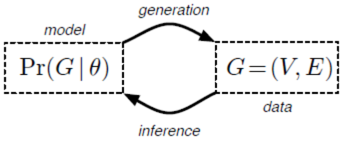
Hálózattudomány

SZTE Informatikai Intézet
Számítógépes Optimalizálás Tanszék
Előadó: London András

6. Előadás

Magas szintű strukturális mintázatok feltárása

- Hálózatokban megjelenő mintázatok modellezése és vizsgálata $\longrightarrow$ véletlen generatív modellek
- $\mathbb{P}(G|\theta)$ – θ kódolja a mintázatot, $\mathbb{P}(G|\theta)$ pedig a feltételes valószínűsége, hogy G gráf tartalmazza azt
- Visszafele irány (inference): adott G valós vagy szintetikus hálózat, határozzuk meg a legvalószínűbb θ-t, amivel ez a hálózat generálódik



ábra

Miért jó?

- Explicit generálhatunk adott hálózatokat (nem egy algoritmus)
- Struktúrával kapcsolatos hipotézisek ellenőrzése
- Modellek „jóságának” ellenőrzése (mennyire közelíti a modell a valóságot?)
- Hiányzó minták vagy jövőbeli struktúrák feltárása

- A 80-as években jelent meg szociológia témájú folyóiratban: Holland, Laskey, and Leinhardt, "Stochastic blockmodels: First steps." *Social Networks*, 5(2), 109–137 (1983)
- Gyakran használják: gépi tanulás, komplex rendszerek vizsgálata, statisztikus fizika
- Léteznek általánosításai irányított és súlyozott gráfokra is

Modell definíció

Az SBM egyszerűen egy $\theta = (k, z, M)$ hármas, ahol

1. k a csoportok (közösségek/ pont osztályok) száma a hálózatban
2. z egy n hosszú vektor, ahol z_i megadja, hogy az i pont melyik csoportba tartozik
3. M egy $k \times k$ blokk mátrix, ahol M_{uv} megadja annak a valószínűségét, hogy egy u csoportbeli és egy v csoportbeli pont kapcsolódik egymáshoz

Megj: először k-t kell fixálni, továbbá az azonos csoportban lévő pontok sztochasztikusan ekvivalensek

Hálózat generálása SBM-mel

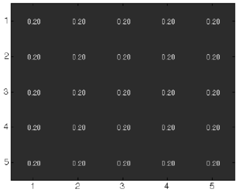
1. Adott (k, z, M) hármas
2. Minden i, j pontpárra dobunk egy érmét: M_{z_i,z_j} valószínűséggel behúzzuk (i, j) élt, $1 - M_{z_i,z_j}$-vel nem

Szemben a $G(n, p)$-vel aminek két paramétere van, és $G(n, \vec{k})$-val aminek $1 + n$, az SBM-nek $1 + n + \binom{k}{2}$ paramétere van $\Longrightarrow$ Lehetőséget ad rengeteg különféle hálózat generálására

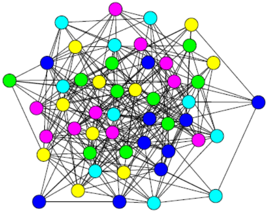
A véletlen gráf

A $G(n, p)$ az SBM egy speciális esete :

- $k = 1$, azaz egy csoport van
- $M \equiv p$
- $M_{z_i, z_j} = p$, mivel $z_i = z_j$ minden pontpárra



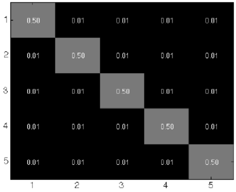
random graph block matrix

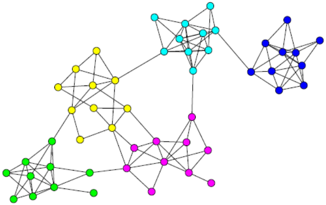


random graph

ábra. A blokkmátrix és a realizált gráf (Forrás : Aaron Clauset, Network Analysis and Modelling course)

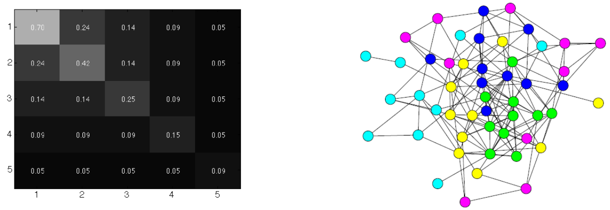
Hálózat közösségszerkezettel





ábra. Gráf közösségszerkezettel (Forrás: Aaron Clauset, Network Analysis and Modelling course)

Mag-periféria szerkezet



core-periphery block matrix   core-periphery network

ábra. Gráf mag-periféria szerkezettel (Forrás: Aaron Clauset, Network Analysis and Modelling course)

Maximum likelihood

Adott egy (valós) hálózat – mi az a blokkmodell, ami legjobban közelíti ezt a struktúrát?

Likelihood:

$$\mathcal{L}(G|M,z) = \prod_{(i,j)\in E} M_{z_i,z_j} \prod_{(i,j)\notin E} 1 - M_{z_i,z_j}$$

$\Longrightarrow$ Válasszuk M-et és z-t úgy, hogy ez maximális legyen

Egy példa

Legyen

- N_u – pontok száma az u csoportban (blokkban)
- $N_{uv} = N_u N_v$ – lehetséges élek száma u és v között
- E_{uv} – élek száma (megfigyelés) u és v között $\Longrightarrow$ becslésünk az élvalószínűségre: $\hat{M}_{uv} = E_{uv}/N_{uv}$

A likelihood függvényünk így

$$\mathcal{L}(G|M,z) \prod_{u,v} M_{uv}^{E_{uv}} (1 - M_{uv})^{N_{uv}-E_{uv}} =$$

$$= \prod_{u,v} \left(\frac{E_{uv}}{N_{uv}}\right)^{E_{uv}} \left(1 - \frac{E_{uv}}{N_{uv}}\right)^{N_{uv}-E_{uv}}$$

Vegyük a logaritmusát, fejtsük ki és a kapott kifejezést – ami csak z-től függ – maximalizáljuk