COMP5425

MULTIMEDIA RETRIEVAL

THE UNIVERSITY OF SYDNEY

Week07

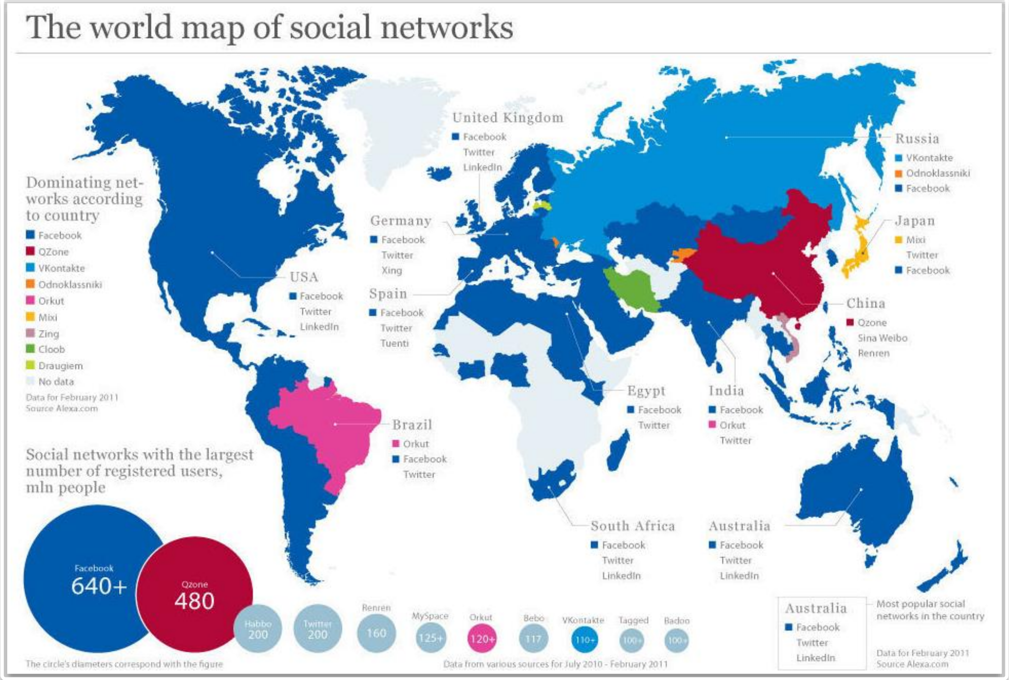
Semester 1, 2025

Social Media

- Background
- Content
  - Text + Multimedia
- Users
  - Profile
  - Context

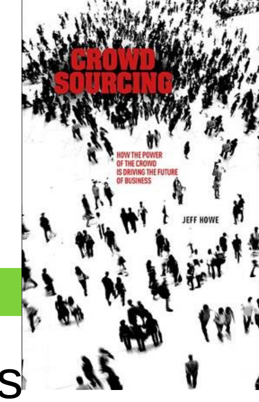
Explosive Worldwide Growth of Social Media

- Network: Facebook, Wechat, WhatsApp
- Blogging/Microblogging: Twitter, LiveJournal, Blogger
- Sharing: Flickr, YouTube, Pinterest, Instagram, TikTok
- News: Digg, Slashdot
- Bookmarking: Delicious
- Knowledge: Wikipedia
- Shopping: Groupon
- Location: Foursquare
- Games: Zynga, Playfish
- Career: LinkedIn

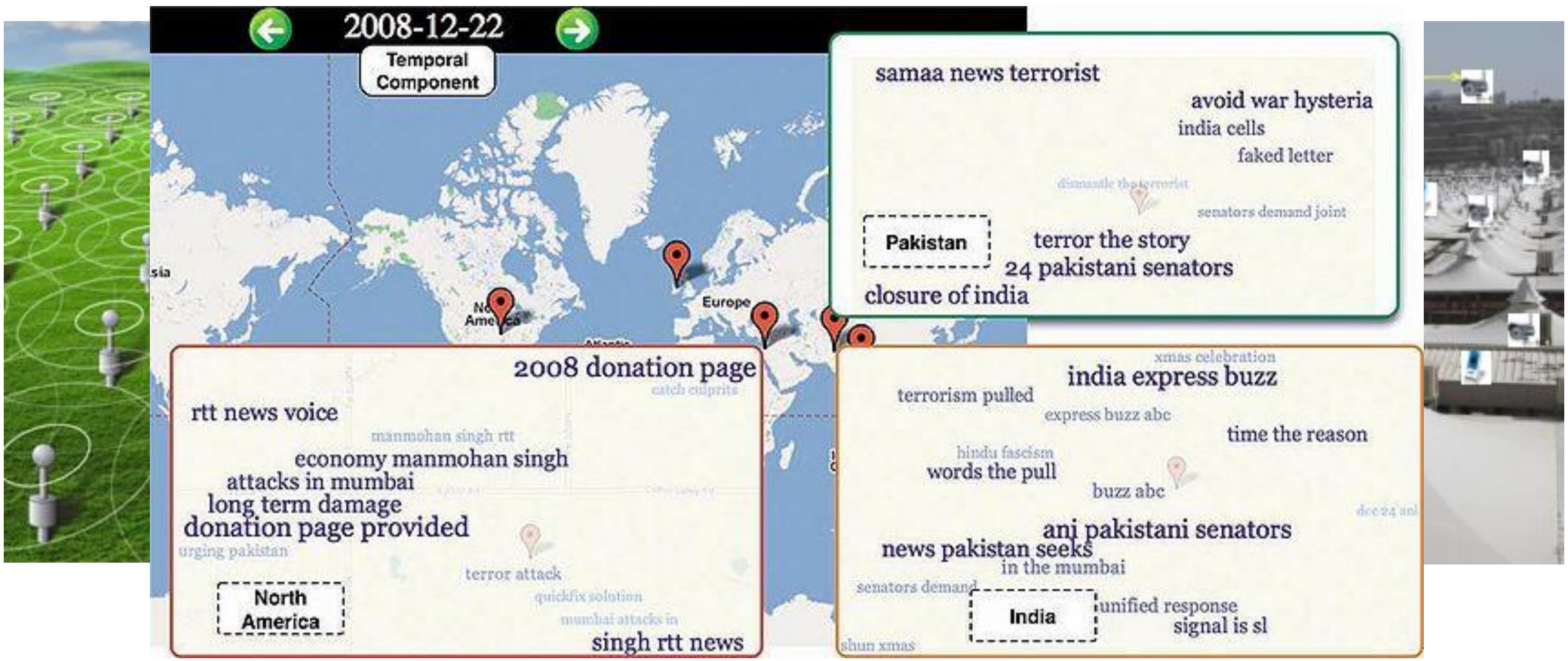


Courtesy of Rong Yan, "Social Networking," CCF 2011.

Social Media as Sensor



- A sensor network consists of spatially distributed autonomous sensors to monitor physical or environmental conditions, such as temperature, sound, light, pressure, etc. and to cooperatively pass their data through the network to a main location.



autonomous. intelligent. active. mobile. no maintenance. multimodal. social. sentiment .

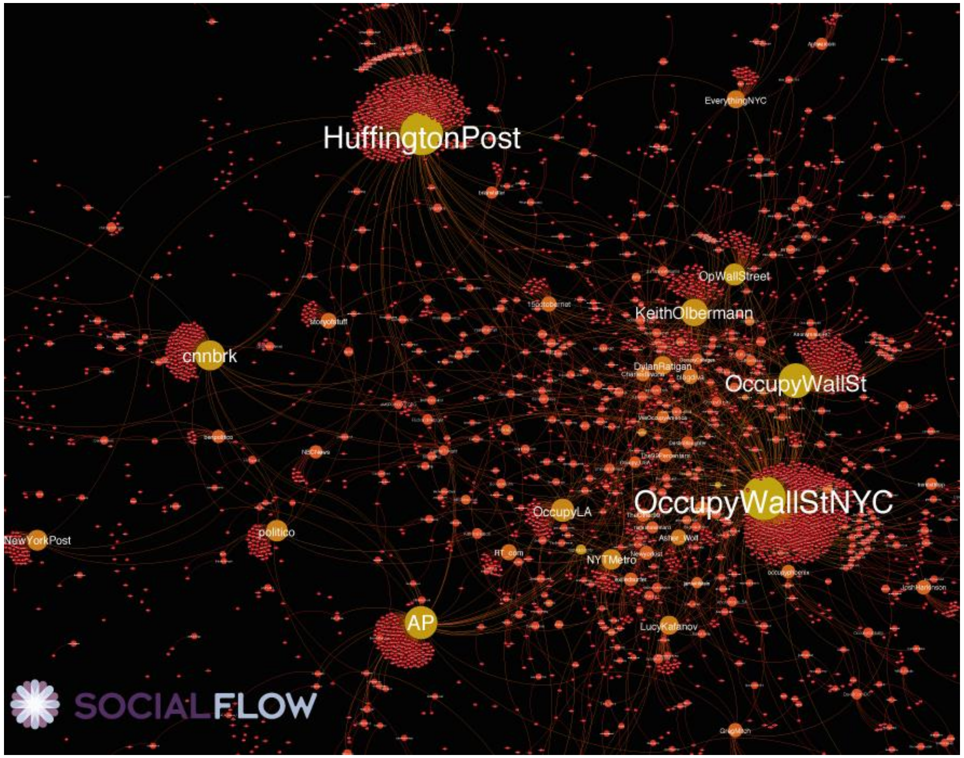
Rich application of social media

- Topic evolution and tracking [Alsumait, ICDM'08; Ramage, WSM'10]

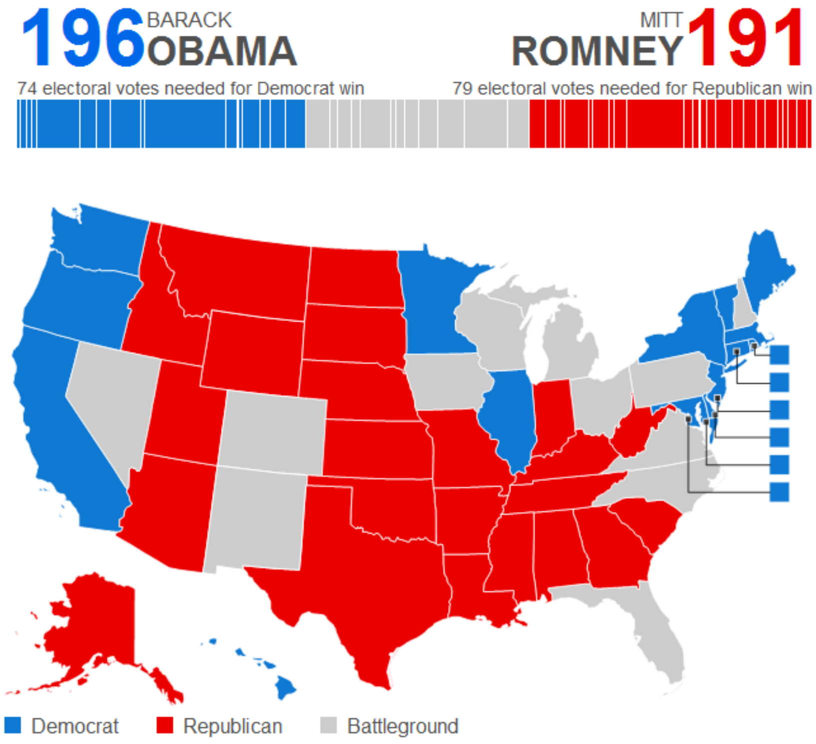
Source: stories.facebook.com

Rich application of social media

- Tracking political activity and campaigns [Jin, MM'10]



#OccupyWallStreet: origin and spread visualized



http://www.bbc.co.uk/news/world-us-canada-19794259

Extensive Research Interests on Social Media

- Computational social science [Lazer, Science'09]
  - … leverage the capacity to collect and analyze data with an unprecedented breadth and depth and scale…
- Twitter as medium and message [Savage, Comm of ACM'11]
  - Twitter data may help answer sociological questions that are otherwise hard to approach…

- IEEE Trans. on Big Data, IEEE Trans. on Computational Social Systems, Special issues in IEEE Intelligent Systems, IEEE Trans on Multimedia, TOMCCAP…
- Special sessions/workshops in international conferences
  - Social Media Workshop at ACM Multimedia and IEEE ICME
  - Web Search and Data Mining (WSDM)
  - AAAI Conference on Weblogs and Social Media (ICWSM)
  - WWW/SIGKDD/SIGIR/ICDM/CIKM

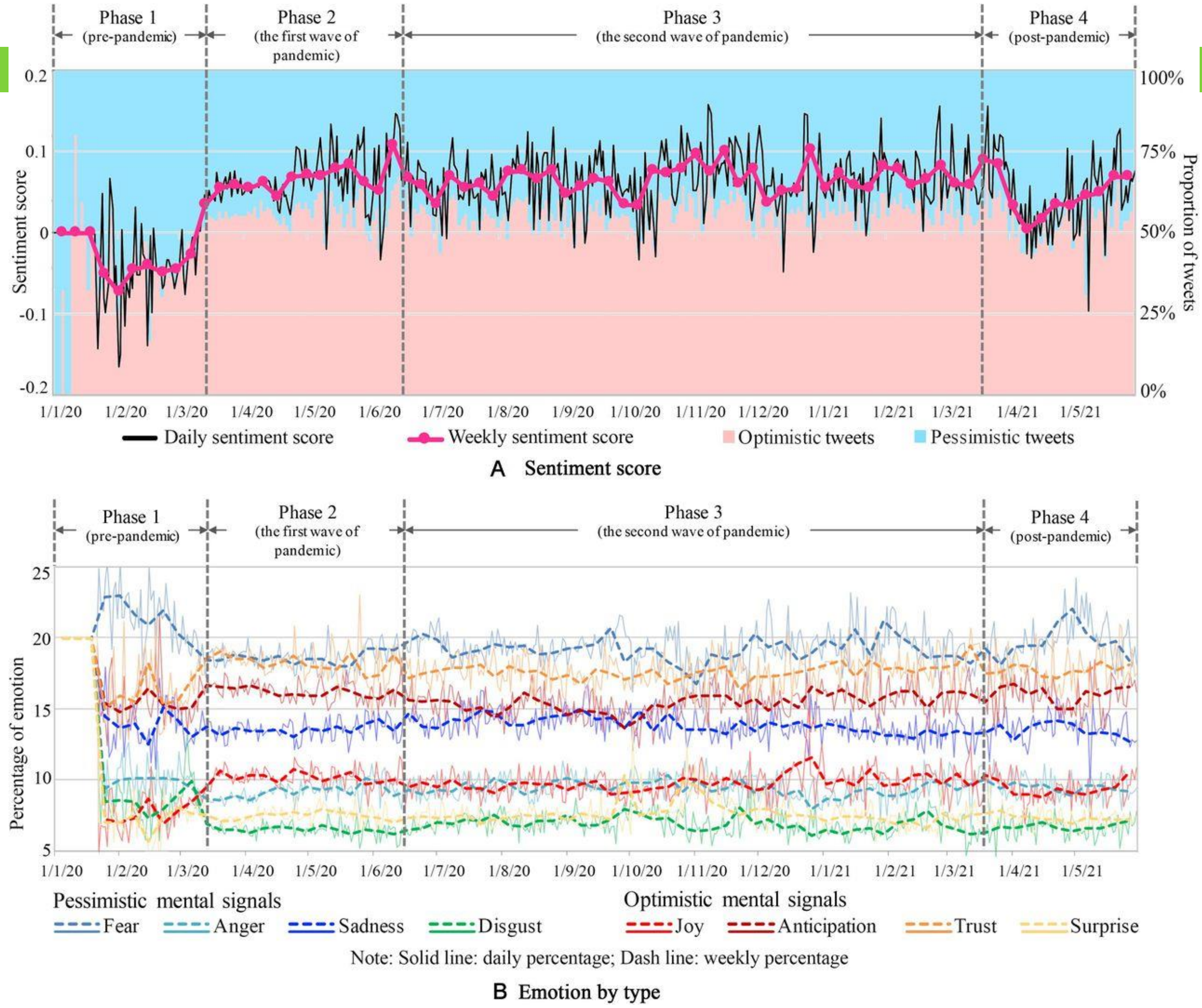
Tracking shifts in Mental Health Signals in COVID-19

- Motivation
  - Widespread problems of psychological distress have been observed in many countries following the outbreak of COVID-19, including Australia.
  - What is lacking from current scholarship is a **national-scale** assessment that tracks the shifts in mental health during the pandemic timeline and across geographic contexts.

- Method
  - Analysing geotagged tweets in Australia
  - Employing machine learning and spatial mapping techniques to classify, measure and map changes in the Australian public's mental health signals, and track their change across the different phases of the pandemic in eight Australian capital cities.

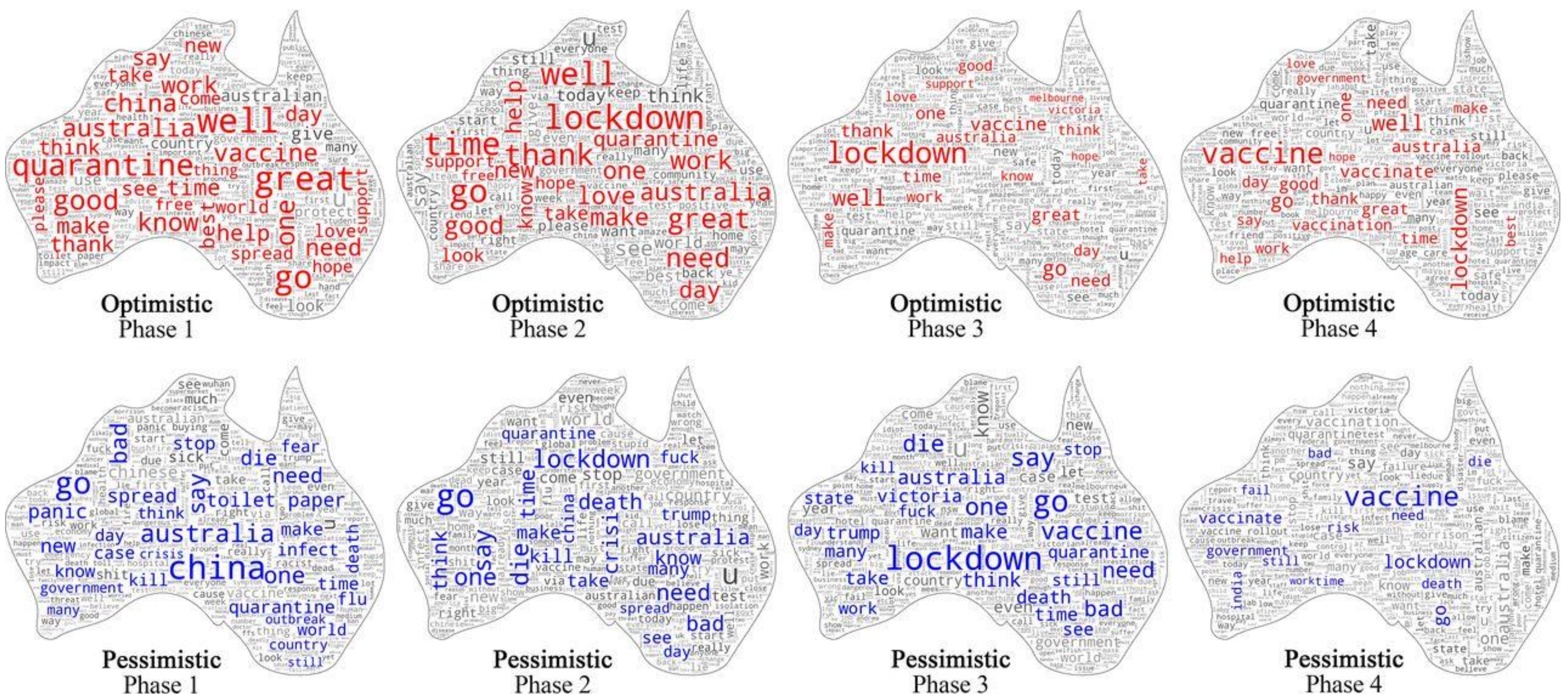
S. Wang, et al. The times, they are a-changin': tracking shifts in mental health signals from early phase to later phase of the COVID-19 pandemic in Australia. BMJ Global Health, 2022;7:e007081.

Tracking shifts in Mental Health Signals in COVID-19



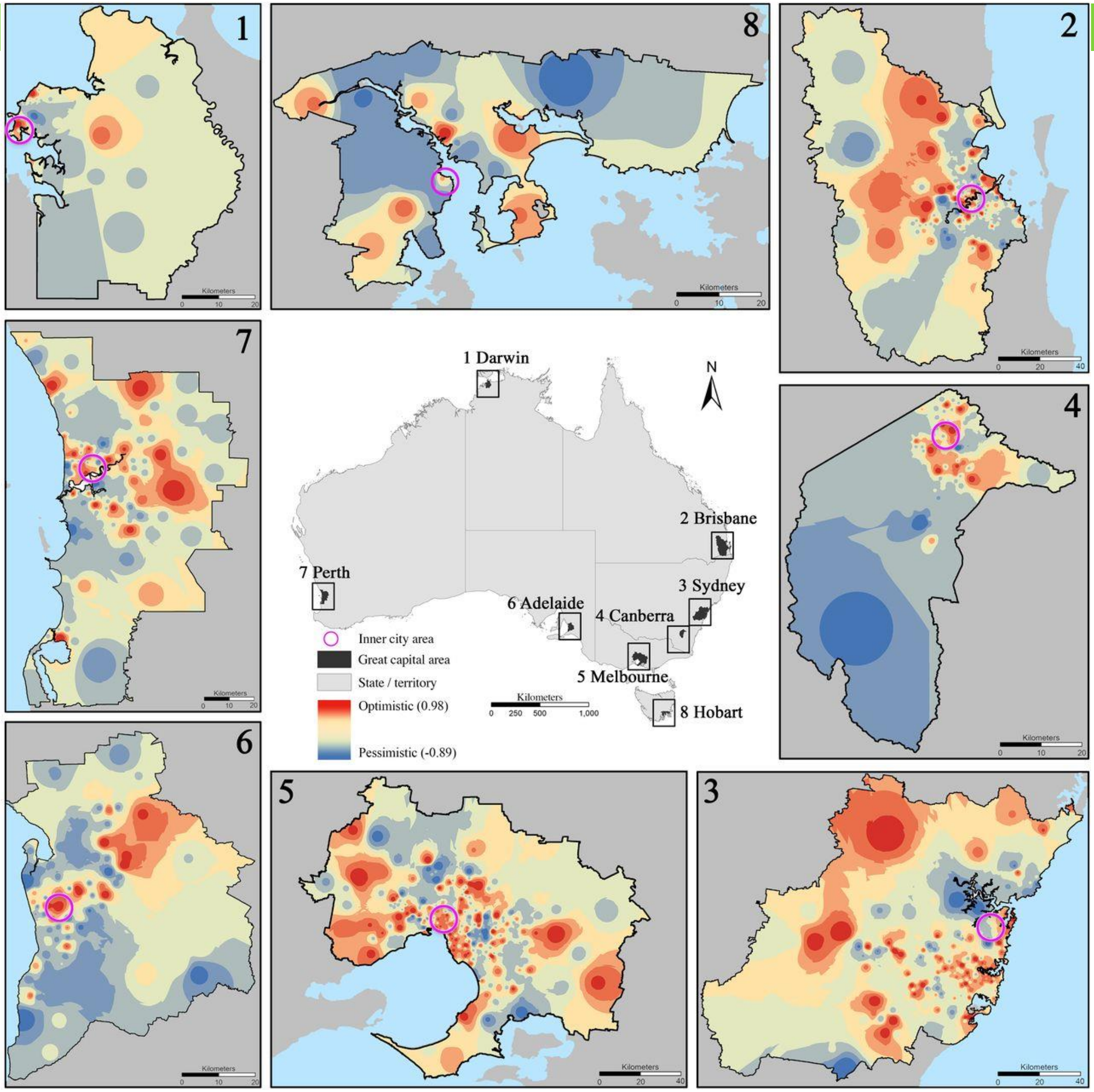
Temporal change of the public's (A) sentiment score and (B) emotion by type over four phases.

Tracking shifts in Mental Health Signals in COVID-19



Keywords potentially related to optimistic and pessimistic mental health signals over four phases.

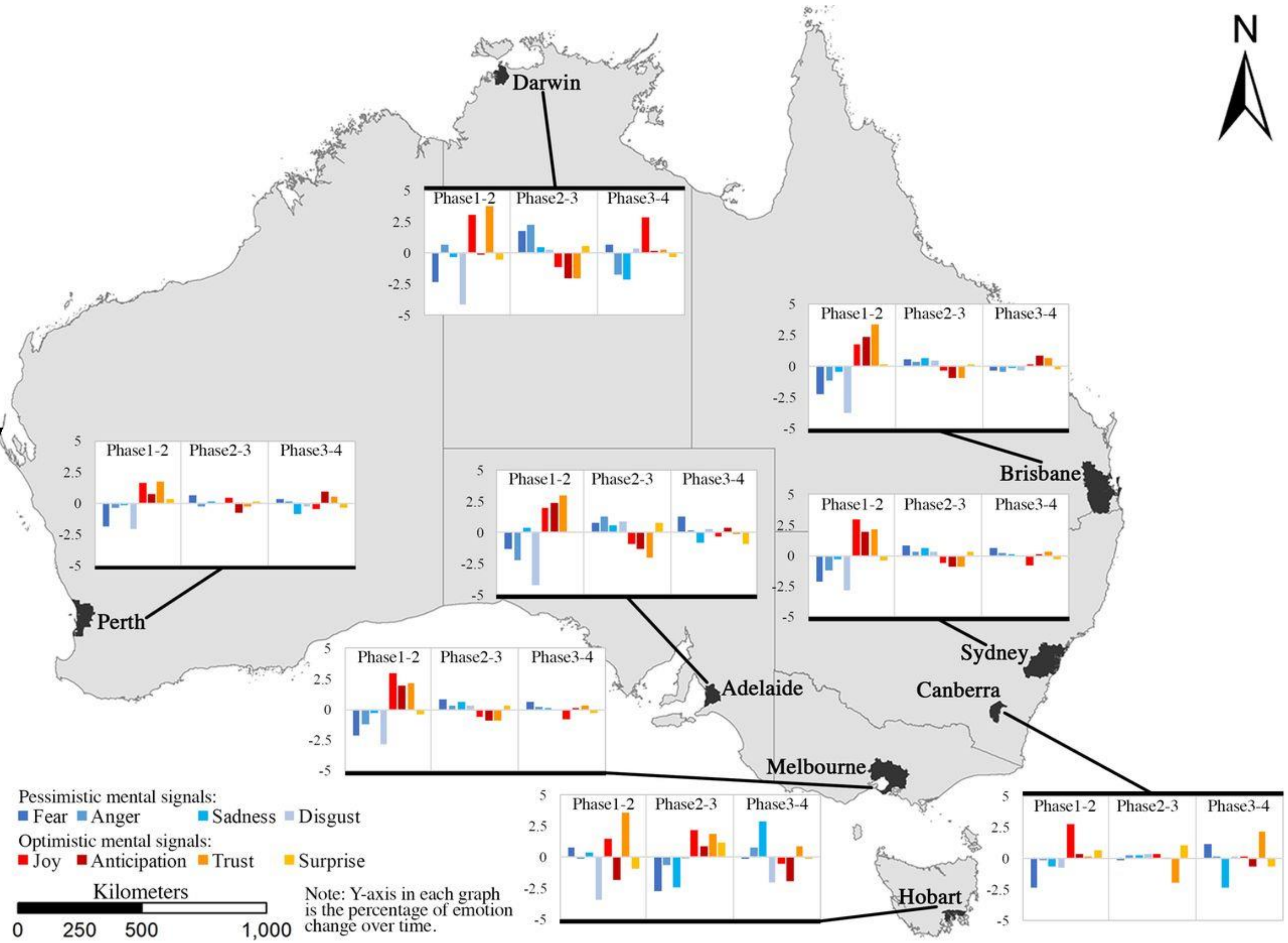
Tracking shifts in Mental Health Signals in COVID-19



Kernel density estimates of optimistic and pessimistic mental health signals in eight capital cities.

Tracking shifts in Mental Health Signals in COVID-19

Change of emotion by type over four phases in eight capital cities



Tracking shifts in Mental Health Signals in COVID-19

- Data Collection
  - 244,406 geotweets were retrieved from the total 860+ million tweets in Australia using AFT-API (Academic Full Track – API)
    - Searching terms: pandemic, epidemic, virus, covid*, coronavirus, corona, and vaccin*
    - Time: 1/1/2020 – 31/05/2021

Tracking shifts in Mental Health Signals in COVID-19

- Sentiment Analysis
  - Valence Aware Dictionary for sEntiment Reasoning (VADER) model
  - Combines lexical features and five simple heuristics and utilizes a human-centric approach via the combination of qualitative analysis and empirical validation using human raters
  - Returns four scores: positive, negative, neutral, and compound
  - https://github.com/cjhutto/vaderSentiment

Tracking shifts in Mental Health Signals in COVID-19

- Emotion Analysis
  - National Research Council Canada Emotional Lexicon (NRCLex)
  - Mental health signals
    - Pessimistic: feat, sadness, anger, disgust
    - Optimistic: joy, anticipation, trust, surprise
  - https://pypi.org/project/NRCLex/

What is social multimedia?

- A group of internet-based applications that build on the ideological and technological foundations of Web 2.0, which allows the creation and exchange of user-generated content (text, image, video, location, time, link)

- Social Multimedia = Multimedia Generated for Social Interactions

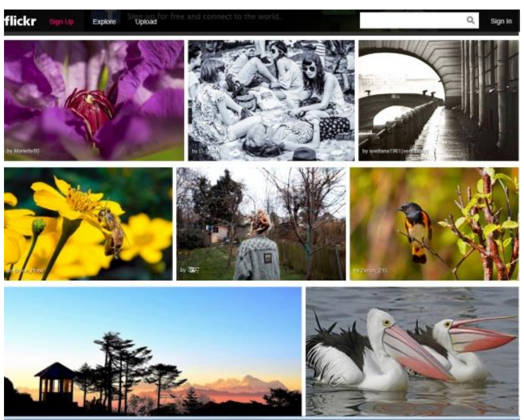
Andreas M. Kaplan, Michael Haenlein (2010). "Users of the world, unite! The challenges and opportunities of social media," Business Horizons, 53(1): 59-68. [Courtesy of Jiebo Luo]

Why multimedia (image/video)?

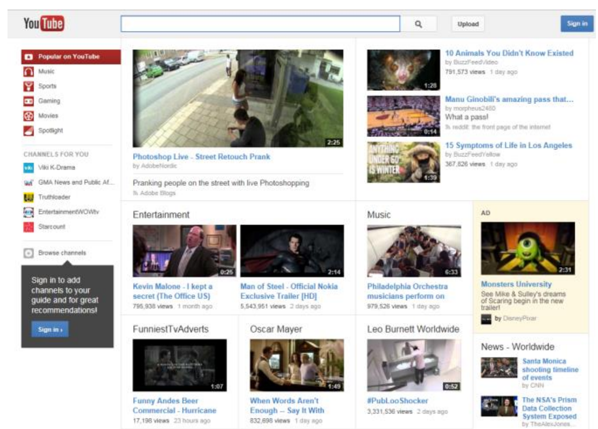
- Pictures and videos are more distinctive than plain texts, and can be comprehended more quickly
- Photography is the only "language" understood in all parts of the world
- "A picture is worth a thousand words"
- 眼见为实，一图胜千言

Social multimedia on the Web

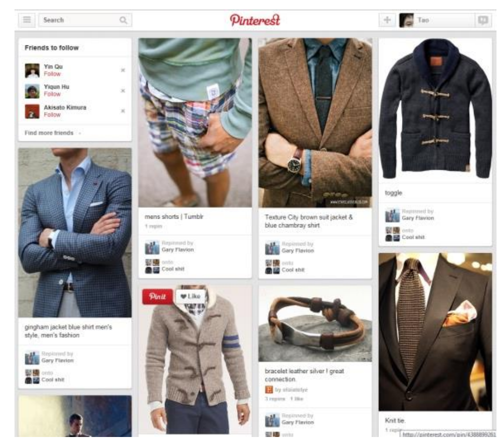
- Explosion of image/video data
  - digital photos, personal photo/video collections, geospatial imagery, broadcast news/sports videos, Wikipedia, social media, etc.
- Data mining can unlock the wealth of information in social multimedia
  - Understanding of users can benefit social multimedia recommendation (context-aware, community-aware, personalized)



50 million of photos uploaded per month (2012)



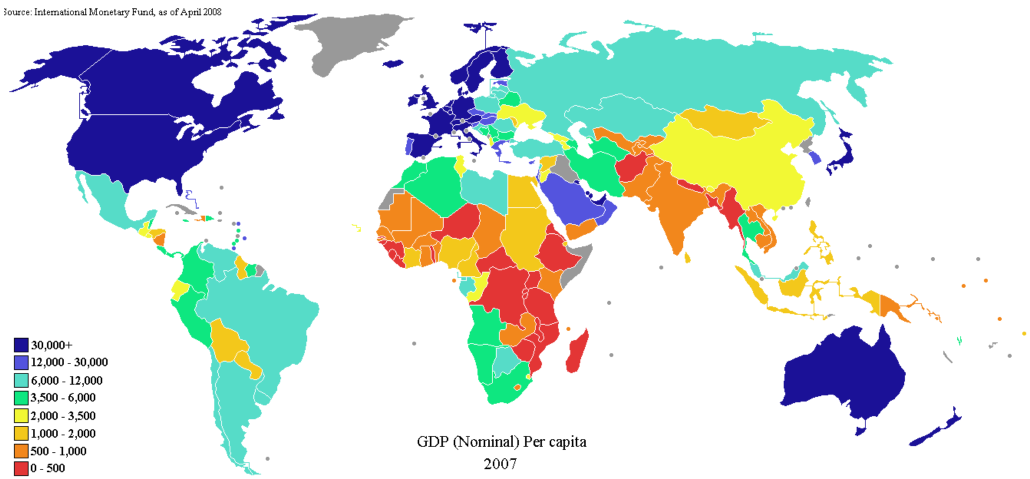
1 billion visits per month, 4 billion hrs watched per month, 72 hrs of video uploaded per min



Pinterest: 48 million users (Feb. 2013)

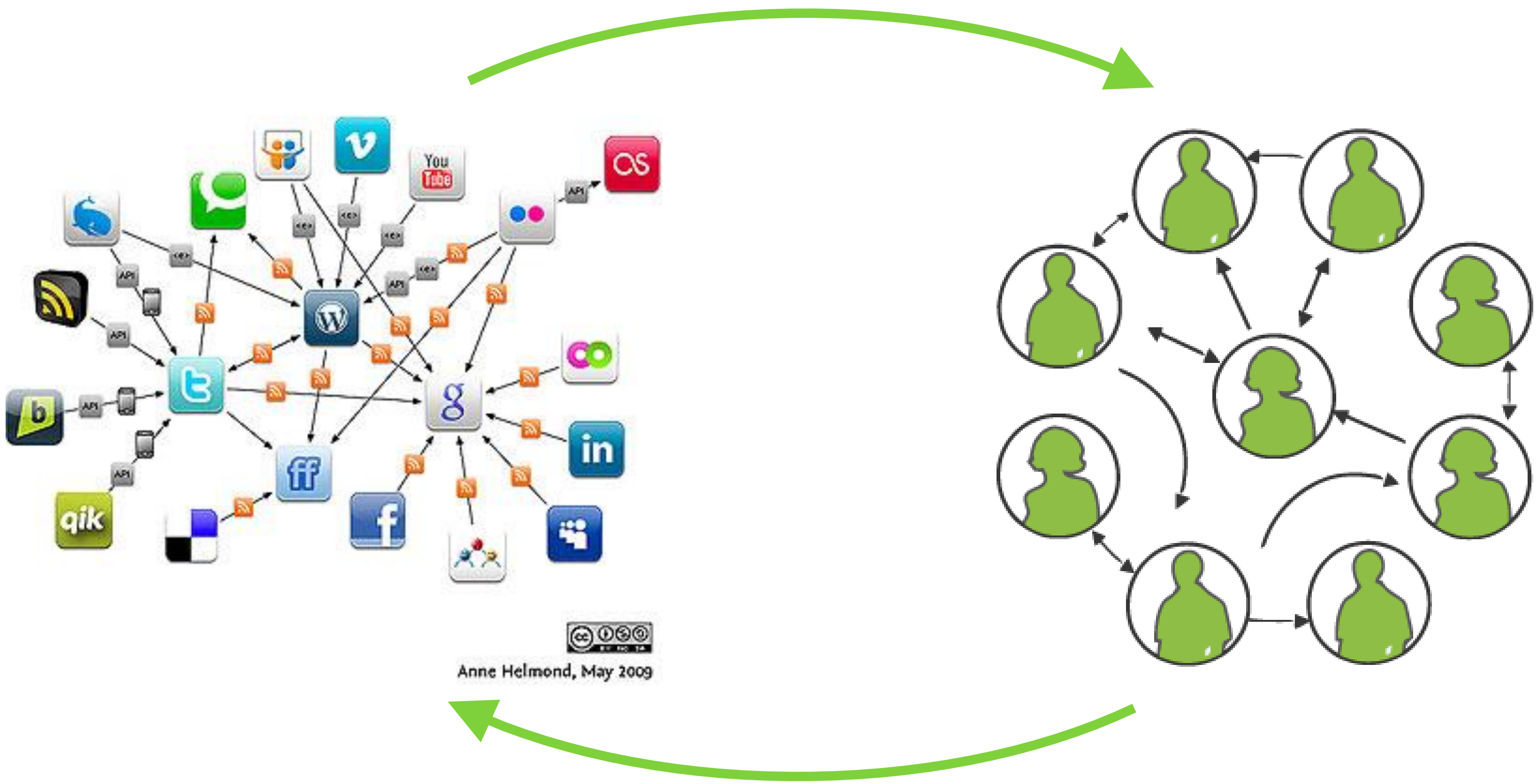
Social multimedia on the Web

- To predict other *human geography* metrics, such as *GDP*, *wealth* by region, *unemployment rate, stock market sentiment*
- To monitor *health/disease*, *environment*, *ecology*, *social unrests*, …



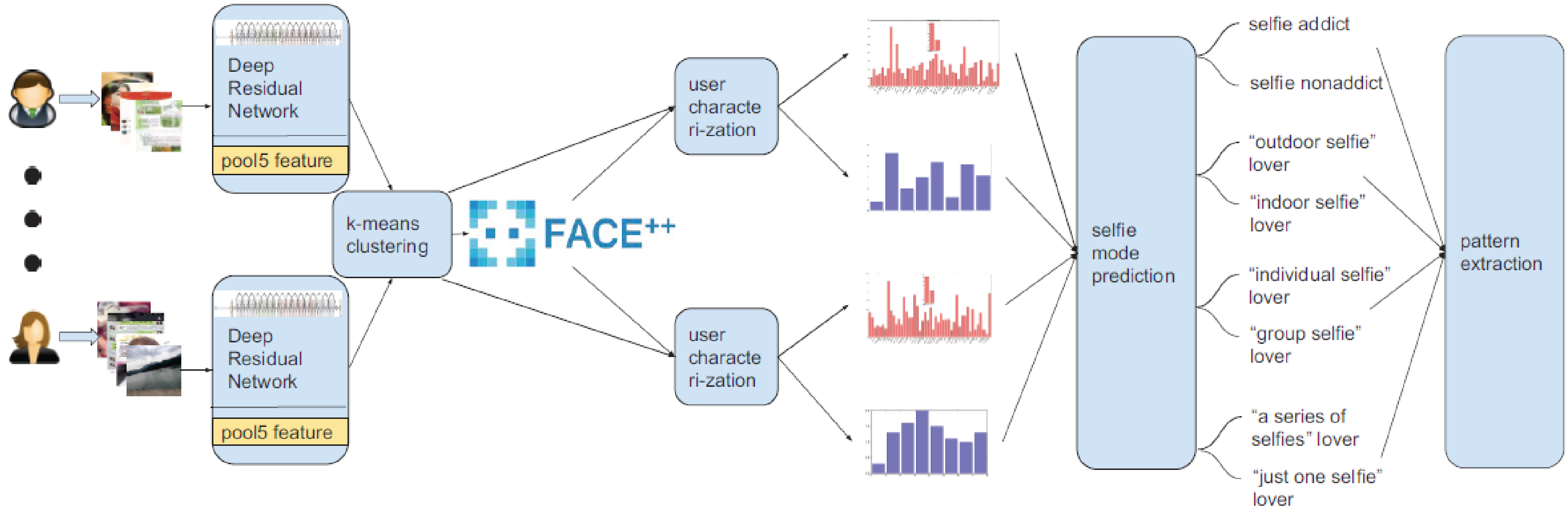
The Role of “Multimedia” for Social Computing

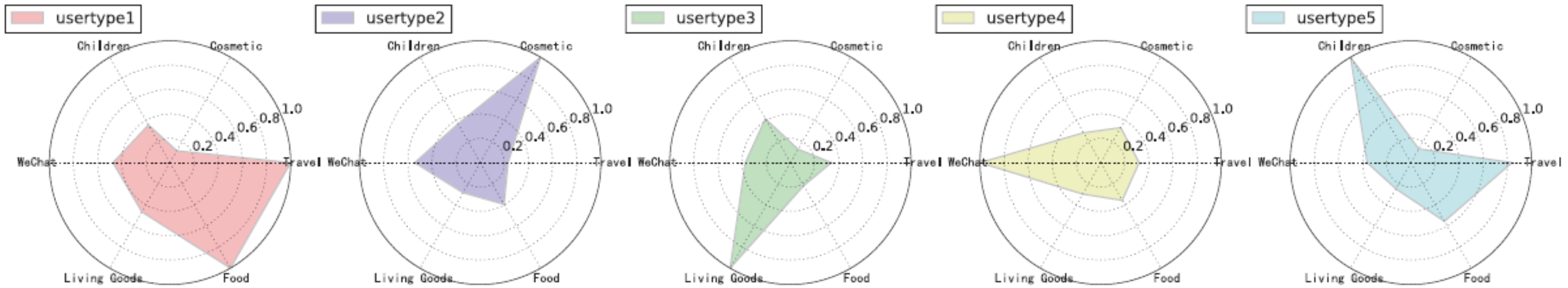
Leverage social context for better understanding multimedia data



Leverage social multimedia data for better understanding users

Mining Personal Patterns from Selfie Behaviors





Sentiment Analysis in Social Media

- Sentiment is arguably the most important signal from social media
  - User connections
  - User preference
- Most existing methods are based on textual information only
  - Comments, reviews, textual tweets, and status updates
- Twitter
  - Easy? Only a limited number of words per tweet
  - Difficult? Lots of noise and little information
- Questions
  - Do users express themselves only using text?
  - Can the emerging multimedia content provide additional useful signals?

Textual Sentiment Analysis

- Dictionary-based approach
  - Lexicon contains a large amount of words with sentiment labels
  - Emoticons, widely used in online social networks
    - Simple and effective in most cases, however fail to capture the rich contextual information

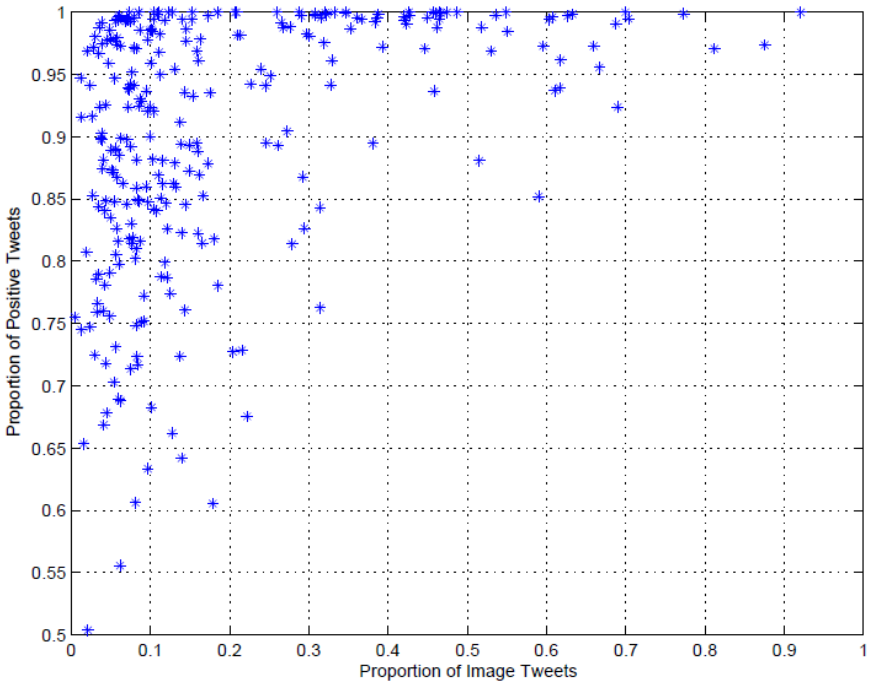
- Semantic analysis
  - Using NLP related techniques to build more robust features
  - Difficult to develop a method that works for all languages

- Use sentiment140 for textual sentiment analysis
  - Using emoticons as auxiliary information
  - Open API available

Source: http://www.sentiment140.com

Image Tweets

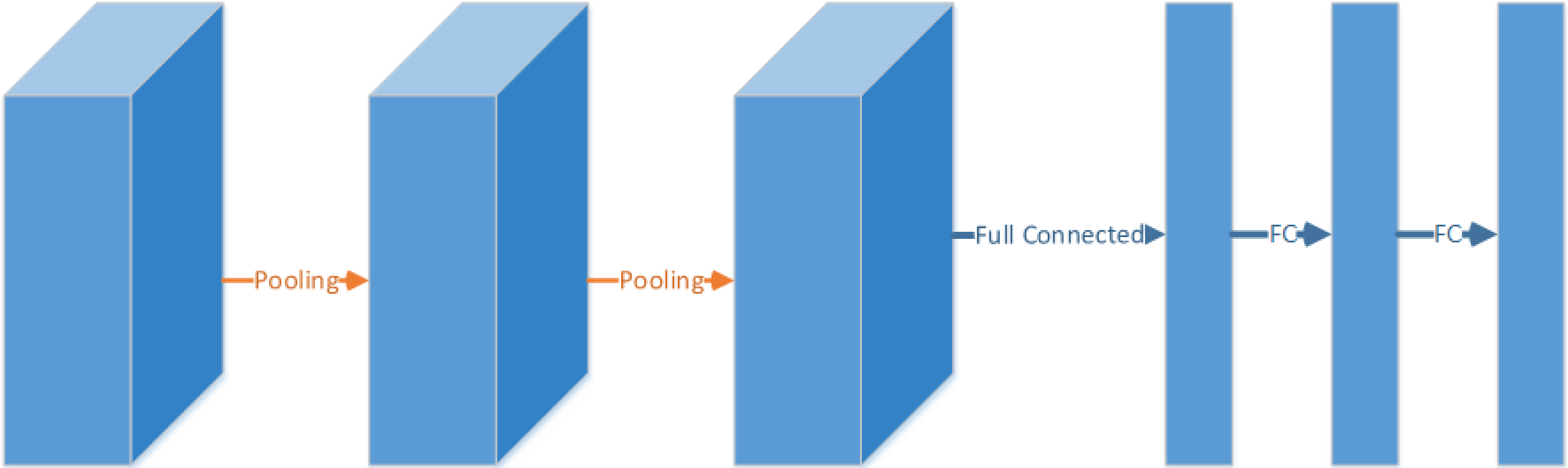
- Image tweets: tweets that contain images
- Different users may prefer different types of tweets



Observation: Users who prefer image tweets tend to have more positive tweets

Deep Learning for Image Sentiment Analysis

Weakly labelled Images



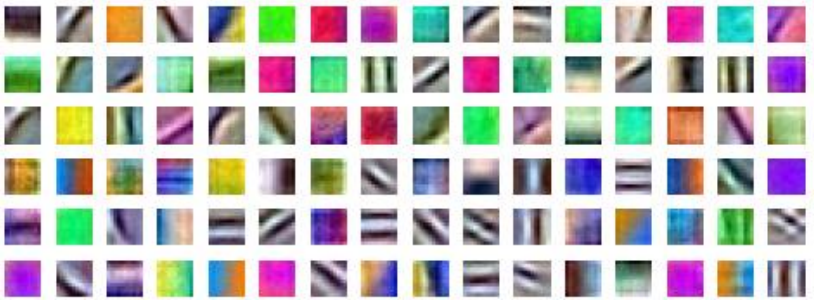
Convolutional Neural Network for Image Sentiment Analysis

- Domain-transfer Learning;
- Boosted Learning using Noisy Labels

Q. You, et al., Visual Sentiment Analysis by Attending on Local Image Regions, AAAI, 2017.
Q. You, et al., Robust Image Sentiment Analysis Using Progressively Trained and Domain Transferred Deep Networks, AAAI, 2015.

Deep Learning for Image Sentiment Analysis

(a) Filters learned from CNN



(b) Filters learned from PCNN

Figure 4: Filters of the first convolutional layer.

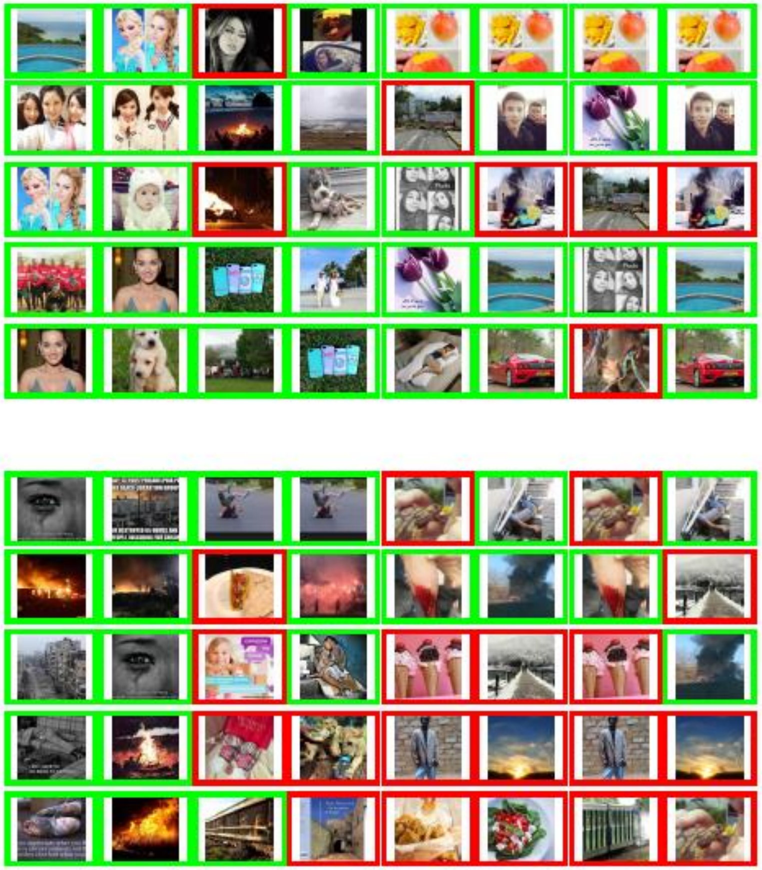


Figure 5: Positive (top block) and Negative (bottom block) examples. Each column shows the negative example images for each algorithm (PCNN, CNN, Sentribute, Sentibank, GCH, LCH, GCH+BoW, LCH+BoWAll). The images are ranked by the prediction score from top to bottom in a decreasing order.

Deep Learning for Image Sentiment Analysis

- Half million weakly labeled Flickr images from Visual Sentiment Ontology
- Two GTX Titan GPUS and 32 GB RAM
- Statistics of the data set

Models	# of training images	# of testing images	# of iterations
CNN	401,739	44,637	300,000
PCNN(fine-tune)	369,828	44,637	100,000

- Performance of CNN and PCNN on the testing

Algorithm	Precision	Recall	F1	Accuracy
CNN	0.714	0.729	0.722	0.718
PCNN	**0.759**	**0.826**	**0.791**	**0.781**

MuMin

- A Large-scale Multilingual Multimodal Fact-Checked Misinformation Social Network Dataset
  - 21 million tweets belonging to 26 thousand Twitter threads over 70 languages
  - https://mumin-dataset.github.io/

Dataset	#Claims	#Threads	#Tweets	#Users	#Articles	#Images	#Languages	%Misinfo
MuMiN-large	12,914	26,048	21,565,018	1,986,354	10,920	6,573	41	94.79%
MuMiN-medium	5,565	10,832	12,659,371	1,150,259	4,212	2,510	37	94.20%
MuMiN-small	2,183	4,344	7,202,506	639,559	1,497	1,036	35	92.71%

D. Nielsen, et al., MuMiN: A Large-Scale Multilingual Multimodal Fact-Checked Misinformation Social Network Dataset, SIGIR, 2022.

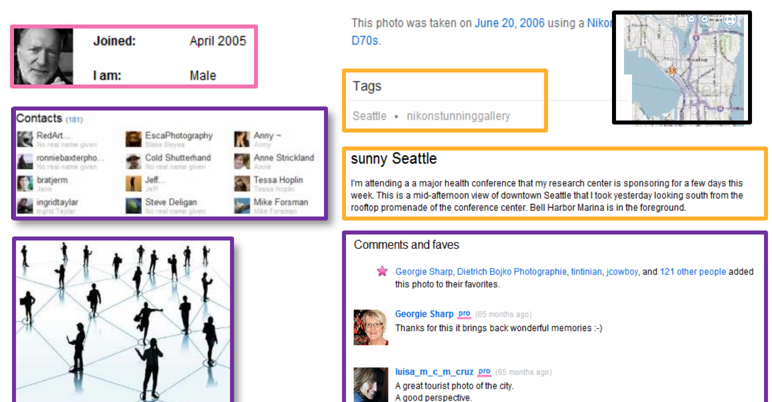
Challenges for understanding users

- Heterogeneous data forms
  - Profile: gender, residence, interest, age…
  - Content: text, image, video, link, trajectory…
  - Context: identity, location, time, status, emotion, weather, comments, views, friends, groups…
  - Behaviors: gesture, commenting, sharing, distributing, …
- Missing attributes
- Large-scale social graph construction
- Real-time sharing/propagation

Heterogeneous data types in social multimedia

Textual data (script, tags, ...)	Visual data	Concepts (auto. recognized)	Audio data	Demographic	Social data	Location

- City
- Sky
- Building
- Sea
- ...

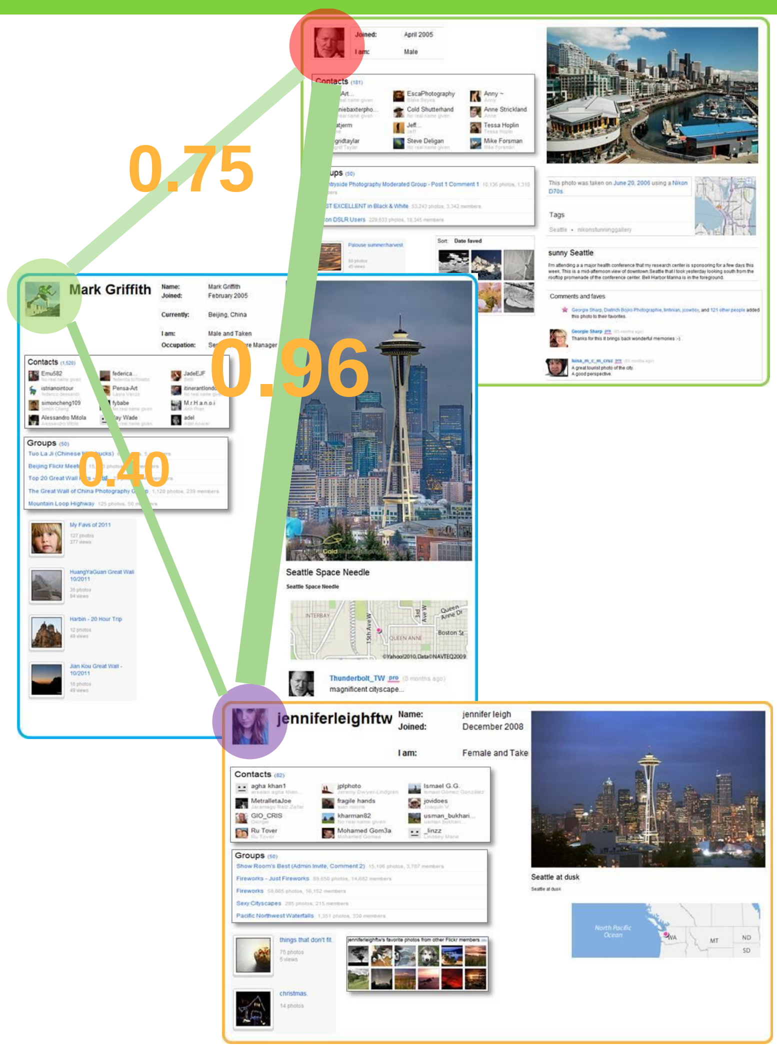


Understanding Users

- Part 1: Understand user profile
  - Community discovery from heterogeneous data [Zhuang, MM'11]
  - Multi-social-graph construction [Yao, WWW'13]
- Part 2: Understand user context
  - Mobile visual localization [Liu, MM'12]
  - Mobile recommendation [Zhuang, Ubicomp'11]
- Part 3: Understand user Interaction
  - O-search [Zhang, MM'11]
  - SocialTransfer: cross-domain social media recommendation [Roy, MM'12]
  - Interactive multimodal mobile visual search [Wang, MM'11]
  - Browse-to-Search [Zhang, MMSP'11/2; Lu, MM'12]

Part 1: Understand user profile

- Social graph from multi-dimensional (m-d) relations
- Explore m-d user relationship
  - mutual comments
  - co-locations
  - similar/related photos
  - common interest groups
- Build a unified social graph
- Infer hidden relationship/interest as a new context



Controversial Language Usage with COVID-19

- Context
  - COVID-19, the official name from WHO
  - However, terms such as Chinese Virus and Wuhan Virus were also used for many reasons.
    - Discrimination, Hate Speeches
    - Developing Racism, Dividing Societies

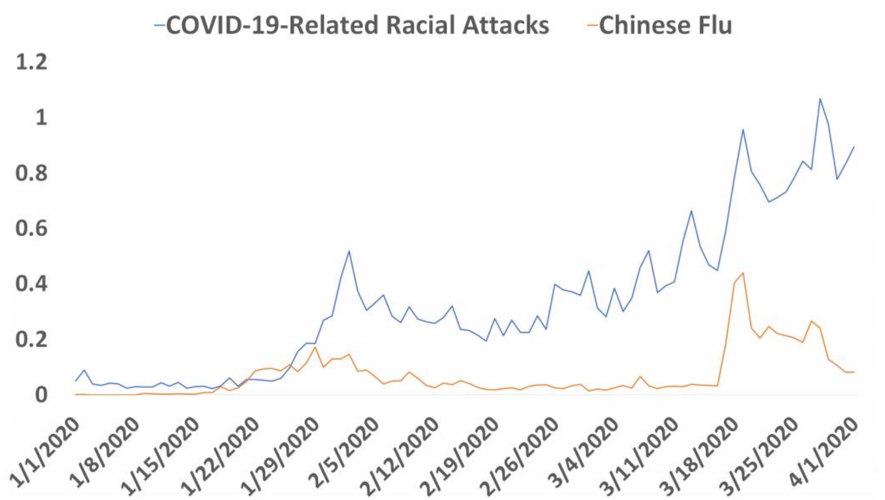


Fig. 1: Density of online media coverage with the "Chinese Flu" term and COVID-19 related racial attacks.

H. Lyu, L. Chen, Y. Wang, and J. Luo. Sense and Sensibility: Characterizing Social Media Users Regarding the Use of Controversial Terms for COVID-19. IEEE Trans. on Big Data, 2021.

Controversial Language Usage with COVID-19

- Question
  - Who are using these controversial terms?
  - Prevent the exponential growth of such a mindset

- Solution
  - A social media and big data study
  - 17 million tweets on Twitter
  - Demographics analysis
    - User level features, political status, geo-locations

H. Lyu, L. Chen, Y. Wang, and J. Luo. Sense and Sensibility: Characterizing Social Media Users Regarding the Use of Controversial Terms for COVID-19. IEEE Trans. on Big Data, 2021.

Controversial Language Usage with COVID-19

- Approach
  - Data Collection and Pre-processing
    - Controversial keywords
      - “chinese virus” and “#ChineseVirus”
    - Non-controversial keywords
      - “corona”, “covid19”, “covid19”, “coronavirus”, “#Corona”, “#Covid 19” and “#coronavirus”
    - 4-day period: March 23-26, 2020
      - 1,125,285 tweets for Controversial Dataset (CD)
      - 16,320,176 tweets for Non-controversial Dataset (ND)

Controversial Language Usage with COVID-19

- Approach
  - Data Collection and Pre-processing
    - Baseline Dataset
      - 7 user-level features were either collected or computed
        - Followers_count, friends_count, statuses_count, favorites_count, listed_count, account_length, verified status
      - 1,125,176 tweets in CD with 593,233 distinct users
      - 1,599,013 tweets in ND with 490,168 distinct users

Controversial Language Usage with COVID-19

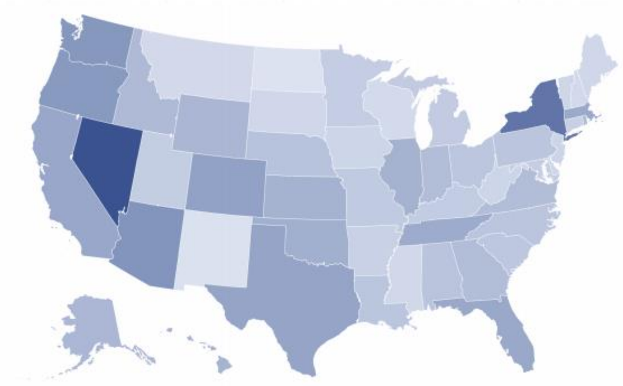
- Approach
  - Data Collection and Pre-processing
    - Demographic Datasets
      - Face++ API to obtain inferred age and gender information by analyzing users' profile images

TABLE I: Composition of Profile Images.

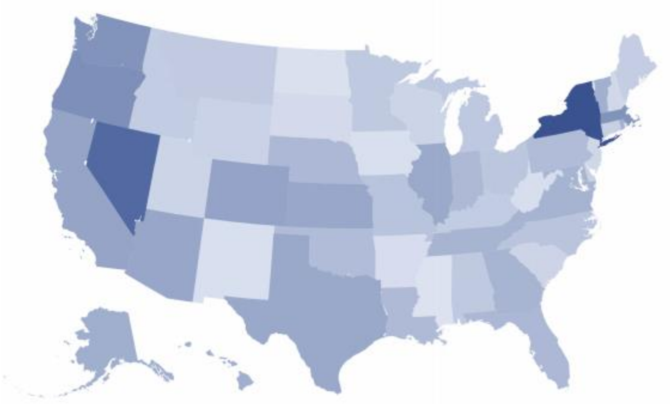
	Controversial	Non-Controversial
One Intelligible Face	47,011	109,718
Multiple Faces	5,596	11,393
Zero Intelligible Face	54,539	96,218
Invalid URL	264,894	187,379
Total	372,040	404,708

Controversial Language Usage with COVID-19

- Approach
  - Data Collection and Pre-processing
    - Geo-location Datasets



(a) Controversial



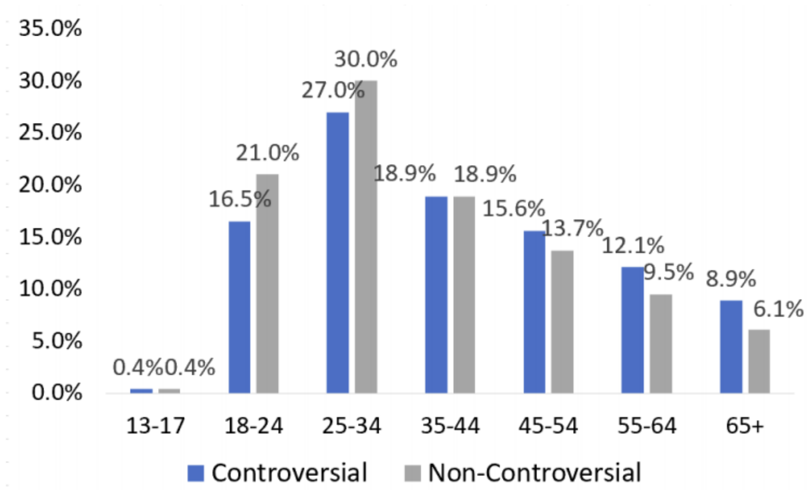
(b) Non-Controversial

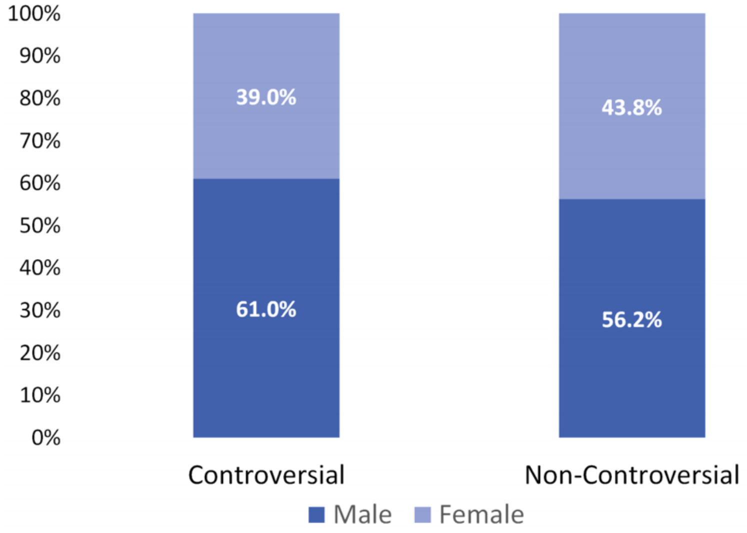
Fig. 2: Distribution of a) controversial and b) non-controversial tweets in the US, by state, and normalized by the state population. No significant differences can be observed. It is interesting to note that New York and Nevada have the highest number of COVID-19 related tweets per capita.

Controversial Language Usage with COVID-19

- Approach
  - Data Collection and Pre-processing
    - Geo-location Datasets
      - Only a very limited number of tweets contain self-reported locations (1.2% of crawled data)
      - Use the user profile location as the source of geo-location, which has a substantially higher percentage of entries in the crawled datasets (16.2% of crawled data).
    - Aggregate Datasets
      - Datasets with both demographic and geo-location features were created.
      - These datasets contain complete attributes that were analyzed in our study, while trading off with the relatively small size, with 5,772 for CD and 12,403 for ND.

Controversial Language Usage with COVID-19





Controversial Language Usage with COVID-19

- Proportion of Political Following Status
  - Most of users do not follow any of these people.
  - The second biggest group of users in both CD and ND corresponds to users who only follow Donald Trump.

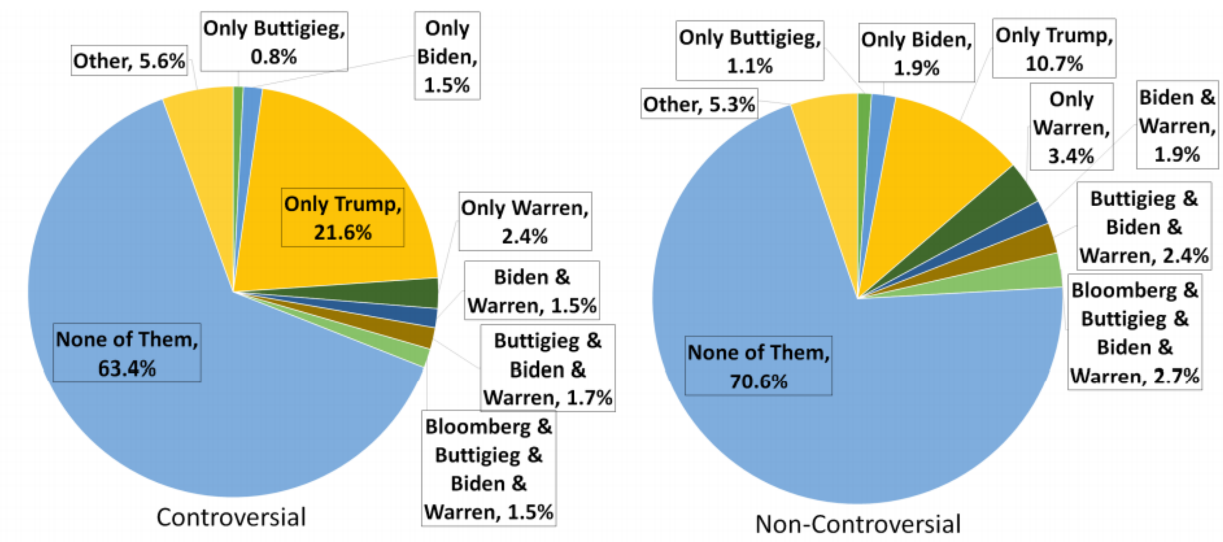


Fig. 5: Proportion of Political Following Status.

Controversial Language Usage with COVID-19

- Geo-location

	Controversial	Non-Controversial
Urban	56.14%	62.48%
Suburban	17.48%	15.58%
Rural	26.38%	21.94%

Controversial Language Usage with COVID-19

- What to do?
  - To predict Twitter users who employ controversial terms in the discussion of COVID-19.

- Investigated six classification models
  - Logistic Regression (Logit)
  - Random Forest (RF)
  - Support Vector Machine (SVM)
  - Stochastic Gradient Descent Classification (SGD)
  - Multi-layer Perception (MLP)
  - XGBoost (XGB)

Controversial Language Usage with COVID-19

Classification performance

Metric	Baseline		
	Logit	XGB	RF
Accuracy	0.625	0.628	0.645
Precision	0.604	0.594	0.606
Recall	0.268	0.315	0.396
F1	0.371	0.411	0.479
AUC-ROC	0.618	0.647	**0.669**

(a) Baseline (N_{CD}=593,233 , N_{ND}=490,168)

Metric	Demographics					
	Logit	XGB	RF	SVM	MLP	SGD
Accuracy	0.584	0.691	0.743	0.510	0.680	0.627
Precision	0.604	0.698	0.735	0.474	0.663	0.665
Recall	0.296	0.589	0.698	0.544	0.631	0.390
F1	0.397	0.639	0.716	0.507	0.647	0.491
AUC-ROC	0.633	0.783	**0.833**	0.699	0.714	0.783

(b) Demographics (N_{CD}=47,011, N_{ND}=109,718)

Metric	Geo-location					
	Logit	XGB	RF	SVM	MLP	SGD
Accuracy	0.611	0.651	0.811	0.543	0.615	0.658
Precision	0.695	0.699	0.757	0.508	0.713	0.702
Recall	0.305	0.436	0.875	0.499	0.504	0.456
F1	0.422	0.537	0.812	0.504	0.410	0.553
AUC-ROC	0.652	0.696	**0.874**	0.696	0.636	0.713

(c) Geo-location (N_{CD}=14,817, N_{ND}=41,118)

Metric	Demographics + Geo					
	Logit	XGB	RF	SVM	MLP	SGD
Accuracy	0.628	0.592	0.644	0.591	0.685	0.658
Precision	0.425	0.402	0.439	0.328	0.495	0.476
Recall	0.524	0.622	0.497	0.293	0.264	0.241
F1	0.469	0.489	0.467	0.310	0.344	0.319
AUC-ROC	**0.624**	0.545	0.604	0.510	0.571	0.560

(d) Aggregate (N_{CD}=5,772, N_{ND}=12,403)

TABLE VI: Classification metrics for the four pairs of datasets. Best AUC-ROC scores are highlighted in each dataset pair.

Controversial Language Usage with COVID-19

Feature Association

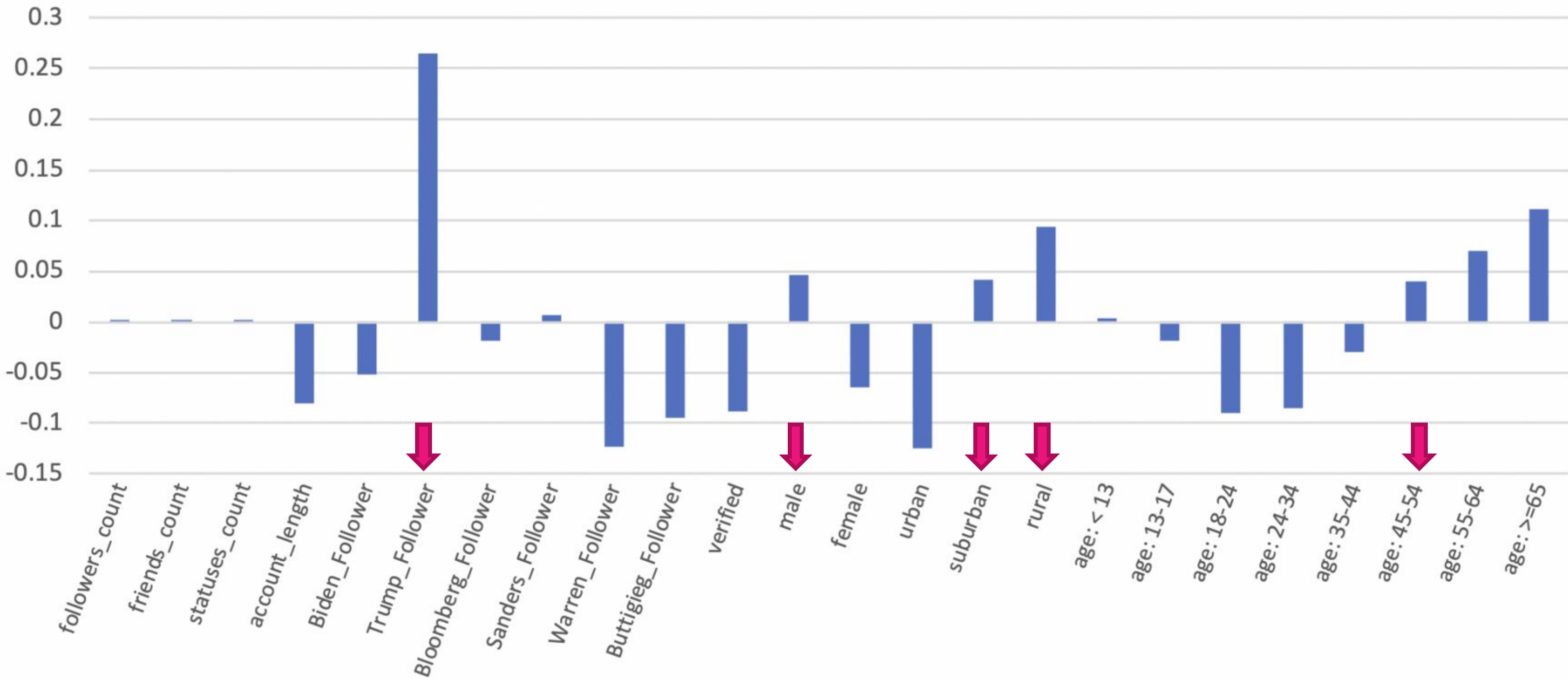


Fig. 6: Logistic regression coefficients for the aggregate dataset. Users who follows Trump, users in rural and suburban regions, male users and users aged over 45 years old tend to use controversial terms, whereas users who follow most top Demogratic presidental candidates, users in urban regions, female users and users aged between 18 to 44 tend to use non-controversial terms. No significant association between account-level attributes (follower, friend and status count) and the use of controversial terms were found.

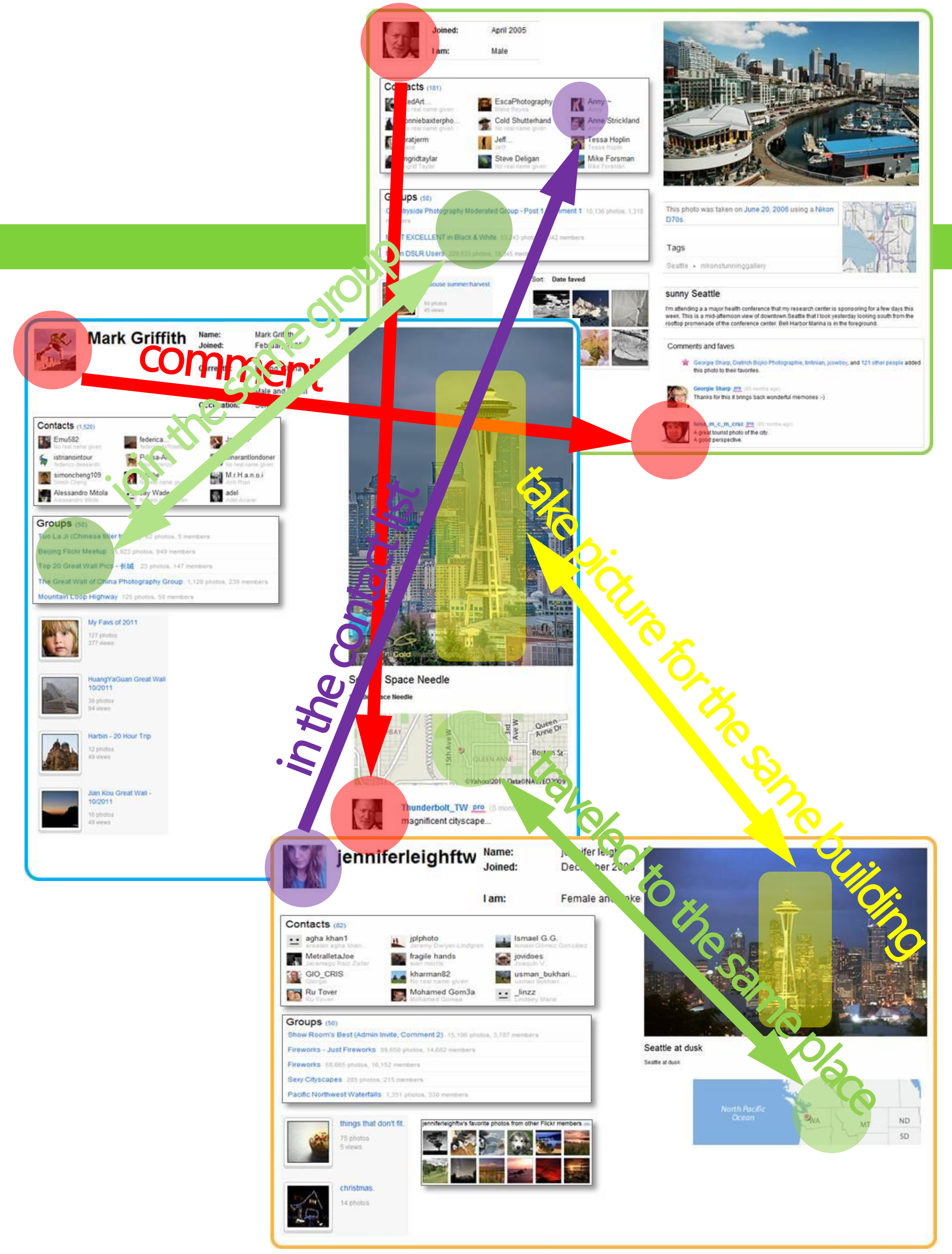
Controversial Language Usage with COVID-19

- Insights
  - Young people tend to use non-controversial terms.
  - Male users is more likely than female users to use controversial terms.
  - Non-verified users are more likely to use controversial terms.
  - More users in the CD group follow Donam Trump than in the ND group.
  - Users living in rural or suburban areas are more likely to use controversial terms.
  -

People may not have "explicit" relationship, but…

They may:

- comment with each other
- travel to the same place
- take or share pictures for the same landmark
- join the same interest group
- be in the contact list



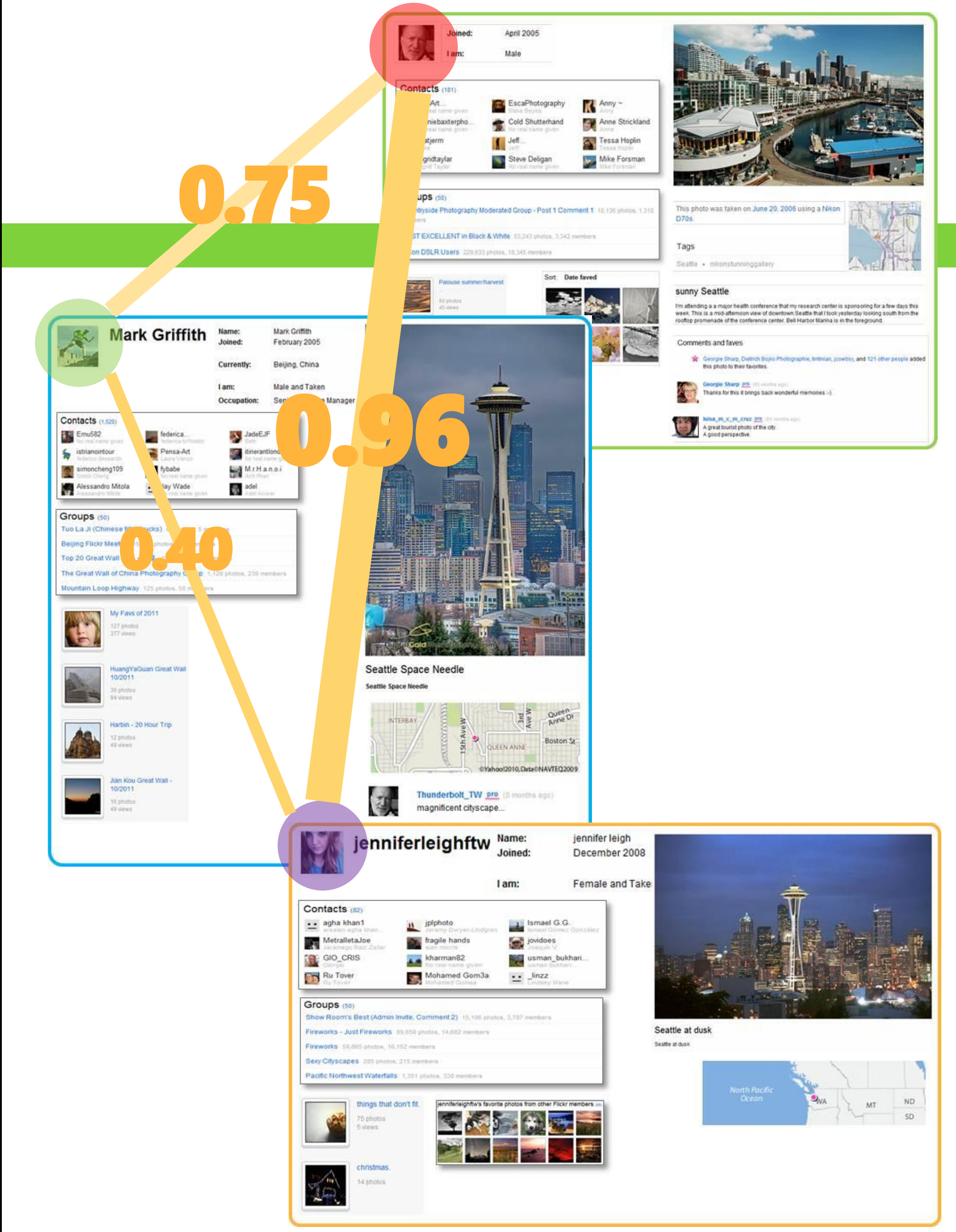
People may not have "explicit" relationship, but…

They may:

- comment with each other
- travel to the same place
- take or share pictures for the same landmark
- join the same interest group
- be in the contact list

This work tries to:

- mine "implicit" social relationship
- explore heterogeneous data and user behaviors



Solutions

- Modeling continuous social relationship

  Learning to model (discriminative & model-free)

- Exploration of heterogeneous data

  (image | tag | location | friend list | interest group | commenting)

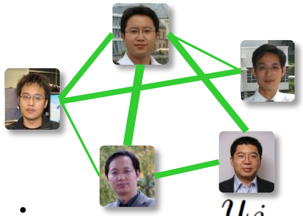
  Multiple kernel learning

- Combination of multimodal features

  Kernel alignment for integration

Approach [Zhuang, MM'11]
model social strength via multiple kernel learning (MKL)

- Assumptions
  - #1: Social strength modeling = Prediction of pairwise relationship



Linear regression model: $f(u_i, u_j) = \mathbf{w}^\top \Phi(u_i, u_j)$
where Φ is pairwise feature mapping $\Phi(\cdot,\cdot): \mathcal{U} \times \mathcal{U} \rightarrow \mathbb{R}^d$

  - #2: Prediction complies with existing (explicit) relationship

Learning-based formulation: loss + regularization

$$\boxed{\min_{\mathbf{w}} \frac{\lambda}{2} \|\mathbf{w}\|_2^2} + \boxed{\sum_{u_i, u_j \in \mathcal{U}} l(y_{ij}, \mathbf{w}^\top \Phi(u_i, u_j))}$$

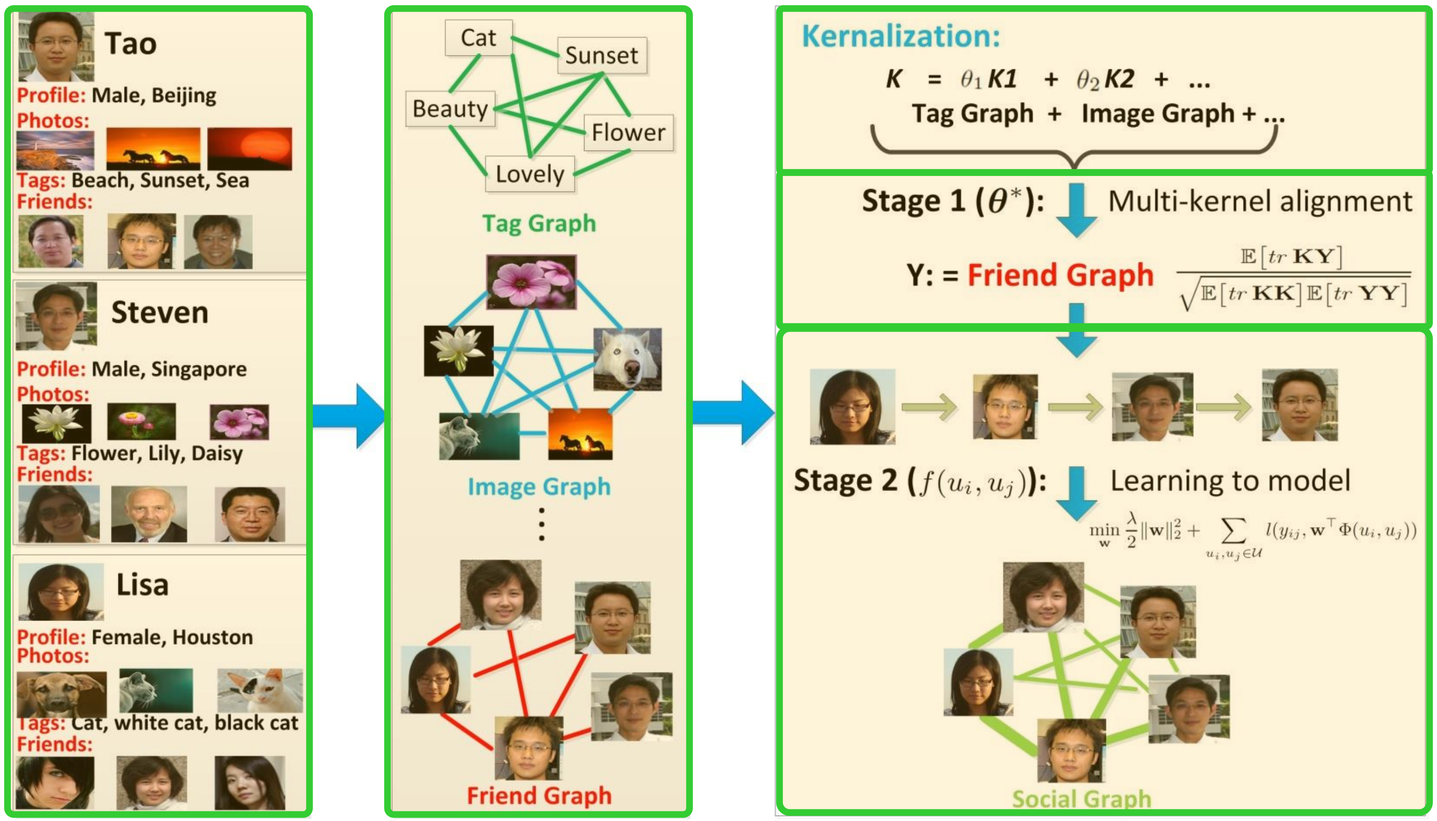
regularization (model complexity)

loss (prediction error)

where the solution is: $\mathbf{w} = \sum \alpha_j \Phi_j(\cdot,\cdot)$

Approach [Zhuang, MM'11]
model social strength via multiple kernel learning (MKL)

- Heterogeneous user data
- Multimodal graph representation
- Kernelized multi-graph over user space
- Multi-kernel alignment (θ^*) to weight kernels
- Learning to model relation: loss + regularization



User Information | Multimodal Graphs | Two-Stage Learning

Experiment

- Data
  - Flickr as the testbed (training : testing = 4 : 1)

#user	#group	#image	#tag	#contact
5,000	109,205	5,001,601	116,372	81,447

- Settings
  - Single kernel vs. uniform combination
  - Multiple kernel learning
- Evaluations
  - Friend/group recommendation
  - Gender/location prediction

Application: social relation/attribute modeling

App #1: interest group recomm.

Social signal	Precision@1
Co-locations	0.37
Similar photos	0.44
Mutual comment	0.46
Uniform fusion	0.48
[illegible]	0.55

App#2: social attributes prediction

- Friend recommendation (70% prec.@3)
- Gender estimation (72% prec.)
- Geo-location estimation (43% prec.)

Understanding Users

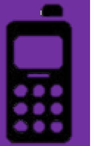
- Part 1: Understand user profile
  - Community discovery from heterogeneous data [Zhuang, MM'11]
  - Multi-social-graph construction [Yao, WWW'13]
- Part 2: Understand user context
  - Mobile visual localization [Liu, MM'12]
  - Mobile recommendation [Zhuang, Ubicomp'11]
- Part 3: Understand user Interaction
  - O-search [Zhang, MM'11]
  - SocialTransfer: cross-domain social media recommendation [Roy, MM'12]
  - Interactive multimodal mobile visual search [Wang, MM'11]
  - Browse-to-Search [Zhang, MMSP'11/2; Lu, MM'12]

80%

of the world's population has a mobile phone

5 Billion

mobile phones in the world



1.08 Billion

smartphone users in the world

50%

of smartphone users access social media services in 2012

33%

photos on social media sites are uploaded from smartphones

$24.5B

annual global revenue for mobile advertising in 2016

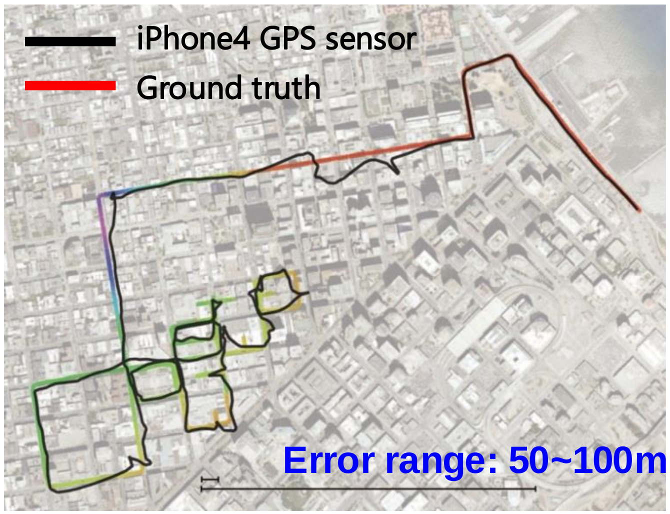
Mobile is driving the adoption of social media!

Robust and accurate mobile visual localization [Liu, MM’12; Xu, ICIMCS’12]

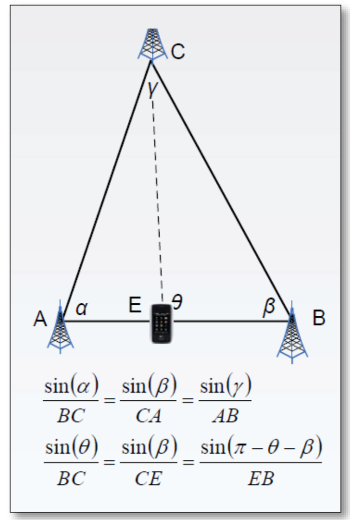
- Comprehensive and accurate sense of geo-context via mobile devices
  - real location of user and scene (error<15m)
  - view direction (error<9°)
  - distance of the target (error<21m), …
- Merits on both scalability and accuracy

User Location: (37.80, -112.41)
View Direction: 39.87°
Distance to Scene: 44.05m

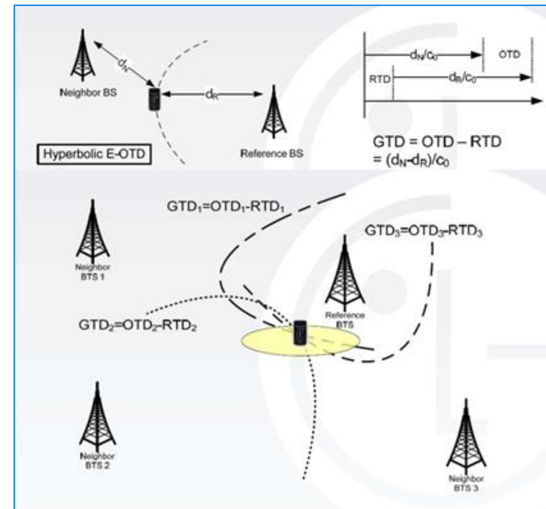
Mobile localization: GPS & wireless signal



GPS



AOA



E-OTD

GPS sensor [Schroth, IEEE SPM'11]

- Track recorded in San Francisco, using iPhone4 at dense areas
- Error range: 50~100m

Angle of arrival (AOA) [Wang, ICC'08]

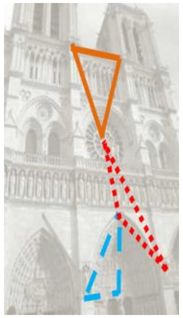
- Triangulation relationship
- Error range: 100~200m

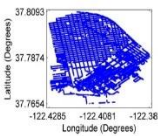
Enhanced Observed Time Difference (E-OTD) [Wang, ICC'08]

- Estimation of time difference of arrival of a signal
- Error range: 50~200m

Mobile localization: vision-based approach

- IM2GPS [Hays, CVPR'08]
  - Find nearest neighbors to estimate location distribution
  - 400 randomly selected queries from 6.5M Flickr photos
  - Locate 25% images within 750km
  - Far from accurate: global feature is not distinctive



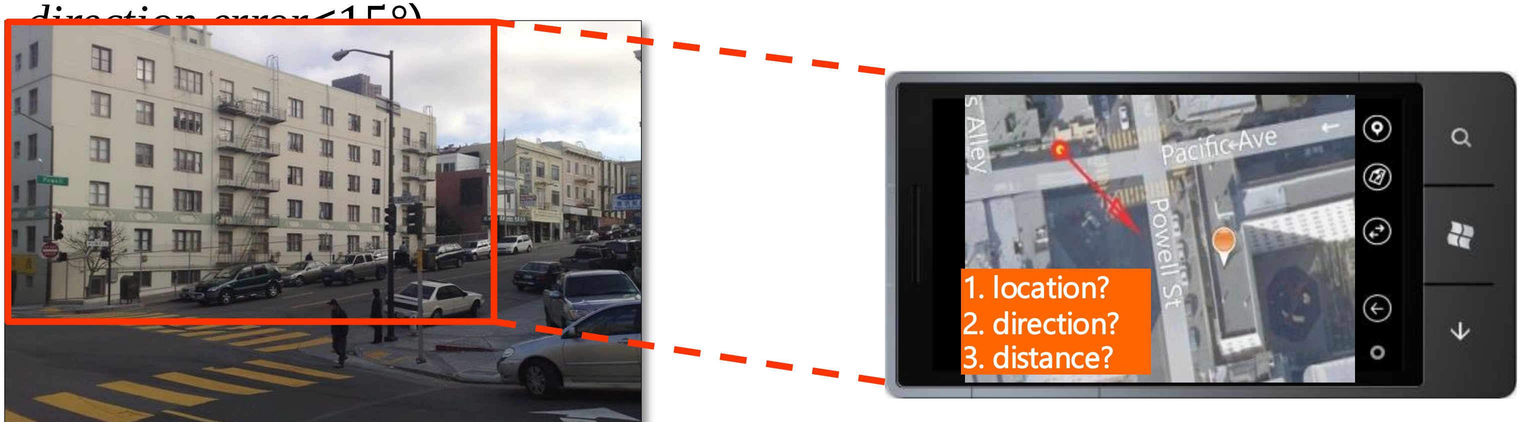


- Local feature for landmark identification [Chen, CVPR' 11; Park, MM'11; Hao, CVPR'12]
  - SIFT vocabulary search, GPS filtering, 3D visual phrase
  - Search similar images: accuracy 80% and recall 75% in SF (1.06M SV)
  - View direction error: 11 degree in NYC
  - Only return nearest images, failure rate ~50% [Yu, MM'11]

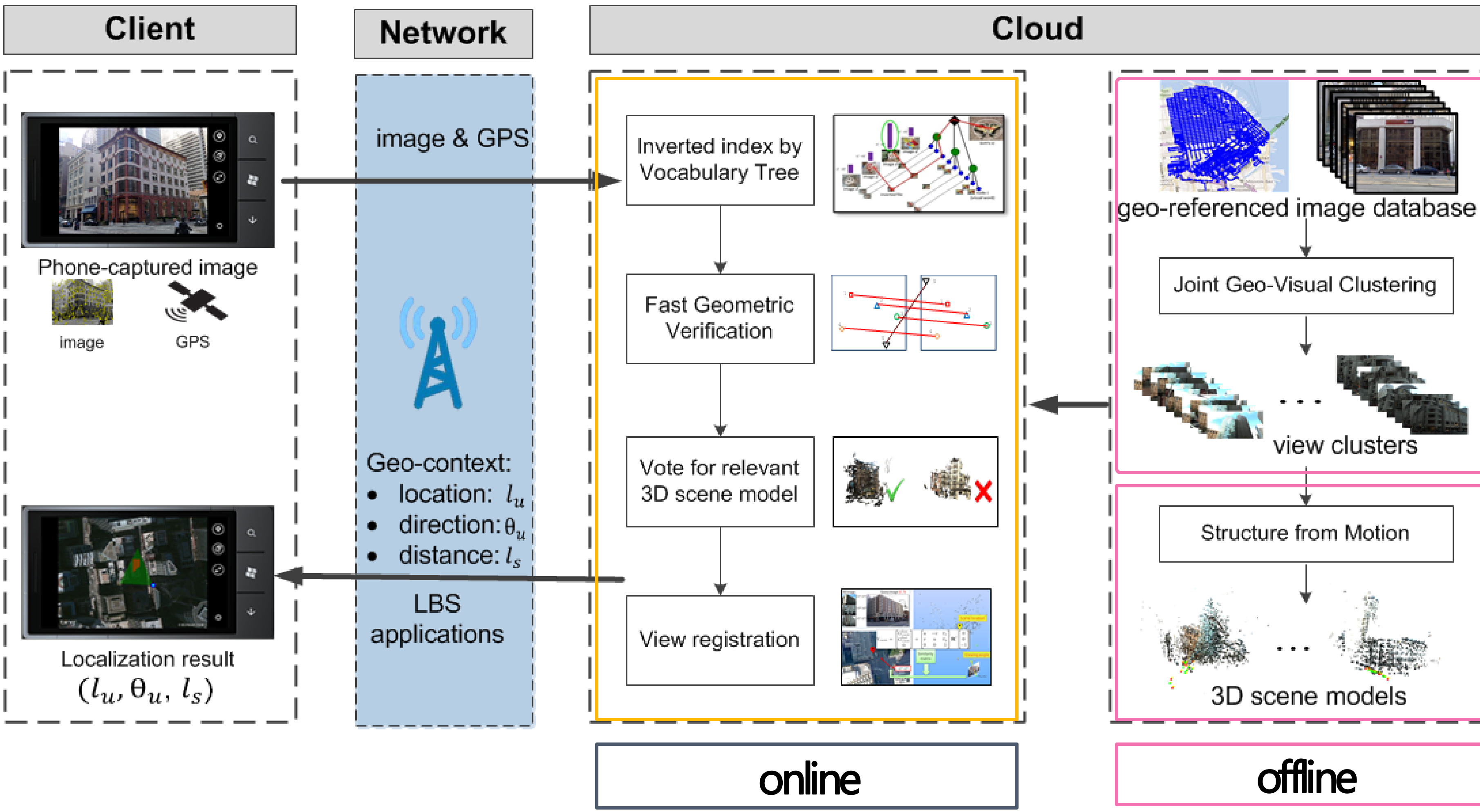
- Matching repeated patterns [Schindler, CVPR'08]
  - 2D-to-3D pattern matching to geo-located planar façade models
  - Localization error: ~6m on 9 façades from 7 buildings
  - Difficult to scale up (rely on existed 3D façade database)

Mobile visual localization

- State-of-the-art
  - GPS sensor: 50~100m (dense areas)
  - Wireless signal: 50~200m (dense areas)
  - Vision-based: incomplete result, not scalable
- This work, we are aiming at
  - Comprehensive location parameters (dense areas): location of user & scene | view direction | distance
  - Tradeoff between scalability and accuracy (e.g., $location\ error < 15m$, $direction\ error < 15°$)

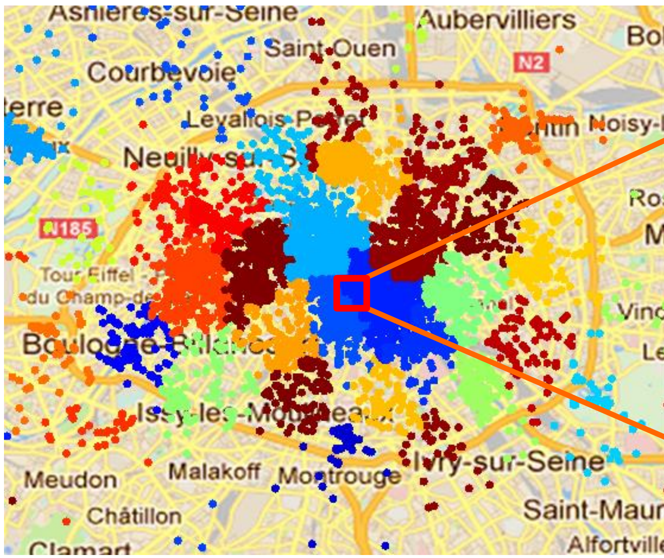


Robust and accurate mobile visual localization



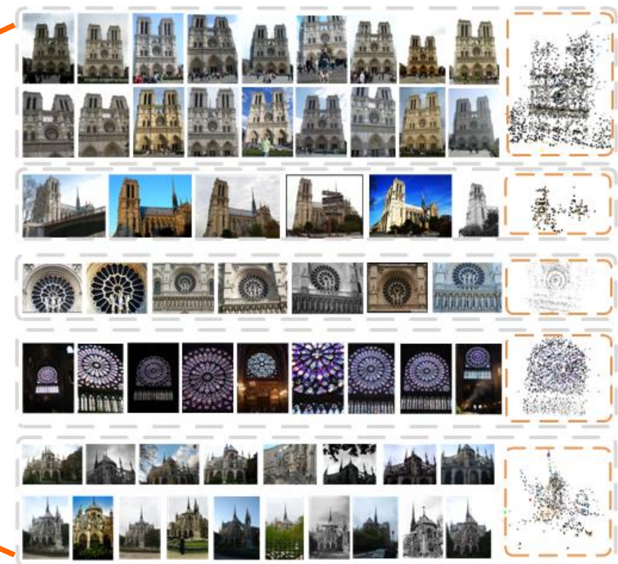
Step 1: geo-visual clustering (offline)

Geographic clusters @ Paris



$$Sim_g(i,j) = \begin{cases} g_{i,j}, & d_{geo}(i,j) < r_g \\ 0, & d_{geo}(i,j) \geq r_g \end{cases}$$

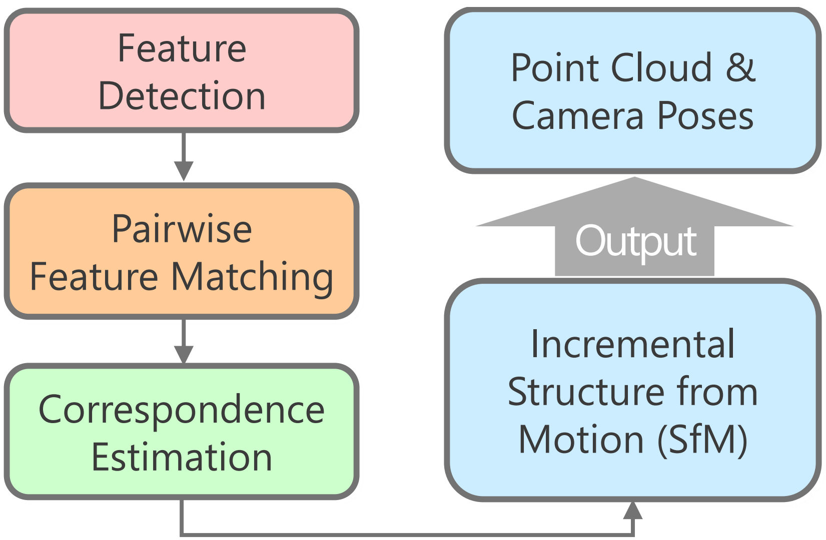
Visual clusters in a geo-cluster



$$Sim_v(i,j) = \begin{cases} v_{i,j}, & v_{i,j} \geq r_v \\ 0, & v_{i,j} < v_g \end{cases}$$

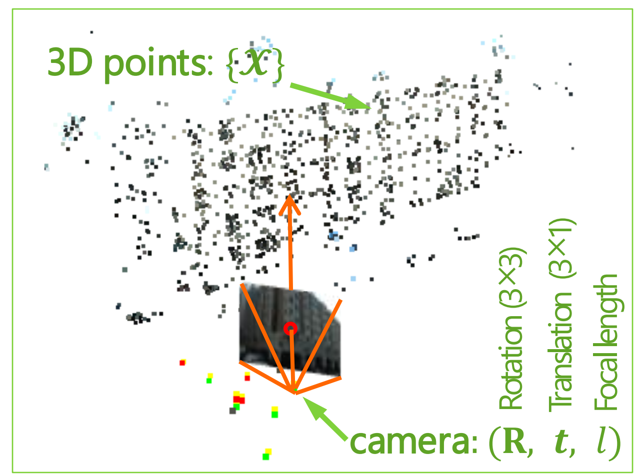
Step 2: scene reconstruction (offline)

- Build a sparse 3D point cloud of the scene
  - Build a 3D point cloud from the unordered candidate image
  - Use "bundler package" developed by [Snavely, SIGGRAPH'06]



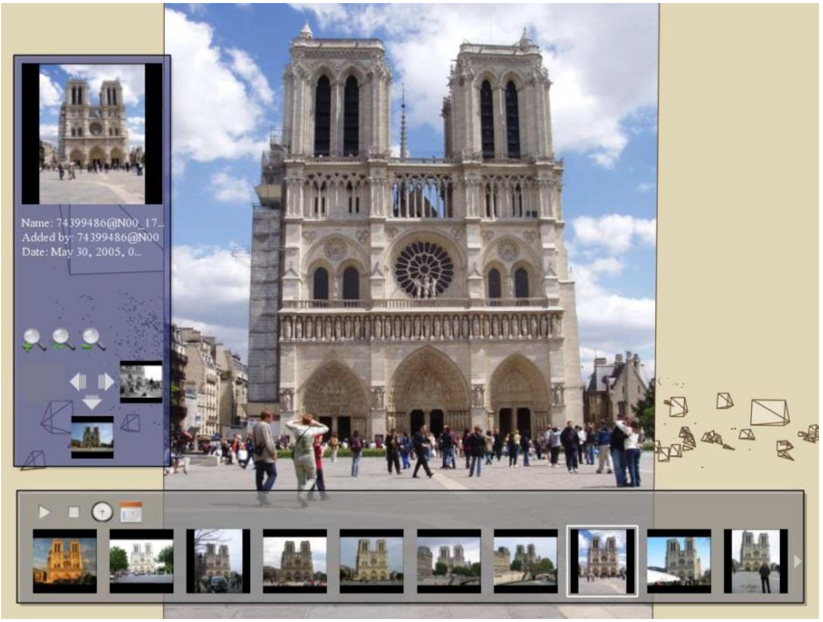
$$\arg\min_{X,\mathbf{R},t,l} \sum_{X^n,\mathbf{R}^m,t^m,l^m} w_{n,m} \parallel x^{n,m} - \widehat{x}^{n,m} \parallel^2$$

$$\lambda \begin{pmatrix} \widehat{x}_u^{n,m} \\ \widehat{x}_v^{n,m} \\ 1 \end{pmatrix} = l^m(\mathbf{R}^m\mathbf{X}^n + \mathbf{t}^m)$$



Reconstructed scene: $\{\mathcal{X}, (\mathbf{R}, \boldsymbol{t}, l)\}$

Step 2: structure-from-motion applications



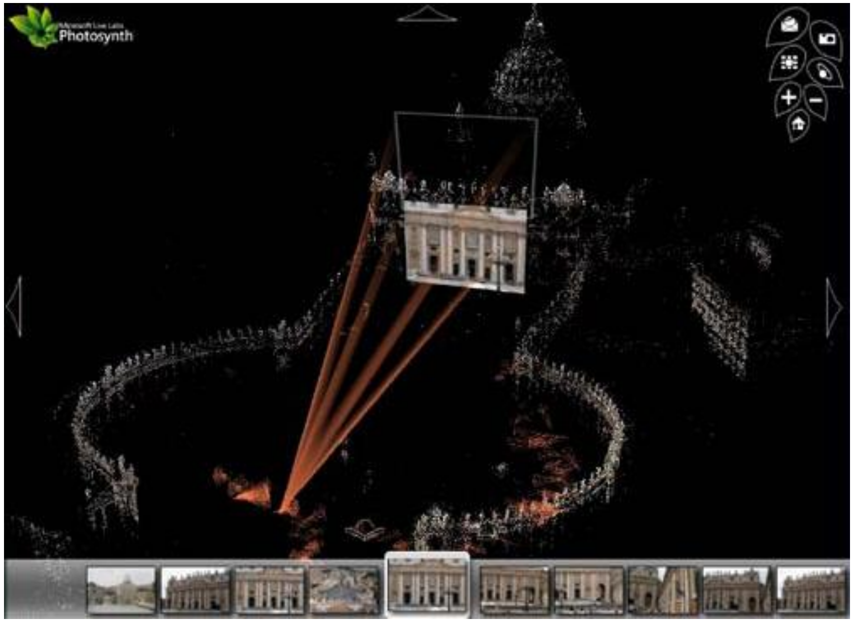
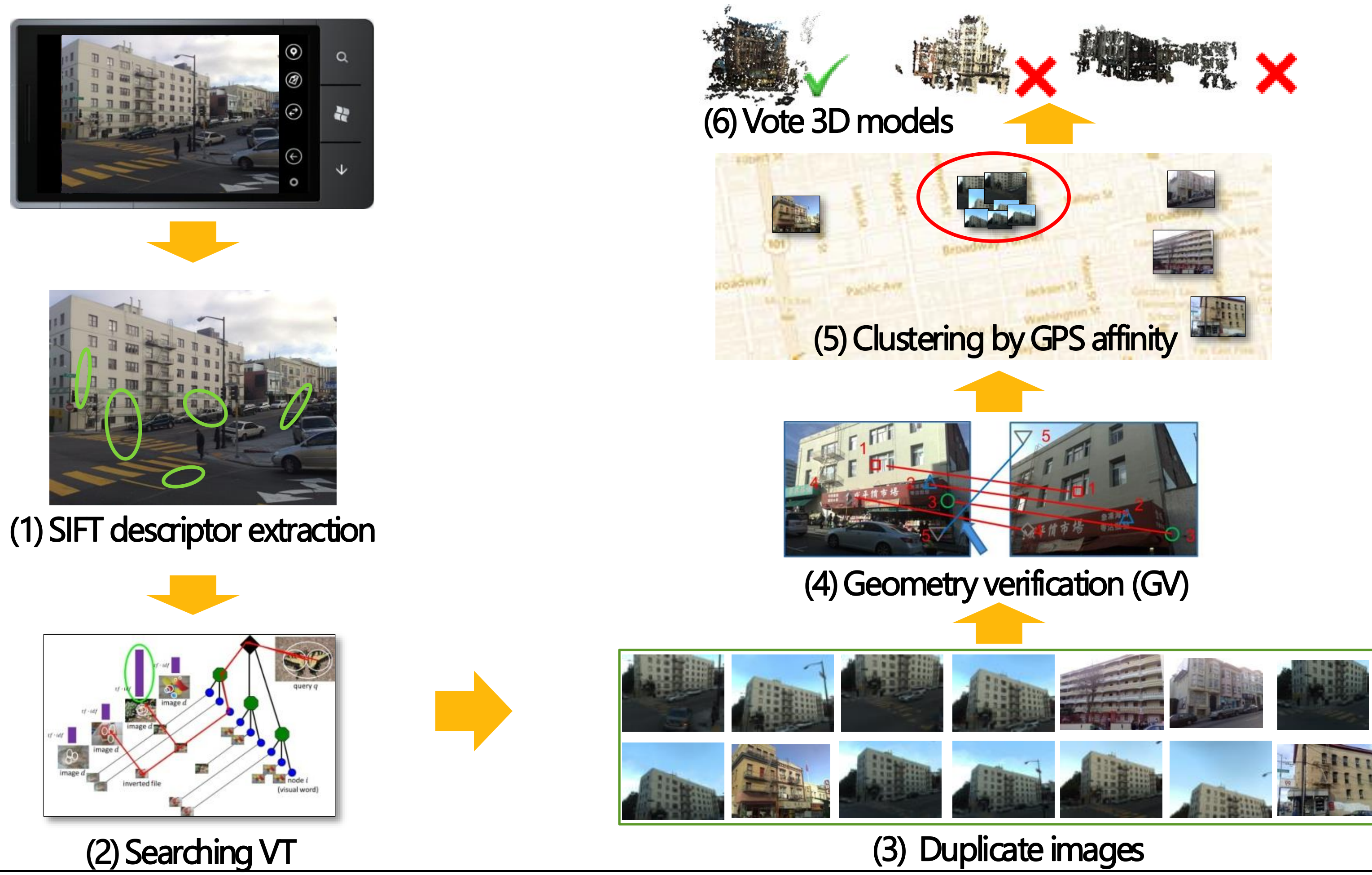


Photo Tourism [Snavely, SIGGRAPH'06]
- Reconstruct 3D points and cameras
- Browse photos in 3D way

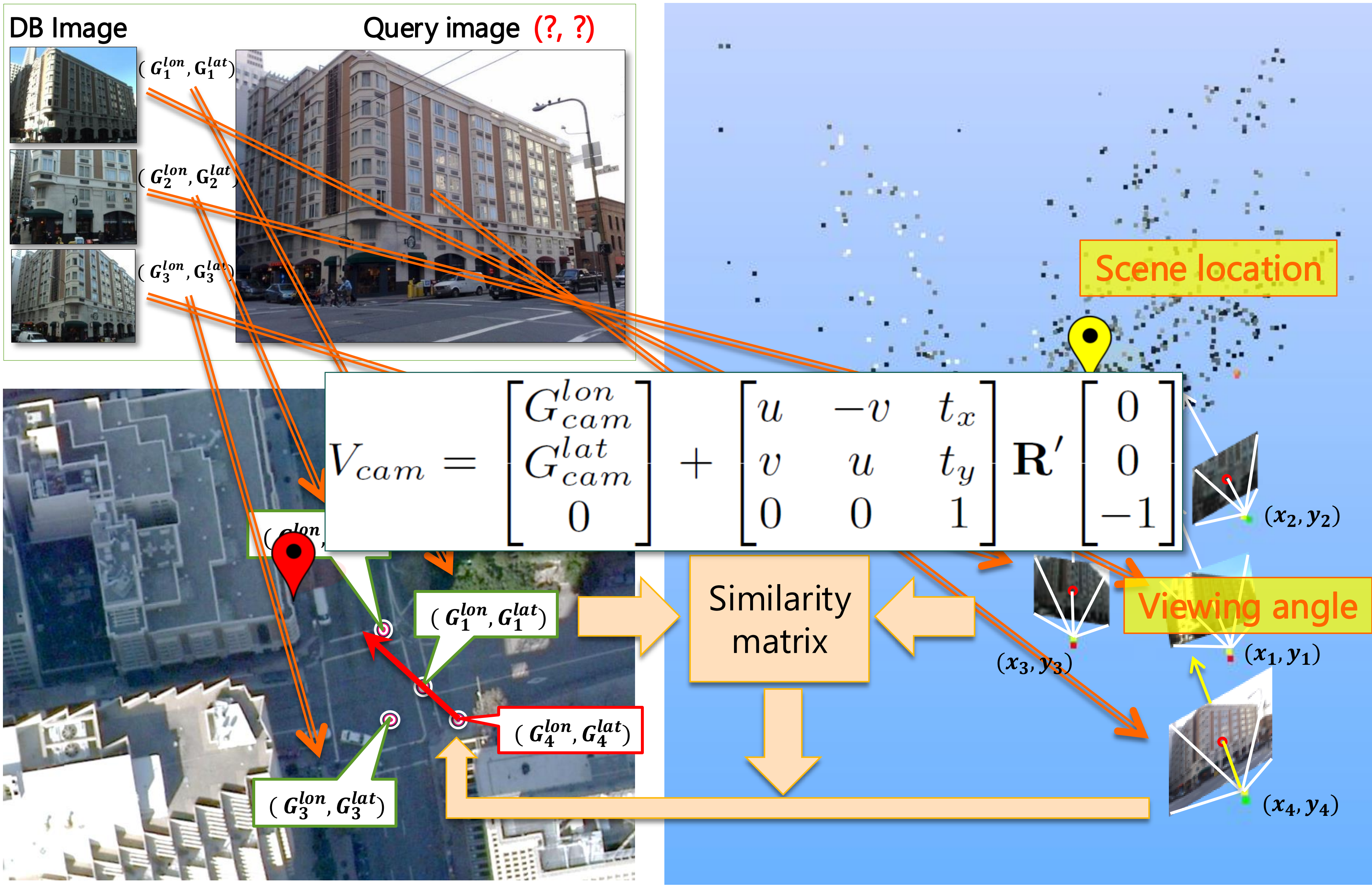
Photosynth [Snavely, IJCV'07]
- Capture and view the world in 3D

Step 3: 3D scene model retrieval (online)

- Online process: searching duplicate images with similar location

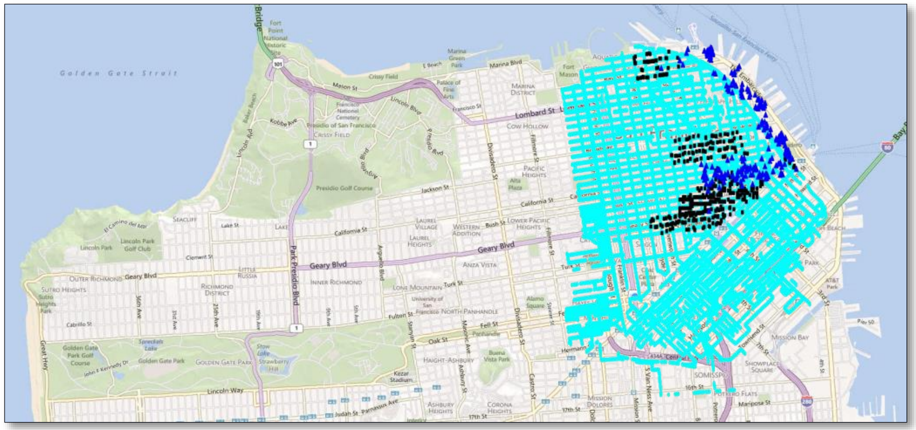


Step 4: geo-context estimation (online)



$$V_{cam} = \begin{bmatrix} G_{cam}^{lon} \\ G_{cam}^{lat} \\ 0 \end{bmatrix} + \begin{bmatrix} u & -v & t_x \\ v & u & t_y \\ 0 & 0 & 1 \end{bmatrix} \mathbf{R}' \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix}$$

Experiments



SV image locations

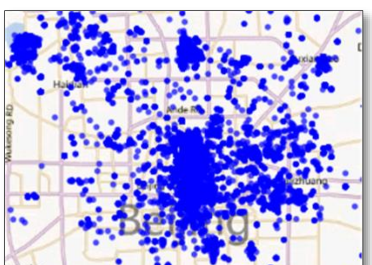
596 query image locations from GPS

207 query image locations simulated

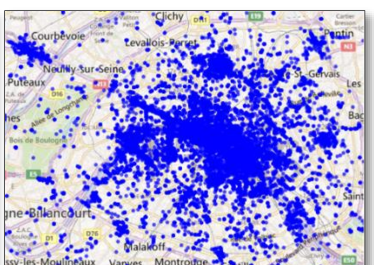
[Chen, CVPR'11] NAVTEQ

- 1.06M street view (SV) images from **San Francisco** [NAV
- 50~150 images per geo-cluster (50m), 15~40 per visual cl
- 803 query images, 283 (35%) queries can sense geo-cont

	Camera location	View direction	Scene location	Phone GPS
Errors	13.79m	8.96°	20.68m	30.34m
Std.	25.84m	11.71°	26.09m	44.33m



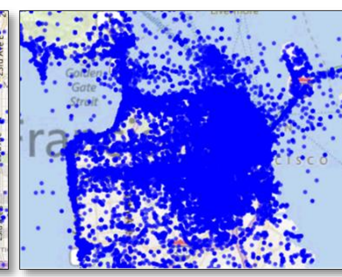
Beijing



Paris

NYC

Seattle



San Francisco

- 0.6M images captured in five cities from **Flickr**
- 50 images per geo-cluster (300m), 3~150 per visual cluster
- 50 query images about hotspots, 37 (74%) queries succeed

	Camera location	View direction	Scene location	Phone GPS
Errors	25.46m	14.55°	32.65m	50.49m
Std.	28.77m	13.88°	37.03m	66.33m

Evaluation of scalability

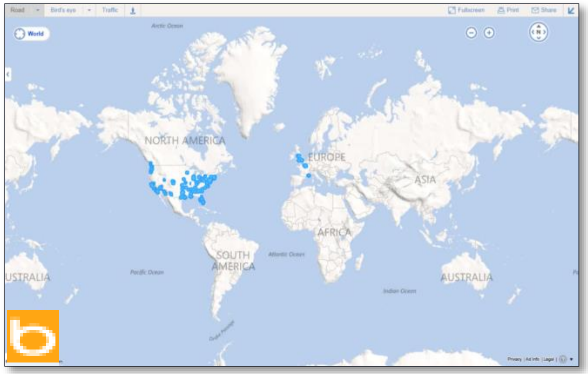
- Time cost

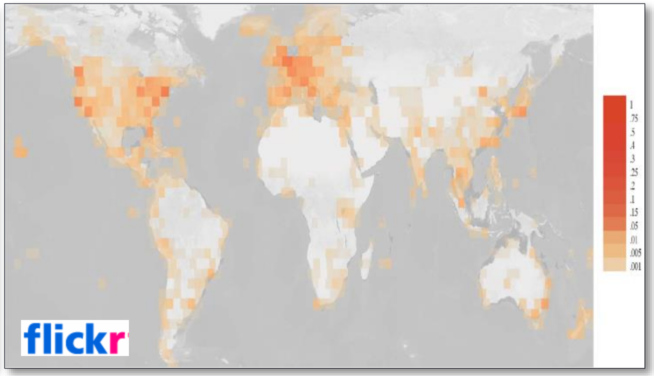
online offline

Operation	Uploading (wifi)	Feature extraction	VT search	Align to 3D model	Registration	Geo-clustering	Visual clustering	Reconstruction
Time (sec)	0.2	0.7~1.2	0.3	2~3	0.1	300	60~600	60~600/model

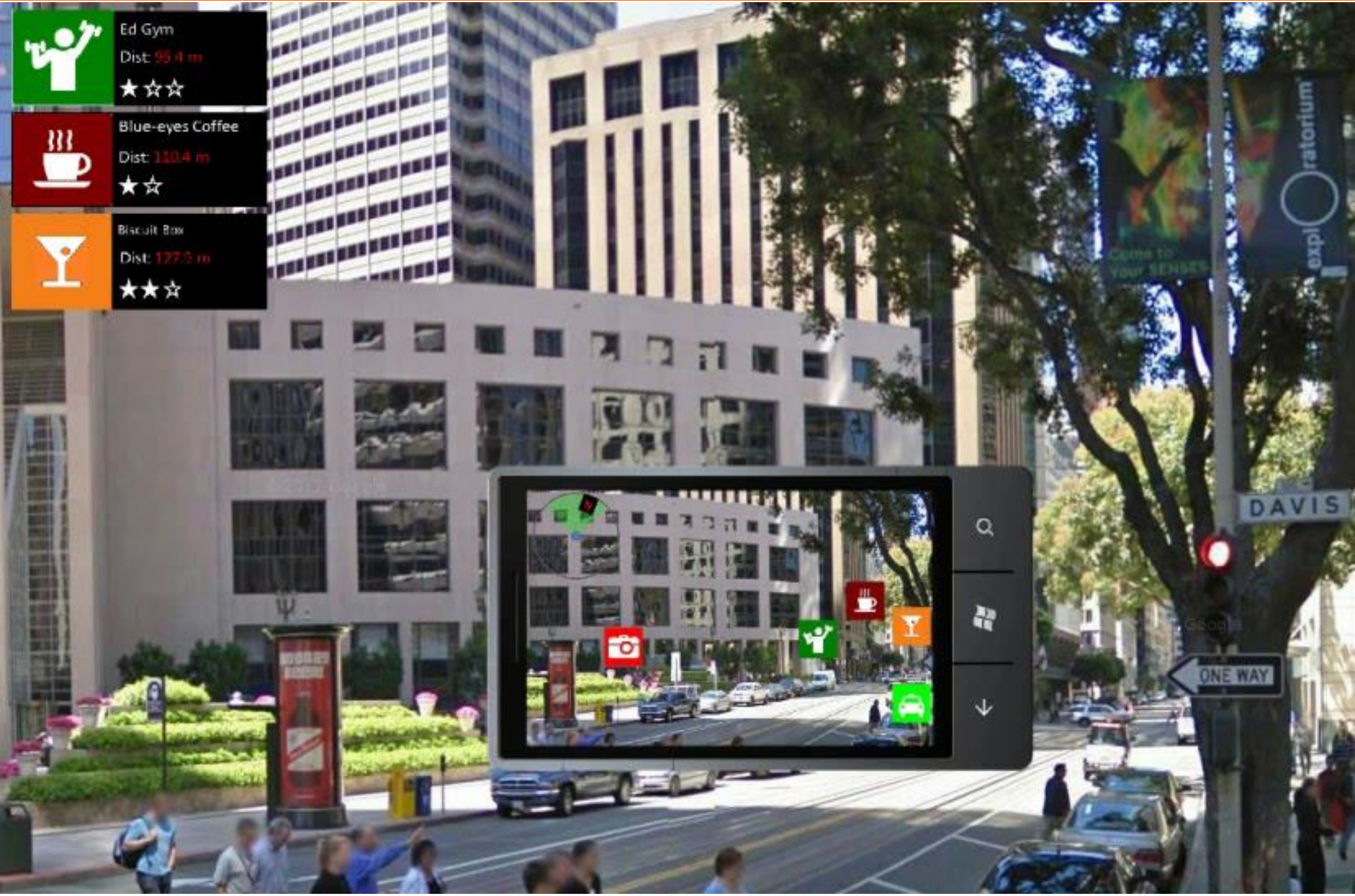
Note: initial positioning takes up to 45 sec for cold start, 35 sec for warm start, and 1-3 sec for hot start [Lehtinen, SoftCOM'08]

- Memory cost
  - 4GB for inverted file (1.06M SF SV images)
  - 20MB for each 3D model with 25 SV images
- Coverage of Street View images



Application 1: On-spot tour guide based on accurate localization



Application 2: Collaborate routing for finding perfect rendezvous

(a) user A

(b) user B

(c) shared map and route

(d) suggested rendezvous

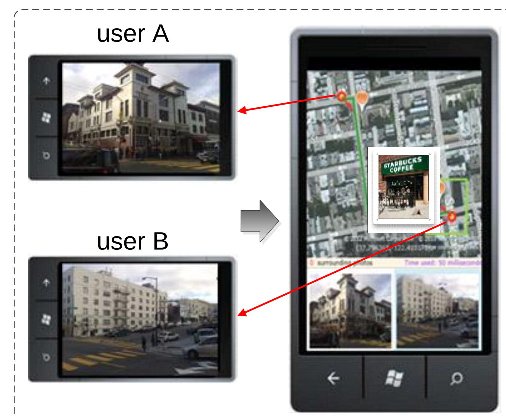
Application 3: Sight-seeing guide

User studies

- 12 Subjects
  - 6 graduate and 6 undergraduate students
- Simulated tasks
  - On-spot tour guide (App 1)
  - Collaborative routing (App 2)
- Summary of user studies

Questions	Score (1-5)
Attractiveness	3.7
Efficiency	3.3
Clarity of Intent	3.8
Practicality	3.8
Ease to use	4.2
Finding Rendezvous	4.4
Preference/Recommendation	4.4

App. 1



App. 2

Discussions

- 35% (283/803) queries can achieve accurate geo-context

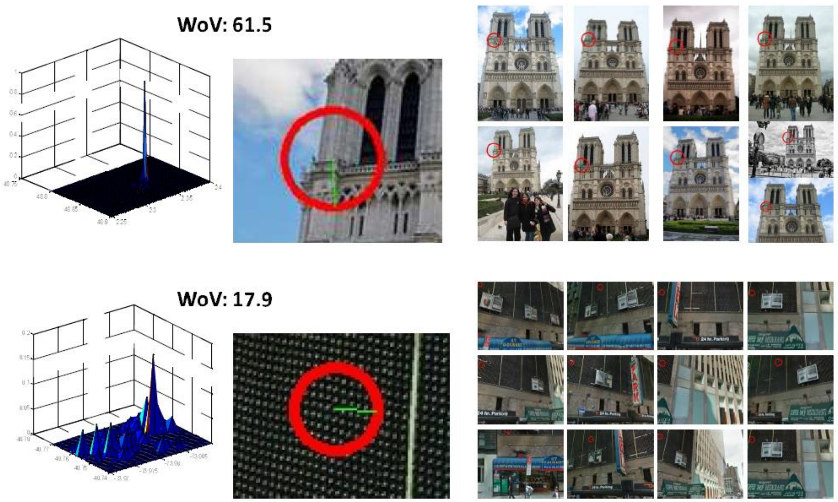
Image matching error-prone due to repetitive patterns

3D models not good enough from low quality images

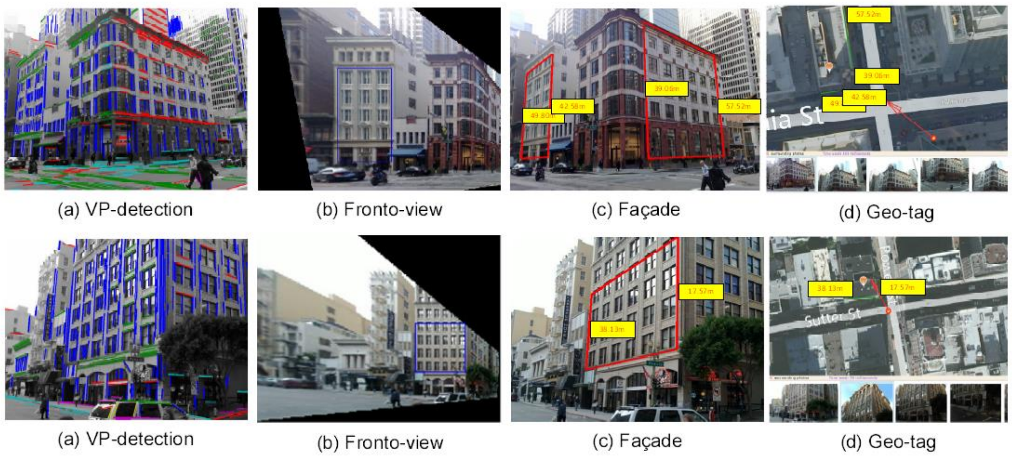
Challenges for mobile visual localization

- Suggesting the best view for scene recognition
- Investigating adding aerial images for localization
- Exploiting multi-view imagery or video for better localization

Location-discriminative codebook [TOMCCAP'13]



Parsing multi-view building façade [MMSJ'13]

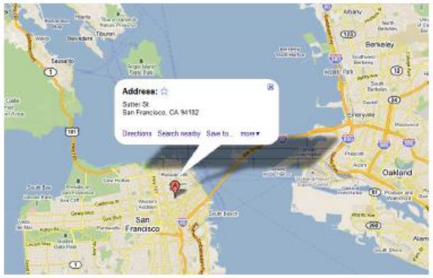


Contextual and personalized POI recommendation

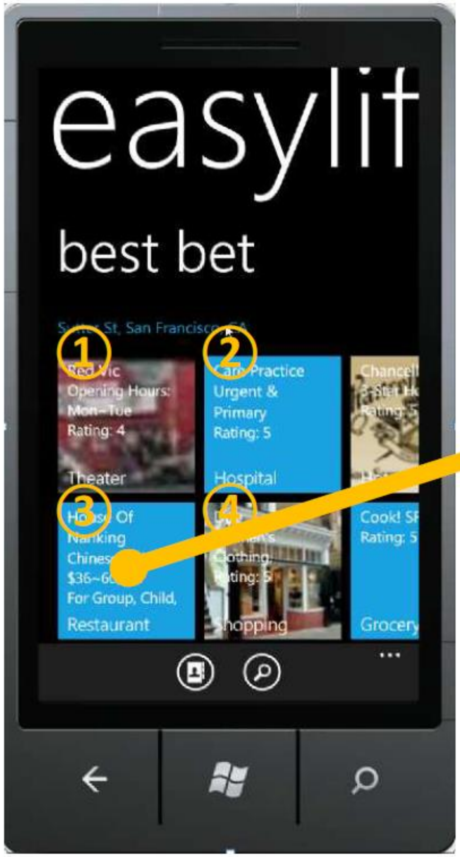
- Recommends local business for mobile phone users
  - Entity types (domains: restaurant, hotel, etc.)
  - Entities within each type (“Beijing roast duck,” “Hilton,” etc.)
- Uses “user+sensor” context for more relevant results
  - Personalized (user: behavior)
  - Contextual (sensor: time, location)
- Improves accuracy of recommendation up to 2.1 times

Scenario

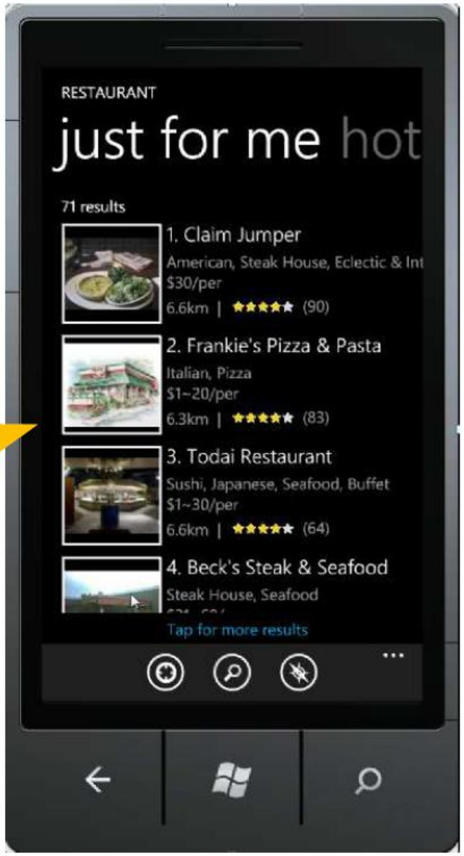
User: Clark
Time: 4pm
Location: Sutter St. San Francisco, CA, US, 94012



(a) user and sensory context



(b) rank of entity types



(c) rank of restaurant entities

Understanding Users

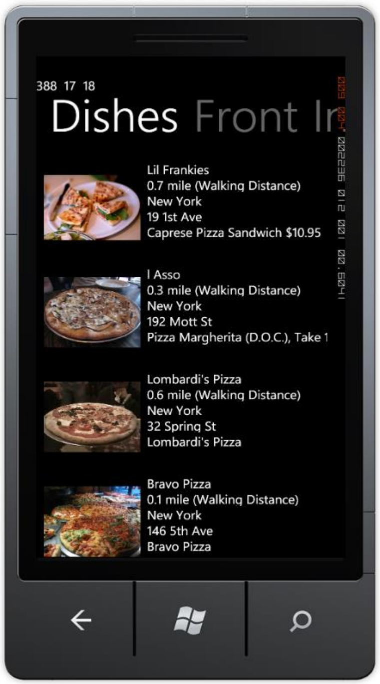
- Part 1: Understand user profile
  - Community discovery from heterogeneous data [Zhuang, MM'11]
  - Multi-social-graph construction [Yao, WWW'13]
- Part 2: Understand user context
  - Mobile visual localization [Liu, MM'12]
  - Mobile recommendation [Zhuang, Ubicomp'11]
- Part 3: Understand user Interaction
  - O-search [Zhang, MM'11]
  - SocialTransfer: cross-domain social media recommendation [Roy, MM'12]
  - Interactive multimodal mobile visual search [Wang, MM'11]
  - Browse-to-Search [Zhang, MMSP'11/2; Lu, MM'12]

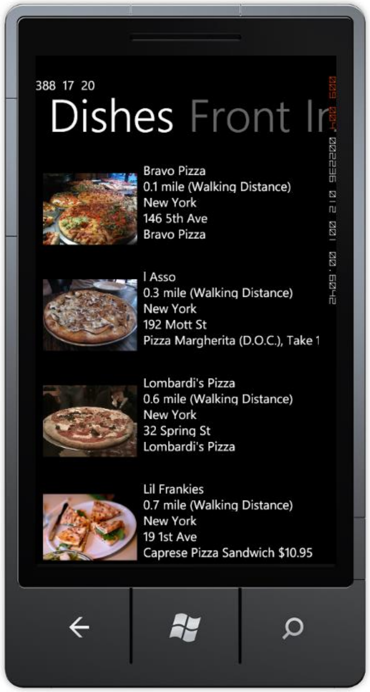
O-search: user interactions w/ mobile [MMSP’11, MM’11, TMM’13]

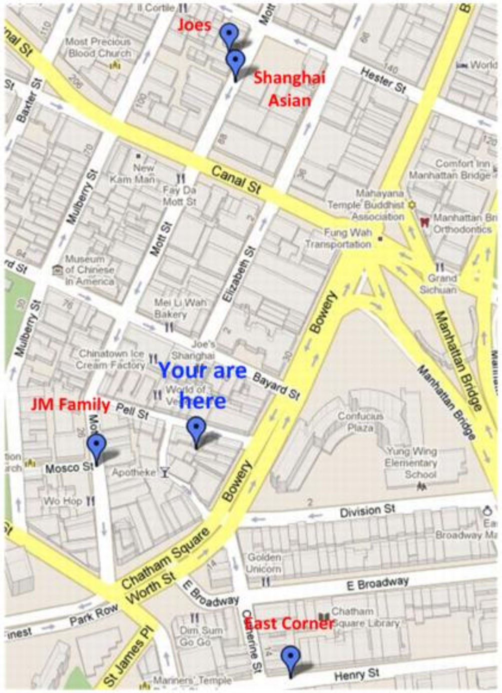
(a) Prototype (b) Crop (c) Line (d) Lasso (e) Tap

- Interactive expression of visual intent via mobile phone
  - “Capture + lasso” as the query
- Enables natural interactions between user and phone-image
  - Tap (pre-segmentation) | Line (rectangle) | Crop | O (lasso)
- Contextual visual recognition on-the-fly
  - Understand what is captured | lassoed| indicated (OCR)
  - Context-embedded Vocabulary Tree (CVT)
- Recognition-based entity recommendation
  - Recommend nearby entities serving the same captured food

Task recommendation





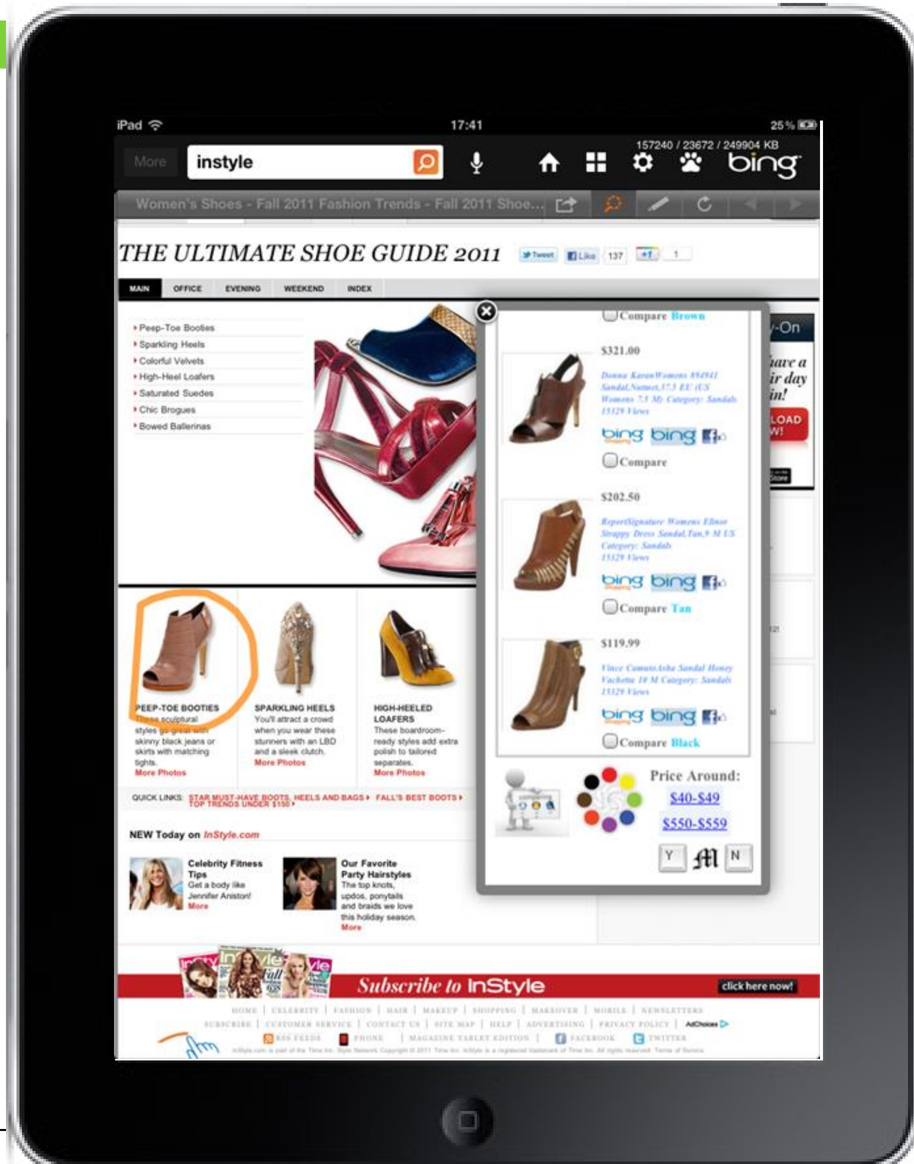


Lassoed image

Top-1 visual search result, "Bleecker Street Pizza at 69 7th Ave S. New York"

Nearby restaurants serving "Pizza" via text keyword search

Browse-to-Search [Lu, MM'12, TOIS'13]



Summary

- Understanding user from social media
  - Understanding user context
  - Understand user (profiling)
  - Understand user interaction

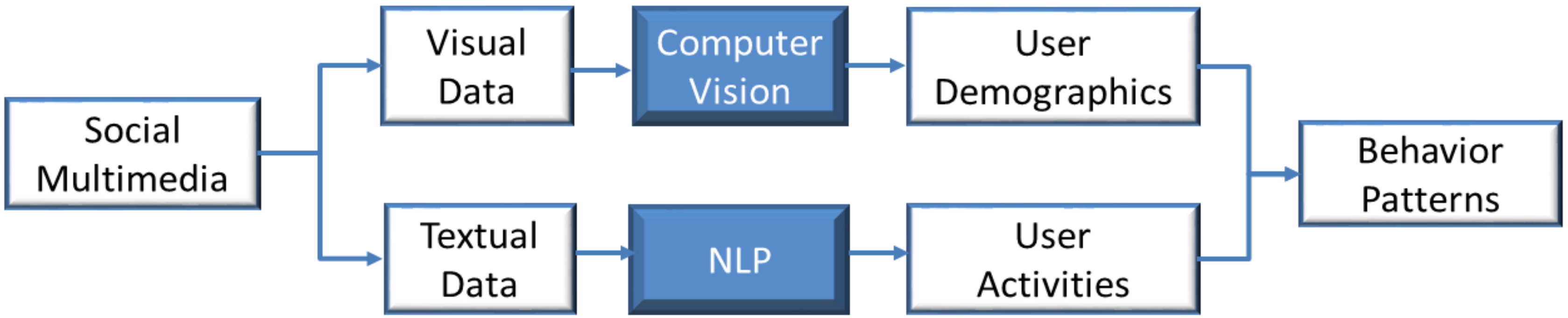
Social Media for Social Good

- Like it or not, the information is out there, the question is whether we use it for social good and what social good we can do.

Social Media for Social Good

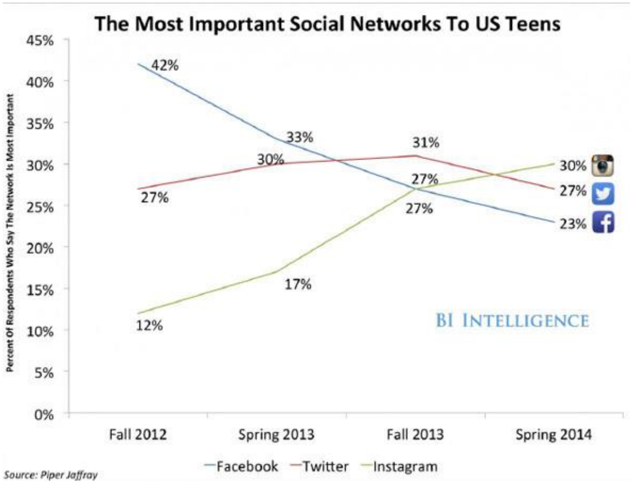
- Under age alcohol
- Drug usage
- Urban sensing
- … …

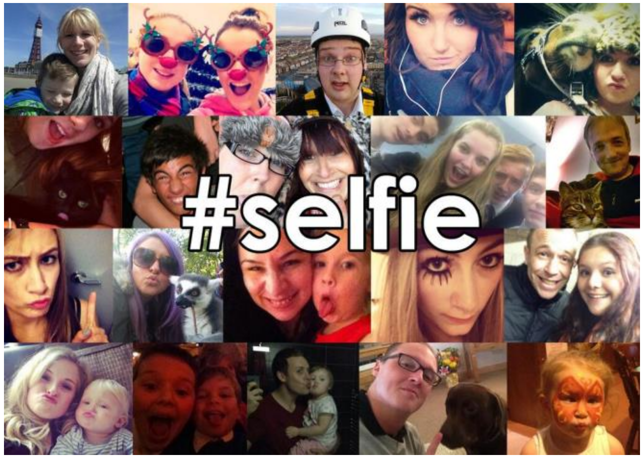
Drinking Behavior

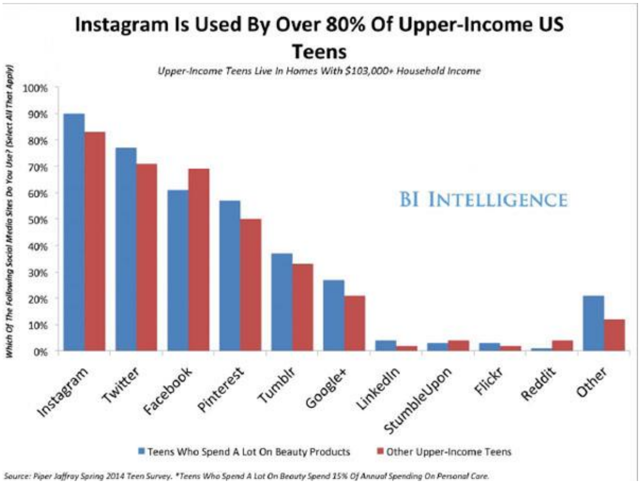


R. Pang, et al., Monitoring adolescent alcohol use via multimodal analysis in social multimedia, IEEE International Conference on Big Data, 2015.

Drinking Behavior





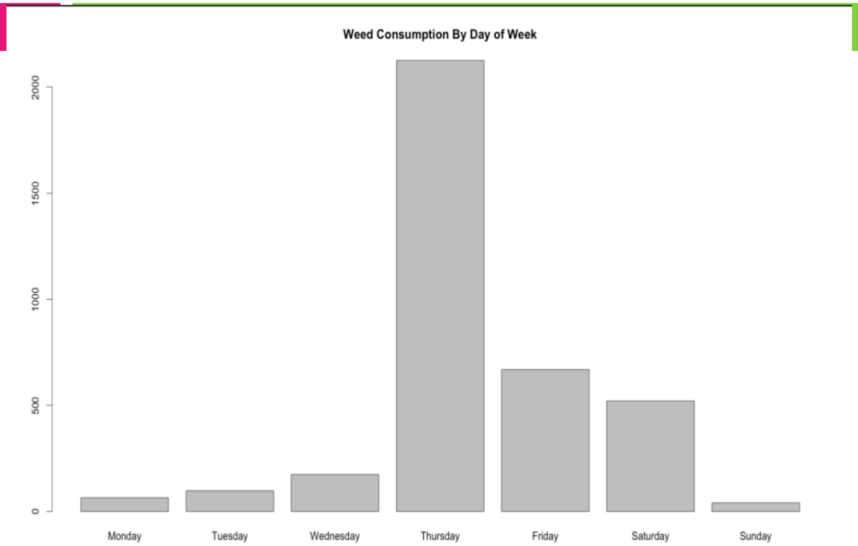


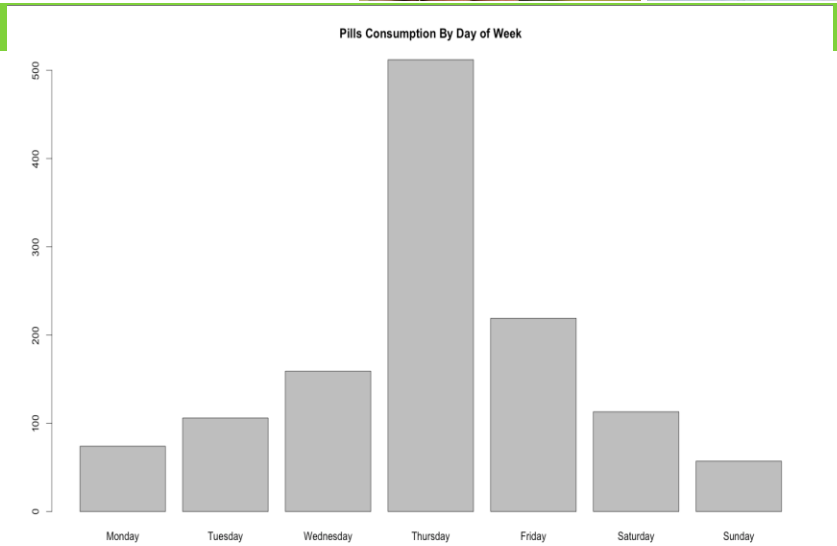
Drinking Behavior

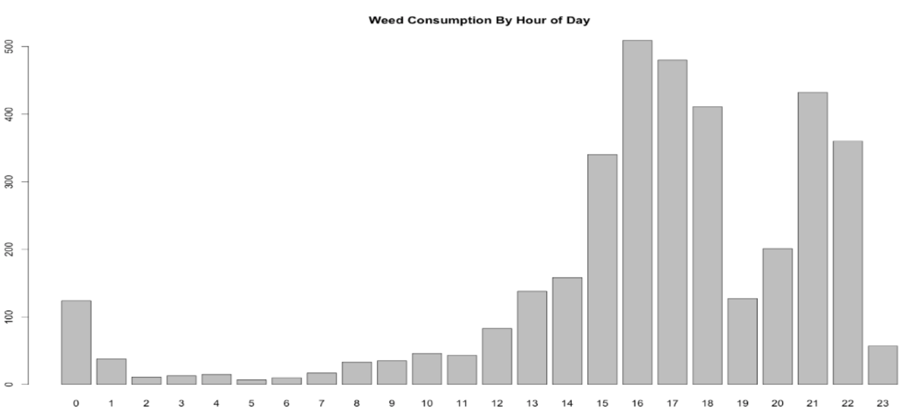
- Youth Exposure to Alcohol Media

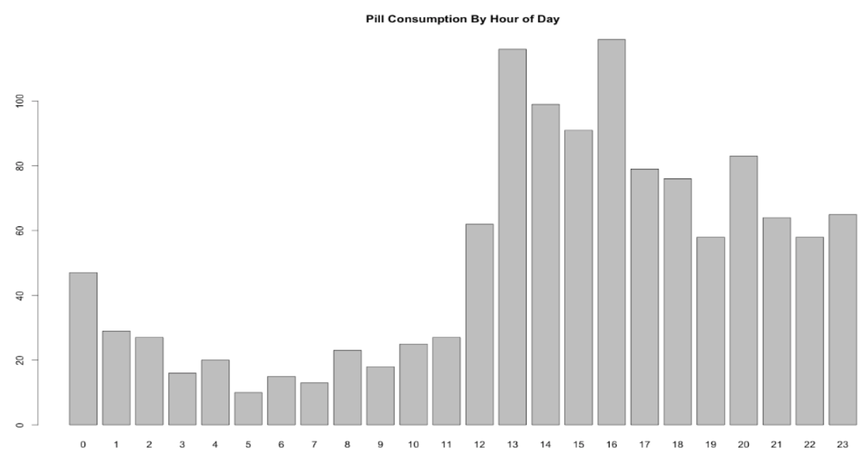
	Vodka 1	Vodka 2	Champagne	Beer 1	Beer 2
Young Male	6.43%	6.79%	6.10%	13.21%	10.95%
Adult Male	29.69%	42.16%	24.27%	**52.41%**	**51.91%**
Young Female	**19.76%**	**15.12%**	19.49%	11.58%	12.17%
Adult Female	44.12%	35.93%	50.14%	22.79%	24.97%

Drug Use Patterns from I









Main Contributors: Yiheng Zhou, Numair Sani, Jiebo Luo

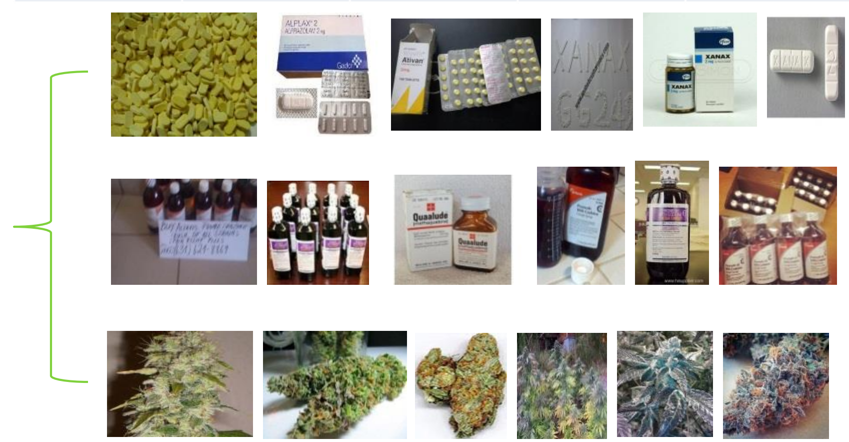
Drug Image Classifcation

- Fine-tuned CNN
  - Starting with the pre-trained VGG Net
  - Fine-tuned CNN features + SVM
    - Using noisy data downloaded from Google
- Fine-tuned data statistics
  - Instagram photos

label	pills	bottle	weed	total	Non-drug
#	2421	1233	675	4329	12253

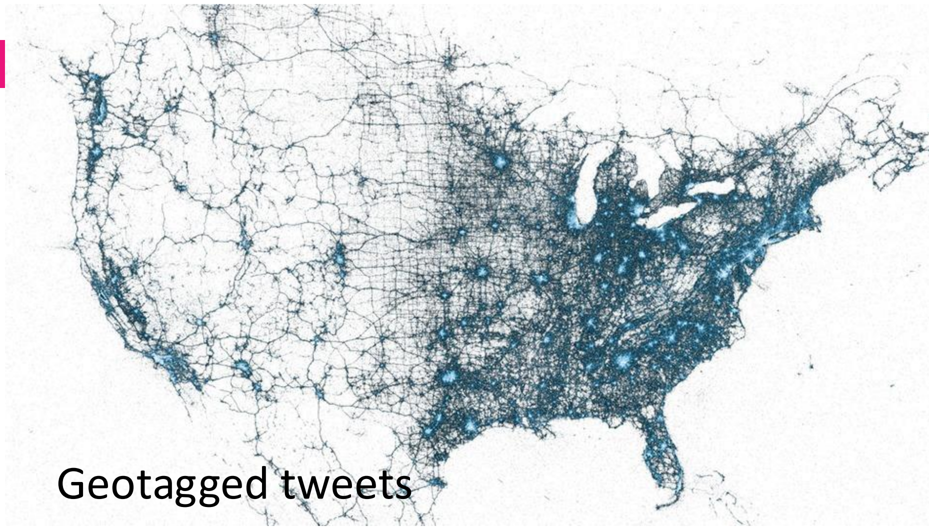
```
pills   2mg,10mg,15mg,30mg,80mg,adderall,alprazolam2mg,alprazolam,ambien
,benzodiazepines,clonazepam,diazapam,dilaudid,drug,ecstasy,Endocet,hydro
codone,illegal drug,illegal pills,klonopin,kpins,Magnacet,madma,methadon
e,molly,narcotics,opiate,opiates,oxycodone,oxycotin,painkillers,Percocet
,percocets,percosets,percs,prozac,roxicodone,roxies,roxycodone,suboxone,
tramadol,trazadone,valium,vicoden,vicodin,vicodine,vicodins,xanax,xanaxb
ars,xanex,xanies,xannies,xans,zannies
bottles actavis,alpharma,codeine,hydrocodonesyrup,hydromet,hydromorphone
,morphine,nosealnodeal,pill bottles,promethazin,promethazineandcodeine,p
romethazine,actavis sale,drug lean,syrup drug
weed    bammer weed,bc bud,blunt,Psilocybin,xanax4sale,marijuana,bud4sal
e,ganja buds,kush,purp,weed4sale,weed
others  heroin,prescriptions
```

	Accuracy	Precision	Recall	F-score
Method 1	91.93%	87.2%	71.5%	0.7822
Method 2	89.05%	71.04%	78.%	0.7458



Main Contributors: Xitong Yang, Meredith McCarron, Lacey Kelly, Jiebo Luo

Urban Sensing



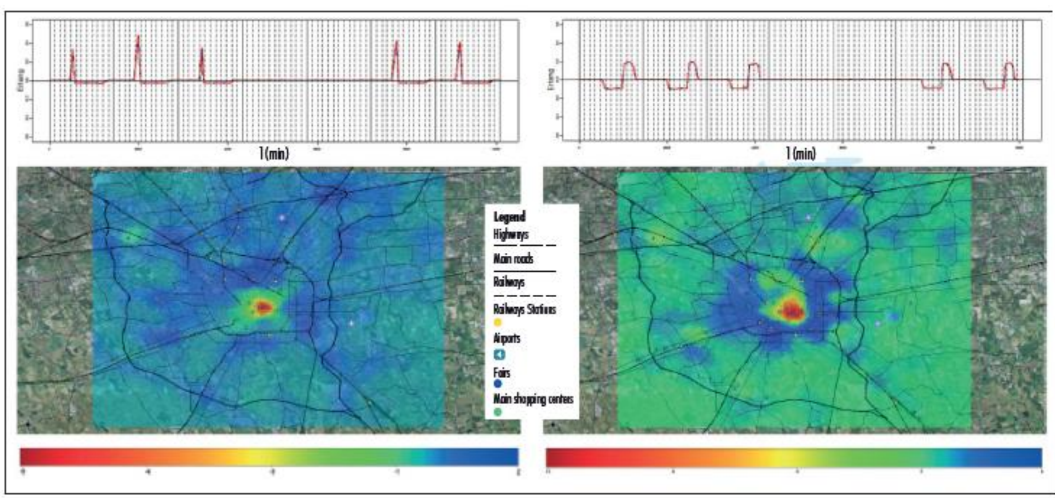
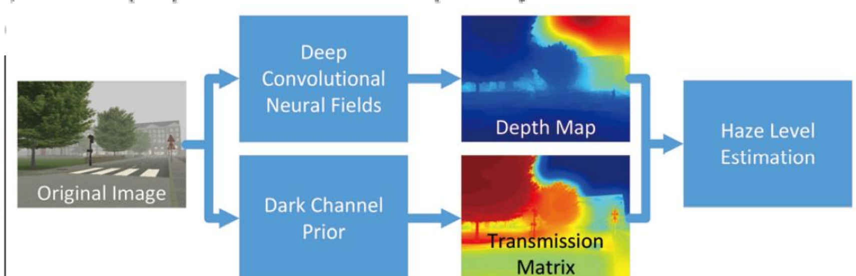


Figure 2 - Daily mobility spaces: morning rush hour map (left), evening rush hour map (right)

Source: DASTU; MOX, 2013.

Morning and evening rush hours



15)

Trends and Open Issues

- Heterogeneous multimedia data
- Heterogeneous information networks
- Dynamic nature of social signals
- Cross-platform and real-time response
- Geospatial-temporal data analysis (trajectory)
- Socio-mobile: social + mobile
- Tighter collaboration between industry and academia
- Continued search for killer applications

Need To Know

- How to understand and characterize users
- How to leverage social media for social good
- Open issues and emerging applications

Feedback to Project Proposal

- + a diverse range of interesting topics
- + professional/academic typesetting, such as using template provided

- - including full name
- - key information is missing at project title
  - no title, do not know the specific task of the project
- - lack of understanding of the Purpose of Abstract or Executive summary
- - business oriented proposal, not academic oriented
- - solutions are not supported with scientific papers
  - no method/solution or generic solution
- - only several pages/too long with imbalanced length
- - misunderstanding of Reflection (not conclusion/summary)

Feedback to Project Proposal

- 1. Typesetting: be professional, pay attention to font, alignment, and consistence, …
- 2. Cover page, student id and unikey, not full names
- 3. Not necessary to follow the exact suggestions from resources on proposal writing. Only take those suitable for this assignment (e.g., creativity and/or significance, challenges). If the same information has been covered in Introduction, no need to repeat in Significance.
- 4. Focus on the project topic and task
  - a) Project title is needed.
  - b) Introduction section needs to provide necessary information, not too lengthy, not too brief, such as what is the problem/task, why it is important for you to work on it or for us to care about it.
  - c) Ordinary topics: course video retrieval with text, video classification, image retrieval
- 5. Related work:
  - a) needs sufficient information on relevant or related studies to inform readers how you know about this topic or field.
  - b) not just list books/papers. You need to actually review it: what they did in relevant to your task.
  - c) you need to categorize them, instead of talking about them one by one.
  - d) techniques vs applications (techniques for detecting birds vs for detecting general objects)

Feedback to Project Proposal

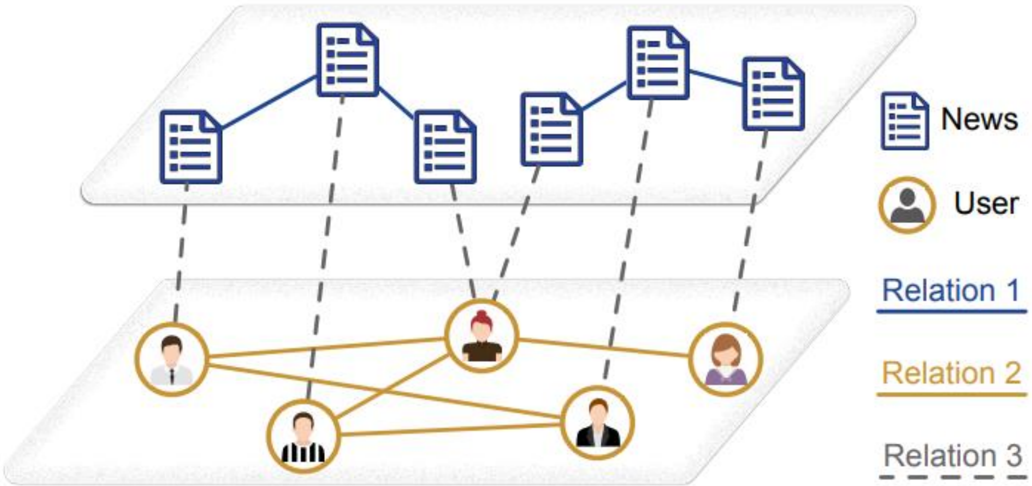
- 6. Solution:
  - a) Needs to provide the basic "recipe" on how you will accomplish the goal or task, such as overview pipeline and key steps.
  - b) Needs to refer to specific papers and explain the steps (e.g., what algorithms do you have in mind, not that I can try this, will also try that); not a list of libraries you will be using, not like a set of steps like a workflow;
  - c) This is multimedia computing task related project, not a general software development project. The focus is on techniqeus. It is not necessary to provide details of development tools and language.
  - d) If you plan for a UI, a mock-up would be better included.
  - e) Evaluation and Dataset: where to download or collect data and how to evaluate
  - g) Find solutions that are closely related to the task. Otherwise, anything coudl be a classificaiton or detection problem. References are too old (refer to venues mentioned in assignment notes).
  - h) Avoid using the techniques of which the source codes are public, such as Faster CNN, and FCOS.
- 7. Plan:
  - a) milestone, links to specific tasks/steps;
  - b) the gantt chart is too small to read; to generic (e.g., design, implementation, building the model, testing); unreasonable plan (e.g., one week for Euclidean distance computation);
  - c) The steps are too detailed, for example more than 2 pages out of 10.
- 8. Reflection on proposal:
  - a) It is about how you structure or organize the proposal according to what principles. It is from a tutor's perspective that you want to teach others on what a good proposal is.
  - b) It is NOT about anything else, such as roles of individual members, how you found your project topic, ...
  - c) Talk about good elements of a proposal. However, I did not see practice in own proposal writings.

Guide to Final Report

- Aim to provide necessary information on what has been accomplished
  - For example, task, solution, and outcome
- You can follow the structure of an academic paper
  - Title, Abstract (Summary of the report)
  - Introduction (E.g.., the task, why it matters, what you plan to do (and why))
  - Related Work (to provide a context of the field)
  - Method (overview illustration/diagram of the solution + details)
  - Experimental Results/Discussions
  - Conclusion (Optional): summary of what you have reported
- Project Video (introduce your project)
  - NOT a recording of your presentation.
  - Do not be too abstract without technical content
- Reflection on Presentation
  - Your understanding to prepare a good presentation

Fake News Detection

- “consuming highly concentrated alcohol could disinfect the body and kill the virus” resulted in more than 5,800 hospitalizations



News relationship

User interaction

Posting

X. Su, et al., Mining User-aware Multi-relations for Fake News Detection in Large Scale Online Social Networks, WSDM, 2023.

Fake News Detection

Fraudster Detection

K. Maity, et al., A Multitask Framework for Sentiment, Emotion and Sarcasm aware Cyberbullying Detection from Multi-modal Code-Mixed Memes, SIGIR, 2022.

Cyberbullying Detection

K. Maity, et al., A Multitask Framework for Sentiment, Emotion and Sarcasm aware Cyberbullying Detection from Multi-modal Code-Mixed Memes, SIGIR, 2022.