Social and Geographic-Aware Multimedia Applications and Technologies (Part II)

Yi Yu
Roger Zimmermann

School of Computing, National University of Singapore

July 14, 2014
Overview of Location-Enabled Multimedia Systems

Location-enabled topics and approaches:
- Geo-fencing
- Map generalization
- User preference profiling
- Correlating preference-aware activity data
- Sensing and discovering knowledge
- Geographic distribution of social events
- Multimedia content diffusion

Overall Image (Part II, 1.5 hours)
Geo-Fence Scenario

❖ A virtual perimeter
  - Coverage of a radio cell or Wi-Fi access point
  - Or manually specified geographic shape
  - Different shapes (e.g., circles, rectangles, polygons)

❖ Basic idea
  - Users enter or exit boundaries of areas → send notification
    - e.g., mobile advertisement, media recommendation
Efficient Geo-Fencing Algorithm

- Pairing points with polygons
  - Scalability
    - Big-data

- ACM GIS Cup 2013
  - Basic spatial predicates (INSIDE, WITHIN)

- A novel geo-fencing algorithm
  - Simple but effective and efficient
  - One of top 2 winners (29 entities)
  - LSH + probing
Geo-Fencing: Points & Polygons

Points
- Multiple instances
  - A unique sequence number
- Moving

Polygons
- Multiple instances
  - A unique sequence number
- Changing

Sequence numbers
- Same space & no overlapping
- A timestamp
Geo-Fencing: Polygon Instances

- Two types
  - With a single out-ring
  - With a single out-ring & multiple inner-rings

- Tips:
  - Inner rings
    - Inside outer-ring
    - Separated from outer-ring
Geo-Fencing: Latest Instance of Polygon

- A largest sequence number less than point sequence number
  - Sequence number of the point is 100, sequence numbers of a polygon are 10, 50, 80, 90, 110, 130.
  - This point compared with polygon with a sequence number of 90.
Geo-Fencing: Polygons, Point, Edges

# edges of 15 polygons (200)

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
<th>13</th>
<th>14</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>285</td>
<td>255</td>
<td>235</td>
<td>196</td>
<td>264</td>
<td>250</td>
<td>240</td>
<td>239</td>
<td>226</td>
<td>226</td>
<td>242</td>
<td>153,15</td>
<td>152,20</td>
<td>250</td>
<td>217</td>
</tr>
</tbody>
</table>
Geo-Fencing: INSIDE

- Crossing number algorithm
  - Inside (number of intersections = odd)
  - Checking each edge
- Exploiting minimum bounding rectangle (point outside MBR is surely outside polygon)
- LSH-based acceleration (point inside MBR)

**MBR**: minimum bounding rectangle
Geo-Fencing: WINTH

- Point outside MBR but within a distance
  - A rectangle centered at the point, edge length=2 * $d_{th}$
    - Non-overlap, point surely not WITHIN a distance
Geo-Fencing: Scalable Framework

- A point inside MBR of a polygon
- Adapt to crossing number algorithm
- A probing scheme to lookup edges

Yi Yu, et al ACM GIS Cup’13

A point inside MBR of a polygon

Adapt to crossing number algorithm

A probing scheme to lookup edges
Geo-Fencing: Point, Polygon, MBR, R-tree

- 15 MBRs, three groups
A separate hash table for each polygon
A fixed number of buckets, N, for each hash table
Hash function
- \( T = \frac{X_{\max} - X_{\min}}{N} \)
- \( \text{HashKey}(x) = \text{int} \left( \frac{(x - X_{\min})}{T} \right) \)
An edge \((x_1, y_1)-(x_2, y_2)\) stored in buckets from key1 to key2
- key1=HashKey(x1)
- key2=HashKey(x2)
Geo-Fencing: WITHIN Detection

Buckets probed for WITHIN $d_{th}$ of Polygon

- **WITHIN in three cases**
  - Inside polygon $P_1$
  - Inside inner ring $P_2$
  - Outside outer ring $P_3$

- **Optimization $P_3$**
  - Range of a point
  - Divide outer area into 4 ranges
  - Only check edges in the same range with the point

Yi Yu, et al. ACM GIS Cup’13

Socially Aware Location-Based Multimedia Systems – Y. Yu & R. Zimmermann
Geo-Fencing: Evaluation Setup

- **Training dataset**
  - Two point files (point500 with 39,289 instances, point1000 with 69,619 instances)
  - Two polygon files (poly10 with 30 instances, poly15 with 40 instances)
  - Ground truth (different combinations of inputs and predicates)

- **Environment**
  - A laptop PC (Intel Core i5 CPU, 64-bit Windows 7)

- **Two predicates**
  - INSIDE & WITHIN 1000 (by kNN, LSH, R-tree, R-tree+LSH)
  - Accuracy & efficiency (4 methods)
Geo-Fencing: Experiment

- 100% accuracy
- Running time for all points and polygons
  - Without system overhead (file I/O, data conversion)
  - Via QueryPerformanceCounter, 100 runs
Part II

Socially Aware Location-Based Multimedia Systems – Y. Yu & R. Zimmermann
Map Generalization

- Producing maps with less detail
- GIS, business mapping applications
- Reducing data without losing general shape of map

http://openstreetmap.us/~migurski/streets-and-routes/
GIS Cup 2014: Map Generalization

- Deadline (August 1st)
- Co-located with ACM SIGSPATIAL GIS 2014
- Task
  - Access to linear geometries along with points
  - Reducing #vertices while preserving topological constraints
Inputs: linear geometries (27) & constraining points (26)
Two Constraints

- Topological relationships between original set of input linear geometries does not change after simplification.

- Relationship between constraining points and linear geometries before and after simplification does not change.

Two Main Algorithms

- Douglas-Peucker algorithm (1973)
- Visvalingam-Whyatt algorithm (1993)
Douglas-Peucker Algorithm & Example (1973)

- Maximum perpendicular distance at $P_4$ is larger than the tolerance
- Remove $P_2$ and $P_3$ as the maximum perpendicular distance at $P_2$ is smaller than the tolerance
- Remove $P_5$, $P_6$, and $P_7$ as the maximum perpendicular distance at $P_7$ is smaller than the tolerance

(1) Initial simplified line
(2) Tentative simplified segment
(3) Final version of simplified line
Pseudo Codes for Douglas–Peucker Algorithm
--using a given threshold epsilon--

function DouglasPeucker(PointList[], epsilon)

    // Find the point with the maximum distance
    dmax = 0, index = 1, end = length(PointList)
    for i = 2 to (end - 1) {
        d = shortestDistanceToSegment(PointList[i], Line(PointList[1], PointList[end]))
        if ( d > dmax ) { index = i, dmax = d }
    }

    if ( dmax > epsilon ) { // Recursive call
        recResults1[] = DouglasPeucker(PointList[1...index], epsilon)
        recResults2[] = DouglasPeucker(PointList[index...end], epsilon)
        // Build the result list
        ResultList[] = {recResults1[1...end-1] recResults2[1...end]}
    } else {
        ResultList[] = {PointList[1], PointList[end]}
    }

return ResultList[]
Visvalingam-Whyatt Algorithm & Example (1993)

- Remove $P_2$ as its effective area is minimal
- Iteration, effective areas of points adjacent to the removed one are recalculated
- Remove $P_6$ as its effective area is minimal

Original line

1. Effective area for $P_2$
2. Effective area for $P_6$
3. Simplified line with six vertices
Pseudo Codes for Visvalingam-Whyatt Algorithm

- Iteratively remove the point with least effective area
- function VisvalingamWhyatt(PointList[], number_to_keep)

```python
ResultList = PointList.clone()
while length(ResultList) > number_to_keep
    // Find the point with least effective area
    minArea = infinity, minIndex = 1
    for i = 2 to i = length(ResultList) - 1
        Comp. effectiveArea[i] of ResultList[i], (triangle by ResultList[i-1, i, i+1])
        if minArea > effectiveArea(i) {
            minArea = effectiveArea(i)
            minIndex = i
        }
    // Remove the point with least effective area
    ResultList.remove(minIndex)
return ResultList
```

Iteratively remove the point with least effective area
Simplification

- Maintain the topological consistency & satisfy point constraint

Before

After
DrawSimResult → GF_PROG = ".\ProgramMultithread_Dist\Simplify"
Do not present our proposed map generalization algorithm in this tutorial.
Part II

Social Networks

Content Sharing

Content Delivery Network

Distributed cache

Geo-fencing

Map generalization

Geo-social behaviors

User preference profiling

Personalized activity logs (e.g., online, physical)

User

User

Physical Location
User Preference Profiling: Background

❖ User modeling
  ▪ An understanding of a user (characteristics, preferences, needs)

❖ Modeling user location history
  ▪ Provide personalized services
  ▪ Geographic data ➔ semantic geo-category
    • (e.g., coordinates ➔ bar, shopping mall)
User Preference Profiling: Background

- Location-aware user profiling (approaches)
  - Term Frequency-Inverse Document Frequency (TF-IDF)
  - Sparse Additive Generative model (SAGE)
  - Latent Dirichlet Allocation (LDA)

- Examples
  - Social check-ins & location preference
  - Check-in patterns & shopping habit
  - Topical diversity, geographical diversity, interest diversity
    - Tweet prediction
    - Location inference
Check-ins & Preference

Location history (document), categories (terms)
User preference hierarchy (TF-IDF)

\[ u.w_c = \frac{\{u.v : v.c = c'\}}{|u.V|} \times \log \frac{|R|}{\{u : c' \in u.C\}} \]

- Number of venue \( c' \) being visited by \( u \)
- Number of users visiting venue \( c' \)

Total similarity between \( a \) and \( b \)

\[ w_1 \times \text{sim}_1(u_a,u_b) + w_2 \times \text{sim}_2(u_a,u_b) \]

Category Name | Number of sub-categories
---|---
Arts & Entertainment | 17
College & University | 23
Food | 78
Great Outdoors | 28
Home, Work, Other | 15
Nightlife Spot | 20
Shop | 45
Travel Spot | 14

(a) Overview of a location-based social network
(b) Detailed location category hierarchy in FourSquare
Check-ins & Habit Pattern: Background

- Top 9 categories or 410 sub-categories
  - Trade area analysis
  - Not effective or too high dimensional

- User profiling
  - Histogram of user check-ins (sub-categories)
  - Latent Dirichlet Allocation (LDA)
    - Assumption
      - document: a mixture of topics
      - Topic: probability of mentioning a word
    - Goal: calculate proportion of documents by examining word distribution
Check-ins & Habit Pattern

- Distribution of topics (users)
  - User (document)
  - Topic (term)
- Store profile: histogram of topics
  - Generated by all customers
Preliminaries in Log Space

- For a term $v$ in a model $\phi$
  - Term frequency is $\beta_v$
  - Log-frequency is $\phi_v = \log \beta_v$
  - Get distribution via normalization
  
  $$p(v \mid \phi) = \frac{\beta_v}{\sum_{v} \beta_v} = \frac{\exp(\phi_v)}{\sum_{v} \exp(\phi_v)}$$

  
  $$p(v \mid \phi^0 + \phi^u + \phi^g) = \frac{\exp(\phi^0_v + \phi^u_v + \phi^g_v)}{\sum_{v} \exp(\phi^0_v + \phi^u_v + \phi^g_v)}$$

  
  $$\beta_v = \beta^0_v \cdot \beta^u_v \cdot \beta^g_v$$

- $\phi^0(\beta^0)$: Basic reference model
- $\phi^u(\beta^u)$: Difference between one model and the reference model
- $\phi^g(\beta^g)$: Difference between another model and the reference model
Sparse Additive Generative Model

- SAGE models the difference from a background distribution in log-space
  - Use $\phi^0$ to denote the background model
  - Other components $\phi^u$, $\phi^g$ used to model the differences from the background model
  - Sparsity-inducing for each specific model (difference of term frequency)
  - Generative facets combination through simple addition in log space

![Diagram](image-url)
A probabilistic model considers
- User preferences, global topics, regional language models, term distribution

\[ \eta_u^u: \text{user dep. distr. over regions} \]
\[ \eta^0: \text{global distr. over regions} \]
\[ \theta^0: \text{global distribution over topics} \]
\[ \theta_u^u: \text{user dependent distribution over topics} \]
\[ p(r|\eta^0 + \eta^u) \]
\[ \theta^r: \text{regional distribution over topics} \]
\[ p(z|\theta^0 + \theta^r + \theta^u) \]
\[ p(w|\phi^0 + \phi^r + \pi) \]
\[ \pi: \text{topic matrix, each row is a distribution over terms} \]
\[ \phi^r: \text{region-dependence of terms} \]
\[ \phi^0: \text{global distribution over terms} \]
Location Prediction via Tweets

- Only 1% of tweets geo-tagged
- Location prediction for new tweet
  - Based on the words used in the tweet and its authors’ information
  - Most proper region is given

\[ p(r|\eta^0 + \eta^u) \]

\[ \eta^0: \text{global distr. over regions} \]
\[ \eta^u: \text{user dep. distr. over regions} \]
\[ \theta^0: \text{global distribution over topics} \]
\[ \theta^r: \text{regional distribution over topics} \]
\[ \theta^u: \text{user dependent distribution over topics} \]

Maximize likelihood

Pick a region \( r \)
Pick a topic \( z \)
User info
Topics
New tweet
Part II

Social Networks

Content Sharing

Content Delivery Network

Distributed cache

Physical Location

Geo-fencing
Map generalization
User preference profiling
Geo-social behaviors
Geo-social connection

User

Correlating preference-aware activities (e.g., online, physical)
Correlating Preference-Aware Activities

- From a user-centric point of view
  - Extract activity data
    - E.g., online presence, physical check-ins
  - Category by semantic concepts
Personalized Video Soundtrack Recommendation

Rajiv, Ratn Shah, Yi Yu, et al. accepted as full paper by ACM MM'14

(1) Raw UGV
(2) Scene Mood Recognition
(3) Soundtrack Recommendation
(4) Personalization
(5) Ready for Sharing

Models $M_G, M_F$

$\text{SVM}^{hmm}$ Training

Training dataset with Geo-tagged video

Visual feat., F
Geo-feat., G

Mood annotation

Listening history

E.g.: Mp3 Files, Tags

Music

Server side

Smart phone app

Offline

Online

External Data Source

foursquare

YouTube

lost.fm

Socially Aware Location-Based Multimedia Systems – Y.Yu & R. Zimmermann
Mood Recognition from Video

Rajiv, Ratn Shah, Yi Yu, et al accepted as full paper by ACM MM’14

A video segment $V$ with geo-sensor info and visual content

- Geo-sensor info $\rightarrow$ geo-category $\rightarrow$ geo feature $G$ $\rightarrow$ scene mood $C_G$
- Visual content $\rightarrow$ visual feature $F$ $\rightarrow$ scene mood $C_F$
- Late fusion of $C_G$ and $C_F$ $\rightarrow$ scene mood $C$
Matching Songs with User Preference

Rajiv, Ratn Shah, Yi Yu, et al. accepted as full paper by ACM MM’14

Scene mood C

Soundtrack retrieval

Initial song list

Personalization by correlating audio features

User specific songs

Mood tag Songs

Music database

With mood-tag as index

Sweet

Funny

Sad

Listening history

last.fm

YouTube

Hollywood
Part II

Socially Aware Location-Based Multimedia Systems – Y. Yu & R. Zimmermann
Sensing knowledge: Background

- Geo-social data
  - User behaviors in physical world
  - Social activities

- Dynamics of user activities

- Locally social impacts
  - Location recommendation
  - Business attractiveness
Participatory Sensing

http://en.wikipedia.org/wiki/Participatory_sensing
http://www.mobilizingcs.org/about/participatory-sensing

- Concept of communities
- Using personal mobile devices and web services to systematically explore interesting aspects
  - E.g., urban computing, public health, cultural identification, mobile multimedia computing
- E.g., Foursquare, Instagram
Sensed Data in Foursquare

- Vast volumes
  - E.g., check ins, venue photos, venue tips
- Valuable knowledge
  - Associated with geographic category (e.g., beach, food)

Foursquare helps you find the perfect places in New York to go with friends:

- **The Metropolitan Museum of Art**
  - 1000 5th Ave New York, NY
  - People talk about: egyptian, american wing, mcqueen exhibit
  - People also say (935 tips):
    - The Wall Street Journal: "Hanging out on the Met steps is a New York tradition, and billionai..."

- **Frying Pan**
  - Bar 205 12th Ave New York, NY
  - People talk about: bucket of coronas, sangria, summer
  - People also say (251 tips):
    - Christina Ray: "Poke around the rusty, creepy depths of the Frying Pan (docked at Pier 66 in..."

- **Eataly NYC**
  - Gourmet Shop 200 5th Ave New York, NY
  - People talk about: gelato, prime rib sandwich, birreria
  - People also say (739 tips):
    - nymag: "In his NY Diet, Jeffrey Steingarten calls Eataly "spectacular" - there are "wonderful,"
Implications of Sensed Data in Foursquare

- User activities in physical world & sharing activities on Internet
  - Personalized information & social behaviors
- Applied to research
  - Location-aware recommendation
    - E.g., providing media advertisement, travel plan
  - Urban computing
    - E.g., providing sustainability and outlook of urban environment, people life quality, city planning, social sciences
Check ins in Foursquare

- An intersection of virtual social networks and physical world
- As of December 2013, 45 million registered users with 5 billion check-ins
  - Spreading the world about their favorite spots

http://en.wikipedia.org/wiki/Foursquare
Photos in Foursquare

Northern Spy Food Co.
611 E 12th St (between Avenue A & B), New York, NY 10009
New American Restaurant, American Restaurant

Business storefronts and interiors (e.g., restaurant) and service contents
Tips in Foursquare

Northern Spy Food Co.
New American Restaurant and American Restaurant
511 E 12th St (btwn Avenue A & B), New York, NY 10009

People talk about:
"... do they have the best kale salad on their brunch menu, but they also keep..." (36 tips)
"Eggs over polenta and biscuits and gravy, get them both!" (35 tips)
"... w apple butter and pork sticky rolls are addictive. Wash it all down w a shandy...." (8 tips)

Kale salad and roasted chicken are musts for dinner
Steven Y - 2 weeks ago

The biscuits are to die for! They come with seasonal preserves and they are yummy.
Alp Ozcelik - 2 weeks ago
Empirical Observation of User Activities in Foursquare: Motivation

Yi Yu et al

Release data and source codes used in this study

<table>
<thead>
<tr>
<th>City</th>
<th>#Venues</th>
<th>#Users</th>
<th>#Check-ins</th>
<th>#Tips</th>
<th>#Photos</th>
</tr>
</thead>
<tbody>
<tr>
<td>LA</td>
<td>63,991</td>
<td>166,922</td>
<td>39,693,415</td>
<td>321,386</td>
<td>906,820</td>
</tr>
<tr>
<td>NYC</td>
<td>126,658</td>
<td>341,545</td>
<td>109,230,334</td>
<td>890,750</td>
<td>1,821,591</td>
</tr>
</tbody>
</table>

- Distribution and relationship
- Observe interesting things behind data
Empirical Observation of User Activities in Foursquare: Data Description

- **A Foursquare venue**
  - A physical location
    - E.g., Union Square Park (Outdoors & Recreation--Park), Times Square (Outdoors & Recreation--Plaza), John F. Kennedy International Airport (Travel & Transport--Airport)

- **Two regions (NYC & Los Angeles)**
  - 10 primary categories
    - Arts & Entertainment, College & University, Event, Food, Nightlife Spot, Outdoor & Recreation, Professional & other places, Residence, Shop & Service, and Travel & Transport

Release data and source codes used in this study

Yi Yu et al
Distribution of tips, photos, and check-ins

1. Arts & Entertainment
2. College & University
3. Event
4. Food
5. Nightlife Spot
6. Outdoor & Recreation
7. Professional & other places
8. Residence
9. Shop & Service
10. Travel & Transport

- Tow similarities:
  - Similar in different regions
  - Similar trend in different categories
Dynamics of Sharing Activity in Foursquare

More popular at weekends than weekdays

More professional activities on weekdays

More photos on weekend

More tips on weekdays

Socially Aware Location-Based Multimedia Systems – Y. Yu & R. Zimmermann
CCDF for #tips, #photos, and #checkins at Venues in Foursquare

- Almost half venues have only one tip while few venues have more than 100 tips.
- A common trend is that only a few venues have a large number of events.

CCDF: Complementary cumulative distribution functions
Distribution of inter-visit time and inter-visit distance

- **Inter-visit time**: time interval between two successive events
  - 50%: tips, 7.3 days, photos, 1.83 days; average: tips, 50.0 days, photos, 17.3 days

- **Inter-visit distance**: distance between two venues successively visited
  - 50%: tips, 3.72 km, photos, 4.03 km; average: tips, 6.10 km, photos, 6.67 km
Initial Observations in Foursquare

- Category dynamics among venue photo sharing, tip sharing, and check ins have analogous geo-temporal rhythms
- Shared venue photos are highly relevant to food
- Users prefer to share photos rather than tips
- Venue photos are more important in promoting Venues
Discussion on This Study of Foursquare

- The potential applications
  - Location recommendation
  - Trip planning
  - Media advertising
  - Urban environment improvement
  - City operation

- Distributing our data and source codes under request of research purposes

- Collecting data from broader regions
  - More detailed usage of sensed data
  - Prediction models
Local Experts of Geo-Category

Iterative model for social expertise discovery

- Hub score of user u’s in category c,
  \[ u_c.h = \sum_{v \in v_c} v_c.a \]
- Authority score of Venue v,
  \[ v_c.a = \sum_{u \in u_c} u_c.h \]
- Iteratively identifying local experts

Multiple users

Scores for categories

Authority nodes (venue)

Hub nodes (user knowledge)

Mutual Inference (HITS)

Jie Bao, et al ACM GIS’12
Competitor Venue Analysis

- **Gravity model for business attractiveness**
  - Probability of venue $V_j$ being visited

\[
P(V_j) = \frac{\sum_{i=1}^{n} a_{ij}}{\sum_{i=1}^{n} \sum_{j=1}^{m} a_{ij}}
\]

- $V_1 \ldots V_m$
- $C_1 a_{11} \ldots a_{1m}$
- $\vdots \vdots \vdots$
- $C_n a_{n1} \ldots a_{nm}$

$a_{ij}$: the number of visits of customer $C_i$ to venue $V_j$
Part II

Social Networks

Content Sharing

Physical Location

User

Geo-fencing
Map generalization
User preference profiling

Geo-social behaviors
Geographic distribution of social events

Distributed cache

Content Delivery Network

Content Sharing

Social Networks

User

Geographic distribution of social events
Personalized activity logs (e.g., online, physical)
Geographic Distribution of Social Events: Background

❖ Geo-social data
  ▪ User behaviors in physical world
  ▪ Geographic reach and interest
    • Geo-category context
    • Multimedia content

❖ Locally social influence and methods
  ▪ View focus and entropy
    • Geographic YouTube video popularity
  ▪ Cumulative distribution function
    • Geographic friendship
YouTube Video Background

- Geographic locality of interest (views)
  - Only watched in a confined, local region

- Measuring local views
  - A highly localized popularity
    - Non-uniform distribution of views
    - Large fraction of views in a few regions
  - A global popularity
    - Uniform distribution of views
View Focus

- Distribution of views of a video over different regions

- View focus: highest fraction of views
  - A high view focus $\rightarrow$ views limited to a specific region

$$F_i = \max_k \frac{V_{ik}}{V_i}$$
View Focus

- **40% YouTube videos**
  - 80% views in a single region

- **Increase of views**
  - More views, watched in more regions
  - View focus linearly decreases with order of magnitude

CDF: cumulative distribution function, Pr(x<th)
View Entropy

- Distribution of views over different regions
- A standard entropy (fraction, probability)
  - Higher value (spread more uniformly)
  - Lower value (focused in few regions)

\[ H_i = - \sum_k p_{ik} \log_2 p_{ik} \]

\[ p_{ik} = \frac{v_{ik}}{V_i} \]

Fraction of video views

PDF

r_1 r_2 r_3 r_4 r_N
View Entropy

- 40% YouTube videos: view entropy lower than 1 bit
- View entropy grows linearly, Reach a plateau (3.5 bits)

CDF: cumulative distribution function
Social Sharing

- Different paths of finding videos
  - Via video sharing website
  - Through social connections
    - A significant impact on video views

- Social ratio
  - Fraction of views due to social sharing
Social Sharing

- **Low level**
  - Video spreading

- **Predominant**
  - **Social diffusion**
    - Constrained by geographic distance
    - Within boundaries of a single geographic region

Anders Brodersen, et al. WWW’12
Social Graph (Check-ins)

- **2-dimensional surface**
  - Great-circle distance
    - Calculated from geographic coordinates

- **Social link**
  - Nodes and links

![Social Graph (Check-ins) Diagram](image)
Friendship Distance (Check-ins)

- Online friendship with geographical locality?
  - Distance between users larger than distance between friends
  - With a short distance two users are more likely to build a social tie

[Diagram showing the Cumulative Distribution Function (CDF) of distance for friends and users compared across different platforms: Brightkite, Foursquare, Gowalla.]
Part II

Social Networks

Content Delivery Network

Content sharing

Geo-social behaviors

Geo-social connection

Map generalization

Geo-fencing

User preference profiling

Personalized activity logs (e.g., online, physical)

User

Physical Location
Multimedia Content Diffusion: Background

- **Public media coverage**
  - Globally popular items
    - Content delivery network

- **Social networking platforms**
  - Spreading through social connections
    - 700 YouTube video links are shared on Twitter every minute

http://www.streamsend.com/
Geographic Social Network

- Given a graph $G = (N, K)$ and geographic location of nodes
  - Place all nodes in a 2D metric space
  - Great-circle distance

$L_i = \frac{1}{k_i} \sum_{j \in \Gamma_i} e^{-\frac{l_{ij}}{\beta}}$

- Node locality
- Distance
- Network scaling factor, e.g., 2000km

how a user’s friends are close to this user
Geographic Followers

- **Cascade of Twitter messages**
  - On the author’s personal page
  - Author’s followers

- **Node locality distribution**
  - 80% users with node locality larger than 0.1
    - A set of followers in a much larger region

![Node locality distribution graph with 80% at node locality 0.1](image)
Cascades of YouTube Links on Twitter

- Cascade over a social network
  - First user shares content
  - Contacts share same content
- Time delay distribution
  - 40% of tweets (a delay of about 15 minutes)
  - 10% (a delay of around 2 minutes)

![Graph showing time delay distribution](image)

Links to videos quickly spread
Geo-Social Content Delivery

- Content delivery network
  - Holding copies of data items
  - Different geographic locations

- Performance of CDNs
  - Geographical distribution of requests
  - Popular on a planetary or in a particular geographic area

![Diagram showing content delivery network with globally popular and locally popular cached content.]
Caching Policies

Basic cache policies

- Least-Recently-Used (LRU): \( P(v) = \text{clock}(v) \)
  - Involve the simple aging effect
- Least-Frequently-Used (LFU): \( P(v) = \text{Freq}(v) \)
  - \( \text{Freq}(v) \) is the number of times video \( v \) has been requested since stored in the cache
- Mixed policy: \( P(v) = \text{clock}(v) + \text{Freq}(v) \)
  - Combines both LRU and LFU features to balance both temporal and popularity effects
Augment Content Caching

P(v) = clock(v) + Freq(v) + extra weight

- Geosocial
  - Extra weight of video v is added by the sum of the node locality values of all the users that have posted a message about it

- Geocascade
  - Extra weight of video v is added by the sum of the node locality values of all the users participating in the item's social cascade
Performance of Hits

- Hit ratio
  - Percentage of requests whose items are found in a server’s cache

- Assumption
  - Every video contained in a tweet is requested by each follower with a certain probability $P$
    - Every request directed to closest server

L: Mixed strategy + Geosocial weight
R: Mixed strategy + with Geocascade weight
Thank You