

Microsoft
Research
10 Asia
year anniversary

Social Computing at MSRA

Natural Language Computing, Microsoft Research Asia

Chin-Yew LIN cyl@microsoft.com

What is Social Computing?

□ Social

□ Living together in communities

- Live → dynamic
- Together → more than one person
- Community → same locality

□ Computing

□ To determine by the use of a computer

□ Social Computing

□ ***To live together in communities using computers***

□ Creating social contexts online via the use of technology

* American Heritage Dictionary

+ <http://www.microsoft.com/sharepoint/capabilities/collaboration/social.aspx>

Microsoft®

Research

Why Social Computing?

- Web is increasingly social
 - ▣ Wikipedia, blogs, Digg, Flickr, del.icio.us, Amazon, Yahoo! Answers, Live QnA, forums, Facebook, MySpace, LinkedIn, ...
- Social web
 - ▣ People + Content + Network
- How do we turn **social web** into **value** for **people**?
 - ▣ Assimilation of knowledge → systems pull from data
 - ▣ Dissimilation of knowledge → systems to people
 - ▣ Elicitation of knowledge → systems pull from people
- ***“Pay attention to that man behind the curtain!”***

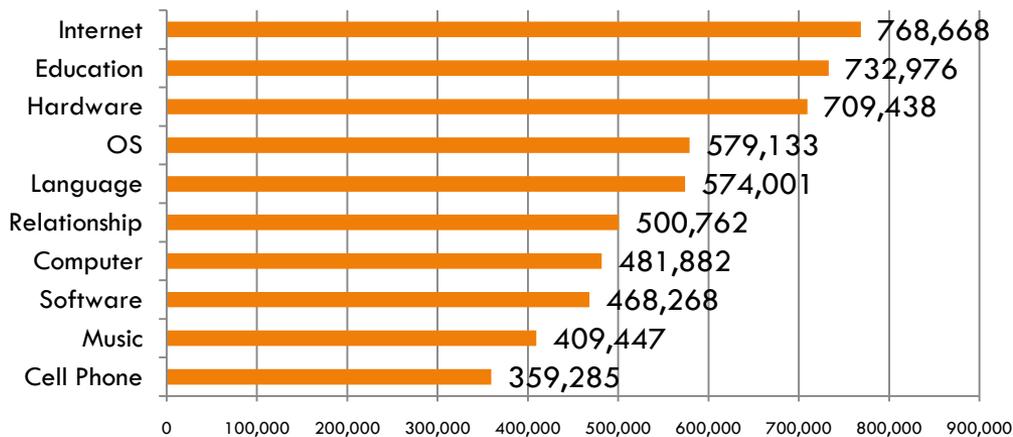
Microsoft®

Research

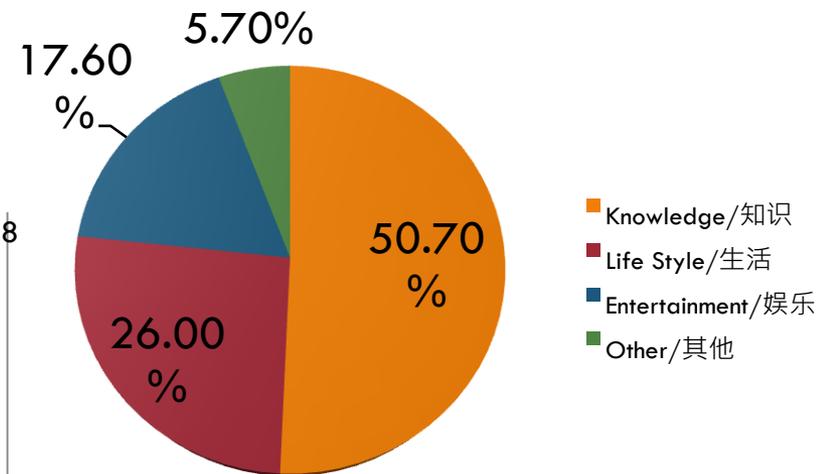
Baidu Zhidao (百度知道)

- 17,012,767 resolved questions in two years' operation
- 8,921,610 are knowledge related
- 96.7% of questions are resolved
- 10,000,000 daily visitors
- 71,308 new questions per day
- 3.14 answers per question

Baidu Zhidao Top 10 Question Types



Baidu Zhidao Question Types Distribution



- <http://www.searchlab.com.cn> (中国人搜索行为研究/User Research Lab of Chinese Search)

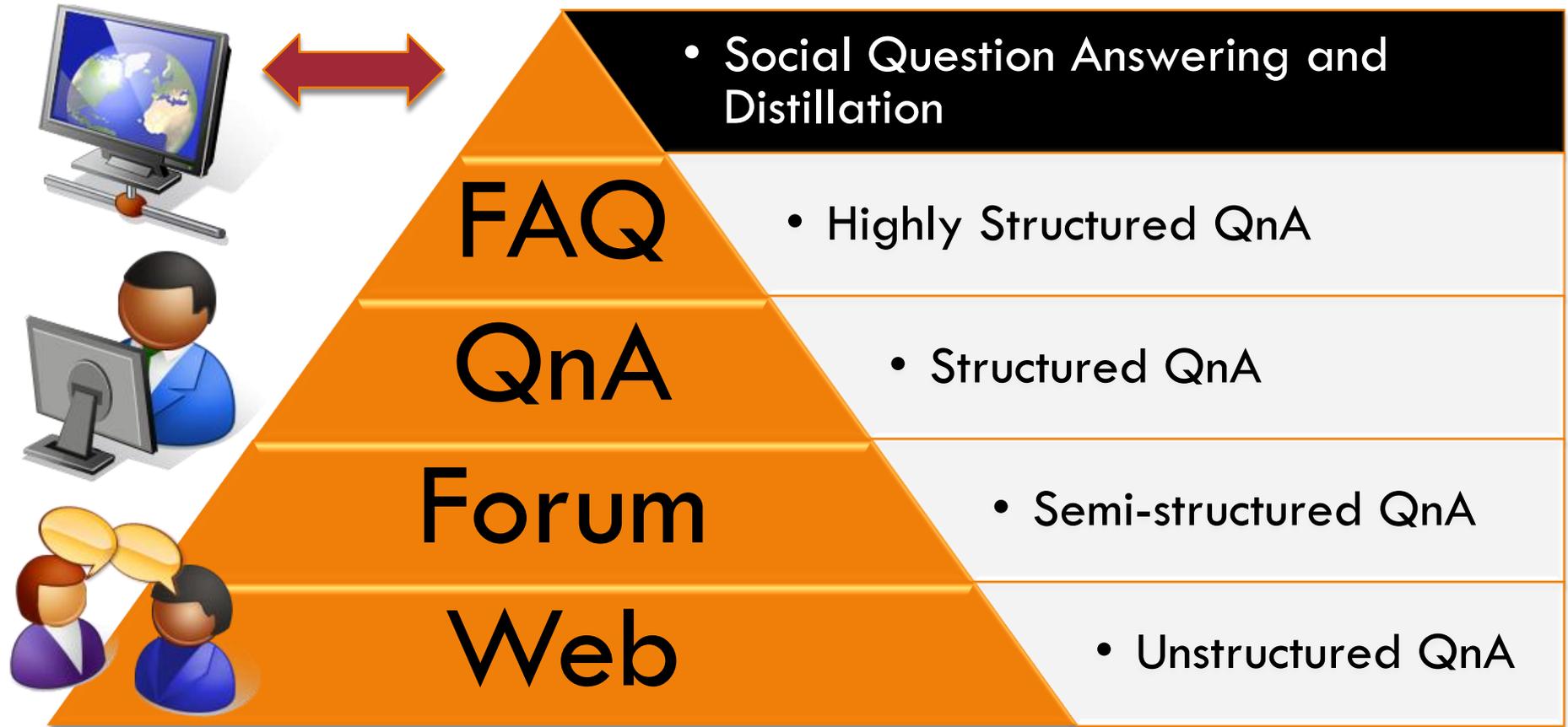
Stickiness of Baidu Zhidao

- 据正望咨询调查，“百度知道”跟搜索的关系非常紧密，而且对搜索黏性的提高有很大帮助，根据其统计，“百度知道”已成为百度的一个核心产品。“百度的用户中有**50%搜索‘知道’**，其用户量已经超过百度贴吧，与其**MP3搜索可相提并论。**”。
- 50% of Baidu users search Baidu Zhidao
- Zhidao search traffic comparable to MP3 search

(<http://news.csdn.net/n/20080425/115453.html>; 04/25/2008)

Social Question
Answering

Knowledge Distillation & Dissemination



GeoLife 2.0

Building Social Networks Using Human Location History



Yu Zheng, Xing Xie and Wei-Ying Ma
WSM, Microsoft Research Asia

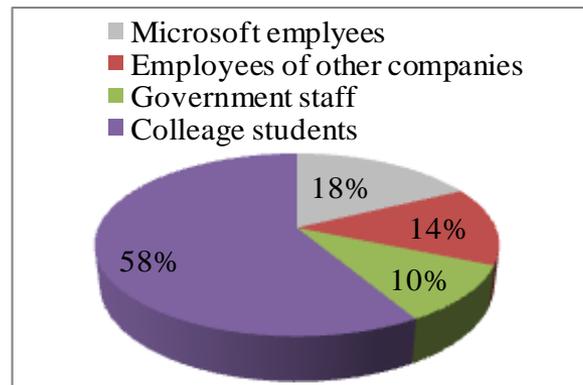
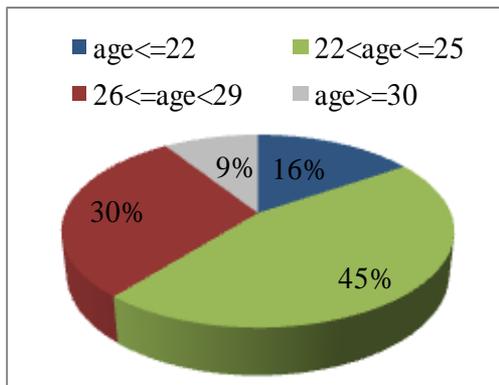
What is GeoLife 2.0?

8

- A **location-history-based GPS-data-driven social-networking** service
 - ▣ Enables people to build connections using their GPS trajectories
 - ▣ Understand a user and a location
 - ▣ Explore the **similarity between users** and the **correlation among locations**

GPS Devices and Users

- 60 devices and 138 users
- From May 2007 to present



A Large-Scale GPS Data Set

- 10+ million GPS points
- 260+ million kilometers
- 36 cities in China and a few cities in the USA, Korea and Japan



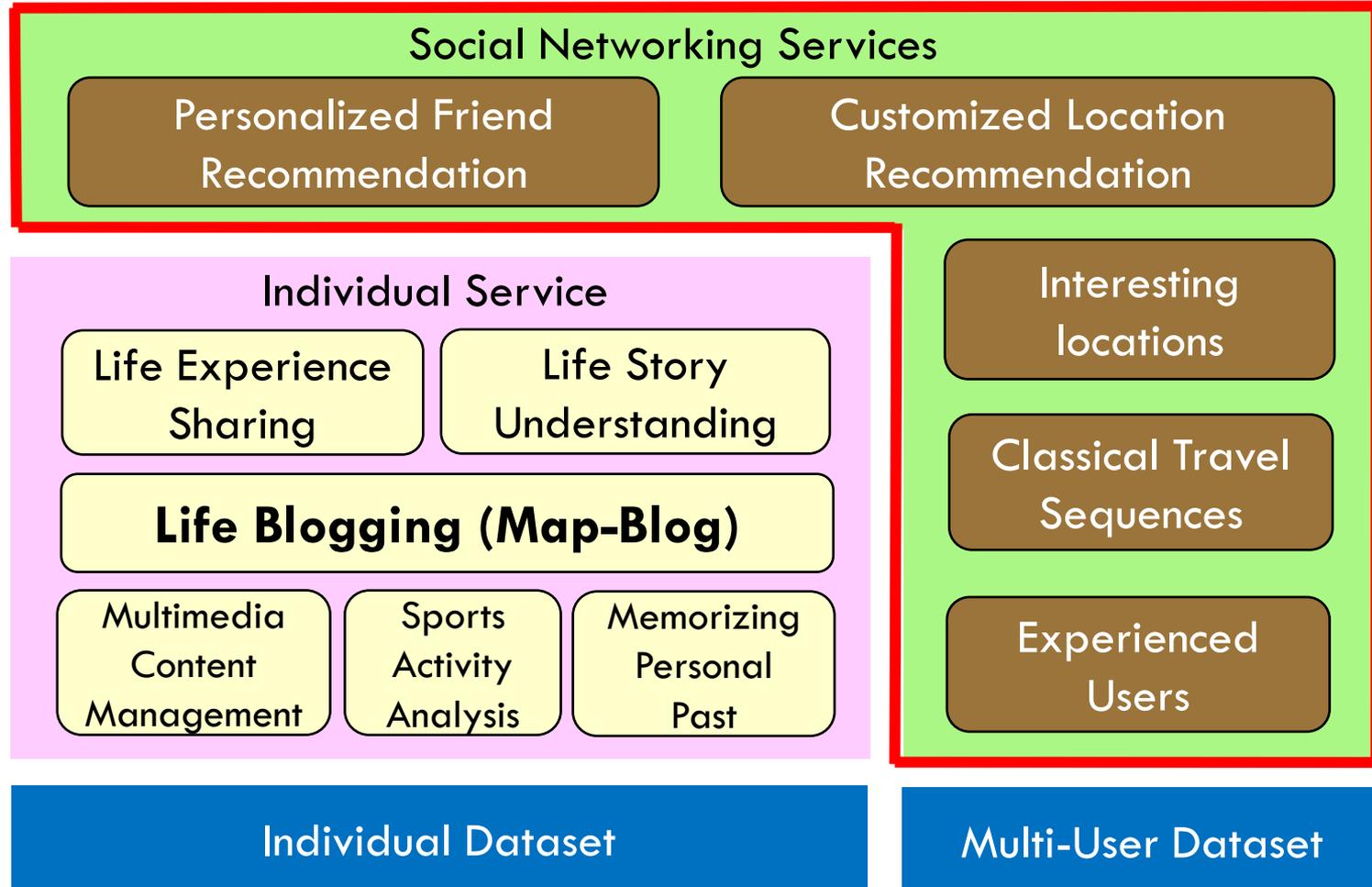
* Bikely: <http://www.bikely.com>

* GPS Track Route Exchange Forum: <http://www.gpsxchange.com>

* GPS Sharing: <http://gpssharing.com>

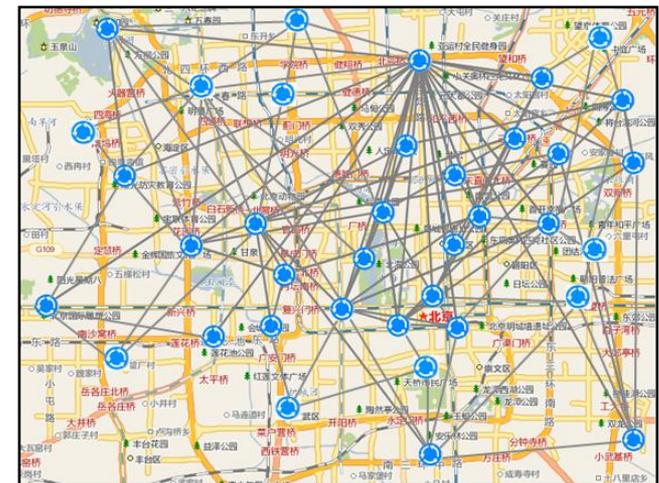
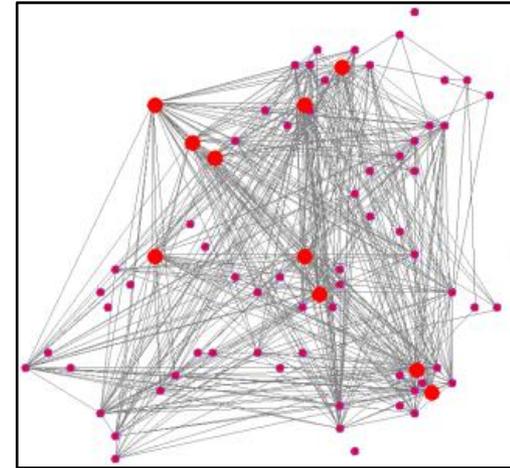
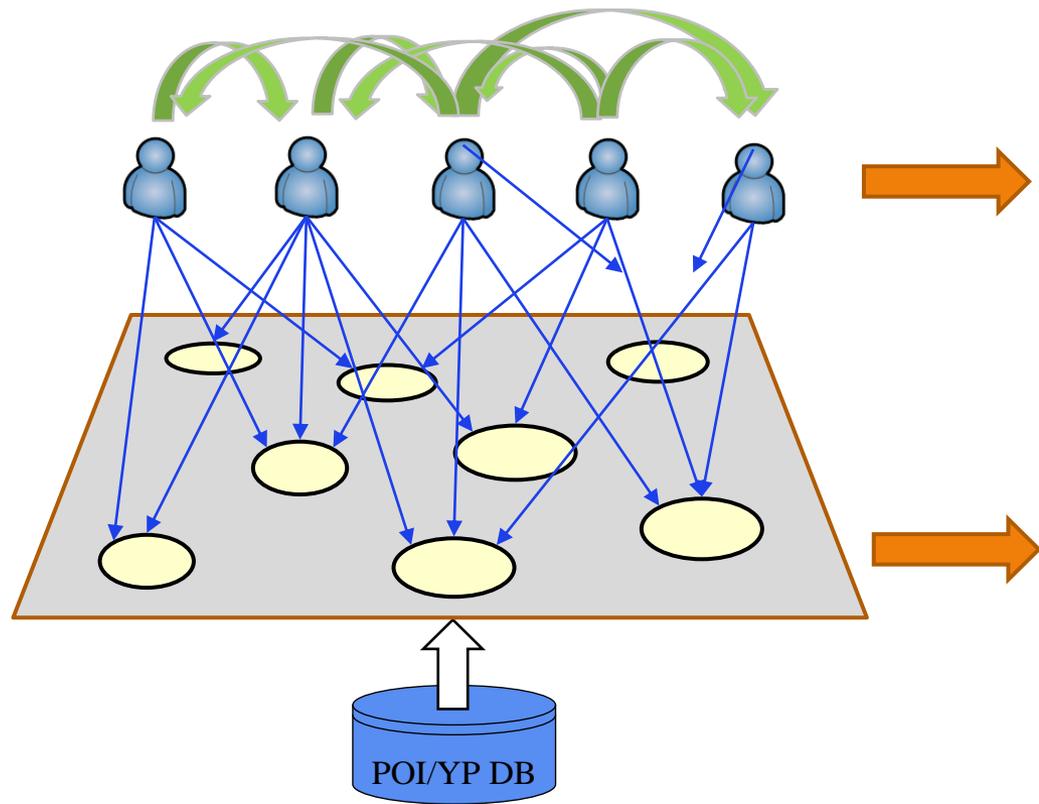
GeoLife Application Architecture

GeoLife 1.0

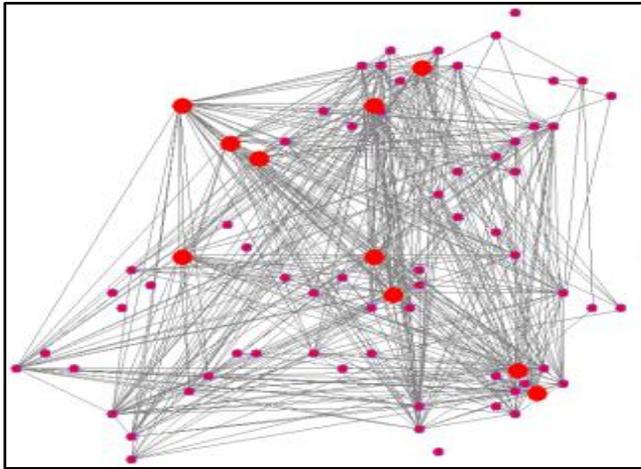


GeoLife 2.0

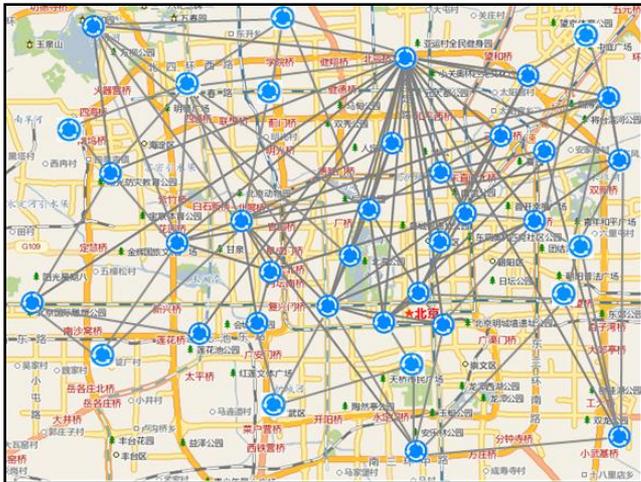
A Location-History-Based Social Network



Applications



- Understanding People
 - ▣ Similar users: Friend recommendation
 - ▣ Experienced users: Travel experts recommendation
 - ▣ Group users: Community discovery



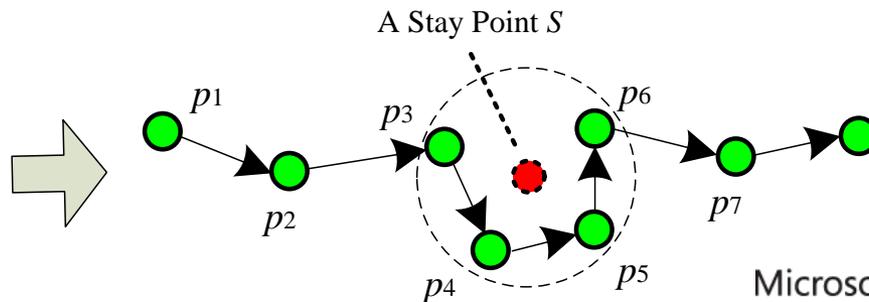
- Understanding Locations
 - ▣ Personalized location recommendation
 - ▣ Mining interesting locations
 - ▣ Detecting classical travel sequences

Preliminary

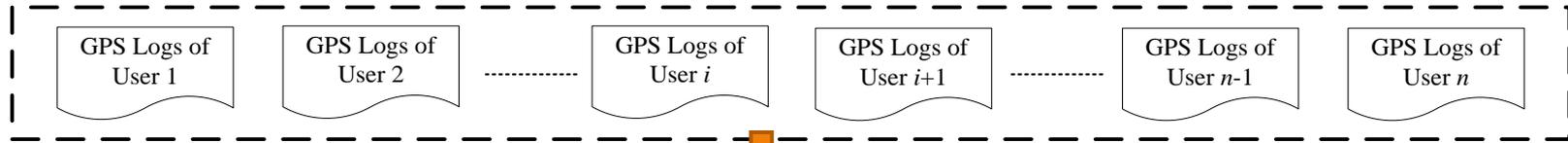
- **GPS logs P and GPS trajectory**
- **Stay points $S = \{s_1, s_2, \dots, s_n\}$.**
 - ▣ Stands for a geo-region where a user has stayed for a while
 - ▣ Carry a semantic meaning beyond a raw GPS point
- **Location history:**
 - ▣ Represented by a sequence of stay points with transition intervals

$$LoCH = (s_1 \xrightarrow{\Delta t_1} s_2 \xrightarrow{\Delta t_2}, \dots, \xrightarrow{\Delta t_{n-1}} s_n)$$

	Latitude,	Longitude,	Time
p_1 :	Lat1,	Lngt1,	T1
p_2 :	Lat2,	Lngt2,	T2
		
p_n :	Latn,	Lngtn,	Tn

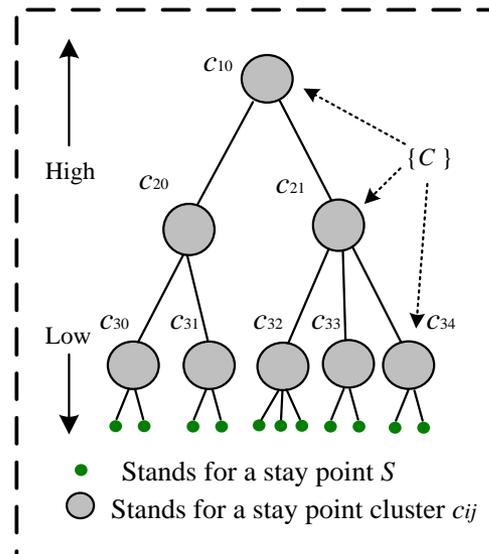
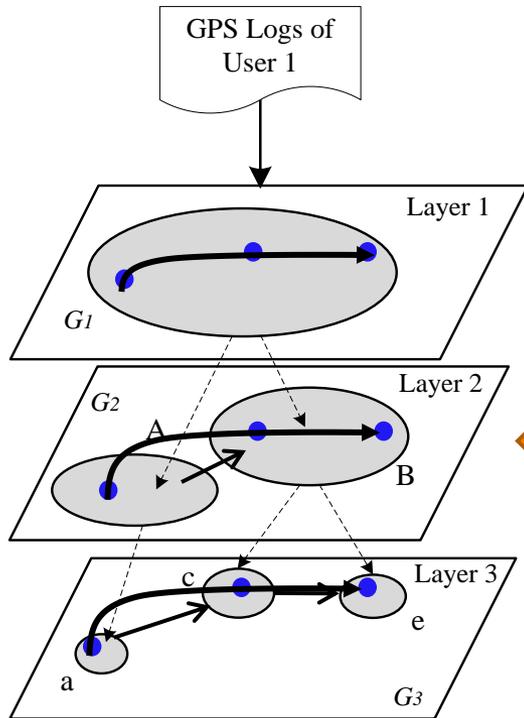
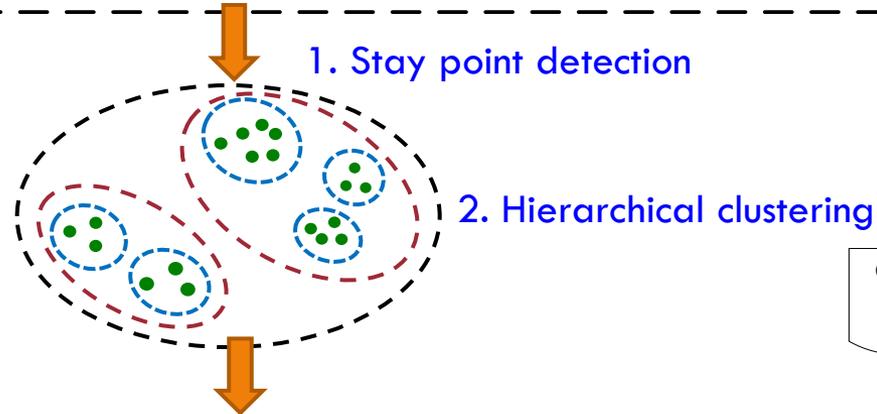


Modeling Individual Location History

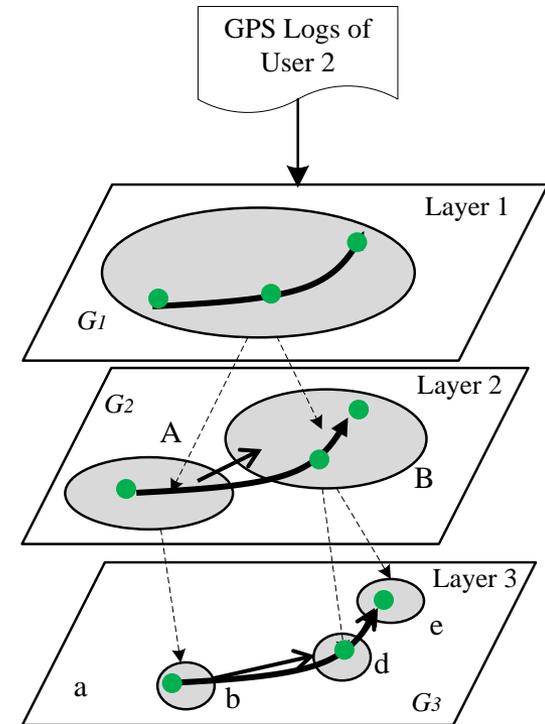


1. Stay point detection

2. Hierarchical clustering

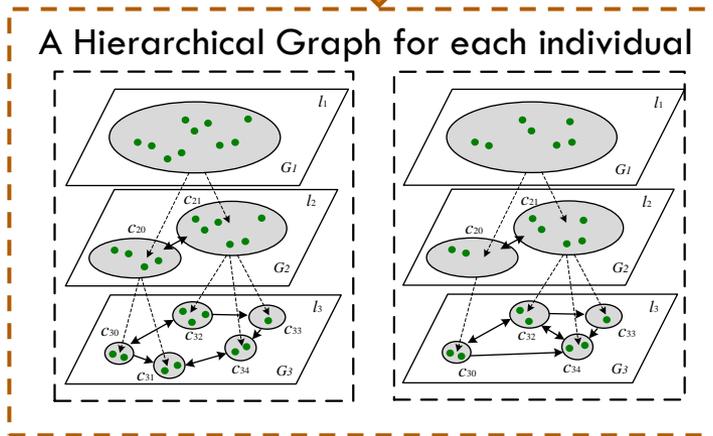
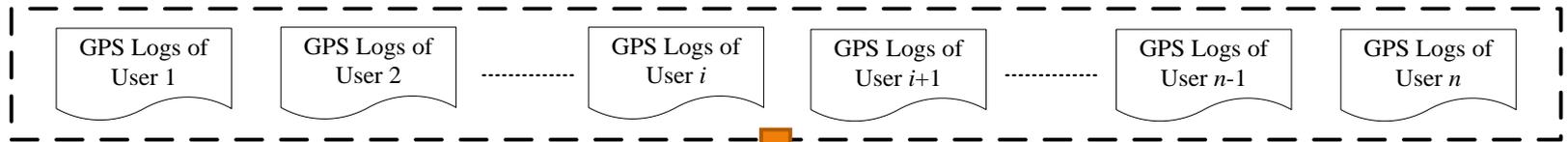


Shared Hierarchical Framework

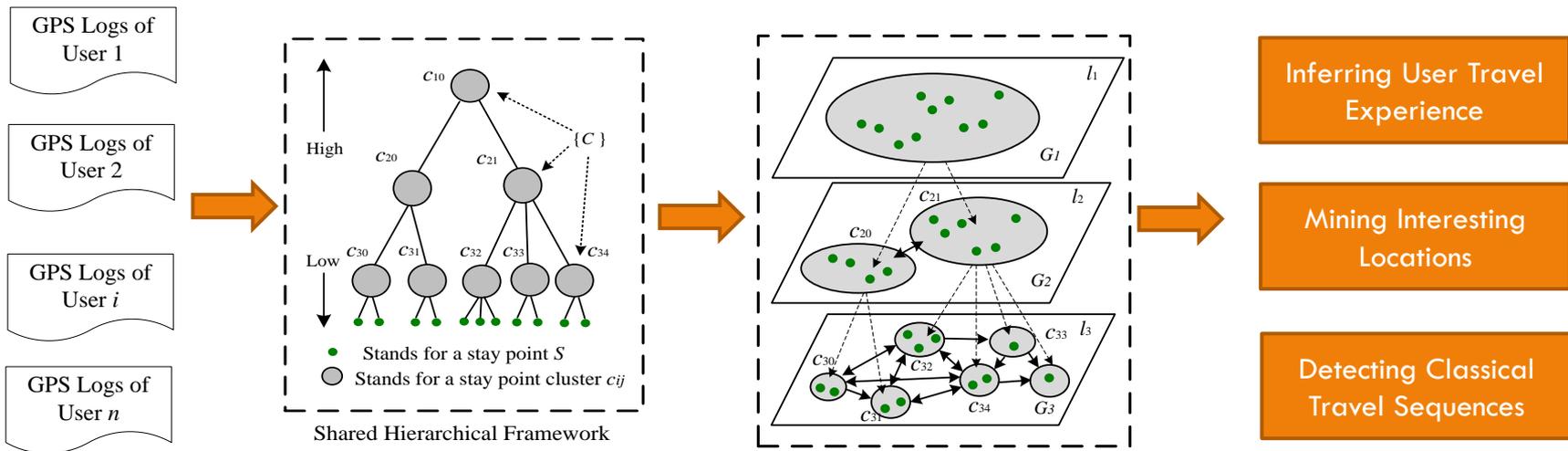


3. Individual graph building

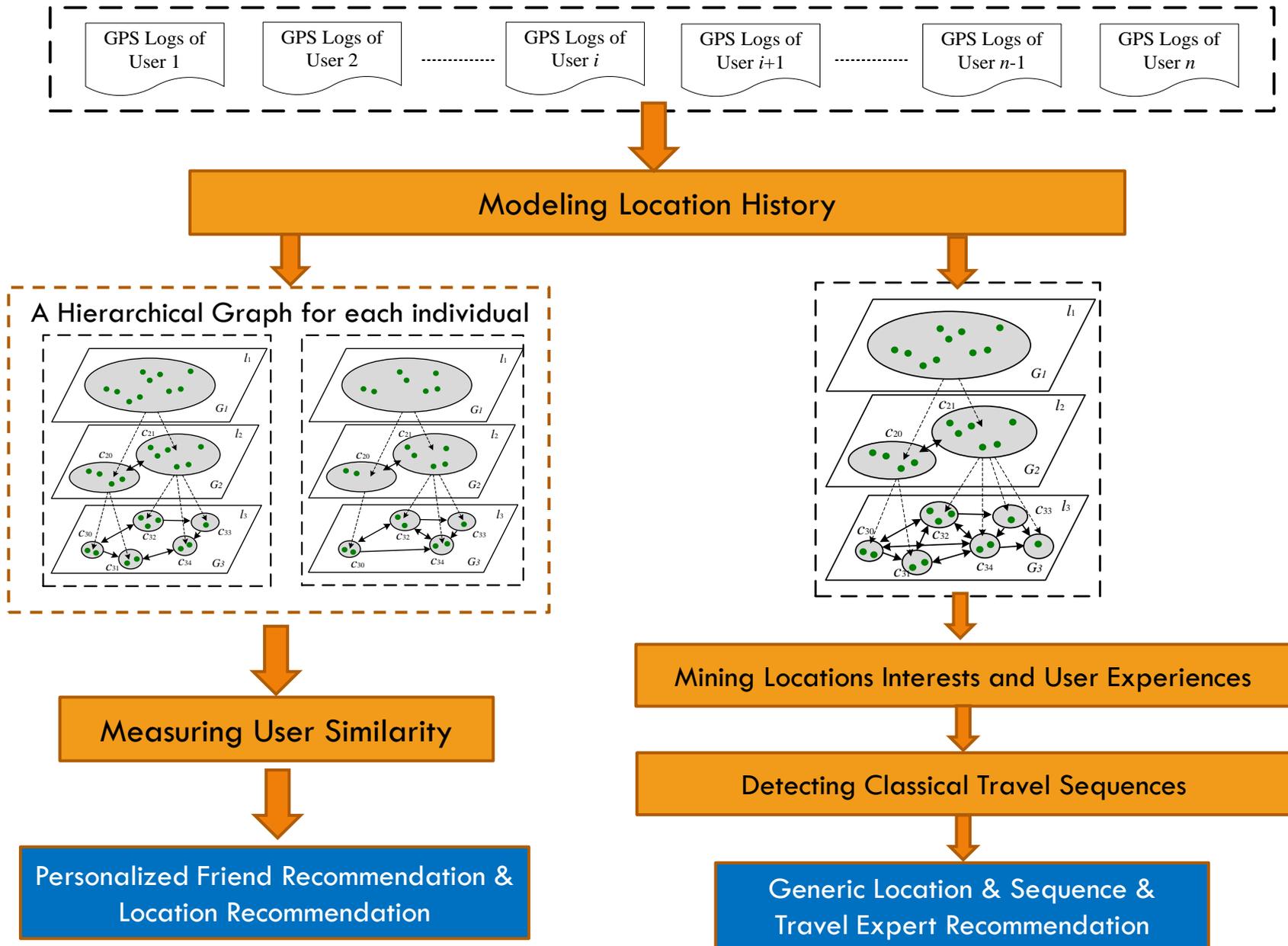
Architecture (1)



Modeling Multiple Users' Location Histories



Architecture (1)



GeoLife: Point of Interest

The screenshot displays the GeoLife web application interface. At the top left is the GeoLife logo, followed by navigation tabs for "Public" and "My GeoLife". Below these are tabs for "Locations", "Users", "Routes", "Navigation", and "Search". The main area is a map of Beijing, China, with numerous blue icons representing user-generated points of interest. A sidebar on the right lists "Interesting Locations" with details on user counts and photos:

- Hai Dian: Users:91 Photos:1160
- San Li Tun: Users:13 Photos:99
- Yuandadu Park: Users:27 Photos:27
- Olympic Park: Users:9 Photos:4
- Happy Valley: Users:7 Photos:29

Below the sidebar is a "Classic Sequences" section showing a list of photo sequences with their respective user counts and travel times:

- Users:18, TravelTimes:87
- Users:16, TravelTimes:80
- Users:15, TravelTimes:70
- Users:11, TravelTimes:69
- Users:10, TravelTimes:60

At the bottom of the map, there is a horizontal strip of small thumbnail images representing various locations.

<http://research.microsoft.com/en-us/projects/geolife/>

Microsoft
Research

GeoLife: Travel Route

Public Route

北京 上海 广州 深圳 天津 重庆
 武汉 西安 南京 杭州 成都 沈阳
 济南 郑州 长春 长沙 苏州 青岛

Hottest Routes in Given Location

ID	Start Time	Mode
1	2007-09-25 22:04:53	
2	2008-06-24 13:50:08	
3	2008-08-03 14:12:44	
4	2008-05-27 14:05:59	
5	2008-05-27 14:05:59	

Latest Routes in Given Location

ID	Start Time	Mode
6	2008-11-24 19:25:39	
7	2008-11-24 18:38:36	
8	2008-11-24 16:54:29	
9	2008-11-24 13:27:29	
10	2008-11-24 13:17:11	

Microsoft Virtual Time: 08:26:14 Speed: 0m/s The route has no tags

Summary

- Social web
 - People + Content + Network
- Turn **social web** into **value** for **people**
 - Assimilation of knowledge
 - Dissimilation of knowledge
 - Elicitation of knowledge
- ***“Pay attention to that man behind the curtain!”***

Reference

- Yu Zheng, Lizhu Zhang, Xing Xie and Wei-Ying Ma, [Mining Interesting Locations and Travel Sequences from GPS Trajectories](#), WWW 2009.
- Yu Zheng, Yukun Chen, Xing Xie and Wei-Ying Ma, [GeoLife 2.0: A Location-Based Social Networking Service](#), MDM 2009 (Demo).
- Yu Zheng, Like Liu, Longhao Wang and Xing Xie, [Learning Transportation Mode from Raw GPS Data for Geographic Application on the Web](#), WWW 2008.
- Quannan Li, Yu Zheng, Xing Xie, Yunkun Chen, Wenyu Liu and Wei-Ying Ma, [Mining User Similarity Based on Location History](#), ACM GIS 2008.
- Yu Zheng, Quannan li, Yukun Chen, Xing Xie and Wei-Ying Ma, [Understanding Mobility Based on GPS Data](#), ACM UbiComp 2008.