

# Jövő Internet közösségi alkalmazások elméleti kérdései

Ispány Márton

Debreceni Egyetem, Informatikai Kar

Információ Technológiai Tanszék

2016. április 29.



# Az előadás vázlata

- \* Városiasodás+infokommunikációs eszközök elterjedése: intelligens városok létrejötte és az urban computing kialakulása
- \* Az intelligens város (Smart City) definíciója
- \* Urban Computing
- \* Alkalmazott technológiák: közösségi érzékelés, közösségi adatgyűjtés, ezek információ technológiai háttere, analitikai modellek

# Bevezetés

- \* Az urbanizáció gyorsuló folyamata egyre több nagyvárost hozott létre, mely modernizálta sok ember életét, viszont számos kihívással is szembesített.
- \* Bár ezek a kihívások korábban megoldhatatlannak tűntek, napjaink szenzor technológiái révén a városi terekben keletkező nagytömegű adat (big data) a nagy léptékű számítási infrastruktúrával együtt lehetővé válik ezek hatékony kezelése.
- \* **Eredmény: élhetőbb, zöldebb, fenntarthatóbb, okosabb városok kialakulása.**
- \* **Multidiszciplináris terület: az informatika találkozik a hagyományos város tudományokkal (város-mérnökség, közlekedéstudomány, közgazdaságtan, környezettudomány, ökológia, szociológia).**

# Motivációk és célok

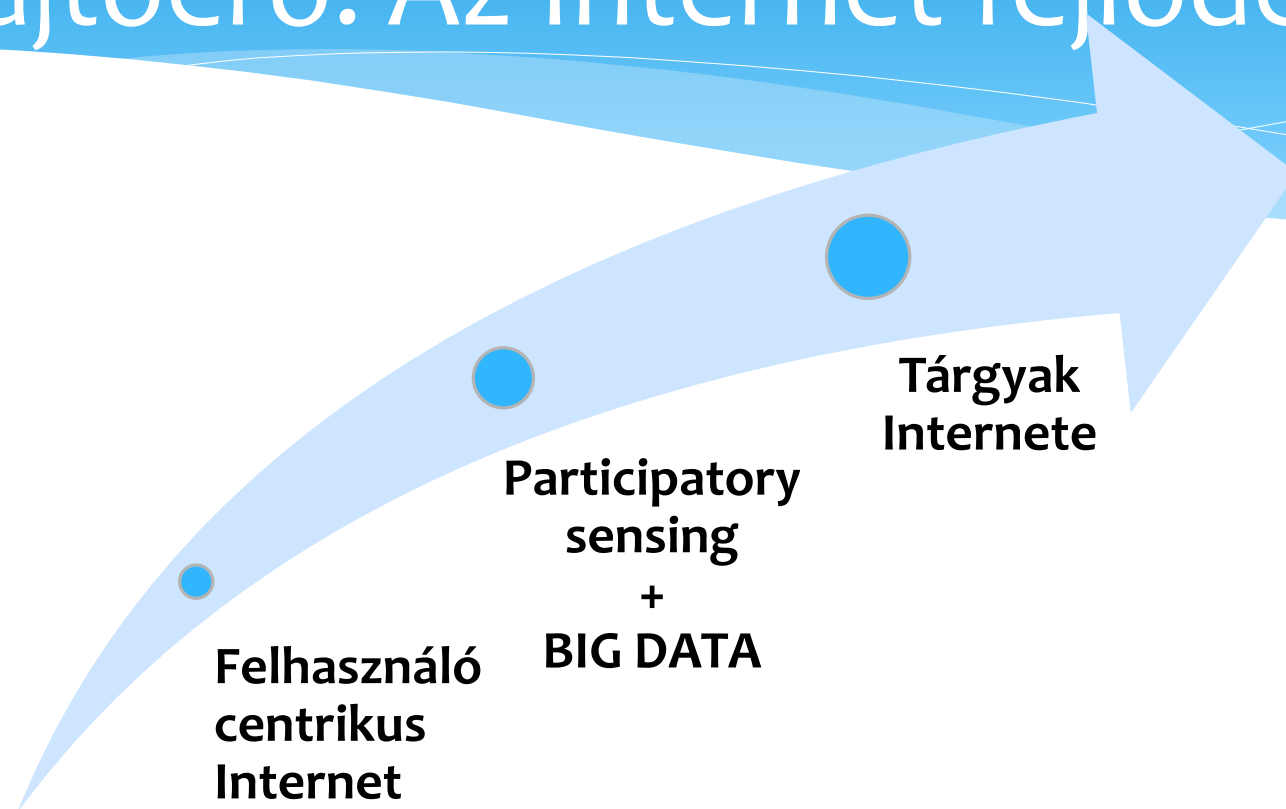


(a) Motivation: Big cities, data and challenges



(b) Goal of urban computing

# Hajtóerő: Az Internet fejlődése



# Smart Cities Manifesto

- \* <http://humansmartcities.eu/join-our-network/manifesto/>
- \* **Követelmények**
- \* **Összekapcsolható (connectable)**  
egymással kapcsolatban lévő eszközök között megosztható az információ
- \* **Elérhető (accessible)**  
minden jelenjen meg a weben
- \* **Mindenütt jelenlevő (ubiquitous)**  
megjelenés weben, mobil és más okos eszközökön
- \* **Közösségi (sociable)**  
közzététel a közösségi hálón
- \* **Megosztható (sharable)**  
az objektumok elérhetőek és címezhetőek
- \* **Látható/kitejesztett (visible/augmented)**  
láthatóvá teszi a rejtett információt

# Az intelligens város keretrendszer

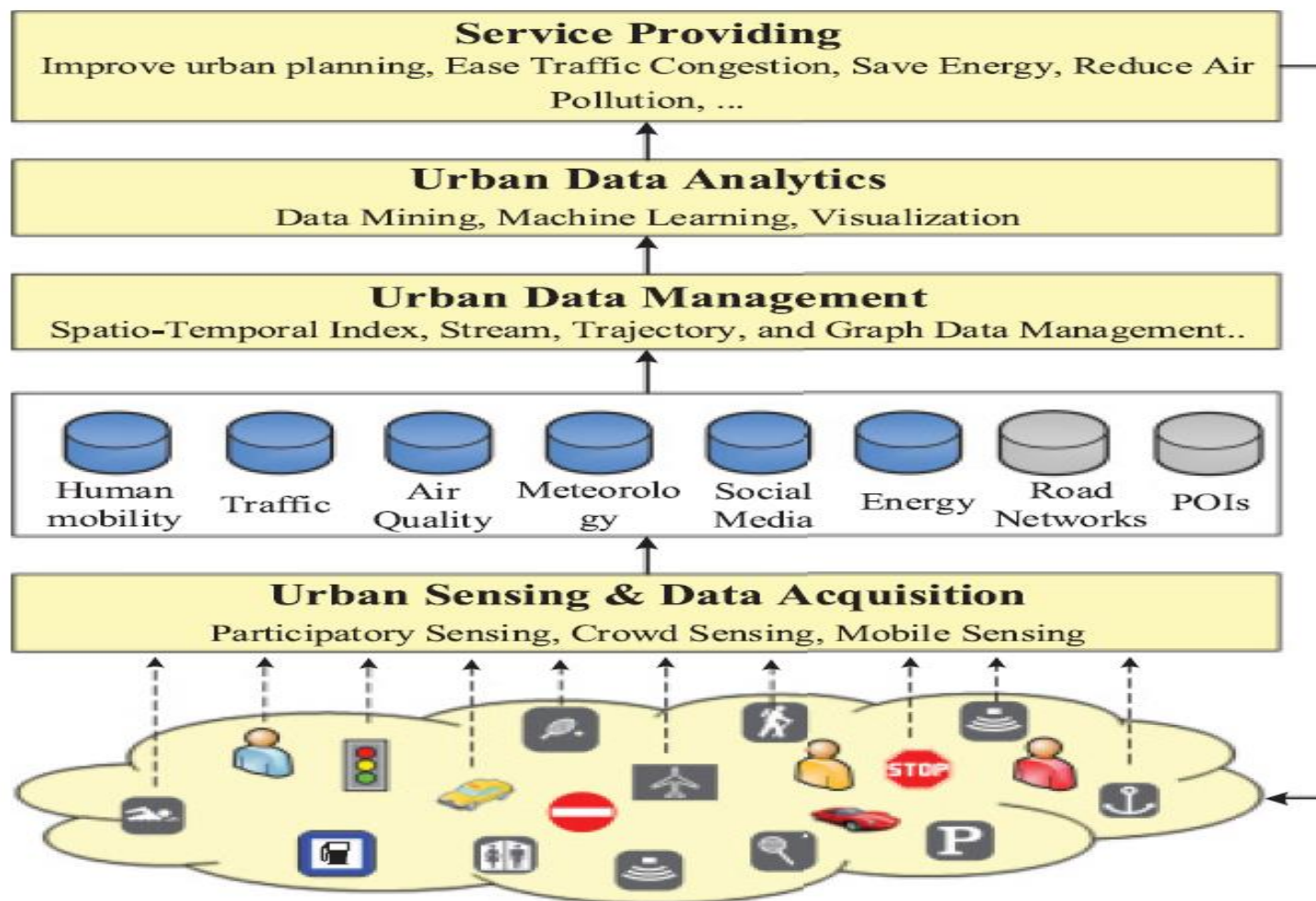
- \* **Urban computing:** az a folyamat, mely során nagytömegű heterogén adatot gyűjtünk össze, egyesítünk, majd elemzünk városi terekben elhelyezett különböző adatforrások révén, hogy a városban keletkező problémákat megoldjuk.
- \* **Adatforrások:** eszközök, szenzorok, járművek, épületek, emberek
- \* **Segít a városi jelenségek megértésében és a városok jövőjének előrejelzésében.**

# Az intelligens város keretrendszer

- \* Városi érzékelés (urban sensing)
- \* Városi adatmenedzsment (urban data management)
- \* Adatelemzés
- \* Szolgáltatás nyújtás



# Az intelligens város keretrendszer



# Kihívások

- \* Városi érzékelés és adatgyűjtés: nemzavaró, állandó gyűjtése az adatoknak városi szinten. Energia-takarékosság, titoktartás. Lazán kontrollált és nem egyenletesen elosztott szenzorok. Nemstruktúrált, implicit, zajos adatok.
- \* Heterogén adatok feldolgozása. Többszörösen megerősített tudás kinyerése. Hatékony és eredményes tanulási képesség. Vizualizáció.
- \* Hibrid rendszerek: a fizikai és a virtuális világ keverése.

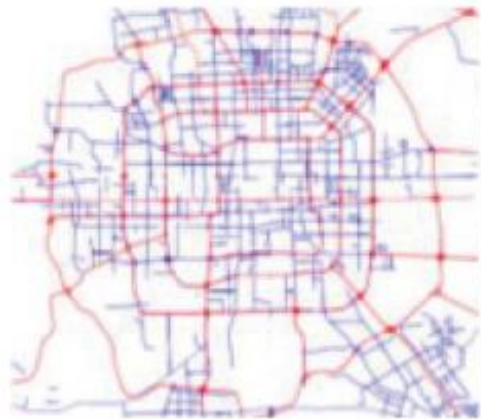
# Városi adatok

- \* Földrajzi adatok. POI- points of interest
- \* Közlekedési adatok. Loop sensors, surveillance cameras, floating cars
- \* Mobiltelefon adatok. CDR – call detail record
- \* Elektronikus jegy és kártya adatok
- \* Környezeti adatok. Különféle kémiai anyagok koncentrációja
- \* Közösségi háló adatok. A felhasználók viselkedése/érdeklődése
- \* Gazdasági adatok. Bankkártya tranzakciók, ingatlan árak
- \* Energia. Járművek fogyasztása
- \* Egészségügy. Betegek száma egy régióban

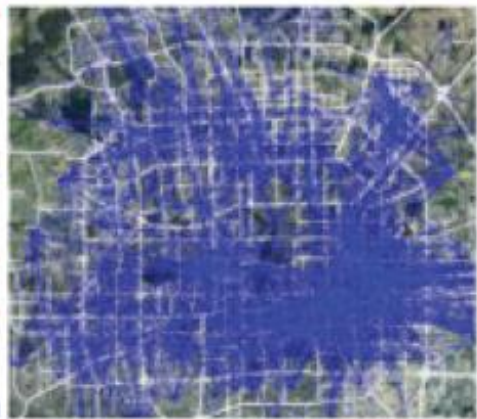
# Intelligens város alkalmazások

- \* Várostervezés. A közlekedési problémák feltárása a taxi trajektóriák vizsgálatával. A város funkcionális régióinak vagy határának a meghatározása.
- \* Közlekedési rendszerek. Vezetési gyakorlat fejlesztése (Vtrack, T-Drive). Taxi szolgáltatás, tömegközlekedés javítása.
- \* Környezet. Légszennyezés, zajszennyezés
- \* Energiafogyasztás (üzemanyag, elektromosság)
- \* Közösségi alkalmazások. Hasonló felhasználók, helyi szakértők, helyszín, esemény ajánlás, a városi élet megértése
- \* Gazdaság. Hol tudunk hasznot termelni?
- \* Biztonság. Közlekedési rendellenességek keresése. Katasztrófa érzékelés, evakuálás

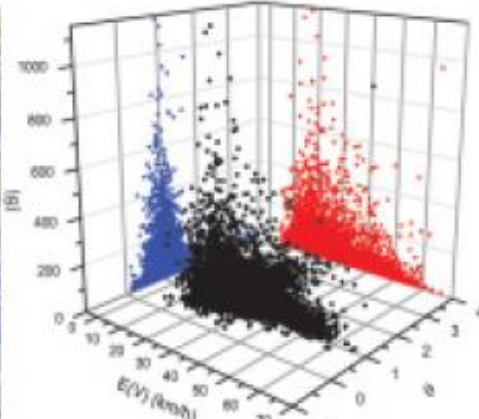
# Közlekedési problémák feltárása



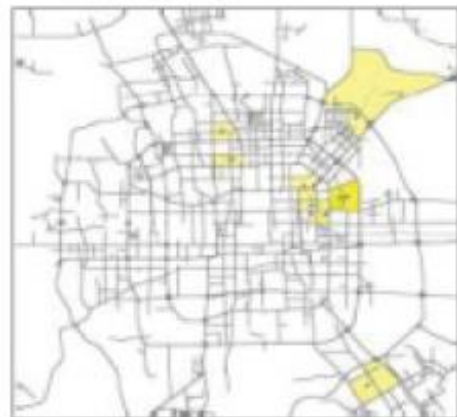
(a) Map segmentation



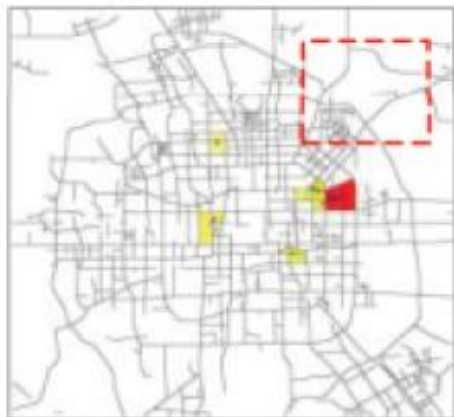
(b) Region graph



(c) Edges in the feature space



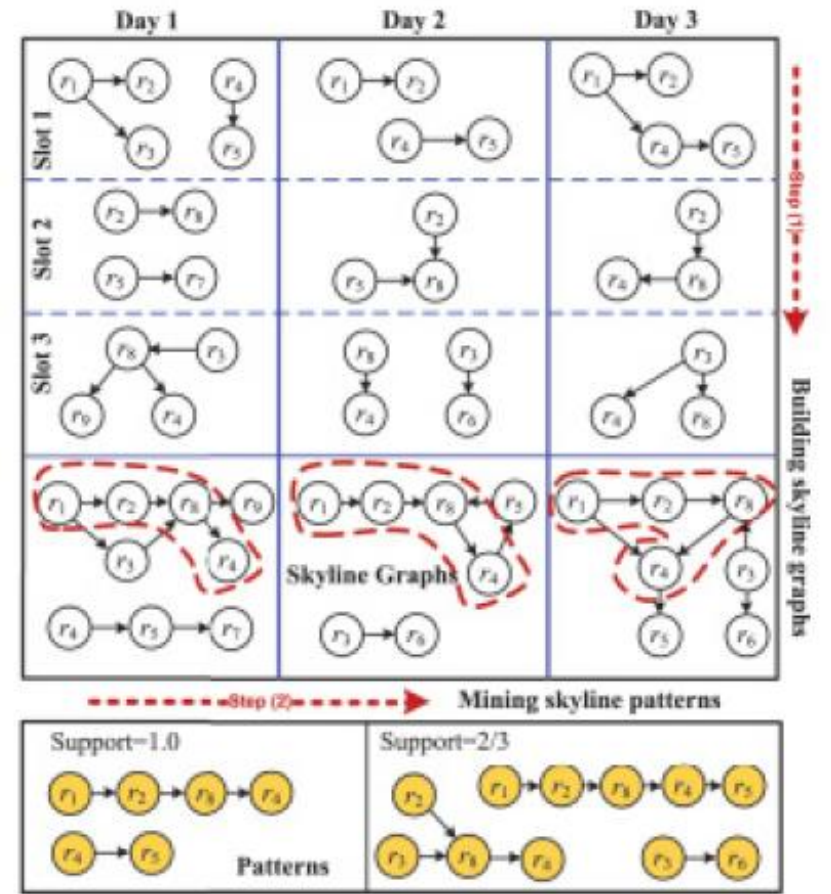
(d) Results of 2010



(e) Results of 2011



(f) A subway line launched

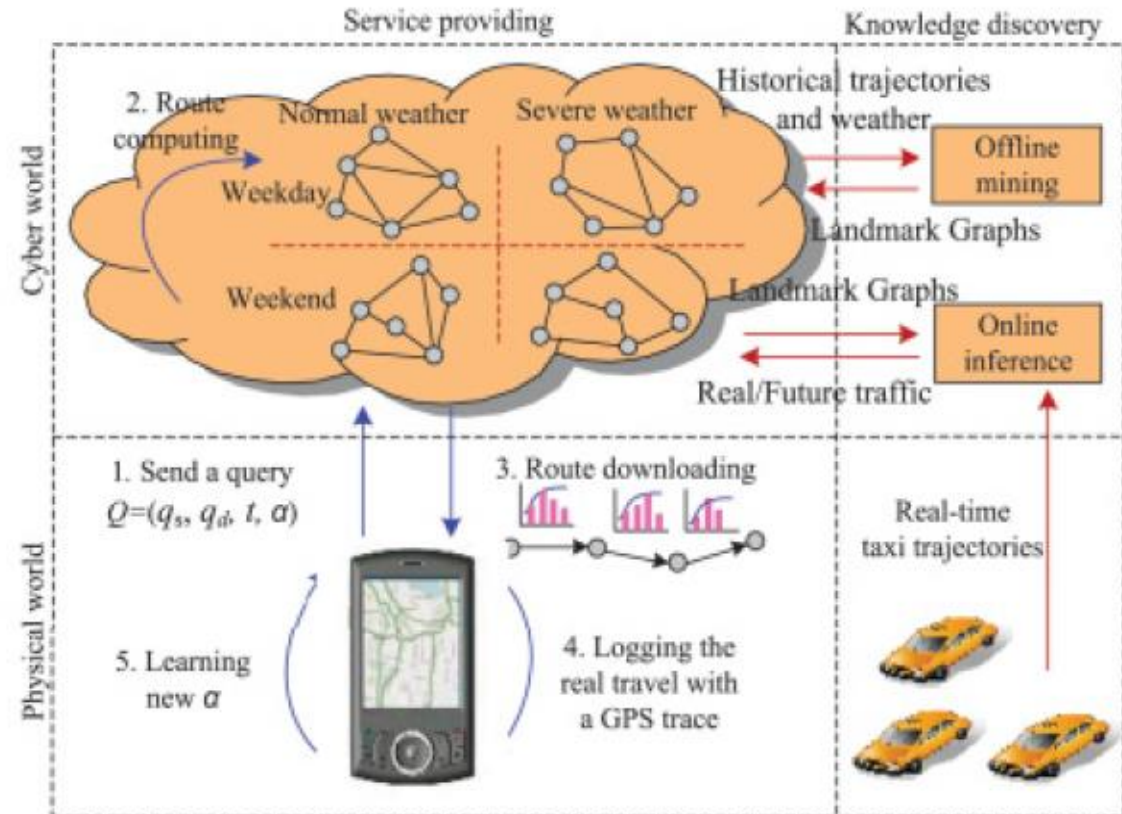


(g) Find problematic designs via graph pattern mining

# T-Drive: navigáció taxik alapján



(a) Landmark graph of Beijing (k=4000)



(b) Framework of T-Drive system

# Környezetvédelem

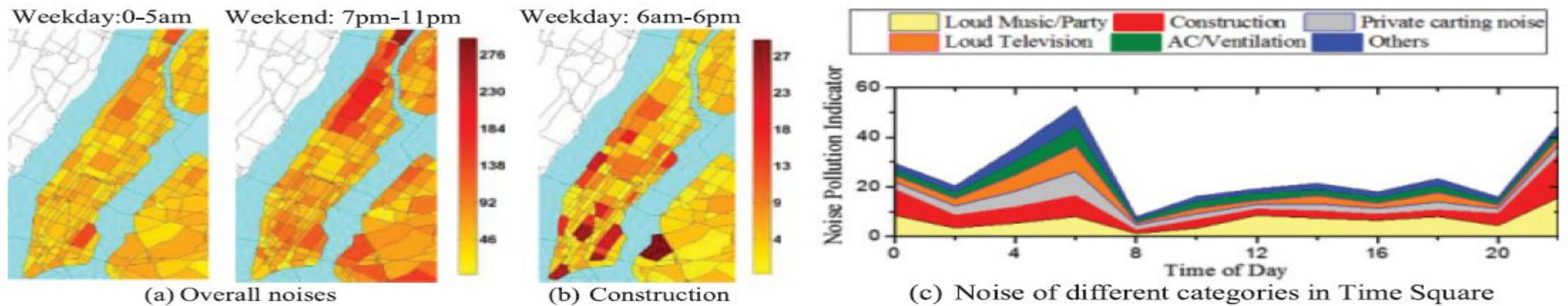


Fig. 9. Diagnosing the noise pollution of New York City.

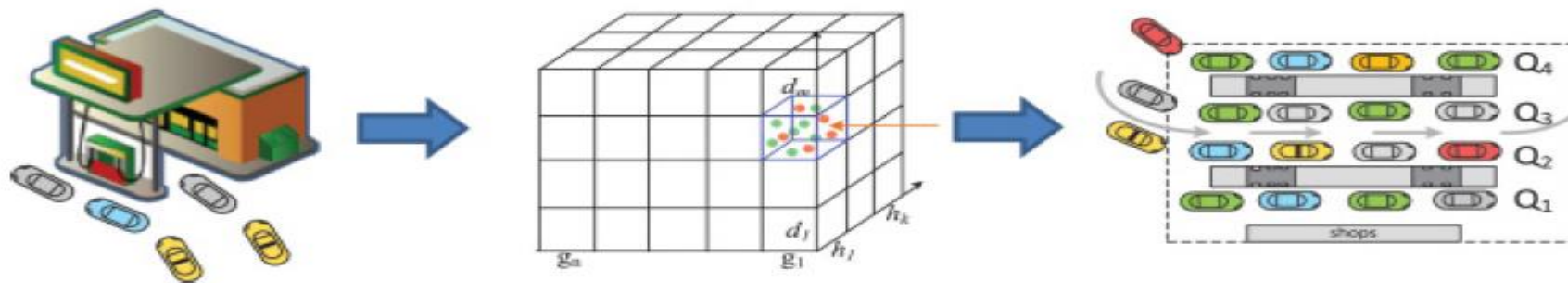


Fig. 10. Crowdsensing urban refueling behavior with GPS-equipped taxis.

# Alkalmazott módszerek

- \* Városi érzékelés (urban sensing)
- \* Városi adatmenedzsment (urban data management)
- \* Tudás-fúzió heterogén adatokon
- \* Ritka adatok kezelése
- \* Városi adatvizualizáció



# Városi érzékelés és adatgyűjtés

- \* Az érzékelő és adatgyűjtő technológiák fejlődésének eredménye nagy tömegű adat a közlekedéstől a levegő minőségig, a közösségi mediától a földrajzi adatokig.
- \* Az adatgyűjtés módszerei:
- \* Hagyományos érzékelés és mérés
- \* Passzív crowd-sensing
- \* Participatory sensing
- \* A különbség az utóbbi kettő között az adatgyűjtésben való részvétel módjában van, a participatory sensingnél az emberek tisztában vannak az adatgyűjtés céljával és tudatában vannak hozzájárulásuk mértékével.

# Passzív crowd-sensing

- \* Magas szintű infrastruktúrák, pl. tömegközlekedési jegyrendszerek vagy mobil hálózatok, a modern élet biztosítására.
- \* Az eredeti célokon túl használhatóak-e ezek a rendszerek a város, közösség stb. viselkedésének vizsgálatára. Cél: segítség a tervezésben (pl. a városok működésének javítása), a felmerülő problémák (pl. a forgalmi torlódások kezelése) megoldása.

# Passzív crowd-sensing: példák

- \* Érzékelés GPS-sel ellátott járműveken keresztül
- \* Tömegközlekedési járművek (busz, taxi) rendelkeznek GPS-sel, pl. biztonsági, szolgáltatás szervezési okok miatt.
- \* Magánszemélygépjárművek kaphatnak GPS-t pl. biztosítási konstrukciók miatt.
- \* Az így gyűjtött GPS koordináták használhatóak a közlekedés nyomon követésére, emberi mobilitási mintázatok feltárására.
- \* Példák: taxi trajektóriák felhasználása okos szolgáltatás tervezésre, valós idejű légszennyeződés mérés, forgalmi anomáliák előrejelzése, várakozási idő becslése buszközlekedésnél, környezet-tudatos vezetés kialakítása.

# Passzív crowd-sensing: példák

- \* Adatgyűjtés tömegközlekedési jegyrendszerrel: RFID alapú kártyák használata a tömegközlekedésben, metró, busz.
- \* Egy utazás egy tranzakciós rekordot indukál az alábbi adatokkal: ár, időbélyeg, hely (állomás, megálló), jármű-ID
- \* Megfelelő feldolgozás esetén a kártyatulajdonos teljes mozgása feltérképezhető, a város-szintű humán mobilitás vizsgálható és ez alapján a közlekedés tervezhető és optimalizálható.

# Passzív crowd-sensing: példák

- \* Adatgyűjtés mobil-hálózaton keresztül: az ilyen hálózatok, tipikusan több is, teljesen lefedik a városokat, az általuk gyűjtött rekordok (pl. a cella-adatok) egyfajta emberi lábnyomatként használhatóak. Széles körben alkalmazzák az emberi mozgás és a közlekedés modellezésében.
- \* Adatgyűjtés a közösségi hálózatot szolgáltatóktól: az ilyen szolgáltatások fejlődése nagy adattömeget (tweet, fotó, videó), generál. Ezek az információk kiegészülnek esetenként hely és további szemantikus információkkal. Felhasználható pl. riasztásra, rendellenesség detektálásra (természeti katasztrófák, balesetek). Az ilyen adatok szintén alkalmasak emberi mobilitás vizsgálatára, tervezésére (turista útvonalak tervezése).

# Participatory sensing

- \* A széles körben elterjedt, szenzorokkal ellátott, hálózatba kötött (Internet kapcsolat) kézi eszközök lehetővé teszik a polgárok aktívabb részvételét az adatgyűjtésben.
- \* Alkalmazások egy új típusát hozta létre, ahol az adatok egy jelentős részét a résztvevők szolgáltatják és a számukra kiajánlott szolgáltatás(ok) ezen adatok összesítésével adódnak.
- \* A participatory sensing két válfaja: közösségi érzékelés (human crowd-sensing), közösségi adatgyűjtés (human crowd-sourcing)

# Közösségi érzékelés (crowd-sensing)

- \* A felhasználók lehetővé teszik az eszközeikbe beágyazott szenzorok adatainak távolról való elérését egy magasabb szintű közösségi cél támogatására.
- \* Megosztott adatok: GPS koordináták, hang (mikrofon), kép (kamera), környezeti adatok mint hőmérséklet, páratartalom, CO<sub>2</sub>, NO<sub>2</sub> (speciális személyi érzékelő kitek, pl. SmartCitizen).
- \* Alkalmazások: valós idejű busz előrejelzés (Tiramisu projekt), navigáció (Waze), városi zaj térkép, környezeti térkép az életminőségről.

# Közösségi érzékelés (crowd-sensing)

- \* Tömeges szenzorok
  - \* Az okostelefonokon nagyon sok szenzor van
  - \* Az embereknél nagyon sok okostelefon van
  - \* Az okostelefon adatokat oszt meg egyszerűen

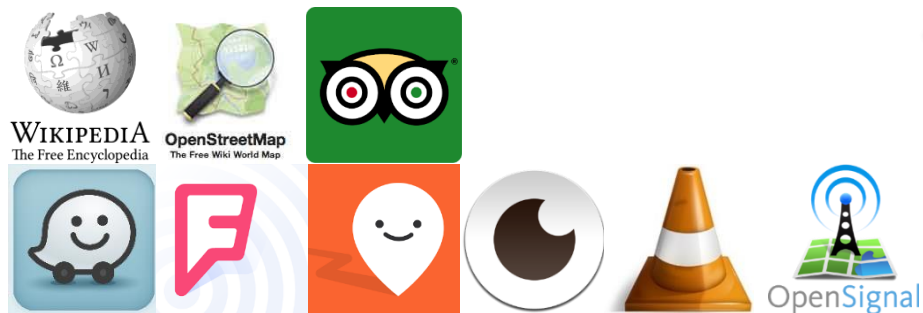
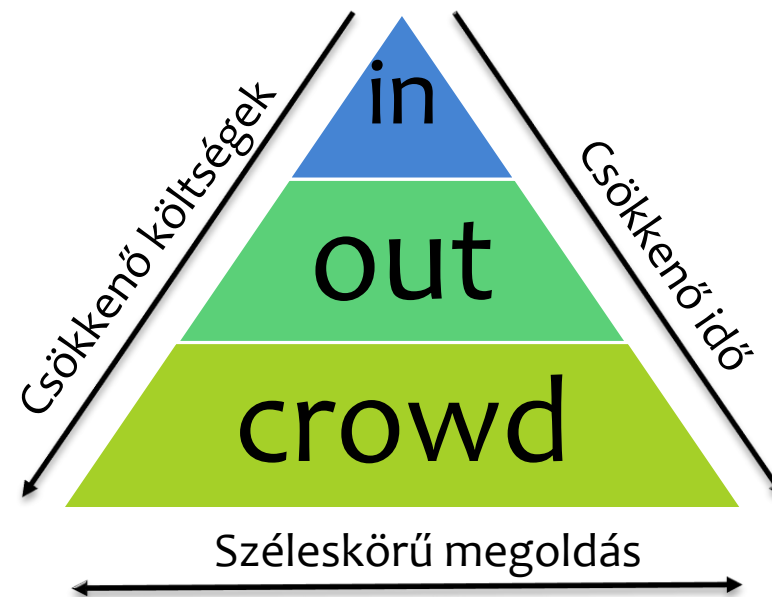
- \* A közösség bevonása az aktív adatgyűjtés folyamatába
  - \* Az ingyeneő felhasználók átalakítása aktív közreműködővé
  - \* Kevés felhasználó is sok hasznos információt hoz





# Közösségi adatgyűjtés (crowd-sourcing)

- \* Közösségi adatgyűjtés (crowd-sourcing), mint szolgáltatásnyújtás
  - \* Olcsó
  - \* Gyors
  - \* Széleskörű
  
- \* Internet platform
  - \* Sok embert elér
  - \* Gyors munkafolyamat



# Közösségi adatgyűjtés (crowd-sourcing)

*“Crowd-sourcing is a type of participative online activity in which a company or organization proposes to a group of individuals of **varying knowledge, heterogeneity, and number**, via an open call, to voluntarily undertake a task.”*



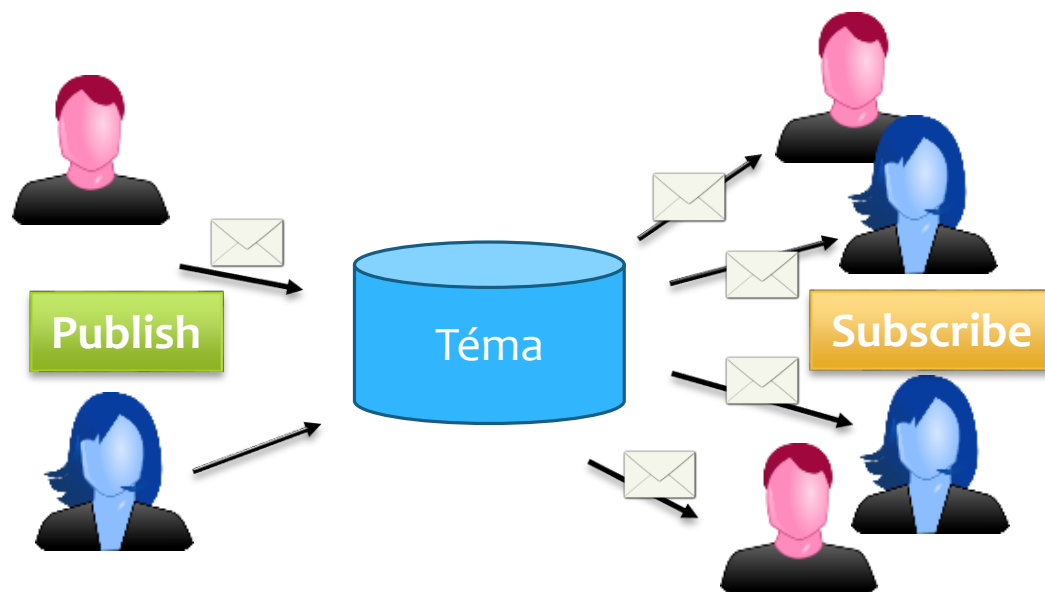
# Közösségi szolgáltatások

- \* Crowdsourcing + crowdsensing
  - \* Az adatok a felhasználóktól jönnek
  - \* A közösség ad a közösségnek
  
- \* Mindenki hasznos, nincsenek ingyenélők
  - \* Lelkes felhasználók, bonyolult feladatok
  - \* Átlagos felhasználók egyszerű feladatok



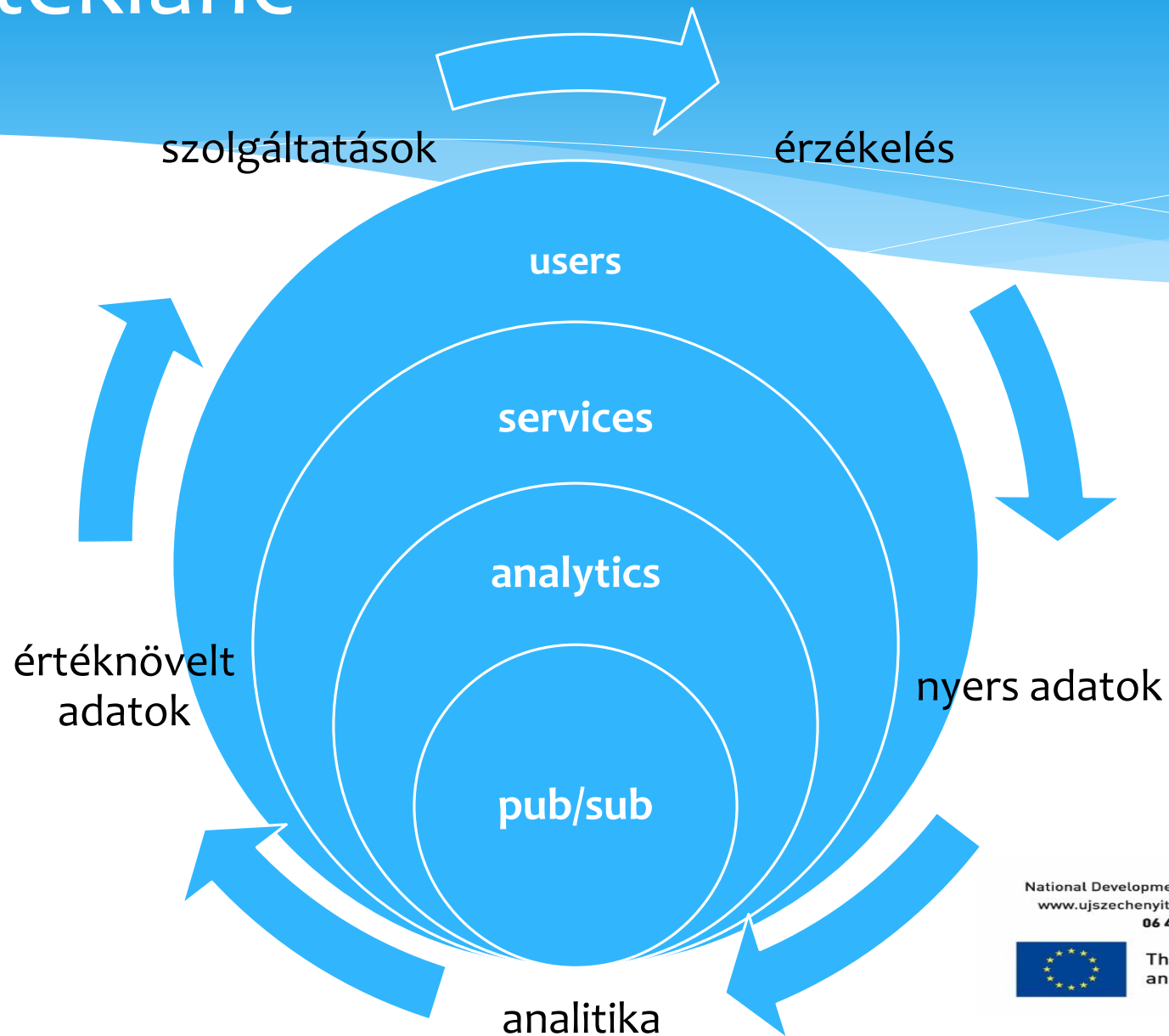
# Hatékony on-line adatmegosztás

- \* XMPP - Extensible Messaging and Presence Protocol
  - \* Publish / Subscribe modell



- \* Alkalmazás
- \* Skálázhatóság
  - \* Felhasználószám
  - \* Adatmennyiség

# Értéklánc



# Városi adatmenedzsment módszerek

- \* A városi környezetben keletkező adatok jellemzői: térbeliség, tér-időbeliség, változékonyság, heterogenitás
- \* Adatfolyam és trajektória menedzsment
- \* Gráf adatok menedzsmentje
- \* Hibrid index szerkezetek

# Adatfolyam és trajektória menedzsment

- \* Adatfolyam-kezelő rendszerek (DSMS): adatfolyamok kezelése és lekérdezése, pl. StreamInsight.
- \* Kihívás: potenciálisan végtelen adatfolyamok kezelése fix memóriában nem véletlen mintavételezéssel.
- \* Megközelítések: tömörítési és ablakolási módszerek

# Adatfolyam és trajektória menedzsment

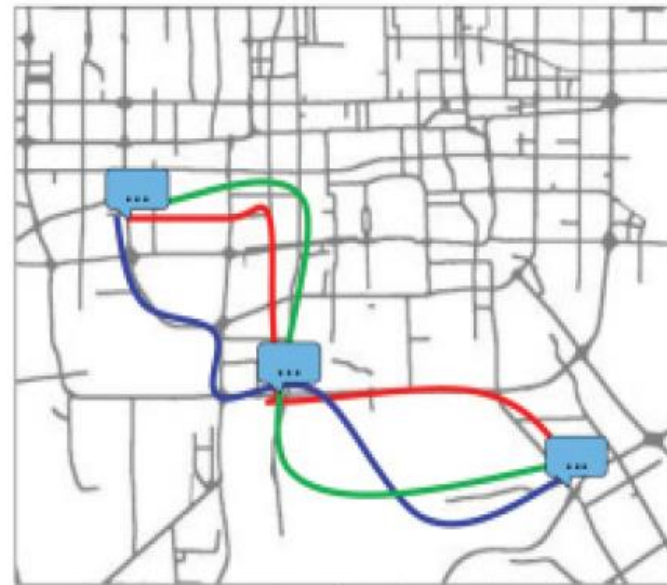
- \* Térbeli trajektória: kronológikusan rendezett térbeli pontok egy sorozata, melyet egy mozgó objektum generál
- \* Trajektória adatok redukciója: mintavételezési frekvencia vs. adattömeg, batch vs. online. PRESS (2014): tér és idő elválasztása
- \* Zajsűrés: kiugró, zajos pontok eltávolítása. Jelfeldolgozási módszerek (Kálmán, particle filter)
- \* Indexelés és lekérdezés: mozgó objektum aktuális helyzetének lekérdezése, 3DR-Tree, MR-Tree
- \* Bizonytalanság kezelése: több lehetséges trajektória, map matching
- \* Trajektória mintázat bányászat: szekvenciális mintázatok bányászata időbélyeggel nehezítve, mozgó objektumok klaszterezése



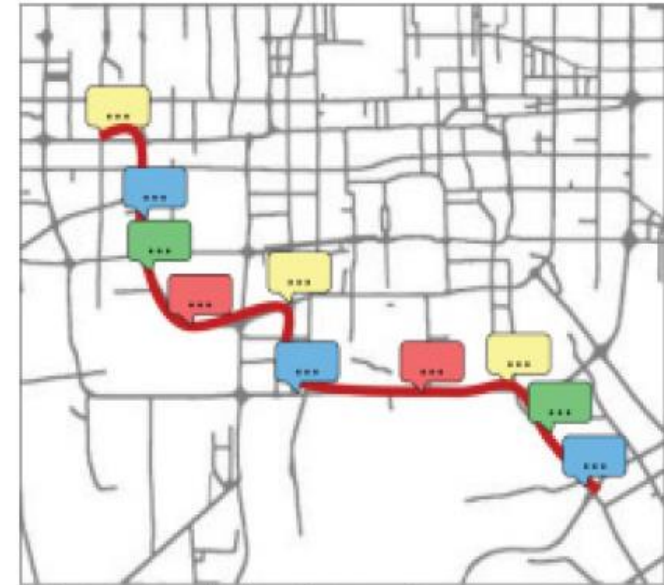
# Példák trajektóriákra



(a) Taxi and bird's traces



(b) A user's check-in data



(c) Check-in data of many users

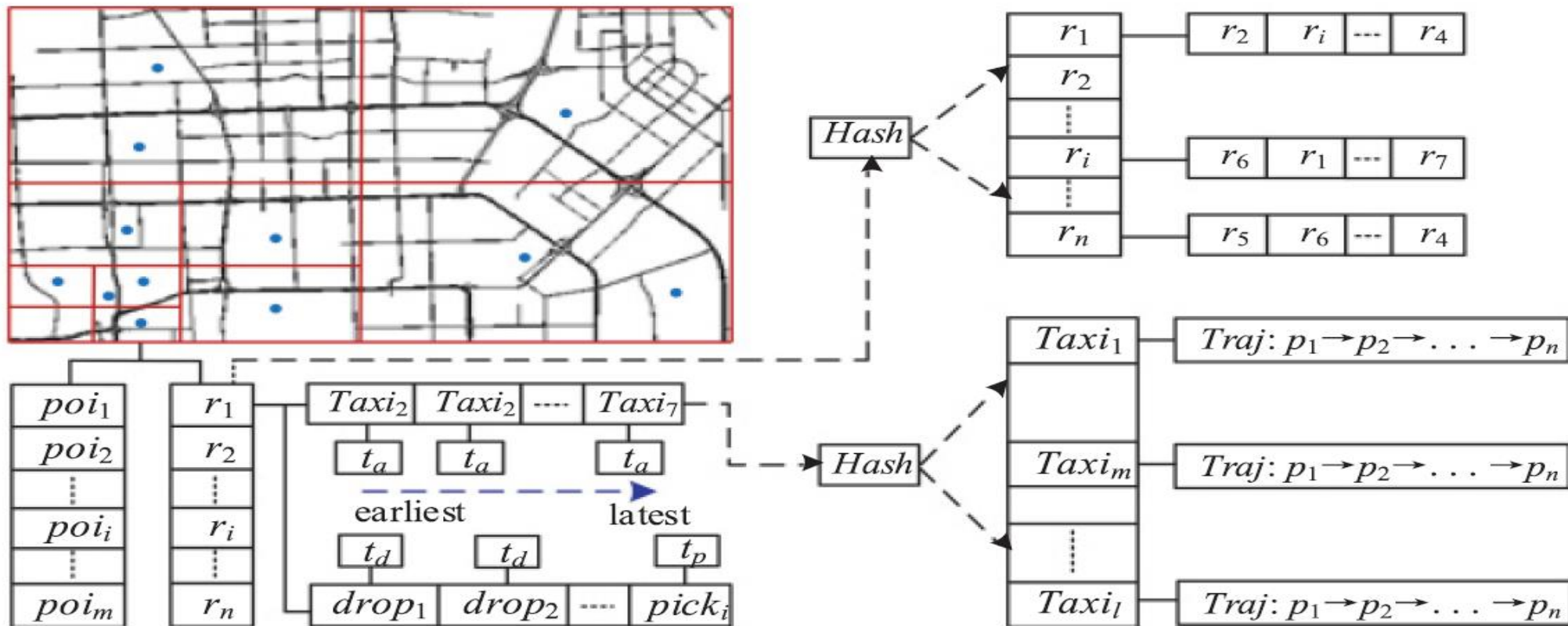
# Gráf adatok menedzsmentje

- \* Gráf típusú adat gyakran fordulnak elő intelligens város feladatokban: út- és metróhálózatok, közösségi és szenzorhálók.
- \* Statikus gráfok: térbeli gráfok
- \* Tér-időbeli gráfok (ST gráfok): az időbeli komponens figyelembevételével, pl. bizonyos utakat időközönként blokkolnak. Különbözik az időben fejlődő gráfoktól (közösségi hálók)
- \* Az ilyen gráfok adatkezelése kidolgozatlan, kutatási terület
- \* Jelenlegi kutatások: részgráf bányászat, időfüggő útvonal tervezés

# Hibrid index szerkezetek

- \* Változatos adatokat kell összefogni és integrálni egy adatbányászati modellbe
- \* Például: POI-k, úthálózatok, közlekedési és emberi mobilitási adatok együttes kezelése
- \* Térbeli indexek, hash táblák, rendezett listák és szomszédsági listák (pl. quad-tree, B-tree alapú indexek)

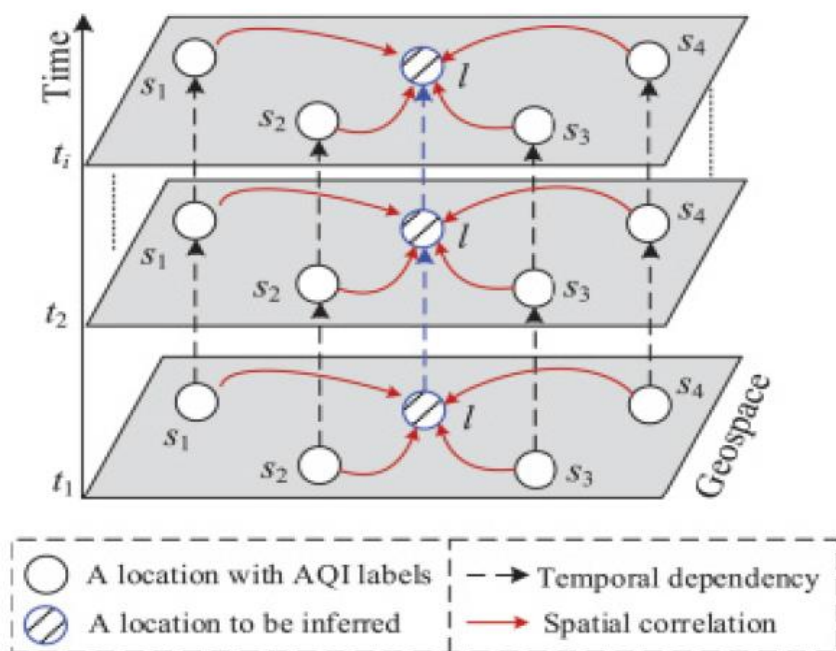
# Példa hibrid indexszerkezetre



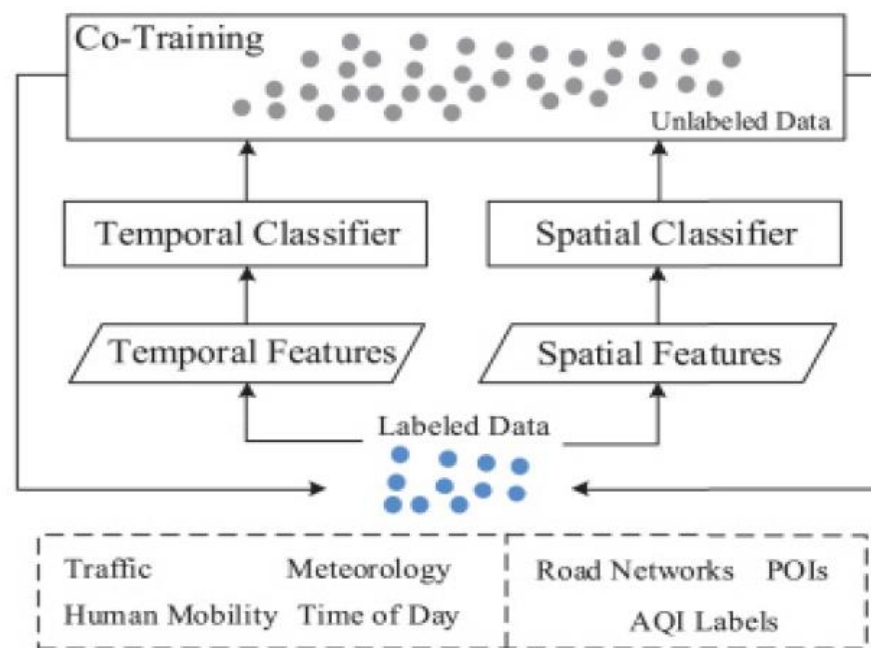
# Tudás-fúzió heterogén adatokon

- \* Olyan technológiák utáni igény, melyek hatékonyan egyesítik a több heterogén adatforrásból kinyert tudást.
- \* Fúzió jellemző szinten (normalizálás), legelterjedtebb
- \* A különböző adatok különböző szinteken való használata: a város régiókra osztása
- \* Egy közös modell szimultán meghajtása részenként különböző adatállományokkal. A legnagyobb hatásfokkal bír.

# Példa tudás-fúzióra



(a) Philosophy of the inference model

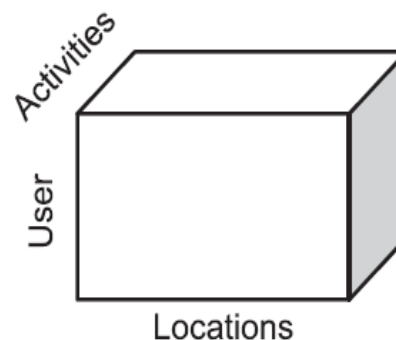


(b) Procedure of co-training

# Ritka adatok kezelése

- \* Hiányzó adatok oka: egy hely alapú közösségi szolgáltatásnál a felhasználó-helyszín mátrix ritka, a tevékenységet harmadikként behozva a kapott tenzor még ritkébb.

$$M = \begin{matrix} & l_1 & l_2 & l_3 & \dots & l_k \\ \begin{matrix} u_1 \\ u_2 \\ \dots \\ u_n \end{matrix} & \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 2 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix} \end{matrix}$$



# Ritka adatok kezelése

- \* Kollaboratív szűrés
- \* Mátrix faktorizáció
- \* Tenzorfelbontás
- \* Félig-felügyelt tanítás, transzformált tanítás



# Kollaboratív szűrés

- \* Collaborative filtering (CF): széles körben alkalmazott modell az ajánló rendszerekben (recommender system)
- \* Alapötlet: hasonló felhasználók hasonló termékeket hasonlóan rangsorolnak
- \* Szerepek az intelligens város alkalmazásokban:  
felhasználók -> sofőrök, utasok, a szolgáltatásra jelentkezők  
termékek -> POI-k (pl. éttermek), útszegmensek, régiók

# Kollaboratív szűrés

$$r_{pi} = \overline{R}_p + d \sum_{u_q \in U'} \text{sim}(u_p, u_q) \times (r_{qi} - \overline{R}_q)$$

$$d = \frac{1}{|U'|} \sum_{u_q \in U'} \text{sim}(u_p, u_q)$$

$$\overline{R}_p = \frac{1}{|S(R_p)|} \sum_{i \in S(R_p)} r_{pi},$$

# Mátrix faktorizáció

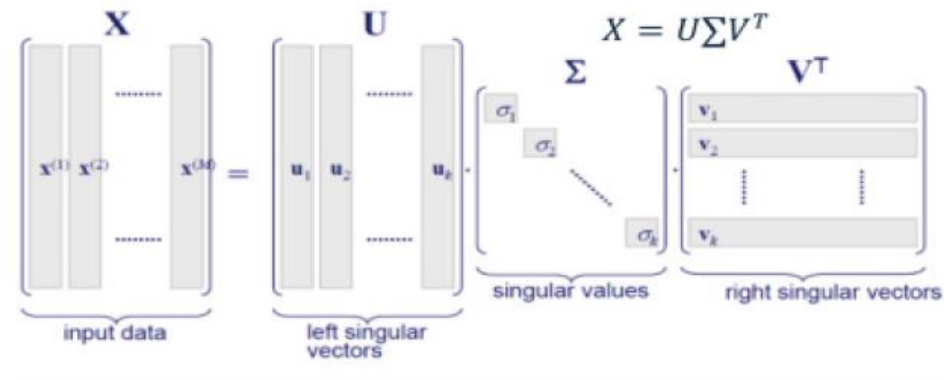


Fig. 20. SVD-based matrix factorization.

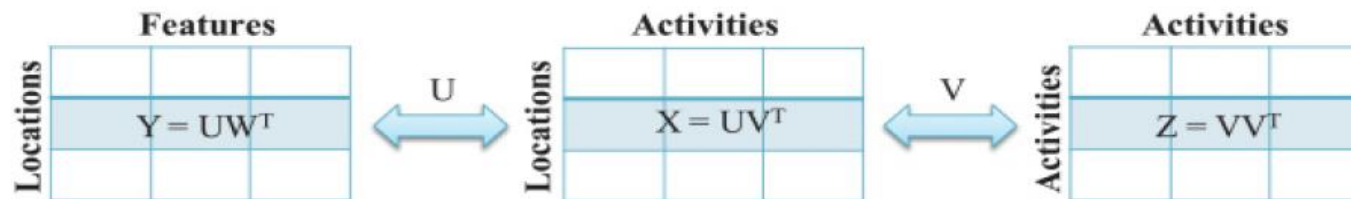
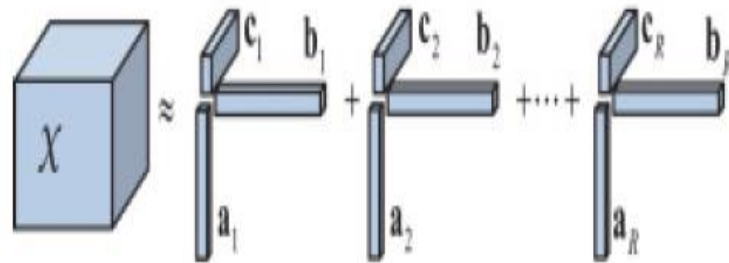
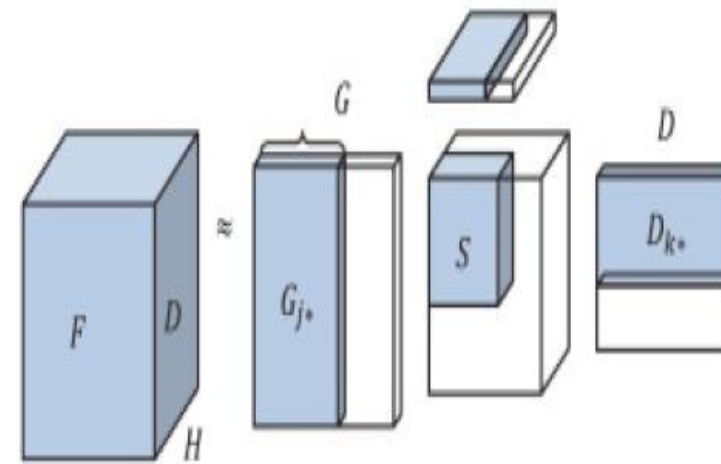


Fig. 21. Coupled matrix factorization for location–activity recommendation.

# Tenzorfelbontás

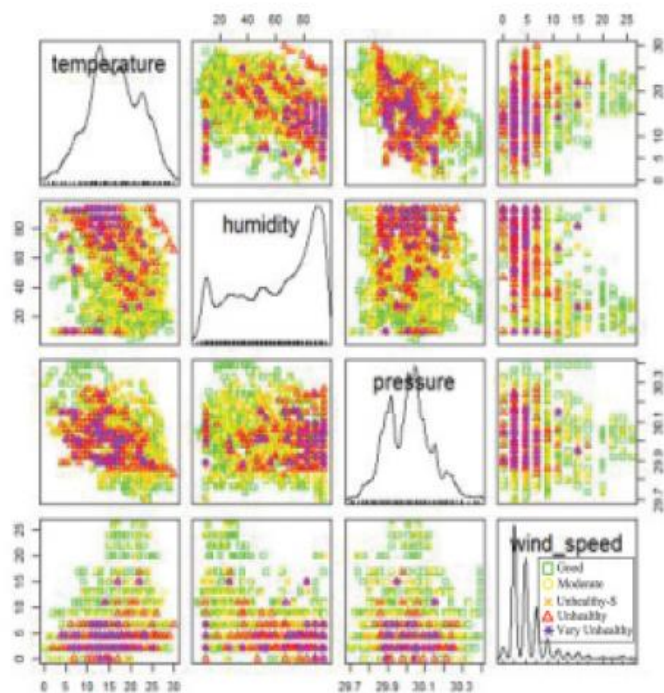


(a) PARAFAC

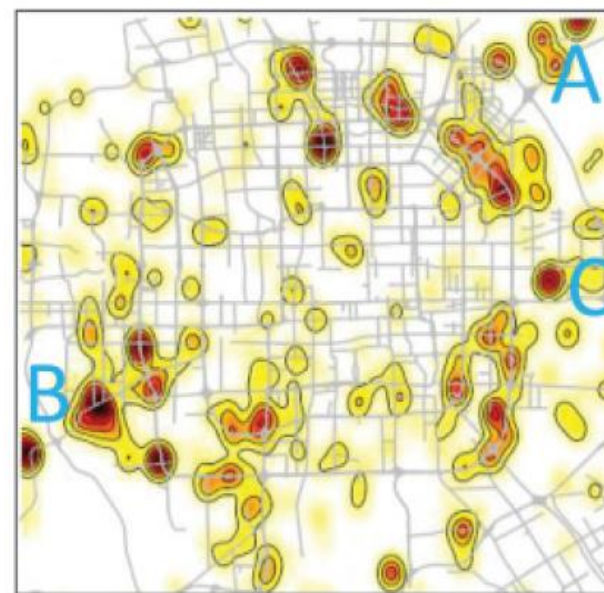


(b) Tucker decomposition

# Vizualizáció



(a) Correlation between meteorology and air quality



(b) Refueling behavior inferred from taxi data

# Jövőbeni kutatási irányok

- \* Kiegyensúlyozott közösségi érzékelés
- \* Ferde adateloszlások kezelése
- \* Multimodális adatforrások menedzselése és indexelése
- \* Tudás-fúzió
- \* Interaktív feltáró adatvizualizáció több adatforrásból
- \* Algoritmus integráció
- \* Beavatkozás alapú elemzés és előrejelzés

# Információforrások

- \* Konferenciák, workshopok: KDD, ICDE, UbiComp, ACM International Workshop on Urban Computing (UrbComp)
- \* Folyóiratok: IEEE Transaction on Knowledge Discovery and Data Engineering,  
ACM Transaction on Intelligent Systems and Technology,  
Personal and Ubiquitous Computing,  
IEEE Pervasive Computing
- \* Adatok: városok open data portáljai, pl. <https://nycopendata.socrata.com/>

# Köszönöm a figyelmet

Irodalom

Yu Zheng, Licia Capra, Ouri Wolfson, Hai Yang, Urban Computing: Concepts, Methodologies, and Applications, ACM Transactions on Intelligent Systems and Technology (TIST) - Special Section on Urban Computing Volume 5 Issue 3, September 2014, Article No. 38