



Cornell CIS

Building a Science Base for the Information Age

Daniel Huttenlocher
Dean



Cornell University
Faculty of Computing and Information Science

CIS Today

- Computer Science Department
 - Ugrad major about 80/yr in Arts&Engr (growing)
 - About 90 Meng/yr and 20 PhD/yr
- Information Science Program
 - Ugrad major about 35/yr in Arts,Engr&CALS
 - New professional masters and about 5 PhD/yr
- Statistical Science Department
 - New ugrad major in Arts
 - Professional masters about 40/yr
- Lots of cross-disciplinary activity: HCI, Social Networks, Machine Learning, Robotics



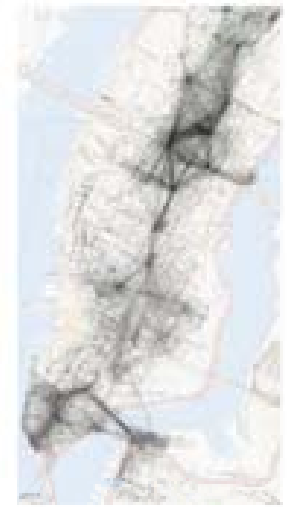
CIS Directions

- Statistics increasingly important in digital age – data mining and machine learning
- Information Science – bridging computing and social sciences
 - Tying CS with Comm, Psych, Soc, Econ, ORIE
 - Nationally ISchools aligning more with CS
- AAP college developing new focus on design
 - Digital media in collaboration with CIS
- Faculty hiring despite tight financial climate
 - Current searches in CS, IS and Stats



New in the Classroom

- Networks, Crowds and Markets
 - David Easley (Econ) and Jon Kleinberg (CS)
 - Sophomore-level distribution with 400 students
 - Pre-publication pdf online
- Interface of online and offline worlds
 - CS1114 robot programming class – version of CS 1110 (fka 100)
 - Rapid prototyping course in Info Sci
- CS 3110 (fka 312 and long ago 212)
 - Map-reduce style functional programming



Trails of Flickr Users
in Manhattan
(Crandall et al. 2009)



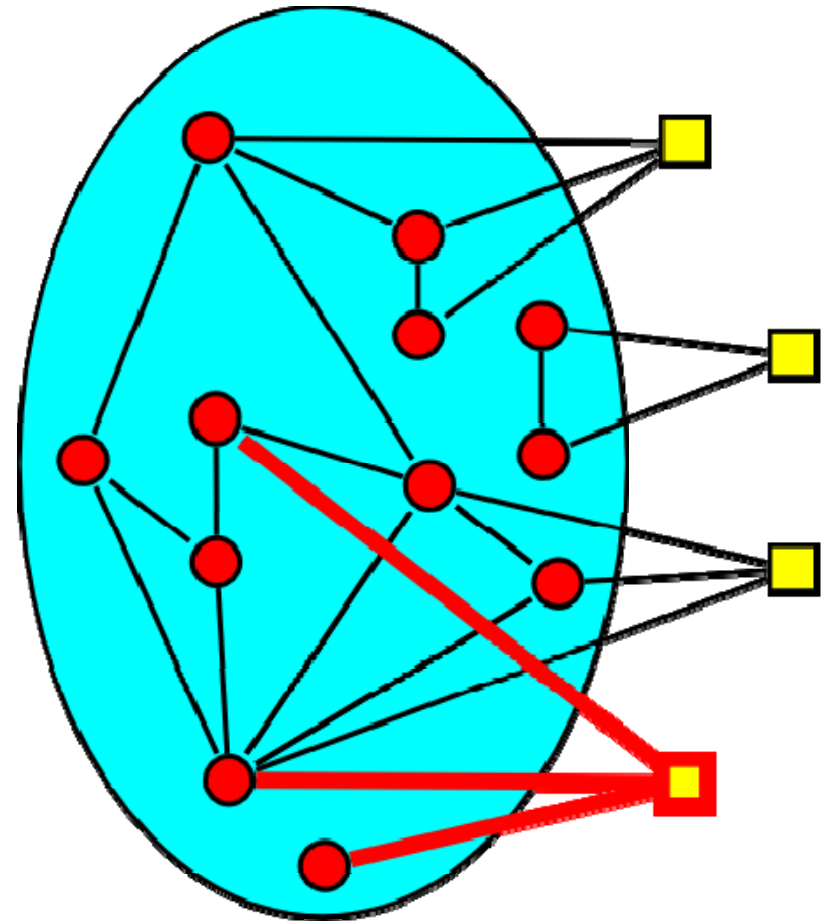
A Taste of Recent Research

- I: Influence in social networks
- II: Meaning of links in social networks – trust /friendship vs. status
- III: Social perception of place – automatically generated “tourist maps”
- Joint work with Lars Backstrom (Facebook), Dan Cosley (Cornell IS), David Crandall (Cornell CS), Jon Kleinberg (Cornell CS), Xiangyang Lan (Goldman), Jure Leskovec (Stanford), Yunpeng Li (Cornell CS & EPFL), Sid Suri (Yahoo)



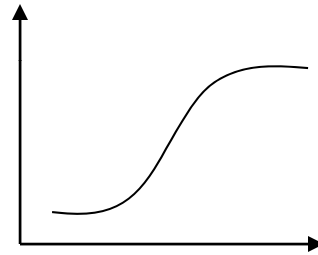
I: Influence in Adopting Activities

- As function of local network
- Red circles represent those in group, yellow squares might join
 - 3 ties vs. 2 ties
- Other structural features
 - E.g. how connected are ties to one another
 - Triads

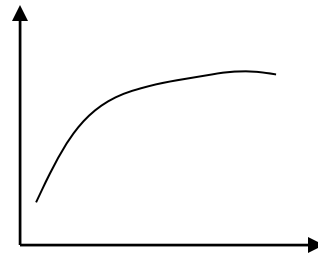


Probability of Adopting New Activity

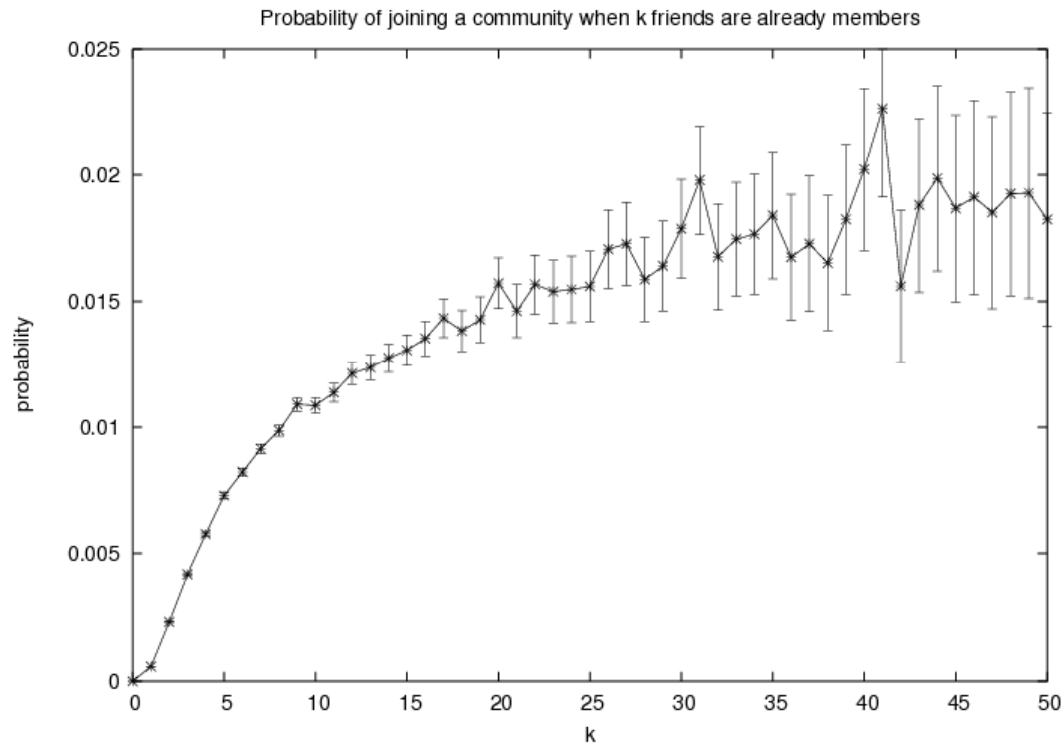
- As function of number of social ties – as opposed to adoption curves based on time
 - S-shaped? Critical mass effect (logistic)



- Concave? Diminishing returns (logarithmic)



Joining LiveJournal Groups



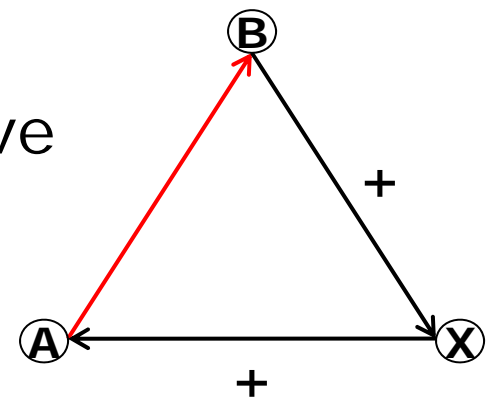
- Supra-linear up to 2, then diminishing returns – “once an accident twice a pattern”
 - 12M users, 250K groups, over 2 month period



II: Meanings of a Link

- Directed edges in network
 - Can indicate relations such as trust and friendship (generally symmetric)
 - Alternatively assessments such as status (generally asymmetric)

- Example of cycle with ++ path
 - Trust/friendship predicts AB positive
 - Friend of friend is friend
 - Status predicts AB negative
 - A higher status than B

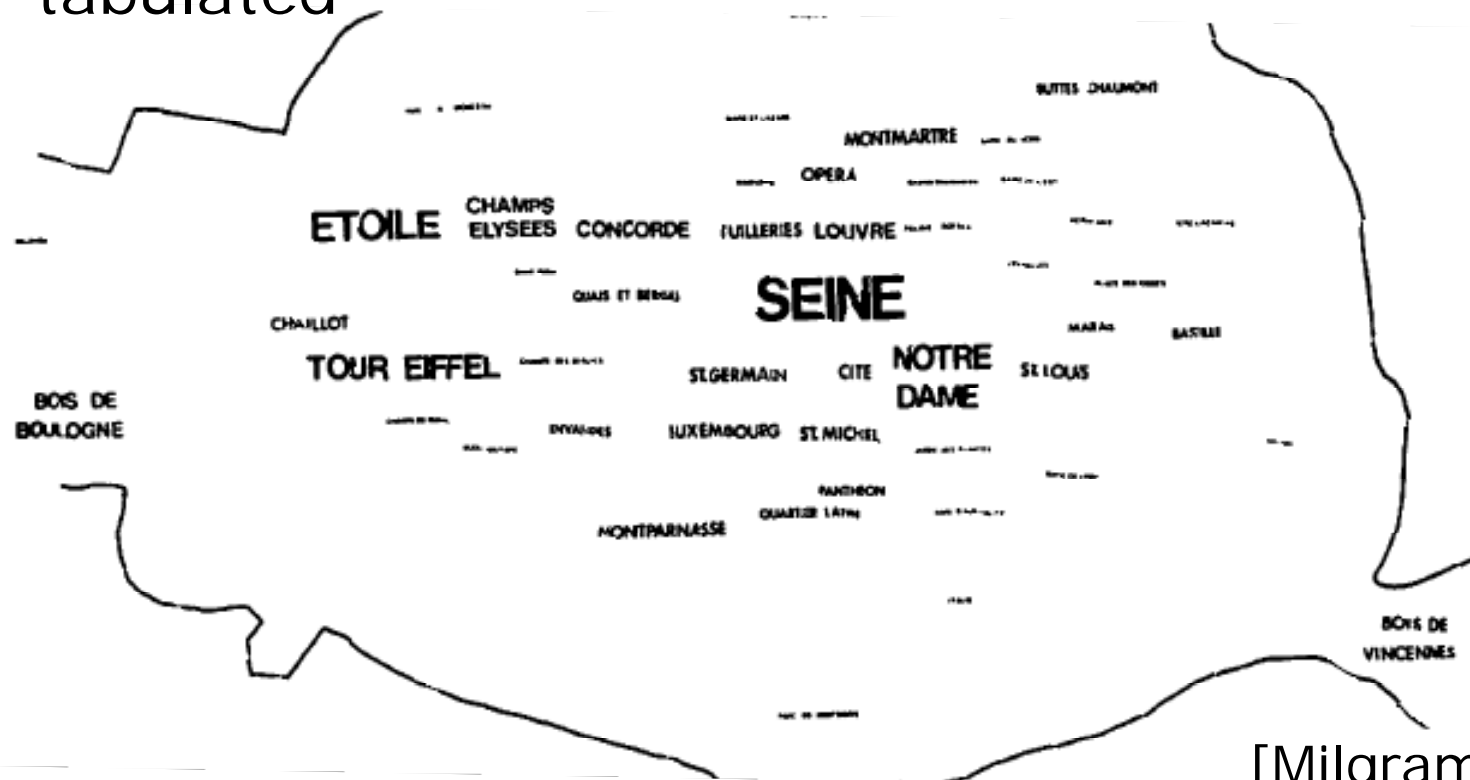


- For Epinions trust network and Wikipedia admin votes, status much better predictor!



Classic Experiment: Hand-Drawn Maps

- 218 subjects each draw map of Paris
 - Total of 4132 elements in maps, hand coded and tabulated

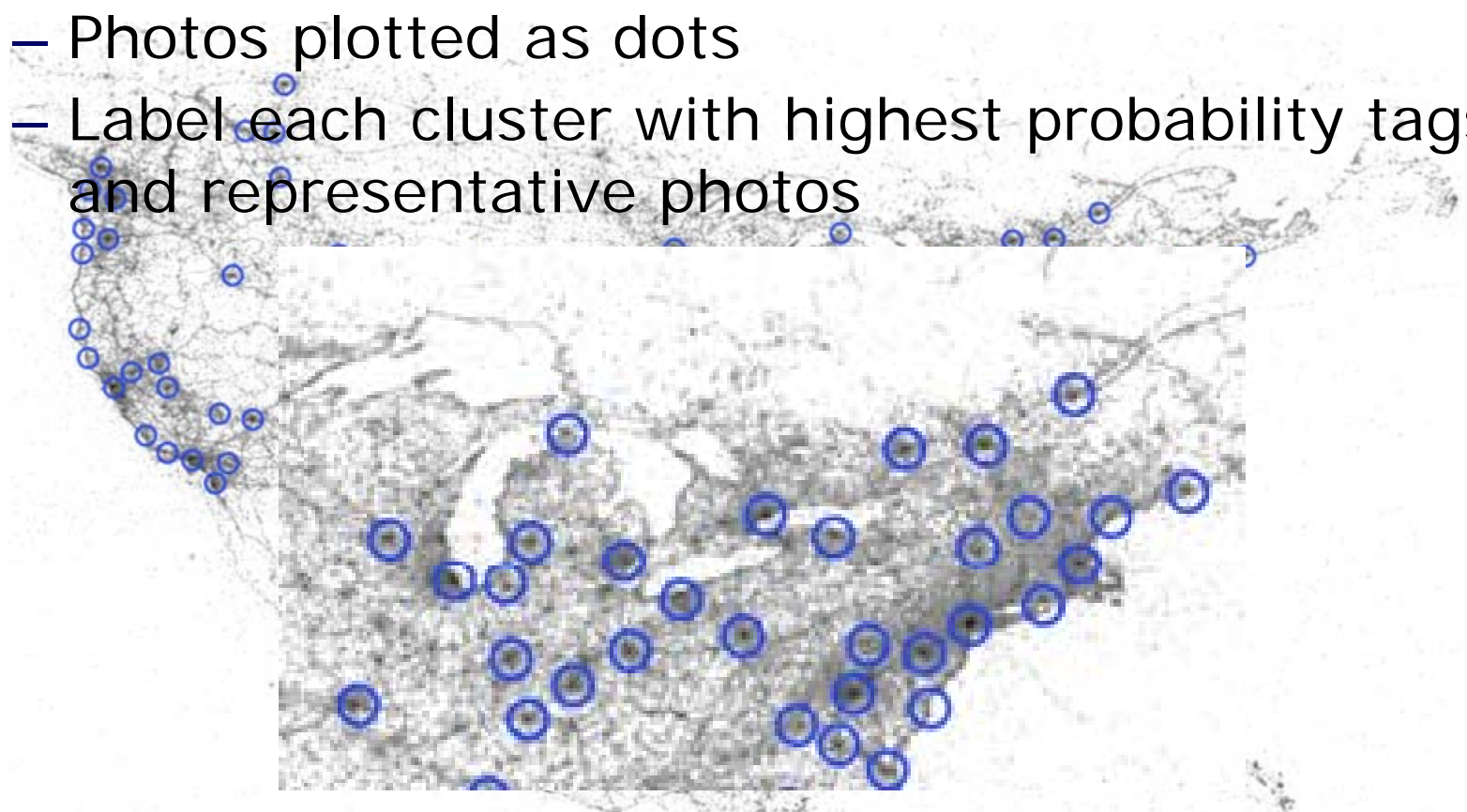


[Milgram76]



Interesting Places from Flickr Photos

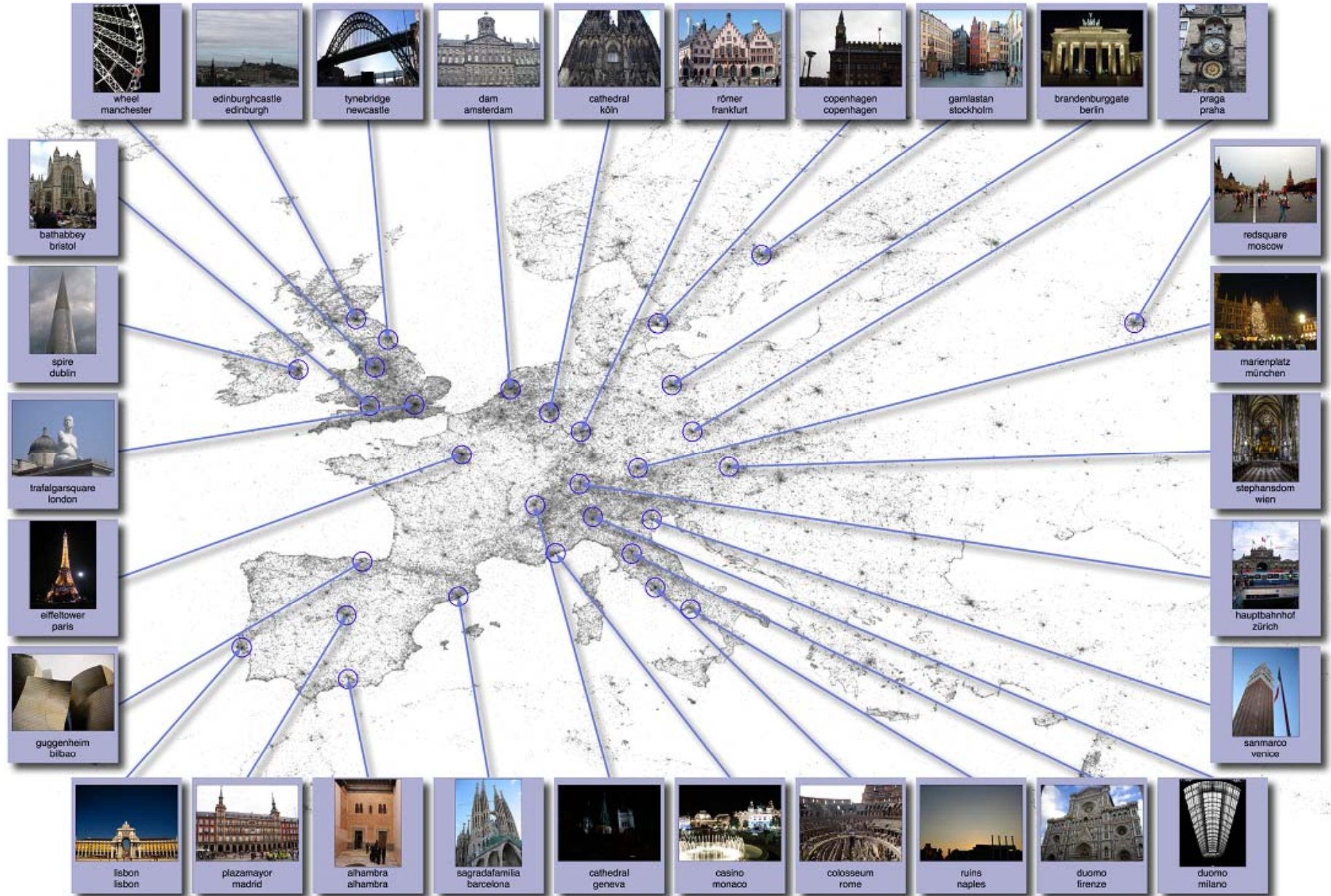
- Top 100 clusters in North America at 100km radius mean shift
 - Photos plotted as dots
 - Label each cluster with highest probability tags and representative photos



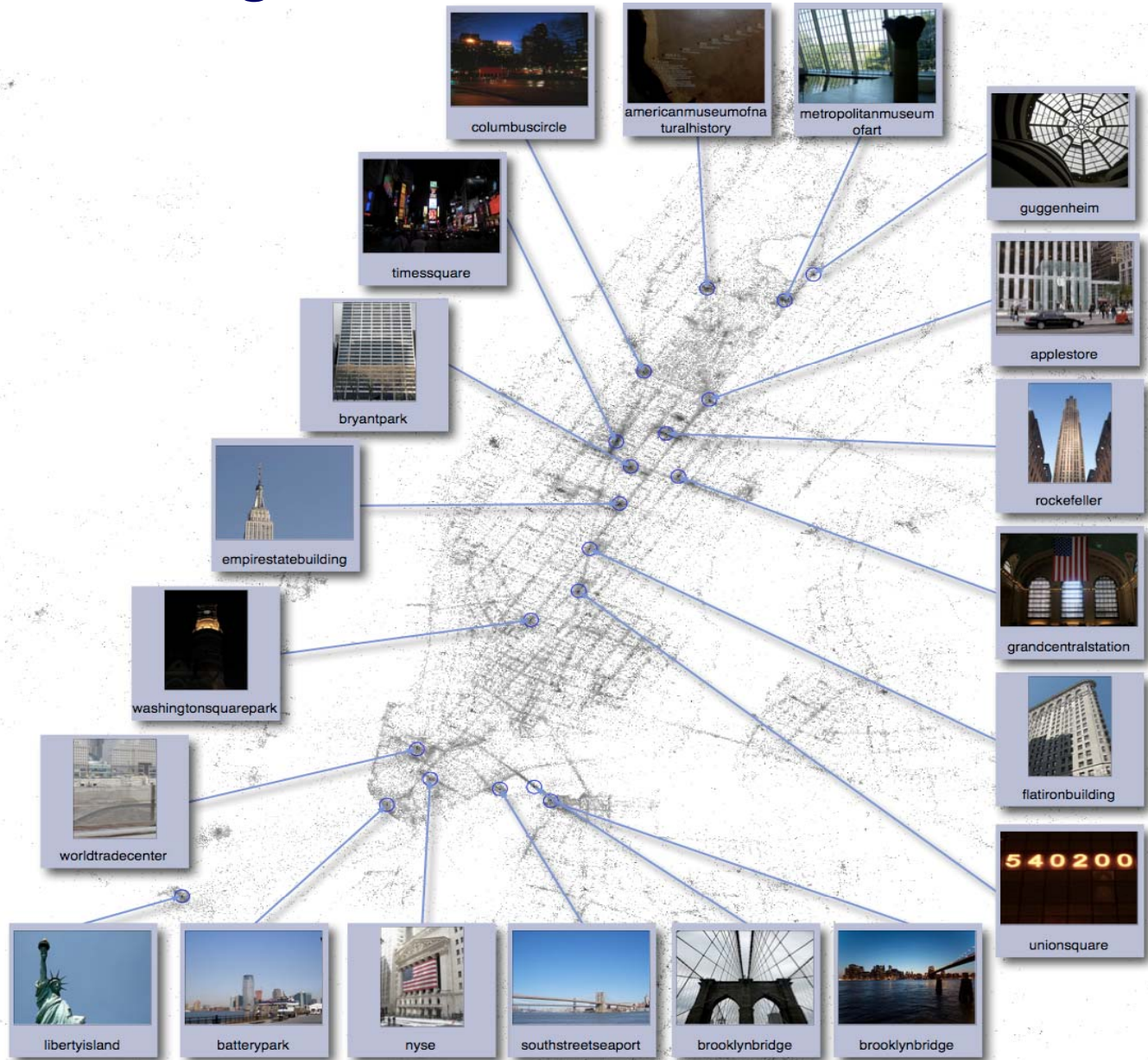
Completely Automatic "Tourist" Maps



Europe



City Scale: Manhattan



Saliency: Places and Events

- Key aspect of photo geotags is that they capture something noteworthy
 - In contrast to always-on location data such as cell phone tracks
- But this high value data can also be very revealing of other (unintended) information
 - People who take photos nearby in time and location generally know each other
 - E.g., if two Flickr users take photos of 6 landmarks within 24 hours and 0.1 degrees (~8km) there is nearly 70% probability they are listed as contacts!
 - Chance only .005% (moreover contacts underreport)



Q&A/Discussion

