Diversified Trajectory Pattern Ranking in Geo-Tagged Social Media

Zhijun Yin[1], Liangliang Cao[1], Jiawei Han[1], Jiebo Luo[2], Thomas Huang[1]

Presenter: Zhao Zhou

Outline

- Motivation
- Problem Formulation
- Framework
- Evaluation
- Conclusion

Motivation

- Social media websites such as Flickr, Facebook host overwhelming amounts of photos.
- In such a media sharing community, images are contributed, tagged, and commented by users all over the world.
- Extra information can be incorporated within social media, such as geographical information captured by GPS devices.

Motivation (Cont.)

flickr from YAHOO!

By andrew mace—
Andrew Mace

This photo was taken on May 9, 2010 in Dumbo, New York, NY, US, using a Nikon D90.

28,163 117 450 26

This photo belongs to

andrew mace—'s photostream (585)

This photo also appears in

- the getty collection (set)
- above ground level (set)
- New York City (set)
- interestingness (set)
- New York City (group)
- Tokina AT-X 124 (12-24 f/4) (group)
- Nikon Capture NX Users Group (group)

Sunset over the Brooklyn Bridge and Lower Manhattan, New York City

Sunset from the Brooklyn-side tower of the Manhattan Bridge (here's the view at night). Clouds to the west blocked any sunset colors, but there was a brief moment where the sky really lit up as the sun peeked through the cloud cover and set right between two towers (so I cranked the aperture to f/18). I bracketed exposures to capture more of the dynamic range of the scene. Exposures blended with Photomatix. I used LiveView to focus, which, if there is enough light, does a better job than phase detection and can control for any back- or front-focusing.

Blended exposures (-2/0/+2) / f/18 / 24mm (36mm=)

Motivation (Cont.)

4.2 million things geotagged this month

Motivation (Cont.)

- Explore the common wisdom in photo sharing community
- Discover trajectory patterns interesting to two kinds of users
  - Some users are interested in the most important trajectory patterns. Ranking
  - Some users are interested in exploring a new place in diverse way. Diversification

Problem Formulation

- Given a collection of geo-tagged photos along with users, locations and timestamps, how to rank the mined trajectory patterns with diversification into consideration.

Framework

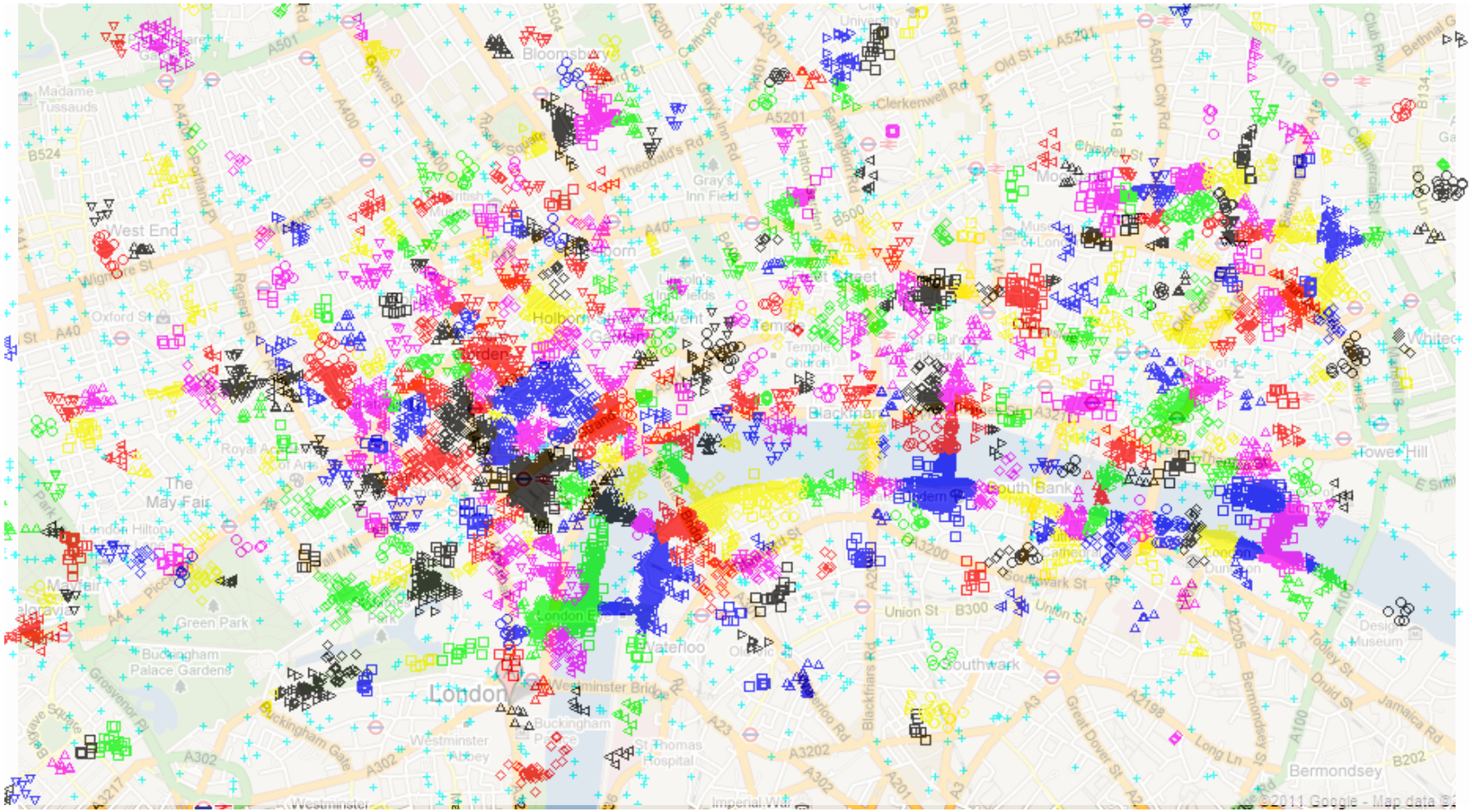
- (1) Extracting trajectory patterns from the geo-tagged photo collection.
- (2) Ranking the trajectory patterns by estimating their importance according to user, location and trajectory relations.
- (3) Diversifying the ranking result to identify the representative trajectory patterns from all the candidates.

Trajectory Pattern Mining

- Since the GPS coordinates of photos are at a very fine granularity, we need to detect locations before extracting trajectory patterns.
- With the detected locations, we can generate the trajectories for each user according to his visiting order of locations during the same day.
- Mine frequent trajectory patterns using sequential pattern mining algorithm.

Location Detection

- Mean-shift algorithm (27974 photos in London)



Location Detection (Cont.)

- Top locations in London and their descriptions. The number in the parentheses is the number of users visiting the place.

londoneye(528), trafalgarsquare(456),
britishmuseum(230), bigben(205), waterloobridge(198),
towerbridge(185), piccadillycircus(182),
royalfestivalhall(175), coventgarden(169),
centrepoint(169), parliamentsquare(150),
cityhall(141), oxfordcircus(138),
lloyds(121), buckinghampalace(107),
naturalhistorymuseum(97), canarywharf(94),
bricklane(91), toweroflondon(91), brighton(90),
embankment(88), soho(80), stpancras(77),
stpaulscathedral(77), leicestersquare(76),
gherkin(75), stjamespark(68), barbican(67),
victoriaandalbertmuseum(64)

Sequential Pattern Mining

- PrefixSpan[1]
- Example (minimum support = 2)

ID	Travel sequence
1	londoneye → bigben → trafalgarsquare
2	londoneye → bigben → downingstreet → trafalgarsquare
3	londoneye → bigben → westminster
4	londoneye → tatemodern → towerbridge
5	londoneye → bigben → tatemodern

  - We can get 3 frequent sequential patterns:
    - *londoneye -> bigben*
    - *londoneye -> bigben -> trafalgarsquare*
    - *londoneye -> tatemodern*

Sequential Pattern Mining (Cont.)

- Top frequent trajectories in London

Trajectory pattern	Frequency
londoneye → bigben	21
bigben → londoneye	19
londoneye → tatemodern	18
londoneye → royalfestivalhall	15
londoneye → trafalgarsquare	14
londoneye → waterloobridge	12
towerbridge → cityhall	12
royalfestivalhall → londoneye	11
tatemodern → londoneye	11
bigben → parliamentsquare	10

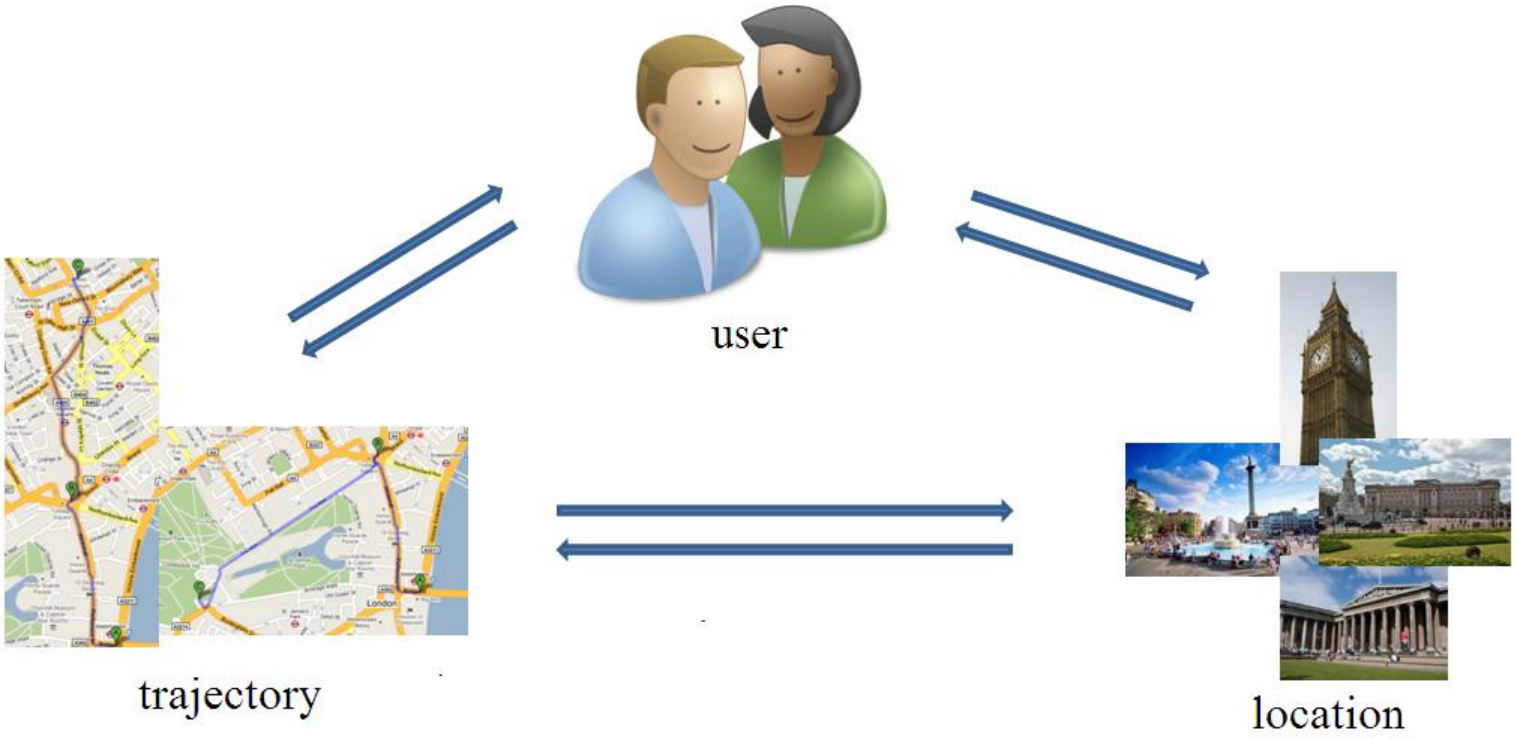
Sequential Pattern Mining (Cont.)

- There are too many trajectory patterns and it is difficult for the users to browse all the candidates.
- Ranking by frequency?
- The top ten trajectory patterns ranked by frequency are of length 2 and not informative.

Ranking

Trajectory Pattern Ranking

- Relationship among user, location and trajectory



Trajectory Pattern Ranking (Cont.)

- A trajectory pattern is important if many important users take it and it contains important locations.
- A user is important if the user takes photos at important locations and visits the important trajectory patterns.
- An location is important if it occurs in one or more important trajectory patterns and many important users take photos at the location.

Trajectory Pattern Ranking (Cont.)

Algorithm 1 Trajectory pattern ranking

Input: M_{TU}, M_{UL}, M_{LT}

Output: A ranked list of tra

1. Initialize $P_T^{(0)}$
2. Iterate
   $P_L = M_{LT} \cdot P_T^{(t)}$ $P_U =$
   $P_T = M_{TU} \cdot P_U$ $P_U = M_{TU}^T \cdot P_T$
   $P_L = M_{UL}^T \cdot P_U$ $P_T^{(t+1)} = M_{LT}^T \cdot P_L$
   $P_T^{(t+1)} = P_T^{(t+1)} / \| P_T^{(t+1)} \|_1$
   until convergence.
3. Output the ranked list of trajectory patterns in the decreasing order of P_T^*, i.e., the converged P_T.

P_T is the eigen vector for M^TM for the largest eigen value, where $M = M_{TU} M_{UL} M_{LT}$. Algorithm 1 is a normalized power iteration method to detect the eigen vector of M^TM for the largest eigen value if the intial P_T is not orthogonal to it.

Trajectory Pattern Ranking

- Top ranked trajectory patterns in London.

Rank	Trajectory pattern	P_T
1	londoneye → bigben → downingstreet → horseguards → trafalgarsquare	0.0037
2	londoneye → bigben → tatemodern	0.0029
3	tatemodern → bigben → londoneye	0.0029
4	londoneye → bigben → parliamentsquare → westminster	0.0028
5	westminster → bigben → downingstreet → horseguards → trafalgarsquare	0.0028
6	royalfestivalhall → londoneye → bigben	0.0027
7	londoneye → royalfestivalhall → tatemodern	0.0027
8	tatemodern → londoneye → royalfestivalhall	0.0027
9	londoneye → tatemodern → towerbridge	0.0027
10	londoneye → towerbridge → tatemodern	0.0027

Trajectory Pattern Ranking (Cont.)

- Top ranked locations in London with normalized P_L scores and frequency.

Location	P_L	# User	Location	P_L	# User
londoneye	0.0157	528	southwark	0.0062	57
trafalgarsquare	0.0125	456	stpaulscathedral	0.0058	77
bigben	0.0121	205	downingstreet	0.0053	52
tatemodern	0.0119	491	horseguards	0.0051	25
royalfestivalhall	0.0093	175	londonbridge	0.0049	37
towerbridge	0.0089	185	embankment	0.0047	23
cityhall	0.0077	141	harrods	0.0047	39
waterloobridge	0.0076	198	toweroflondon	0.0046	91
parliamentsquare	0.0075	150	naturalhistorymuseum	0.0046	97
piccadillycircus	0.0074	182	monument	0.0046	59
britishmuseum	0.0074	230	victoriaandalbertmuseum	0.0045	64
gherkin	0.0073	75	bank	0.0044	63
lloyds	0.0070	121	royalacademy	0.0040	34
coventgarden	0.0070	169	oxfordstreet	0.0040	51
buckinghampalace	0.0064	107	bloomsbury	0.0038	27

Trajectory Pattern Diversification

- The result in top ranked trajectories illustrates the popular routes together with important sites such as *londoneye, bigben*, and *tatemodern*.
- However, it is highly biased in only a few regions.
  - Trajectory 1 (*londoneye -> bigben -> downingstreet -> horseguards -> trafalgarsquare*)
  - Trajectory 5 (*westminster -> bigben -> downingstreet -> horseguards ->trafalgarsquare*)

Diversification

Trajectory Pattern Diversification (Cont.)

- Similar trajectory patterns need to be aggregated together.
- Good exemplars of trajectory patterns need to be selected.
- Those trajectories patterns ranked highly in our ranking algorithm should get higher priority to be exemplars.

Trajectory Pattern Diversification (Cont.)

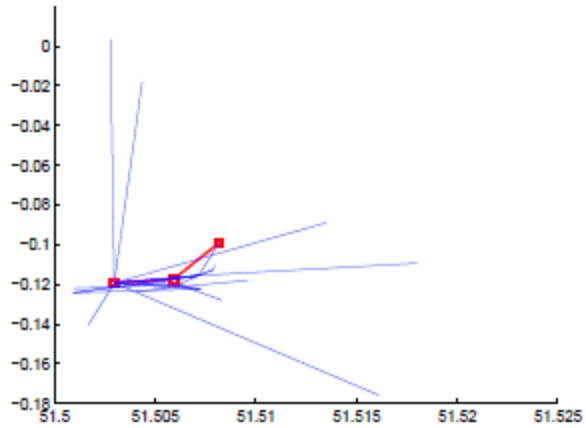
- We define the similarity between two trajectories based on longest common subsequence (*LCSS*).
- The similarity measure *LCSS(i, j)* can be viewed as how well trajectory *i* represents trajectory *j*.
- Suppose trajectory *i* is represented by an exemplar trajectory *r(i)*, we can see that trajectory *i* becomes an exemplar if *r(i) = i*.
- The optimal set of exemplars corresponds to the ones for which the sum of similarities of each point to its exemplar is maximized.

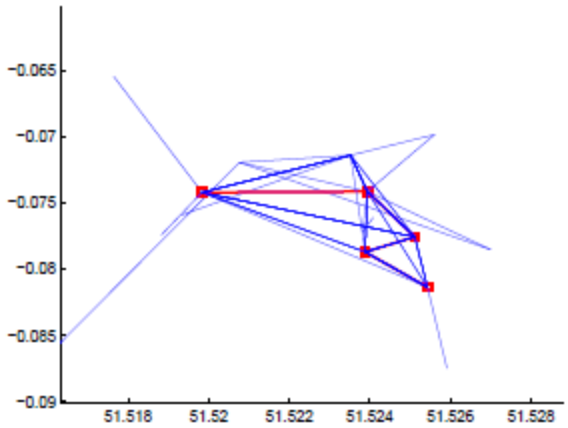
Trajectory Pattern Diversification (Cont.)

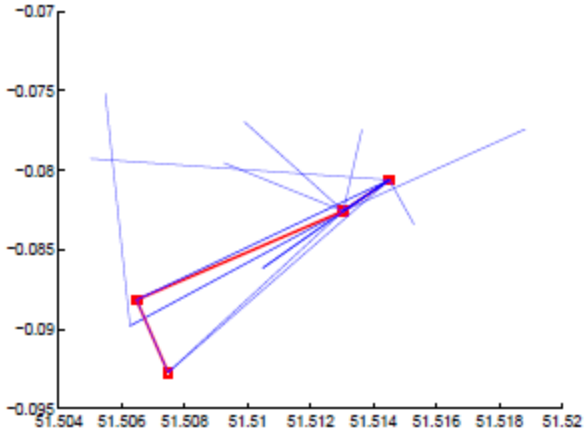
- There are several ways of searching for the optimal exemplars such as vertex substitution heuristic p-median search and affinity propagation.
- Frey and Dueck's affinity propagation[2]: it considers all data points as potential exemplars and iteratively exchanges messages between data points until it finds a good solution with a set of exemplars.
- To incorporate the information of ranking results, we can give higher ranked trajectories larger self-similarity scores in message passing.

Trajectory Pattern Diversification (Cont.)

- Exemplars examples:





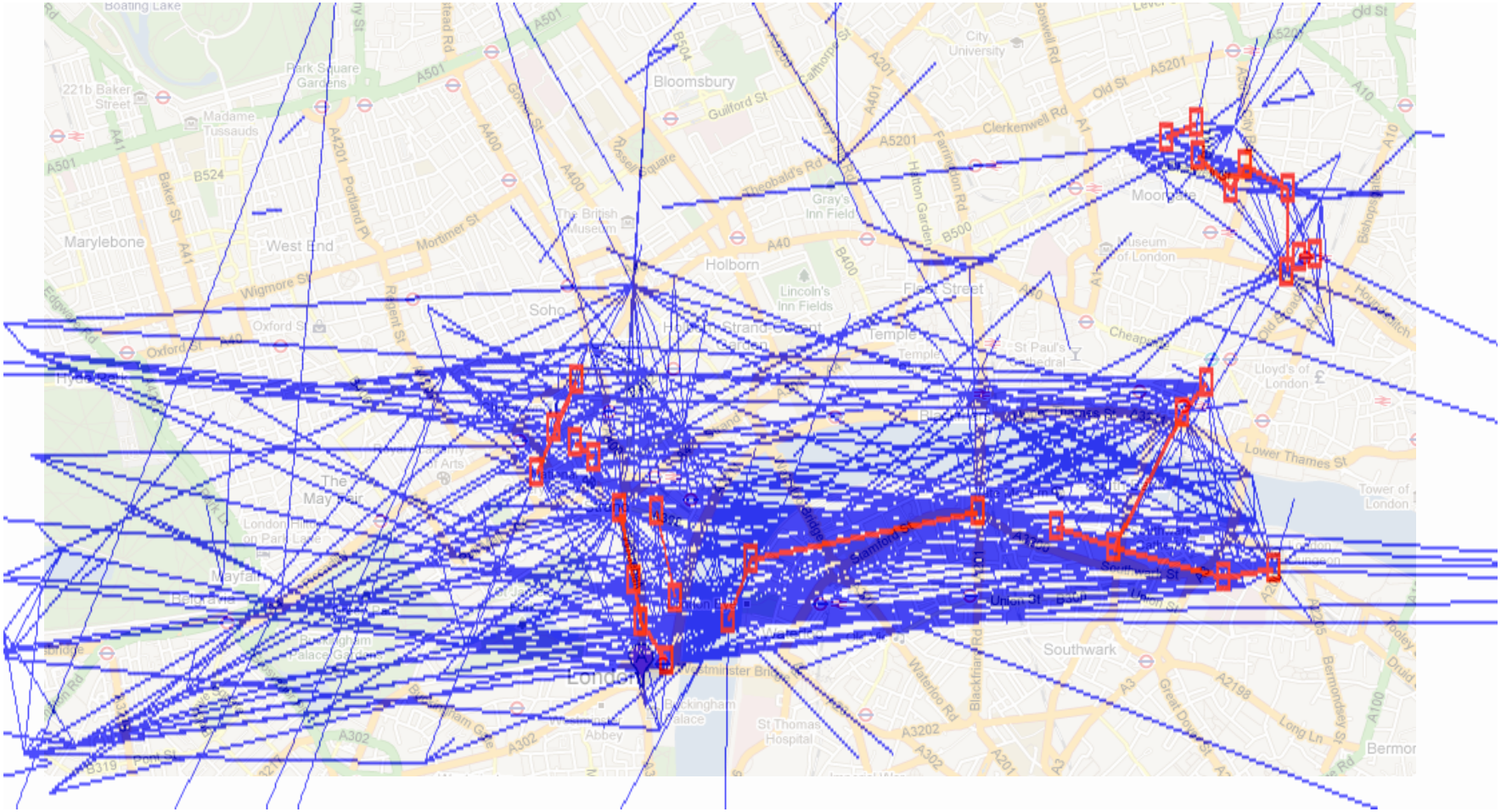


Trajectory Pattern Diversification (Cont.)

- Trajectory pattern diversification results in London.

Rank	Tourist route pattern
1	bigben → downingstreet → horseguards → trafalgarsquare
2	spitalfields → shoreditch(1) → shoreditch(2) → shoreditch(3) → shoreditch(4)
3	charingcross → londoneye
4	bricklane(1) → bricklane(2)
5	londoneye → royalfestivalhall → tatemodern
6	oldstreet(1) → oldstreet(2)
7	piccadillycircus → soho → oldcomptonstreet
8	londonbridge → cityhall → towerbridge
9	gherkin → lloyds → londonbridge → southwark
10	leicestersquare → chinatown

Trajectory Pattern Diversification (Cont.)



Evaluation

- Data Sets
  - We crawled images with GPS records using Flickr API (http://www.flickr.com/services/api/)

City	Photo	User	Loc	Traj	Pat
Barcelona	7764	1799	201	1320	189
Berlin	5401	1242	167	965	136
Chicago	7418	1589	219	1343	164
DC	5091	1368	173	826	123
Los Angeles	7336	2149	208	1095	105
London	27974	5455	883	4712	1202
Madrid	4014	1072	146	651	131
Milan	5200	1239	127	813	44
New York	16365	3836	568	2692	549
Paris	10826	2699	357	1923	620
Rome	6143	1609	158	1086	193
San Francisco	15841	3047	578	2631	245

Evaluation (Cont.)

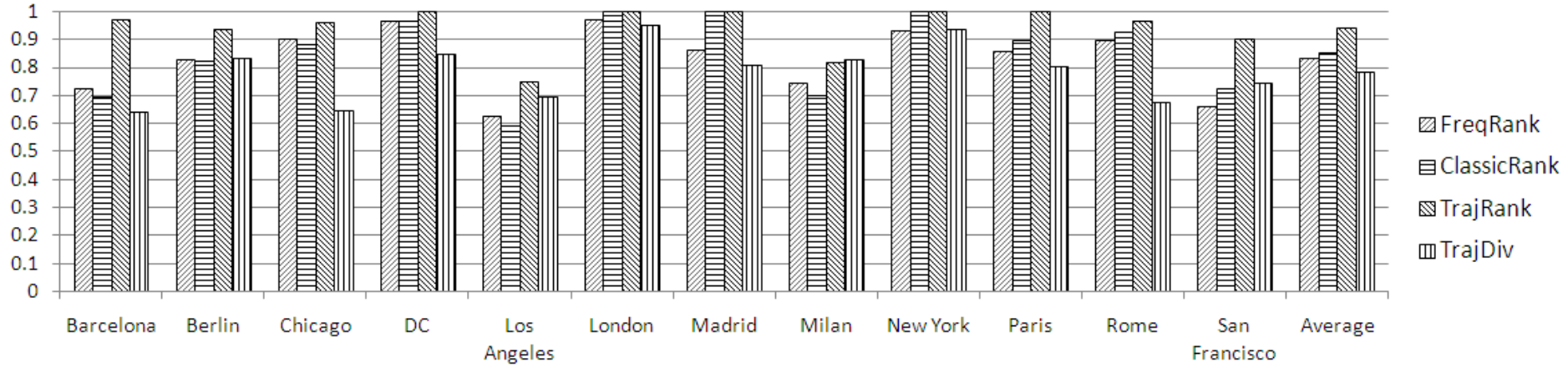
- Compared methods
  - FreqRank: Rank trajectory patterns by sequential pattern frequency
  - ClassicRank: The method used in [3] to mine classic travel sequences. The classical score of a sequence is the integration of the sum of hub scores of the users, the authority scores of the locations.
  - TrajRank: Trajectory pattern ranking
  - TrajDiv: Trajectory pattern diversification

Evaluation (Cont.)

- Measures
  - NDCG (normalized discounted cumulative gain)
    - highly interesting (2), interesting (1), not interesting (0)
  - Location Coverage
    - The number of covered locations in the top results.
  - Trajectory Coverage
    - The summation of the edit distance of each trajectory pattern in the dataset to the closest one in the top result.
    - The score is normalized by the summation of the edit distance of each trajectory pattern to the closest one in the dataset.

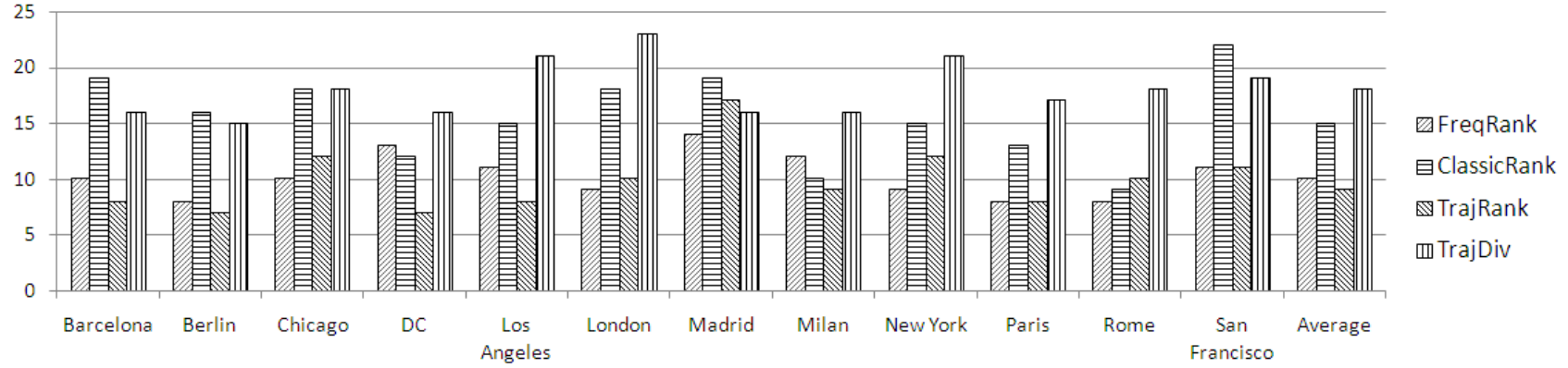
Evaluation (Cont.)

- NDCG

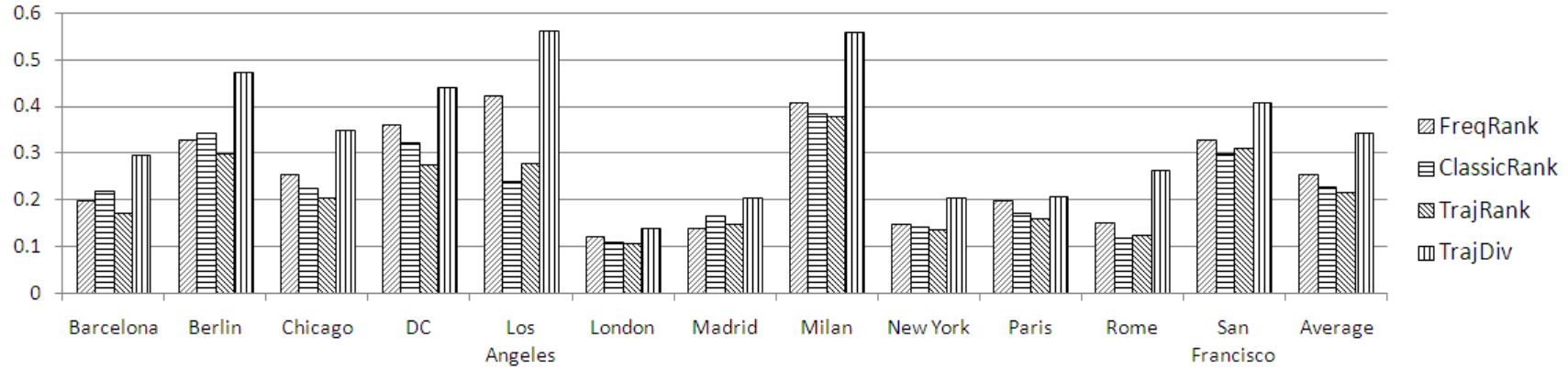


Evaluation (Cont.)

- Location Coverage



- Trajectory Coverage



Evaluation (Cont.)

- London
  - *londoneye -> bigben -> downingstreet -> horseguards -> trafalgarsquare*

Evaluation (Cont.)

- Location recommendation based on current trajectory in London

Current trajectory	Recommended next destination
londoneye	bigben, tatemodern, trafalgarsquare, southbank, parliamentsquare, towerbridge, piccadillycircus, buckinghampalace
londoneye → bigben	downingstreet, horseguards, trafalgarsquare, parliamentsquare
londoneye → bigben → downingstreet	horseguards, trafalgarsquare
londoneye → tatemodern	southbank, towerbridge, piccadillycircus
lodoneye→trafalgarsquare	buckinghampalace

Conclusion

- We studied the problem of trajectory pattern ranking and diversification based on geo-tagged social media.
- We extracted trajectory patterns from geo-tagged photos using sequential pattern mining and proposed a ranking strategy that considers the relationships among user, location and trajectory.
- To diversify the ranking results, we used an exemplar-based algorithm to discover the representative trajectory patterns.
- We tested our methods on the photos of 12 different cities from Flickr and demonstrated their effectiveness.

Thanks!

Reference

- [1] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto, Q. Chen, U. Dayal, and M. Hsu. Mining sequential patterns by pattern-growth: The prexspan approach. IEEE Trans. Knowl. Data Eng., 16(11):1424-1440, 2004.
- [2] B. J. Frey and D. Dueck. Clustering by passing messages between data points. Science, 315:972-976, 2007.
- [3] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma. Mining interesting locations and travel sequences from gps trajectories. In WWW, pages 791-800, 2009.