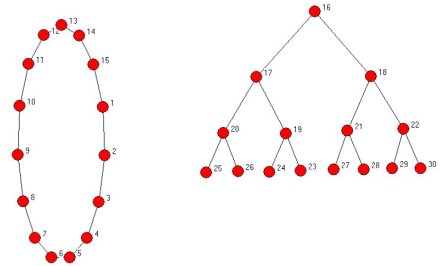
Hálózattudomány

SZTE Informatikai Intézet
Számítógépes Optimalizálás Tanszék
Előadó: London András

2. Előadás

Átmérő

- ℓ_{ij} – a legrövidebb út a hálózatban i és j pont között
- $\Delta = \max_{i,j} \ell_{ij}$ – átmérő: az összes legrövidebb út közül a legnagyobb



ábra. mennyi egy N pontú kör és egy N pontú bináris fa átmérője?

Átlagos úthossz

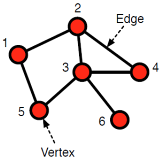
- A legrövidebb utak hosszának átlaga

$$\langle \ell \rangle = \frac{1}{\binom{n}{2}} \sum_{i,j} \ell_{ij}$$

- Valós hálózatokban miért érdekes, milyen információt ad?
- Algoritmusok a kiszámításhoz?

Fokszámeloszlás

- $A = [a_{ij}] \in \mathbb{R}^{n \times n}$ – G szomszédsági mátrixa:
- i pont foka: $k_i = \sum_{j=1}^{n} a_{ij}$
- Fokszámeloszlás: $\mathbb{P}$(egy véletlenül választott pont foka k)



k	$\Pr(k)$
1	1/6
2	3/6
3	1/6
4	1/6

- Miért érdekes egy hálózat fokszámeloszlása?
- Milyen fokszámeloszlást követnek a valós hálózatok?

$\longrightarrow$ kulcsfontosságú fogalom, a későbbiekben részletesen tárgyaljuk.

Melyek a hálózat „fontos" pontjai?

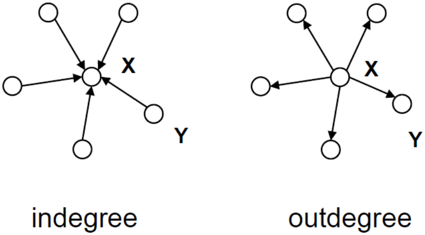
- Strukturális tulajdonság szempontjából, például
  - magas fokszámú
  - a „központban" van
  - valamilyen dinamikus folyamat szempontjából fontos (pl. fertőzés terjedés, véletlen bolyongás)

$\Longrightarrow$ Centralitás

- „Minél centrálisabb annál fontosabb, minél kevésbé centrális annál kevésbé fontos"
- De hogyan „mérjük" a centralitást?

Fokszám centralitás

- **Nagyobb fokszám → fontosabb pont**
- $k_i = \sum_{j=1}^{n} a_{ij}$; irányított: $k_i^{be} = \sum_{j=1}^{n} a_{ji}$, $k_i^{ki} = \sum_{j=1}^{n} a_{ij}$



ábra. Be- és kifok centralitás.

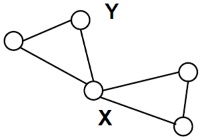
Betweenness (köztiség) centralitás

- Két pont milyen messze van egymástól, **ha át kell menni egy kijelölt harmadik ponton**

$$BC(k) = \sum_{i \neq k \neq j} \frac{\sigma_{ij}(k)}{\sigma_{ij}},$$

ahol σ_{ij} a i és j közötti legrövidebb utak száma, $\sigma_{ij}(k)$ pedig azon legrövidebb $i - j$ utak száma, melyek átmennek k-n

- Brandes algoritmusa: $O(nm)$ futási idejű BC számító algoritmus (m a gráf éleinek száma) ű



ábra. Mennyi X és Y betweennes értéke?

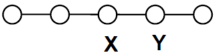
Closeness (közelség) centralitás

- Mennyire van a „központban” egy pont $\rightarrow$ átlagosan **milyen hosszúak egy pontból induló legrövidebb utak** a hálózat többi pontjába

$$C(i) = \frac{n-1}{\sum_{i \neq j} \ell_{ij}},$$

ahol ℓ_{ij} az i és j közti legrövidebb út hossza.

- Számolás: Floyd-Warshall algoritmus



ábra. Mennyi X és Y closeness értéke?

Harmonikus centralitás

Két probléma a closeness-szel:

- a valós hálózatok átmérője általában kicsi $\rightarrow$ a $C(i)$ értékek szűk tartományban változnak
- nem összefüggő hálózat esetén nem számolható

Harmonikus centralitás

$$C^h(i) = \frac{1}{n-1} \sum_{j \neq i} \frac{1}{\ell_{ij}},$$

ahol $\ell_{ij} = \infty$, ha nincs $i - j$ út.

Sajátérték centralitás

- Alapötlet: **nem minden szomszéd egyforma súllyal számít** a centralitás kiszámításánál
- Rekurzív formula:

$$x_i^{(t+1)} = \sum_{j=1}^{n} w_{ij} x_j^{(t)}$$

„Minél fontosabb a szomszéd, annál jobban járul hozzá az adott pont fontosságához"

- Mátrix formában:

$$A\mathbf{x} = \lambda_1 \mathbf{x},$$

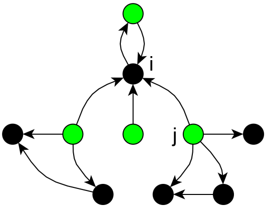
ahol λ_1 az A mátrixhoz tartozó legnagyobb sajátérték (ld. Perron-Frobenius tétel)

PageRank

- **Mi a helyzet ha a gráf nem összefüggő?** $\Longrightarrow$ „Véletlen szörföző”, ld. Google keresőmotor [1]
- Rekurzió:

$$PR(i) = \frac{1-\lambda}{n} + \lambda \sum_{j \in N^+(i)} \frac{PR(j)}{k^{ki}(j)},$$

ahol $\lambda \in [0,1]$ paraméter (ugró faktor), $N^+(i)$ az i pont „be-szomszédsága”



[1] Brin & Page, *Computer networks and ISDN systems*, 1998

Egy kis lineáris algebra?

A PageRank rekurziót vektoregyenlet formában felírva:

$$\mathbf{PR} = \mathbf{PR}R = \mathbf{PR}(\lambda P + (1-\lambda)U)$$

Ezt átalakítva

$$\begin{aligned}\mathbf{PR} &= \mathbf{PR}R = \mathbf{PR}(\lambda P + (1-\lambda)U) = \lambda\mathbf{PR}P + (1-\lambda)\mathbf{PR}U = \\ &= \lambda\mathbf{PR}P + (1-\lambda)\mathbf{PR}\mathbb{1}\mathbb{1}^T\frac{1}{N} = \lambda\mathbf{PR}P + (1-\lambda)\mathbb{1}^T\frac{1}{N}\end{aligned}$$

felhasználva, hogy $U = \mathbb{1}\mathbb{1}^T\frac{1}{N}$ és $\mathbf{PR}\mathbb{1} = 1$. Innen kapjuk, hogy

$$\mathbf{PR} = \frac{1-\lambda}{N}\mathbb{1}(I-\lambda P)^{-1} = \frac{1-\lambda}{N}\mathbb{1}\sum_{n=0}^{\infty}(\lambda P)^n$$

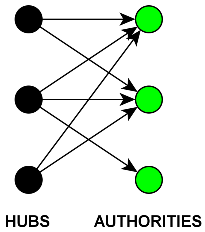
PageRank algoritmus

Input G irányított gráf
Output PageRank értek vektora

1: Initialize $\mathbf{PR}_0 = \frac{\lambda}{N}\mathbb{1}$
2: $k = 1$
3: **repeat**
4: $\quad \mathbf{PR}_{k+1} := \frac{\lambda}{N}\mathbb{1} + \lambda A D^{-1}\mathbf{PR}_k$
5: $\quad k = k + 1$
6: **until** $||\mathbf{PR}_{k+1} - \mathbf{PR}_k||_1$
7: return $\mathbf{PR}_{k+1}$

HITS (Hyperlink Induced Topic Search)

- Kleinberg fejlesztette ki[2], az eredeti PageRank "finomított" változata
- A gráf pontjainak rangsorolásánál megkülönböztet ún. Hub, illetve Authority pontokat
  - Jó Authority pont, amibe sok link mutat
  - Jó Hub az, amiből sok link megy jó Authority pont felé



[2] Kleiberg, *Journal of the ACM* ,1999

HITS algoritmus

Input G irányított gráf
Output a pontok hub és authority értékei

Kezdetben minden pont értéke 1
repeat
  for all hub $i \in H$ **do**
    $h_i = \sum_{j \in F(i)} a_j$ $\{F(i)$: *azon pontok, melyekből megy él* i*-be*$\}$
  end for
  for all authority $i \in A$ **do**
    $a_i = \sum_{j \in B(i)} h_j$ $\{B(i)$: *azon pontok, melyekbe megy él* i*-ből*$\}$
  end for
until konvergál
Normálás

Néhány ingyenes program

Hálózat vizualizácó és elemzés

- Cytoscape (GUI)
- Gephi (GUI)
- iGraph (R, C++, Python)

Feladat:

- egy hálózat (pl. a Zachary-féle karate klub) pontjainak centralitásvizsgálata és vizualizáció.
- gondoljuk át mátrix egyenlet formában a PageRanket és HITS-et.

További olvasnivaló

- Jackson könyv 2. fejezet