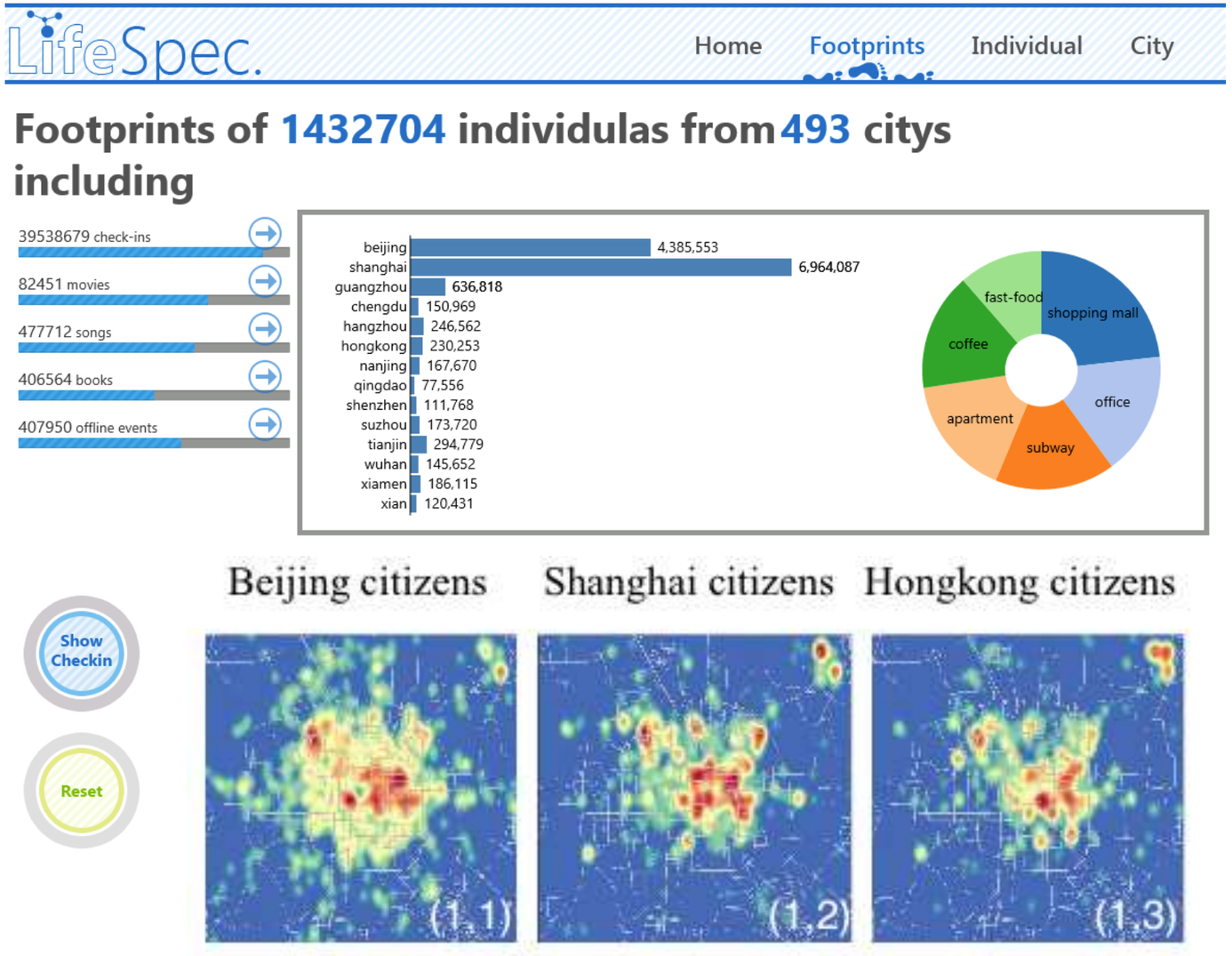
LifeSpec.

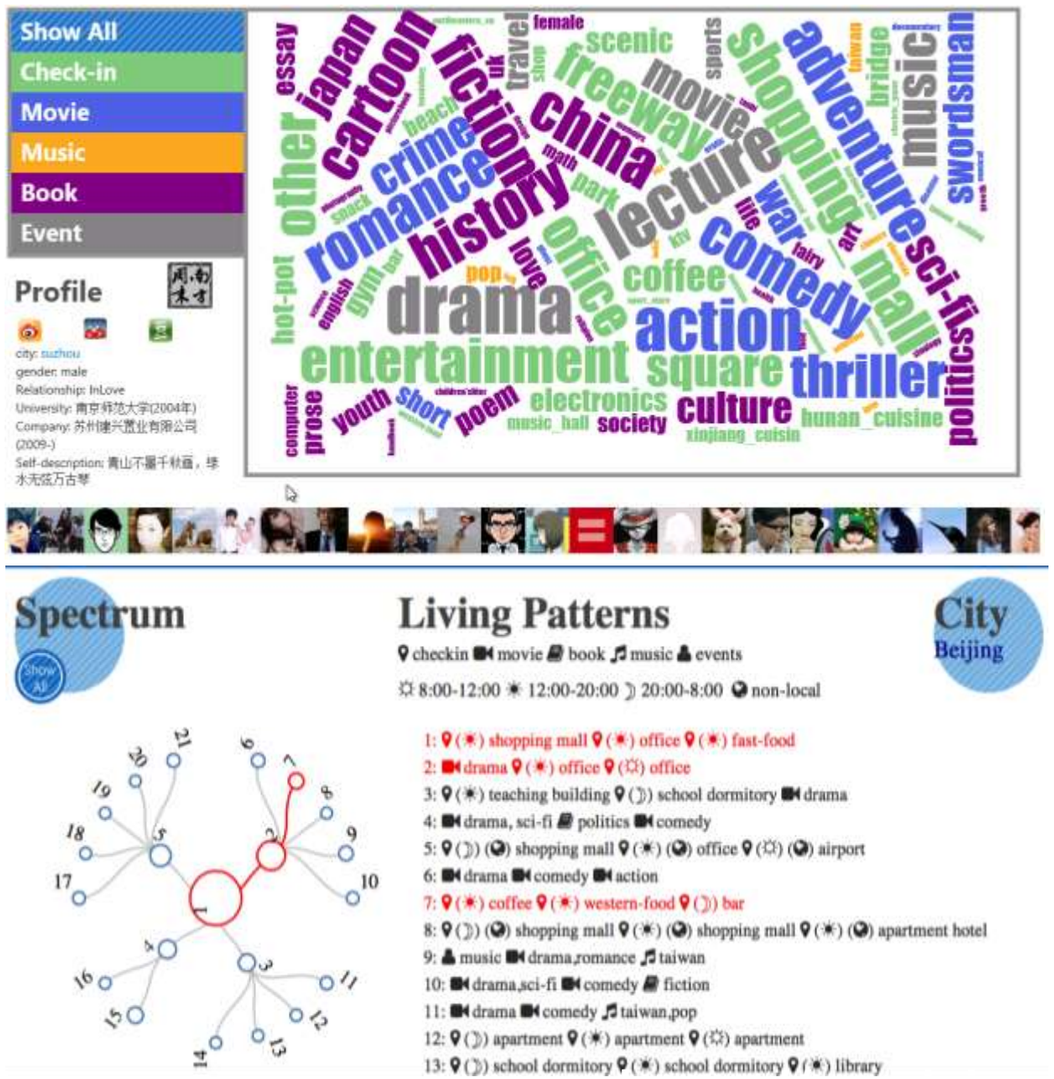
Exploring the Spectrum of Urban Lifestyles

Nicholas Yuan[1], Fuzheng Zhang[1], Defu Lian[1], Kevin Zheng[2], Siyu Yu[3], and **Xing Xie**[1]

[1]Microsoft Research Asia

[2]The University of Queensland, [3]University of California, Berkeley

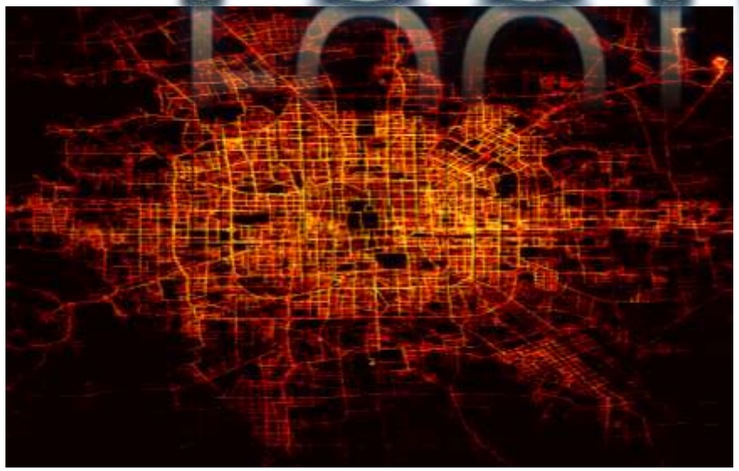


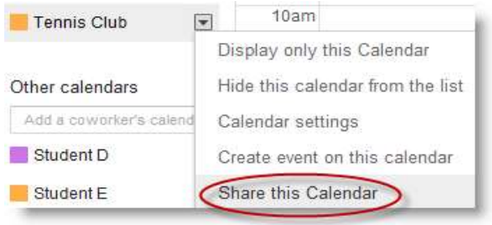


- Lifestyle specification of an individual
- Lifestyle spectrum of a group of individuals

LifeSpec.

FOOTPRINT





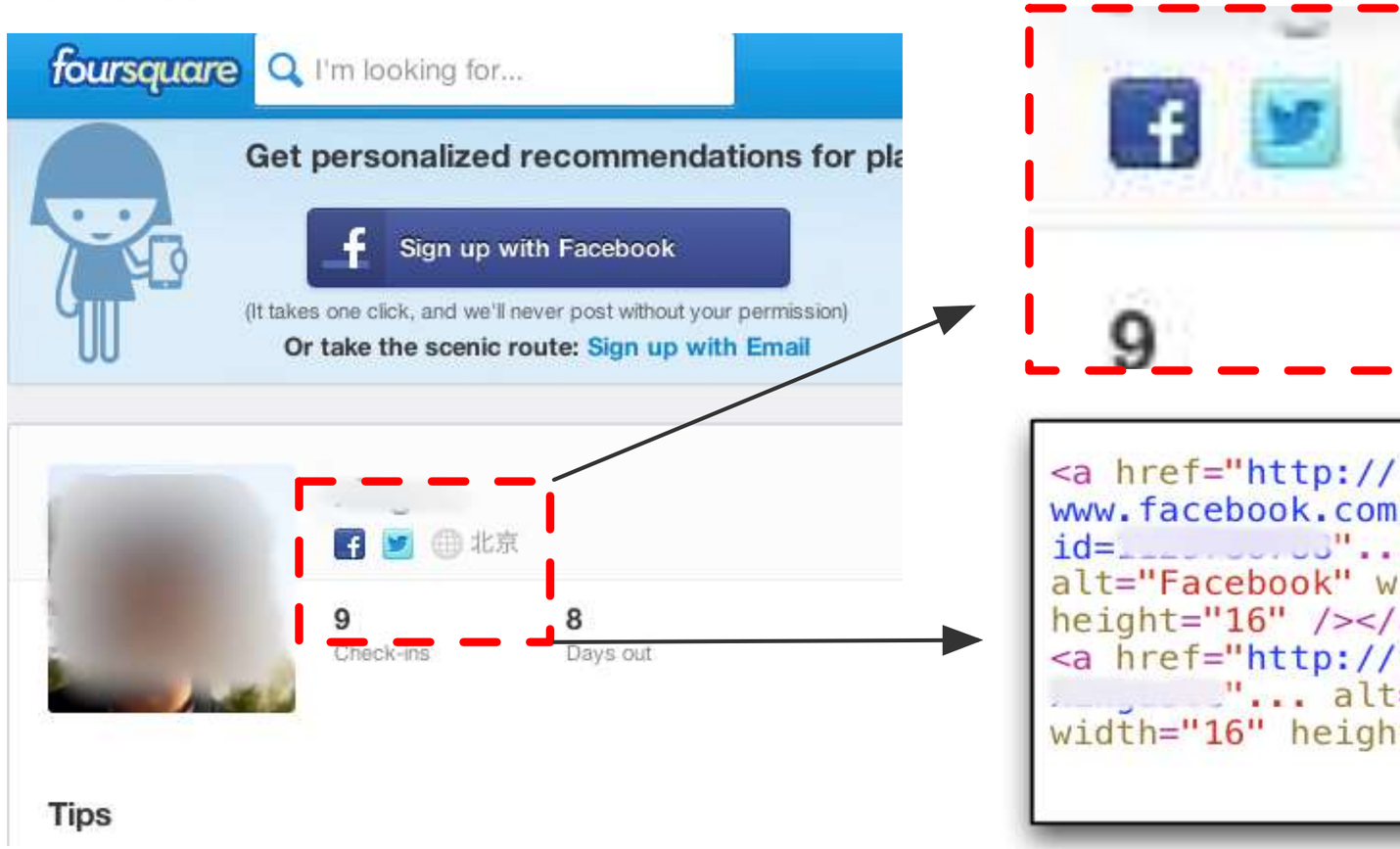
A Day in Your Life

Heterogeneous Online Networks

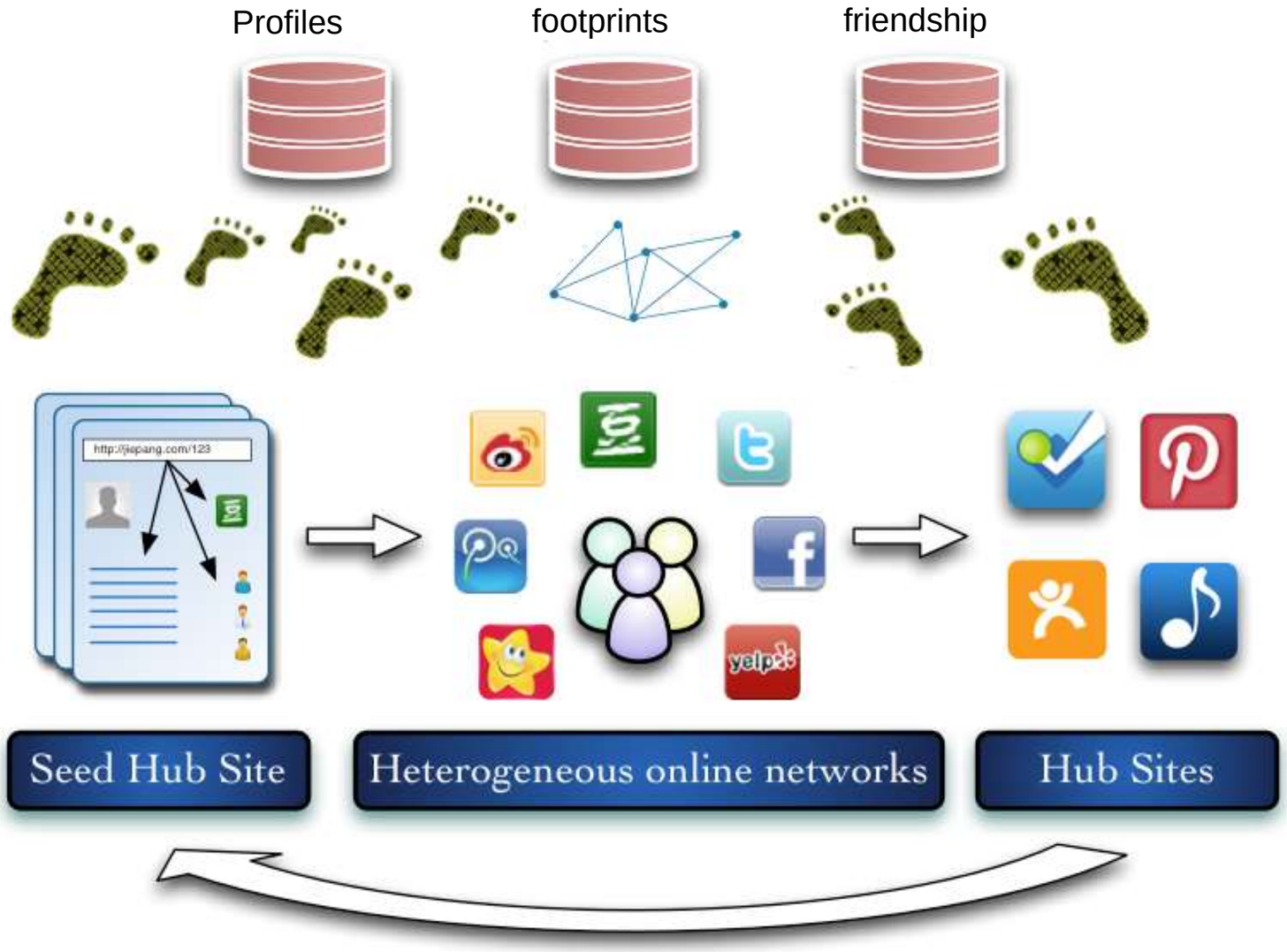
Self-Disclosure

Cross-Domain Posting

Hub Sites



ICONNECT Algorithm



Algorithm 1: IdentifyConnectedUsers (ICONNECT)

Input: seed hub site h
Output: users U, where each user $u \in U$ is a set of profile pages; footprints $\mathcal{F}$; social links $\mathcal{L}$, hub sites H

$H \leftarrow \varnothing$; H.add(h);
$\mathcal{F} \leftarrow \varnothing$; $\mathcal{L} \leftarrow \varnothing$;
A queue storing unvisited profile pages $P_U \leftarrow h$.getAllUsers();
A set of visited profile pages $P_V \leftarrow \varnothing$;
while $P_U.length > 0$ **do**
  $p \leftarrow P_U$.dequeue();
  $D \leftarrow \varnothing$, $P_O \leftarrow \varnothing$;
  $(D, P_O) \leftarrow$ `VisitProfile`$(p, P_V, D, P_O, \mathcal{F}, \mathcal{L})$;
  Merge $\{u \in U | P_O \cap u \neq \varnothing\}$ to u';
  $P_V \leftarrow P_V \cup D$; $P_U \leftarrow P_U \setminus D$;
  if $u' = \varnothing$ **then** $u' \leftarrow U$.createUser();
  foreach $d \in D$ **do**
    u'.add(d);
    $H_0 \leftarrow$ hub sites obtained from $\mathcal{F}(d)$;
    P_U.enqueue($\mathcal{L}(p)$.getProfiles()$\setminus(P_V \cup P_U)$);
    foreach $h' \in H_0 \setminus H$ **do**
      P_U.enqueue(h'.getAllUsers()$\setminus(P_V \cup P_U)$);
    $H \leftarrow H_0 \cup H$;
return $U, \mathcal{F}, \mathcal{L}, H$

Procedure VisitProfile($p, P_V, D, P_O, \mathcal{F}, \mathcal{L}$)

Input: profile page p, visited pages P_V, newly discovered pages D, previously visited pages $P_O \subset P_V$, footprints $\mathcal{F}$, social links$\mathcal{L}$
Output: D, P_O

D.add(p);
$\mathcal{F}(p) \leftarrow$ footprints obtained from the account of p;
$\mathcal{L}(p) \leftarrow$ social links obtained from the account of p;
$P' \leftarrow$ other profile pages displayed on p or $\mathcal{F}(p)$, *directly connected* with p;
foreach $p' \in P'$ **and** $p' \notin D$ **do**
  if $p' \in P_V$ **then** P_O.add(p');
  else `VisitProfile` $(p', P_V, D, P_O, \mathcal{F}, \mathcal{L})$;
return (D, P_O)

Relational Hierarchical LDA (RH-LDA)

checkin movie book music events

8:00-12:00 12:00-20:00 20:00-8:00 non-local

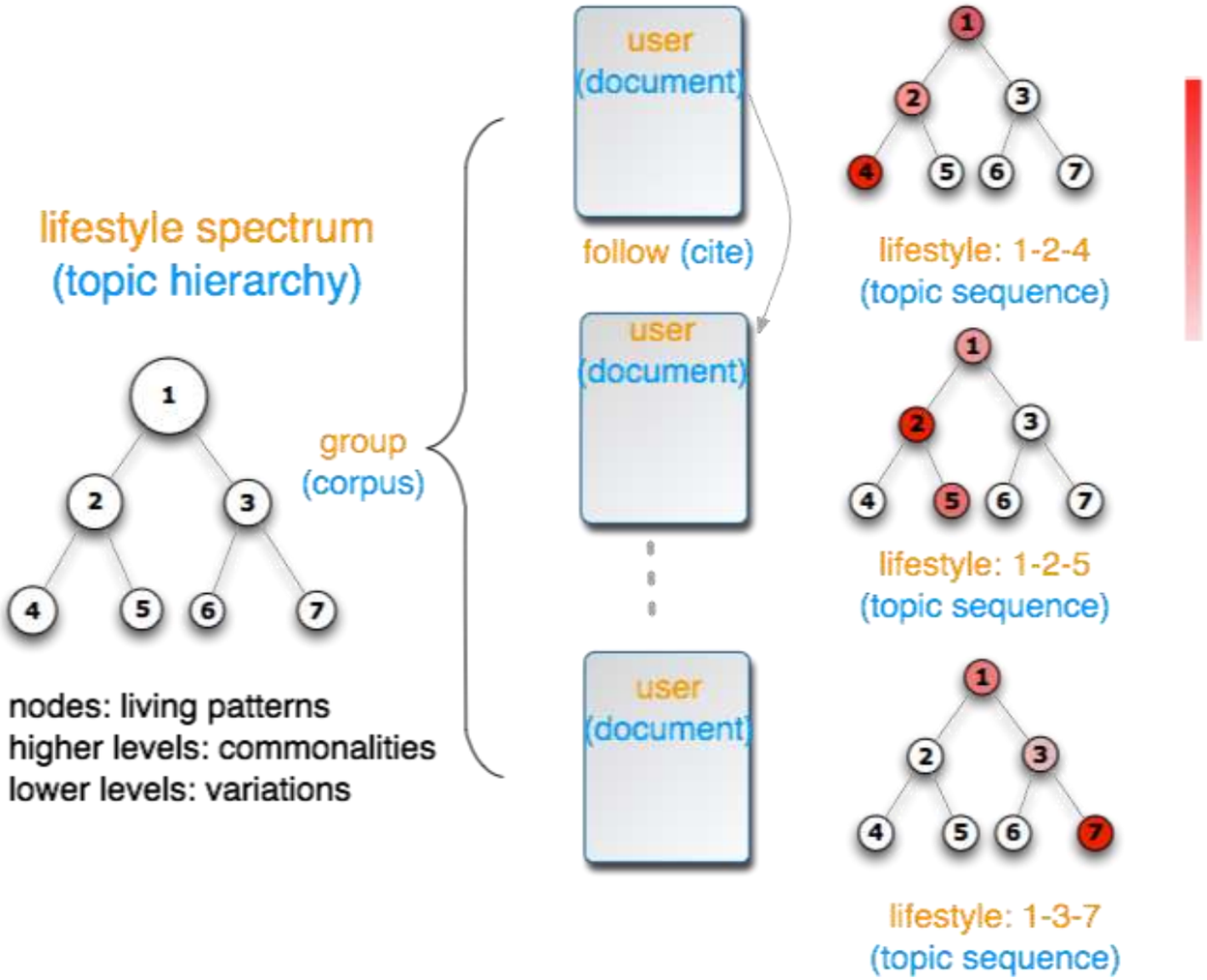
footprint (word): combination of domain specific tags (category)

(☀) shopping mall drama, sci-fi taiwan,pop lecture

living pattern (topic): frequently co-occurring footprints

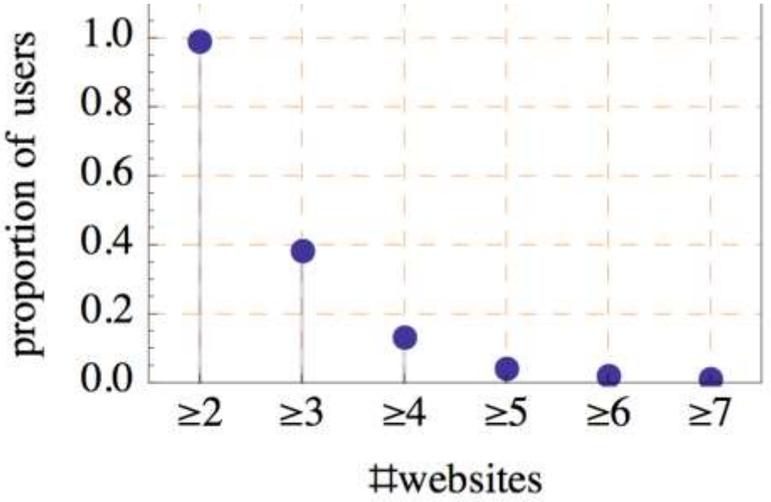
(☀) shopping mall + taiwan,pop + ( )) bar

lifestyle spectrum: tree-structured topic hierarchy



Data

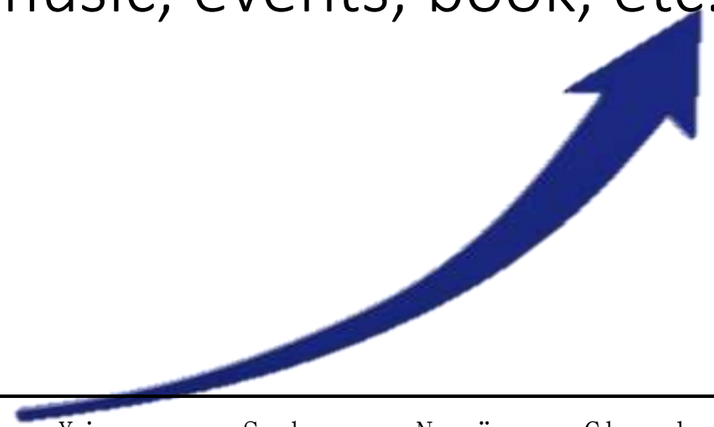
- 4 (major) networks: Jiepang, Weibo, Douban, Dianping
- 1.4M+ unique (deterministically identified) users accounts
- Heterogeneous footprints: tweets, photos, check-ins, movies, books, music, offline events, online purchase history, etc.
- Rich user profiles integrated from different sites (publicly available)



Age	Gender	Residence	Relationship	Occupation	College	High School	Self description	…

Data

- 53 million footprints (check-in, movie, music, events, book, etc.)
- 3 million social links
- 39 million check-ins



	city	Shanghai	Beijing	Guangzhou	Tianjin	Hangzhou	Hongkong	Xiamen	Suzhou	Nanjing	Chengdu	Wuhan	Xian
	users	417,681	162,764	53,089	15,490	34,322	12,599	10,123	19,673	21,558	23,372	20,975	15,261
Footprints	check-in	25,178,189	5,898,447	1,092,138	392,943	619,219	424,650	369,231	560,274	414,202	327,634	321,646	229,678
	movie	1,661,214	1,466,479	171,789	118,775	238,721	57,003	70,172	89,706	174,664	191,042	166,337	123,223
	music	766,165	737,254	85,953	60,658	103,936	30,313	29,716	39,701	82,513	88,426	76,316	62,876
	book	402,318	387,138	51,913	28,188	57,835	18,117	18,516	19,521	44,345	42,241	44,804	28,435
	event	609,076	803,158	101,246	52,133	78,587	18,277	20,889	27,400	46,788	66,640	44,764	72,902
	total	28,616,962	9,292,476	1,503,039	652,697	1,098,298	548,360	508,524	736,602	762,512	715,983	653,867	517,114

Table 1. Summarization of collected footprints for different cities (partially presented due to page limit).

Mobility Footprints

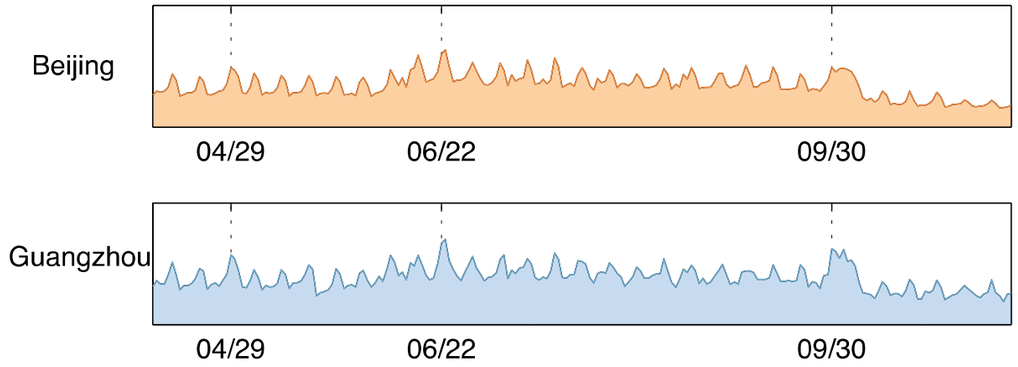
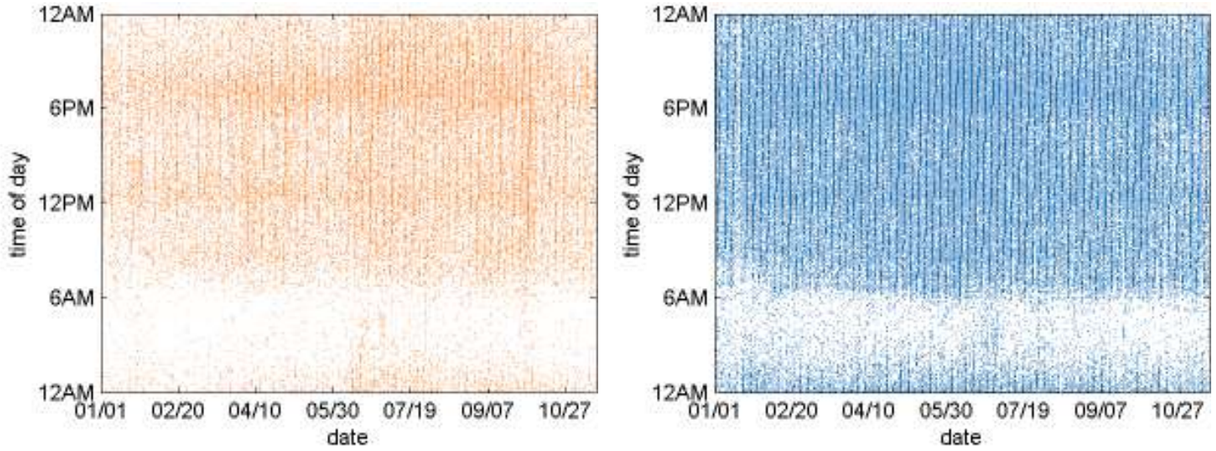


Figure 5. Daily trends of total check-ins in different cities.



(a) Beijing (b) Guangzhou

Figure 6. Diurnal distribution of 1000 sampled users' check-ins.

Rank	City	Bedtime
排名	城市名称	就寝时间
1	广州	23:08
2	东莞	22:53
3	厦门	22:37
4	西安	22:37
5	天津	22:32
6	长沙	22:29
7	重庆	22:25
8	武汉	22:24
9	郑州	22:19
10	成都	22:18
11	北京	22:15
12	昆明	22:10
13	南昌	22:09
14	淮南	22:08
15	上海	22:08
16	杭州	22:04
17	沈阳	22:04
18	西宁	22:03
19	南京	22:02
20	齐齐哈尔	21:50

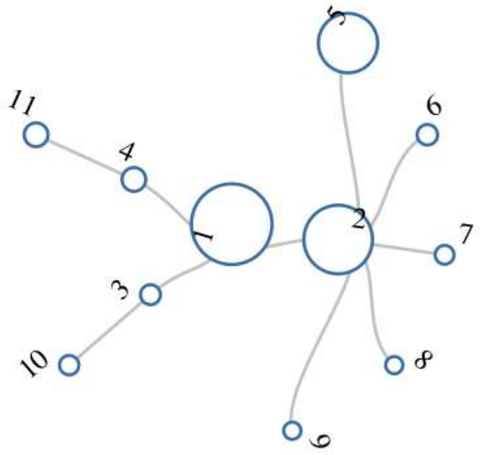
Guangzhou

Beijing

Average bedtime of different cities
(1M respondents, led by Chinese Medical Doctor Association, announced in world sleep day 2013)

checkin movie book music events

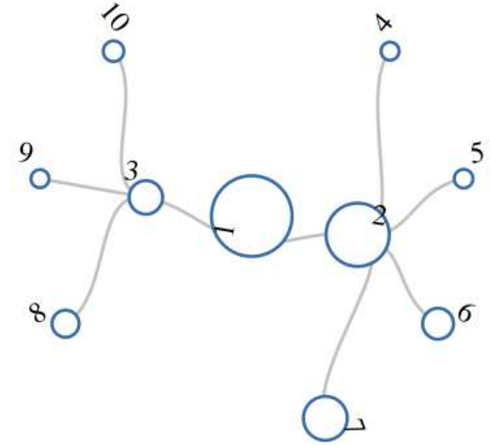
8:00-12:00 12:00-20:00 20:00-8:00 non-local



a) financial practitioners

1: economics (night) (non-local) apartment hotel (sun) (non-local) shopping mall
2: (sun) japanese cuisine (sun) fast-food lecture
3: (sun) hot-pot (sun) bar (sun) snack
4: (sun) snack (sun) fast-food (sun) japanese cuisine
5: drama,romance drama,comedy drama,action
6: (morning) bank (sun) bank (sun) subway
7: fiction,hongkong fiction,love mystery,japan
8: folk,indie indie,folk drama
9: (sun) car-4s fiction,society cartoon,philosophy
10: music drama,romance drama,comedy
11: (night) (non-local) scenic (sun) (non-local) airport (sun) (non-local) office

a') financial practitioners

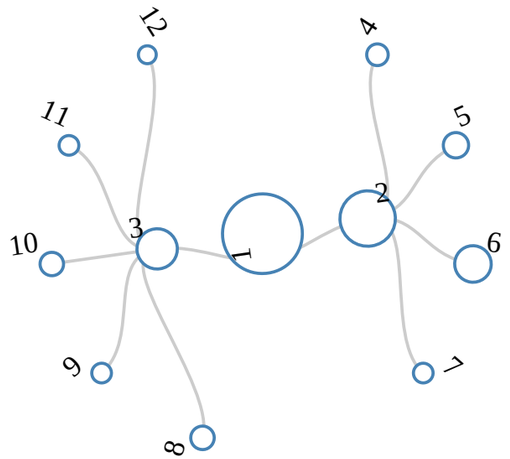


b) software practitioners

1: computer programing,computer movie
2: drama,romance drama,comedy taiwan,pop
3: ux,design fiction,foreignliterature fiction,chineseliterature
4: mystery,japan comedy,action cartoon,mystery
5: taiwan,pop music chineserock,rock
6: (night) apartment (sun) office (night) (non-local) apartment
7: drama,romance drama,action drama,comedy
8: lecture music get-together
9: programing,computer algorithm,computer drama,suspense
10: drama,romance music taiwan,pop

b') software practitioners

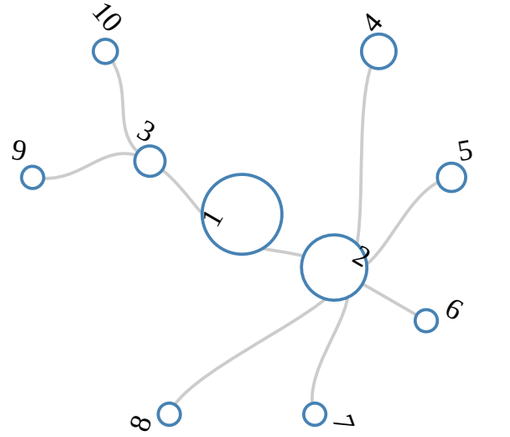
checkin movie book music events ☼ 8:00-12:00 ☀ 12:00-20:00 ☽ 20:00-8:00 non-local



c) students/graduates from School A

1: drama,romance drama,comedy drama,action
2: (☀) school canteen (☀) snack (☀) train station
3: (☀) office (☀) apartment (☼) office
4: taiwan,pop action,sci-fi movie
5: (☀) ( ) airport (☽) ( ) apartment hotel (☀) ( ) apartment hotel
6: (☼) library (☼) school canteen (☀) teaching building
7: japan,jpop mystery,japan jpop,japan
8: drama,romance pop,western exhibition
9: history,chinesehistory mystery,japan action,sci-fi
10: (☀) fast-food (☀) apartment hotel (☽) institute
11: music investment,finance get-together
12: ost,japan cartoon folk,inland

c') students/graduates from School A

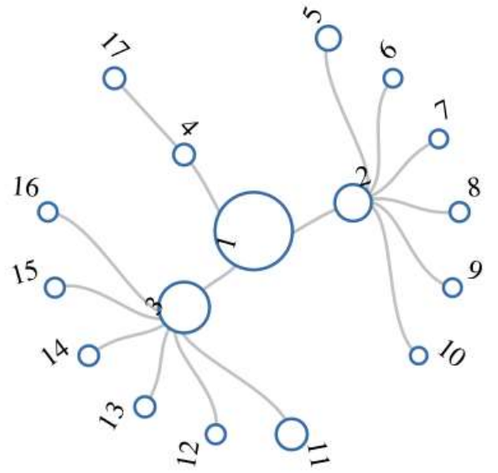


d) students/graduates from School B

1: (☀) coffee (☀) western-food (☽) bar
2: drama,romance drama,comedy taiwan,indie
3: drama,romance drama,comedy drama,action
4: music movie get-together
5: drama,action action,sci-fi action,thriller
6: britpop,uk (☽) institute chineserock,inland
7: fiction,romantic jazz,western japaneseliterature,japan
8: folk,inland chineserock,rock taiwan,pop
9: (☀) freeway (☽) private place (☽) freeway
10: movie comedy,romance pop,western

d') students/graduates from School B

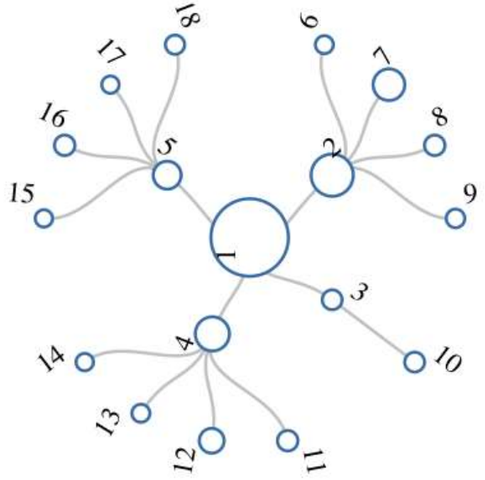
checkin movie book music events 8:00-12:00 12:00-20:00 20:00-8:00 non-local



c) the post-90s generation

1: drama,romance drama,comedy comedy,romance
2: (☀) coffee (☀) (🌍) train station (☽) fast-food
3: (☀) institute (☀) coffee (☼) institute
4: hongkong,essay (☀) (🌍) coffee (☽) electric game
11: (☀) subway (☀) shopping mall (☀) fast-food
12: (☀) teaching building (☼) teaching building music
13: exhibition fiction,youth lecture
14: music taiwan,indie folk,inland
15: korea,kpop taiwan,pop music

c') the post-90s generation

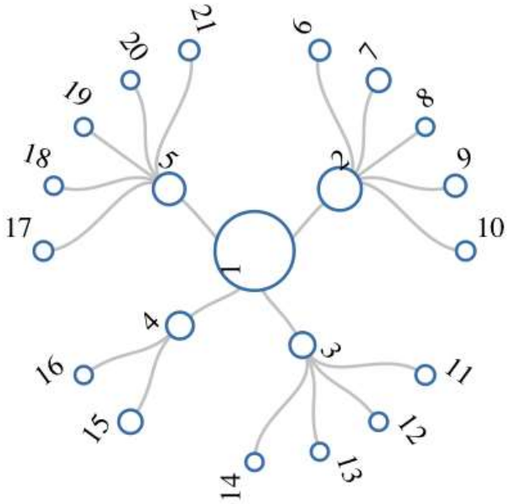


d) the post-80s generation

1: drama,romance drama,comedy drama,action
2: (☀) shopping mall (☀) hot-pot (☀) japanese cuisine
3: (☽) shopping mall (☼) office (☀) snack
4: get-together lecture travel
5: action,thriller (☽) apartment cartoon,suspense
6: (☽) ktv (☼) office (☀) coffee
7: (☀) office (☀) sichuan cuisine (☼) office
8: (☽) apartment (☀) commercial building (☼) commercial building
9: picturebook,taiwan picturebook,cartoon fiction,uk

d') the post-80s generation

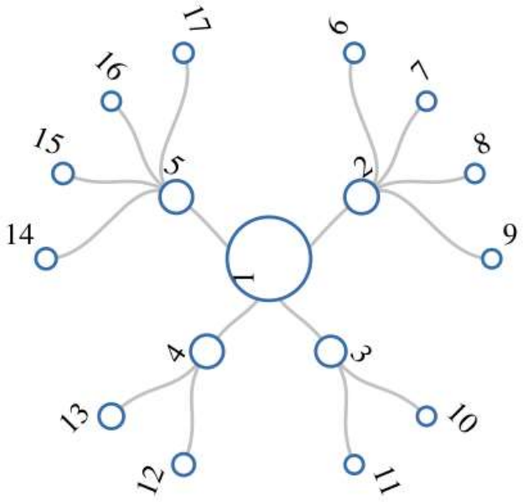
checkin movie book music events 8:00-12:00 12:00-20:00 20:00-8:00 non-local



a) Beijing

1: checkin (12:00-20:00) shopping mall checkin (12:00-20:00) office checkin (12:00-20:00) fast-food
2: movie drama checkin (12:00-20:00) office checkin (8:00-12:00) office
3: checkin (12:00-20:00) teaching building checkin (20:00-8:00) school dormitory movie drama
4: movie drama, sci-fi book politics movie comedy
5: checkin (20:00-8:00) (non-local) shopping mall checkin (12:00-20:00) (non-local) office checkin (8:00-12:00) (non-local) airport
6: movie drama movie comedy movie action
7: checkin (12:00-20:00) coffee checkin (12:00-20:00) western-food checkin (20:00-8:00) bar
8: checkin (20:00-8:00) (non-local) shopping mall checkin (12:00-20:00) (non-local) shopping mall checkin (12:00-20:00) (non-local) apartment hotel
9: music movie drama,romance music taiwan

a') Beijing

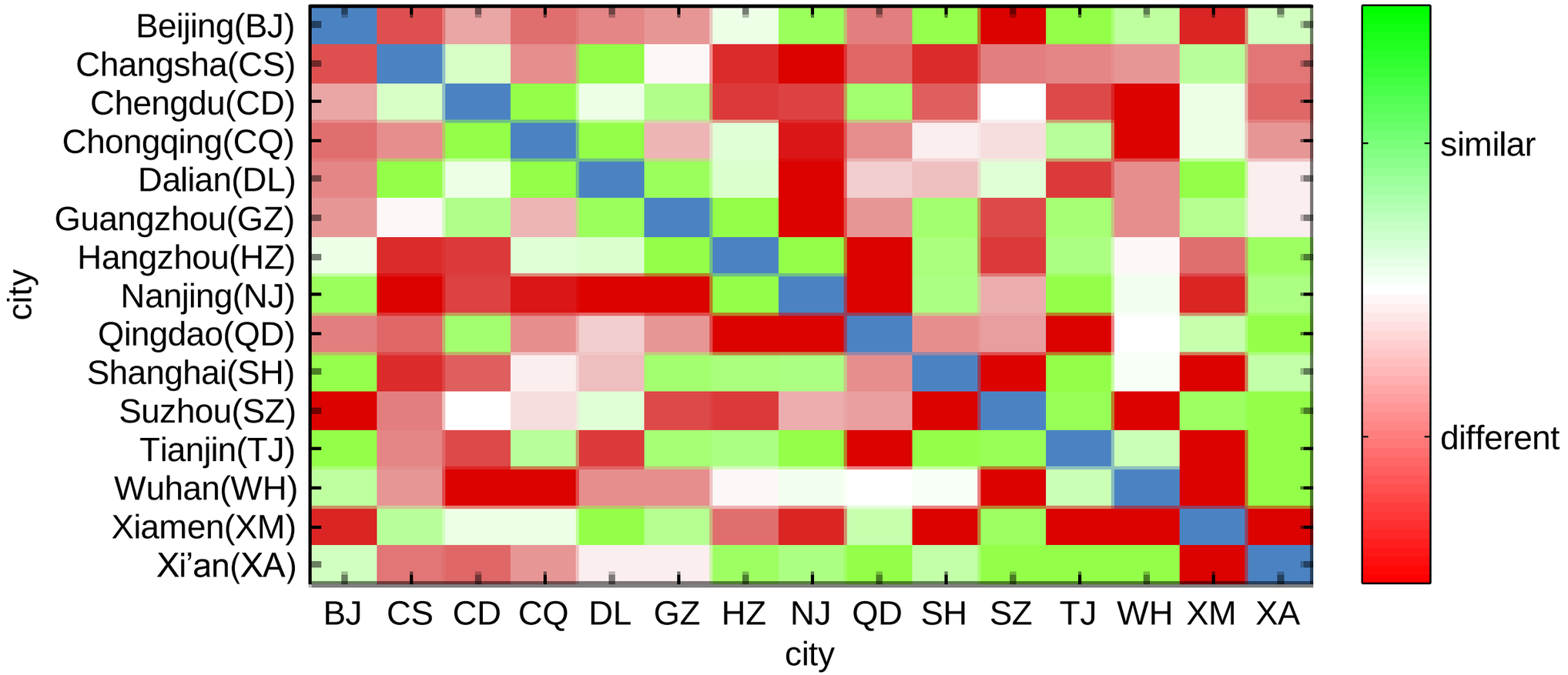


b) Hangzhou

1: checkin (12:00-20:00) jiangzhe cuisine checkin (12:00-20:00) fast-food checkin (12:00-20:00) coffee
2: movie drama,comedy checkin (12:00-20:00) (non-local) coffee checkin (12:00-20:00) (non-local) western-food
3: checkin (12:00-20:00) (non-local) train station checkin (12:00-20:00) (non-local) shopping mall checkin (8:00-12:00) (non-local) outdoors
4: movie drama book fiction music taiwan
5: checkin (12:00-20:00) (non-local) shopping mall checkin (12:00-20:00) (non-local) snack checkin (12:00-20:00) (non-local) coffee
6: music music get-together movie drama,romance
7: checkin (12:00-20:00) office checkin (12:00-20:00) supermarket checkin (12:00-20:00) fast-food
8: movie drama,action book history book fiction
9: checkin (12:00-20:00) library checkin (8:00-12:00) library checkin (12:00-20:00) scenic

b') Hangzhou

Similarity Matrix of Lifestyle Spectrum



Intergroup difference and intragroup variations (place of residence)

Table 2: Average recognition ratio

Method	RTM	hLDA	RH-LDA
Check-in	0.361	0.500	0.667
+Movie	0.389	0.556	0.694
++Music	0.444	0.583	0.722
+++Book	0.472	0.639	0.806
++++Events	0.472	0.667	0.833

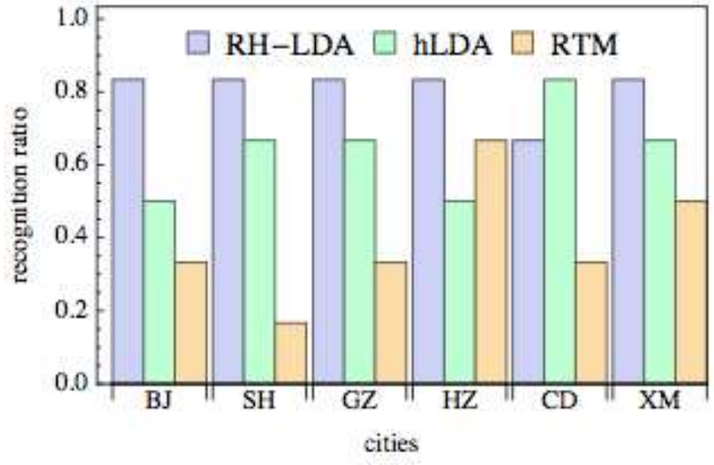


Figure 12: Recognition ratio.

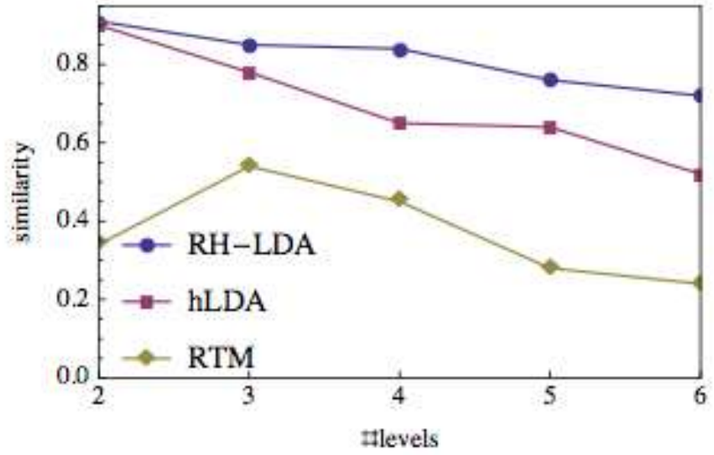


Figure 13: Jaccard Similarity.

36 participants
aged 21-45
20 males 16 females
6 for each city
8+years

Train models for 6 cities
Select based on lifestyles
RR: correctly inferred/all

Collect social data
Pick a path in the spectrum
Measured by JS

Mobility Footprints

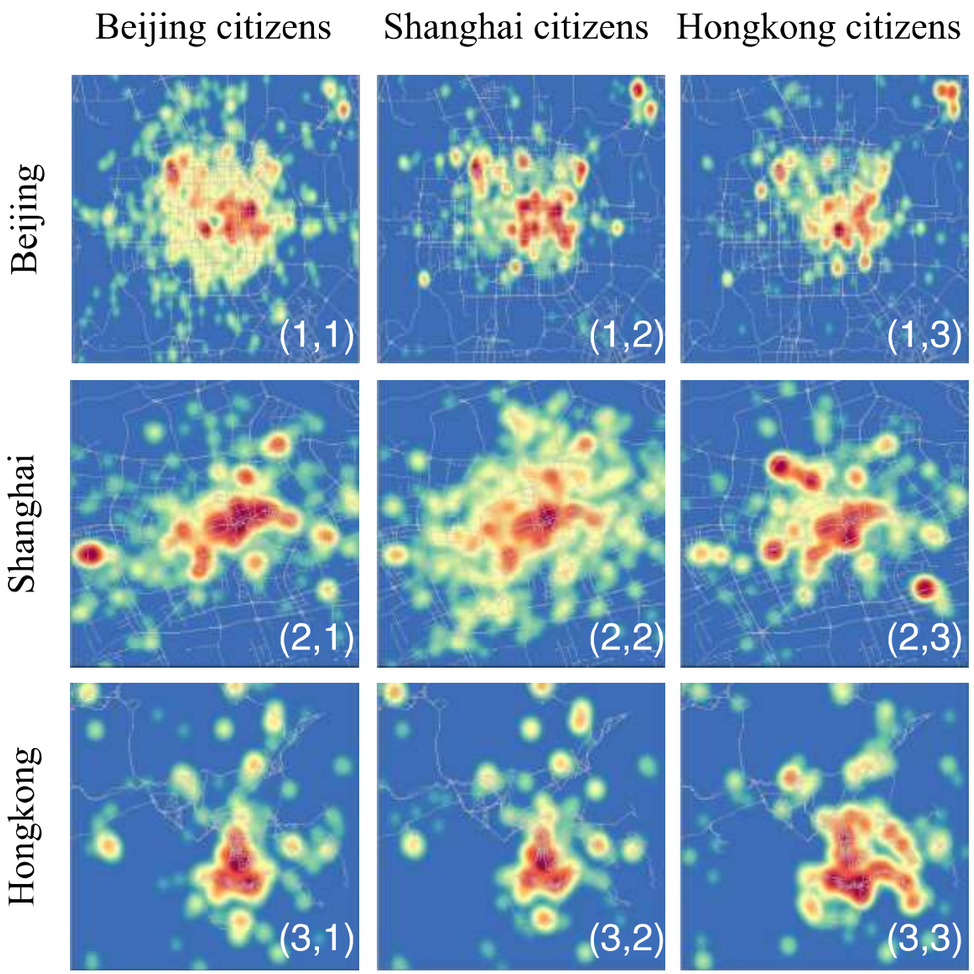


Figure 7. Check-in density distribution of 3 cities showing where people check-in in each other's cities.

LifeSpec as A Data Repository

- Scientific Research
  - Sociology
  - Economics
  - Public health

- Industry Applications
  - Recommendation
  - Advertising

Summary

- Computational framework for exploring urban lifestyles
- IConnect: identifying connected user accounts based on self-disclosure
- RH-LDA: relational-hierarchical model for summarizing lifestyles
- LifeSpec as a data platform for scientific and industry applications

Thanks!

Xing Xie

Microsoft Research Asia