

The Use of Social Observatories for Research on Public Events, User Mobility and Wellness

### **CHUA** Tat-Seng

KITHCT Chair Professor School of Computing, National University of Singapore

## OUTLINE

- Rich Social Media Contents
- The Live Social Observatory
- Analysis of Large-Scale Social Media Content
- User Demography & Mobility
- Wellness
- Summary



# The New Information Age

- The Internet has revolutionized the way information is created, disseminated and consumed
  - Emergence of huge amount of end-user generated data, especially in social networks



Greater connectivity leads to huge amount of live info



• Internet has also rapidly gone mobile, permitting access from anywhere



## Users at Center of Social Media Environment



# Review of Social Media Platforms -1

twitter

- Social Network platforms
  - Three major platforms: Private, Professional, Public

Linked in .

Media Sharing platforms

facebook

YouTube, Instagram, Flickr, ..., Periscope, Vine



- Social Messaging Platforms
  - WhatsApp, LINE, Wechat, SnapChat, ...





# Visual-based Platform: SnapChat



- Snapchat: founded in 2011; Sharing 10-sec video moments that erases itself after certain period; 70% of users <24 yrs old (~100M active users); user base will be bigger than Twitter soon
- Temporal Real Time Messages: people decide how long others view their photos/videos
- Share live stories with friends
- "Discover" news from top sources, eg CNN.
- Vine: 6-sec loop-videos





## Periscope

- Similar to SnapChat in supporting live broadcast from mobile phones
- Users can broadcast and watch any other live broadcasts ۲
- Why it is hot??
  - Support mobile-based broadcast
  - Track events live
  - Interactive
  - Integrate with Twitter Social Network



**Broadcast Video** LIVE from Anywhere

# **Review of Social Media Platforms -2**

Location-based Services

foursquare

4Square, Yelp, Gowalla,..





Social Curation Platforms: Pinterest ...

yelpe



Recent popular apps are all image/ video based..



# Is Social Media a New Phenomenon?

- Social Interactions are inherent in human nature, only the mode changes
- In fact Selfie is now common place,..



But Selfie happened in ancient times too.
Like this painting in Hermitge Museum <sup>(C)</sup>



## OUTLINE

- Rich Social Media Contents
- The Live Social Observatory
- Analysis of Large-Scale Social Media Content
- User Demography & Mobility
- Wellness
- Summary



# **Functions of Social Media Platforms**

#### From users' perspective

- Communications: sharing, interacting, keeping up-to-date.. with friends
- Expression: air/project views
- Self-preservation: Wellness, exercise, self-improvements
- Local communities
- From platform providers' perspective
  - Attract and retain users; enrich contents; monetization
  - Offer innovative/fun services; improve user engagement
  - Understand users; co-viewing and co-creation of contents with users
- Lead to 3 key sources of live info streams:
  - Spontaneous User-Generated Contents (UGC)
  - Device-Generated Contents (DGC)
  - Structured Database Contents (SDC)



# **Characteristics of Social Media**



Publication	Expressing Dis Dis Dis Dis Dis Dis Dis Dis	Cussion Constant Constan	Search Niche classmates b00mPg nglockette Elimitational mylife Ecologi	Networking BtoB N Linked C Plano co XING X	Nobile Tools Tools Ning KickApps Tyscom Collective &
Aggregation	PLURK	Social Pla	atforms		
Conter You flickr lost-frr vineo Smoswer iLike Marcon Photo Music	nt magnelia OScrbd i Greddit ^sldcShare	SKYROCK Constants	A Life at windows Live Concold Same Zorpia.com	Social Games <b>ZYNG3</b> Social Sames <b>ZYNG3</b> Social Sames <b>Social Sames</b> <b>Social Sames</b>	Casual Games confector CAFecor CAFecor Casual Cas
Crowdstorm feedocct II Crowdstorm feedocct II Crowdstorm stylehive Stylehive Recommand* Suggestions	brars inter Ilshelfari Share Addresses	Place events socializ: Events trip		MOG MOG MARCON MARCON MARCON MARCON MOG MOG MOG MOG MOG MOG MOG MOG	MMORPG MMORPG
	Sharing			Gaming	

#### Broadcast Media Filter->Publish

#### Social Media Publish->Filter



## NExT Research Center (http://next.comp.nus.edu.sg)

- NExT: NUS-Tsinghua Center on Extreme Search
  - Initiated in May 2010 with a 5-Year Grant of S\$10 million from MDA/NRF of Singapore
  - Among the first to examine big data analytics
- Vision: To access live information that is not easily available on the current Web



## Overview of Research From Unstructured Info into Structured Knowledge



# The Architecture of a Social Observatory System



#### **Back-End Tech Highlights**

- Fully automated
- Cloud-based
  - 80 VMs
- Big data
  - >100 TB, 2.6 billion records
- Live + Batch Processing
  - Storm + Hadoop + Hbase
- Distributed
  - Crawlers, Database, File Storage, Processing
- Cross indexing
  - Elasticsearch index for text
  - Own hash-based image index
  - Graph index

## Platform as a Service: Indian Elections Sub-Events Around an Entity (Modi)

 vesenze	Narendra Modi 🌉 🔸 Week:201411 👻				👤 🝷 chua
Home	SUB TOPICS				
Timeline					
Sub Topics	Rin Seat Give Ticket Sir	Bip.Follow.Tweet.		Ur.1	Fake.Sir.
Entity Graph	Djp,Seat,Give,Ticket,Si	Twitter,Rss		Mo	dus,Frm
Trending Words	sir please also give ticket to BJP candidate from sitamarhi (Bihar) th birth place of maa ianaki Idon't aive this seat 2 ally	e there is not even a single symbol		thanx	sir fr ur areat
Sentiment		of BJP(lotus) in your twitter cover		word f	
Influence Users		follow "Vyakti pooja"		again:	
Hot Zone					
System Management 🔹				-371	
		Bjp,Varanasi,Contest,Poll,S eat	Kejriwal,Arvind, Gujarat,Question ,Development	Media,Aap,C ongress,Say, Kejriwal	Win,Electio Wish,Sir,Gu rat
	Sir,India,Pm,Wish,Hope	BJP's Narendra Modi to contest from MM Joshi's Varanasi seat in Uttar	A Gujarati Replies to the 16 Questions asked by Arvind	Who exposed Narendra Modi? Media -	happy holi to you sir and wi you all the be
	sir,I hope u r the next PM of India sir.	Bjp,Seat,Varanasi,Contest, Candidate	Kejriwal to Narendra Modi	No Congress - No AAP - YES	for upcoming lok sabha elections. Yo
		BJP's prime ministerial candidate Narendra Modi to contest from	India,Pm,Candidat n	e,Party,Electio	Speech,Kno Support,Bjp hackeray
	Vote, Congress, Bjp, Party, Time MY vote for mr narendra modi bjp, aap has become paap, and congress is self made secular, rest partys another face of CONGRESS	Rally,Address,Shri,Delhi,O disha Today WATCH LIVE: Shri Narendra	Rahul,Gandhi,Guja on	ırat,Turf,Educati	Raj Thackeray supports Narendra Modi, but his Sena may hurt BJP

## Platform as a Service: Indian Elections Details of a Sub-Event (Modi)

Li	iveSenze	Narendra Modi 🧕 - Week:201411 -	<b>1</b> - <b>1</b>
	E	A Home /	
-	Home		
Ξ	Timeline		
÷	Sub Topics	(i) govt,gujarat,state,need,india 2707	
£ĝ£	Entity Graph		
 	Trending Words	E Topic Detail	Images
υŴ	Sentiment	NSUI UP West: @ narendramodi alloted 8 lakh sg. m land to L&T @ Re1 per sg. m whose actual price is Rs950 per sg. m.	
2	Influence Users	State loss 800000xRs949. Why?	
9	Hot Zone	Shrishail s k: @ narendramodi farmer dipend n agricultu it dipend n tree, trees r in forest.if forest II burn, farmer II	
°	System Management 🛛 👻		
		Narendra Modi For PM: RT @ narendramodi: SSC & HSC exams begin tomorrow in Guj. Best wishes to my young friends. Am sure your efforts will bring great results	
		Waseem: 376561 sq mtr land to raheja@Rs457 per sq mtr while air force asked to pay 1100 /sq mtr # GujratDictatorshipModel	
		AMITKUMAR.BHIMANI: @ narendramodi Dear sir , kindly look into the matter regarding provide exemption form B.ed for MscIT&CA students	
		AMITKUMAR.BHIMANI: @ narendramodi Dear sir , kindly look into the matter regarding provide exemption form B.ed for MscIT&CA students	
		AMITKUMAR.BHIMANI: @ narendramodi Dear sir , kindly look into the matter regarding provide exemption form B.ed for MscIT&CA students	GUIARATOISTINATIONINOSI
		MBS: @ narendramodi alloted 8 lakh sq. m land to L&T @ ₹1 per sq. m whose actual price is ₹950 per sq. m. State loss 800000x₹949. Why?	
		Cyber Activist: @ narendramodi alloted 8 lakh sq. m land to L&T @ ₹1 per sq. m whose actual price is ₹950 per sq. m. State loss 800000x₹949. Why?	

## Platform as a Service: Indian Elections Relations between Entities (Modi)



## Platform as a Service: Indian Elections From Data to Event to Reports







## Insights Gained from Social Media Analytics

# Events, topics, Users and their relationships



# Insights Gained from Indian Election Data: Example 1

February 14, 2014, the number of mentions of Arvid Kejriwal (AK) reaches the peak (26,152 mentions)



- Cause:
  - AK resigns from Chief Minister's position after suffering defeat in the Assembly on the Jan Lokpal (anti-corruption bill)
- Sentiments:
  - The ratio of positive sentiments to negative sentiments stays the same
  - Positive sentiments: encouraging AK
  - Negative sentiments: expressing disappointment towards AK



# Insights Gained from Indian Election Data: Example 2

April 12, 2014, the number of mentions for Rahul Gandhi (RG) reaches the peak before the release of election result



- Cause:
  - RG second televised interview
- Sentiments:
  - Positive sentiments: sarcasm towards RG
  - Negative sentiments: Criticism towards RG
- Hot tweets (most retweeted tweets):
  - sarcasm and attack towards RG

## Example 2: Indian Election on Twitter The Negative Sentiments





### Example 2: Indian Election on Twitter The Positive Sentiments





## Insights Gained from Indian Election Data: Example 3

February 22, 2014 - April 13, 2014, "China" has been a continuous keyword mentioned by the users who mentioned Narendra Modi (NM)

- Other candidates never publicly spoke against China as NM did
  - Tweets that mentions other candidates barely mentions "China"



# **Example 3: Indian Election on Twitter**



## Key Technologies -1 (In English and Chinese)

1) Intelligence: Reliable Data gathering Strategy

- Employ multi-faceted approach to gather representative data
- Our strategy ensures high relevance, coverage and diversity



Known Account Crawler Key-user Crawler

Hot location Crawler

Image Object Crawler

### 2) Deep NLP

- Key Phrase Extraction
- Named Entity Extraction
- Sentiment Analysis
- Text Classification

The New York Philharmonic Orchestra will make a	
historic trip to North Korea in February, it has announced.	Person
Dominique de Villepin a été nommé Premier ministre	Location
ce <mark>mardi</mark> en fin de matinée par Jacques Chirac.	Organization
The orchestra's president and executive director, Zarin	Data
Mehta said it would play in the capital Pyongyang	Date
on February 26. In August, the reclusive communist	Time
country's Ministry of Culture sent an invitation to	Title
the orchestra at Lincoln Center in Manhattan.	
朝鲜外务省发言人11月1日在平壤宣布,朝鲜将重返六方会谈,	但前提条件是朝鲜与
美国在六方会谈框架内讨论解除美国对朝鲜金融制裁问题。针	对朝鲜方面的动向。

美国在六方会谈框架内讨论解除美国对朝鲜金融制裁问题。针对朝鲜方面的动向, 各方均表示欢迎。<mark>美联社11月1日</mark>报道说: "长期以来一直拒绝与<mark>平壤进行直接对话</mark> 的美国总统布什<mark>认为,各方达成一致、同意恢复六方会谈应归功于中国</mark>的斡旋。

## Key Technologies -2 (In English and Chinese)

#### 3) Noise Removal

Identify and remove spam messages and bot users

#### 4) Live Event Detection & Tracking

- Clustering and filtering
- Event classification (into emerging & evolving events)
- 5) Predictive Analytics:
  - Detect viral messages and viral events before they become viral
  - Detect influential/ active User





## Key Technologies -3 (In English and Chinese)

#### 6) Prescriptive Analytics:

- Convert data into reports/ Actions
- Towards prescribing (current and future) actions to improve user engagements
- From Data -> Analysis -> Predictive -> Prescriptive Analytics
  - Towards analysis of "why" and "how"...
  - Work on characterization of problem, and solution to alleviate the problems



## OUTLINE

- Rich Social Media Contents
- The Live Social Observatory
- Analysis of Large-Scale Social Media Content
- User Demography & Mobility
- Wellness
- Summary



## Large Scale MM Content Analysis (ICMR 2014, ACM MM 2014)

- Information is becoming increasingly multimedia
  - >30% of microblogs have images/videos
  - Over 40% of such microblogs do not have relevant text descriptions
  - Purely text based approach is inadequate
- Key problems: How to identify images relevant to entities/topics without relevant texts
  - O Huge amount of social-tagged mm content available
  - O Earge semantic & intension gaps in mm contents
  - Classifiers to detection certain classes of visual concepts, logos are effective







# **Key Evolutions in Recent Years**

- Large-scale media resources available
- Social Media contents shifted from text to images/videos
- Research on integrating data, users and social network
- Multi-source, multi-task fusion
- Many applications ...

• Era of deep learning, improved content representation and multimodal fusion



## Large-Scale Social Image Annotation Mining Curated and Weibo Images

- Focus on analyzing social images from different types of social networks
  - Classifying curated images from Pinterests Fashion, Food and Travel domains
  - Identify brand images from Weibo Product and Logo Detection



## Deep Fusion for Complex Video Event Classification



**DeepFusion Network** 

**Discovered Semantic Correlations** 

Results on 2 benchmark datasets (MED'11 & CCV) demonstrate its superior performance as compared to state-of-the-arts methods

### Content Representation via Embedding Learning Features from Images, Tags & Users

#### **Deep Semantic Relation Embedding Architecture**



mAP% of the proposed features (last two) as compared to state-of-the art on benchmarks

NUSWIDE     38.6     32.7     41.3     42.9       CCV     62.5     58.4     67.2     67.8	Dataset/Method	DeCAF	ICMAE	Ours-fc7	Ours-4kSc
CCV 62.5 58.4 67.2 67.8	NUSWIDE	38.6	32.7	41.3	42.9
	CCV	62.5	58.4	67.2	67.8

### Content Representation via Embedding Examples: Image-Image Embedding




### Content Representation via Embedding Examples: Word-Image Embedding





### Content Representation via Embedding Examples: Image-Word Embedding



beach, girl, ocean, sand, water, playing, white, bikini, chair



soccer, field, game, players, zijn, groen, running, roelof, januari, ball



bike, girl, street, yellow, car, white, woman, boy, man, dog



wedding, wearing, dress, white, girl, red, black, pink, dressed, flower



### Content Representation via Embedding Examples: Word-Word Embedding

Dog	foster, sleeping, service, cute, cat, puppy, stray, freshfields, spaniel, pug	Girl	dress, cute, hat, pink, boy, portrait, baby, shirt, birthday, blonde
Iorning + Drink	drink, champagne, sky, juice, drinks, coffee, foggy, cold, soda, sun	Evening + Drink	evening, drink, wine, beer, bartender, alcohol, beverage, martini, liquor, bottle

Table 1: Spearman correlation of 971 word-pair similarities computed by different methods and MEN human judgements. Our method learned from SBU is very close to Word2Vec learned from Google News.

Method	Word2Vec	S-CNN	Ours
MEN	0.65	0.36	0.64



### Content Representation via Embedding Examples: More Interesting Embeddings



### Content Representation via Embedding Examples: Features can also cross modality



- 1. Three boys in a forest clearing look at something off-camera.
- 2. Two long-haired dogs racing down a forest path.
- 3. The parachutist has a bird flying along with him.
- 4. Three people are walking on a green area down a path between a forest of trees.
- 5. Two children looking at horses through a small fence.



- 1. A baby boy in a car seat is asleep and crying.
- 2. A young girl sits on a seat and looks at a train pamphlet.
- 3. A young boy is buckled into a car seat making a silly face.
- 4. There is a little boy in red shorts asleep in his car seat holding a cookie.
- 5. A young girl drawing with a yellow crayon wearing her seat belt.



- 1. The parachutist has a bird flying along with him
- 2. A young boy goes through large rocks in the mountains.
- 3. The brown mountains have snow on them and there are clouds in the sky.
- 4. Skiiers are walking through the snow near the mountains.
- 5. A man is skiing down a mountain with another large mountain behind him and a river between the mountains.



### Content Representation via Embedding Media User Embedding (Cui P. 2015)



### Framework for Representational Learning



### Multimodal Location Estimation with ambient audio analysis

### POI Estimation

- GPS trace of users, and messages, if any
- Utilize audio signals to estimate venue categories
- Essential for micro-videos in social media sites

### Smart Ambience Sound Analytics (SASA) Framework



## OUTLINE

- Rich Social Media Contents
- The Live Social Observatory
- Analysis of Large-Scale Social Media Content
- User Demography & Mobility
- Wellness
- Summary



# What is Mobility?

- Mobility contemporary paradigm, which explores various types of people movement.
  - The quality or state of being mobile
  - (Sociology) movement within or between social classes and occupations
- Why mobility?
  - Urban planning: understand the city and optimize services
  - Mobile applications and recommendations: study the user and offer services



## If welocitity can describe people



# **Multi-Source Social Network Data**

#### More people use multiple sites

% of internet users who use the following number of social networking sites (sites measured include: Facebook, Twitter, Instagram, Pinterest and LinkedIn), 2013 vs. 2014



More than 50% of online-active adults use more than one social network in their daily life

Pew Research Center's Internet Project September Combined Omnibus Survey, September 11-14 & September 18-21, 2014. N=1,597.

#### PEW RESEARCH CENTER

### Multi-Source Multi-Task Learning User profiling: Mobility + Demography

- Utilize multi-source social media data, & if available, mobile phone traces
- Many applications





# Multiple Sources Describing User from Multiple Views



### **Text Representation**

- Linguistic features
  - LIWC
  - User Topics
- Heuristic features
  - Writing behavior

An efficient and effective method for studying the various emotional, cognitive, structural, and process components present in individuals' verbal and written speech samples. Can be highly related to one's demography.

### A text analysis software.



#### Dictionary

#### Word category



### **Text Representation**



- Linguistic features
  - LIWC
  - User Topics
- Heuristic features
  - Writing behavior

Users of similar gender and age may talk about similar topics e.g. female users – about shopping, male – about cars; youth – about school while elderly – about health.



## **Text Representation**



- Linguistic features
  - LIWC
  - User Topics
- Heuristic features
  Writing behavior

As we mention from our research – user's writing behavioral patterns are highly correlated with e.g. age (individuals from 10 – 20 years old are making two times less grammatical errors than 20 -30 years old individuals)

Feature name	Description		
Number of hash tags	Number of hash tags mentioned in message		
Number of slang words	Number of slang words one use in his tweets. We calculate number of slang words / tweet and compute average slang usage		
Number of URLs	Number of URL's one usually use in his/her tweets		
Number of user mentions	Number of user mentions – may represent one's social activity		
Number of repeated chars	Number of repeated characters in one tweets (e.g. noooooooo, wahhhhhh)		
Number of emotion words	Number of words that are marked with not – neutral emotion score in Sentiment WordNet		
Number of emoticons	Number of common emoticons from Wikipedia article		
Average sentiment level	Module of average sentiment level of tweet obtained from Sentiment WordNet		
Average sentiment score	Average sentiment level of tweet obtained from Sentiment WordNet		
Number of misspellings	Number of misspellings fixed by Microsoft Word spell checker		
Number Of Mistakes	Number of words that contains mistake but cannot be fixed by Microsoft Word spell checker		
Number of rejected tweets	Number of tweets where 70% of words either not in English or cannot be fixed by Microsoft Word spell specker		

## **Location Representation**



- Location features
  - Location semantics
  - Location topics

Venue semantics such as venue categories can be related to users demography. E.g. individuals who tent to visit night clubs are usually belong to 10 – 20 or 20 – 30 years old age groups.



We map all 4Square check-ins to 4Square categories from category hierarchy.

For case when user performed check-ins in two restaurants and airport but did not perform check-ins in other venues:

	<i>C</i> <sub>1</sub>		<b>Cat</b> <sub>restaurant</sub>		Cat <sub>airport</sub>	•••	Cat <sub>n</sub>
U <sub>1</sub>	0	0	2	0	1	0	0
•••	*	*	*	*	*	*	*
<b>U</b> <sub>n</sub>	*	*	*	*	*	*	*

### **Image Representation**



- Image features
  - Image concept learning

Extracted image concepts may represents user interests and be related to one's demography. For example female user may take pictures of flowers, food, while male – of cars or buildings.





\* The concept learning Tool was provided by Lab of Media Search LMS. It was evaluated based on 5 LSVRC2012 competition dataset and performed with average accuracy @10 - 0.637

# **Multi-Source Multimodal Analysis**

### TEXT Features:



- Linguistic features: LIWC; User Topics
- Heuristic features: Writing behavior

### LOCATON Features:

- Location semantics
- Location topics



#### A text analysis software.





Dictionary

Word category

#### IMAGE Features:

Image concepts







- EXERCISE Sensor Features:
  - Exercise vectors, categories

### **Ensemble Learning**



## **NUS-MSS:** Dataset Description

	Singapore	London	New York
y	11,732 K	2,973K	5,264K
	366K	127К	305K
	263K	65K	231K
t i i	7,023	5,503	7,957

Accuracy (F1)		Age Groups	
Gender	Age	<20	
0.878	0.509	20-30 30-40	
		>40	NEX

# Application: Individual Mobility Analysis



Polygons – personal mobility sectors



# Application: Mobility Analysis for Business



Polygons – businesses' trade areas



# A Generalized Multi-Source Multi-Task Learning Model



• Graph-guided fused lasso

Lasso, it controls the sparsity.Learn task-specific features.

- Learn task-sharing features
- Capture relatedness among tasks

#### Many Applications:

- Many aspects of user profiling
- User volunteerism tendency
- Health Modeling



## OUTLINE

- Rich Social Media Contents
- The Live Social Observatory
- Analysis of Large-Scale Social Media Content
- User Demography & Mobility
- Wellness
- Summary



# **Problems with Chronic Disease**





# Key Therapeutic Methods





### What We Know from Social Media: Lifestyle and Health in Social Media -1



## What We Know from Social Media: Lifestyle and Health in Social Media -2



### From Social Media to Wellness Convert Social Media Posts to Activity Timeline



# **Example of the Constructed Timeline**



Figure 1: An example personal health event timeline constructed from user published tweets. Note that each tweet can include information corresponding to different events with different details. Event from different categories is represented by different colors.



# **Key Categories in Activity Timeline**

Table 1: Illustration of hierarchial taxonomies and health events and their exemplar tweets.

Event	Sub Event	Example	
	Meals	Dinner just salad	
Diet	Alcoholic Beverages	Too much drink in party	
Dict	Non alashalia Payaragas	Talking about hot chocolates, I might	
	Non-alcoholie Develages	just go and make myself one :D	
	Snacks	found Taylor's pretzels in my backpack	
	Shacks	and I'm so happy wow	
	Fruit	almost eat all the strawberries	
	Others	Eat 20g carbs and go fo running	
	Walking	20 mins walk around office	
	Running	after 1 hour run #bgnow 130	
Exercise	Biking	I just finished 1 hour biking	
	Swimming	BGnow 95, thanks swimming pool	
	Others	Shopping and having a little dinner URL	
	Tests and Examinations	#BGnow 100	
Health & wellness	Symptoms and Complains	Feel too much Fatigue	
	Treatment and Medicine	Insulin injection	
Not a Health Event Other events		Proud father moment, My daughter was accepted to #Harvard	

## Event Detection Results & Current Work

• We have 3 main categories and 14 events. The average result for each categories are:

Event	Precision	Recall	F1-measure
Food & Nutrition	89.57	82.69	85.99
Exercise & Activities	92.23	84.30	88.07
Health & Wellness	74.65	80.71	77.56
Average	87.15	82.57	84.79



### From Social Media to Wellness Multi-Source Data for Wellness

- Social media data is just one aspect (for only socially active users)
- What about other sources of data?
- Mobile health or exercise data (personalized)
  Environmental data
  Wellness (medical) knowledge



KB Source

### From Social Media to Wellness Wellness Advisory System for End-Users

### **ClealthSenze** Wellness Assistant for graceful lifestyles



Advisory System: based on 6 critical illnesses



- For end users advisory system
- For doctors a D2D system

#### Key Research

- Wellness behavioral profiles for groups
- Prediction of critical diseases
- Prescriptive helps for better life style
# OUTLINE

- Rich Social Media Contents
- The Live Social Observatory
- Analysis of Large-Scale Social Media Content
- User Demography & Mobility
- Wellness
- Summary



# **Research on Privacy Preserving**

#### Privacy issues

- Online: user identity (location, query etc)
- 1<sup>st</sup> and 2<sup>nd</sup> order privacy
- Privacy is personal
- Help identify privacy sensitive sentences and documents
- Differential privacy for datasets
- Anti-Privacy Systems
  - We work on building user profiles with privacy sensitive info
  - Prescriptive measures to preserve privacy
  - Help users be aware of privacy issues



### Summary

- We have built a live social observatory system:
  - Analyze large scale live UGCs, DGCs and SDCs
  - Monitor events and happenings in city to help users lead better life
  - Derive demography, mobility & wellness knowledge for "some users"
  - Just social media data is insufficient, need to also integrate (structured) data from many sources
- Build academic collaborative social observatory system:
  Support academic and collaborative research
- Meet Challenges for next generation social media networks
  - More video
  - More noise/spam
  - More reactive and prescriptive



# **Current Work**

- Collaborative Social Observatory systems
- User Profiling and Privacy research
- Research on social (micro-) videos
- Research on Predictive and Prescriptive Analytics
- Wellness system for research and deployment at regional or country level
  - to support rural healthcare and general wellness of people
- Other applications: Smart city



## THANKS

#### Visit our Web Observatory: http:///WWW.NEXTCENTER.ORG/

