

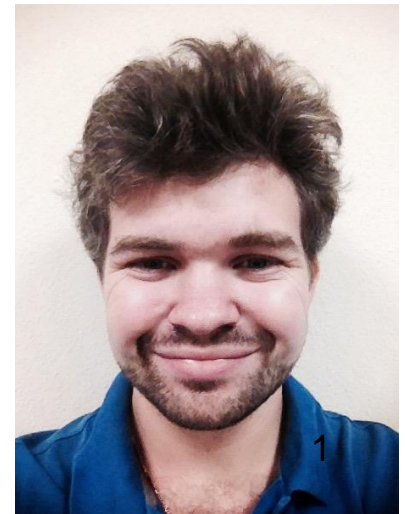
Social Media Computing

Lecture 1: Introduction

Lecturer: Aleksandr Farseev

E-mail: farseev@u.nus.edu

Slides: <http://farseev.com/ainlfruct.html>



Outline

- **Age of Social Media**
- **Analysis of Social Media**
- **Challenges in Social Media**

The New Information Age

- The Internet has revolutionized the way information is created, disseminated and consumed
 - Mixture of info available has **changed from purely text, to include mm data, and live media**
 - Emergence of **huge amount of end-user generated data**, especially in social networks

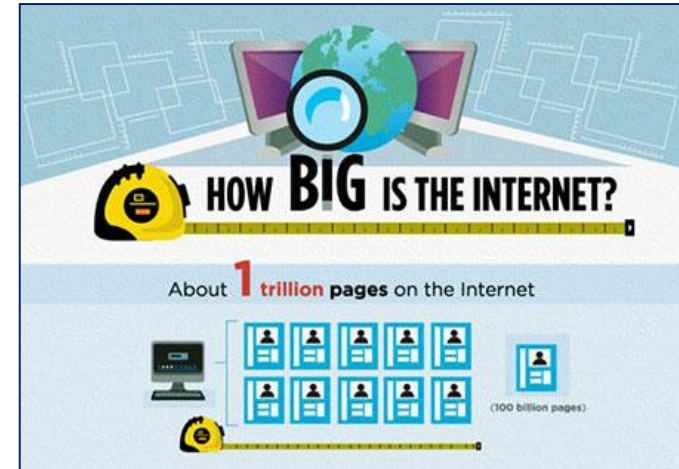


- Greater connectivity leads to **huge amount of live info**
- **Internet has also rapidly gone mobile**, permitting access from anywhere



How Big is Internet?

- The size is about **1 Trillion pages**, but
~ **4.83 billion pages are indexed**
(<http://worldwidewebsize.com>, Jan 2014)
- Studies claimed that the **deep web**
(the dynamic pages) is ~500 times
larger than the indexed
 - This does not include the huge amount of
(real-time) forum and social postings
- Like deep Web, the amount of info available in live and
social Web is huge



Traditional Media

Communication Media:

One-One



Broadcast Media:

One-Many

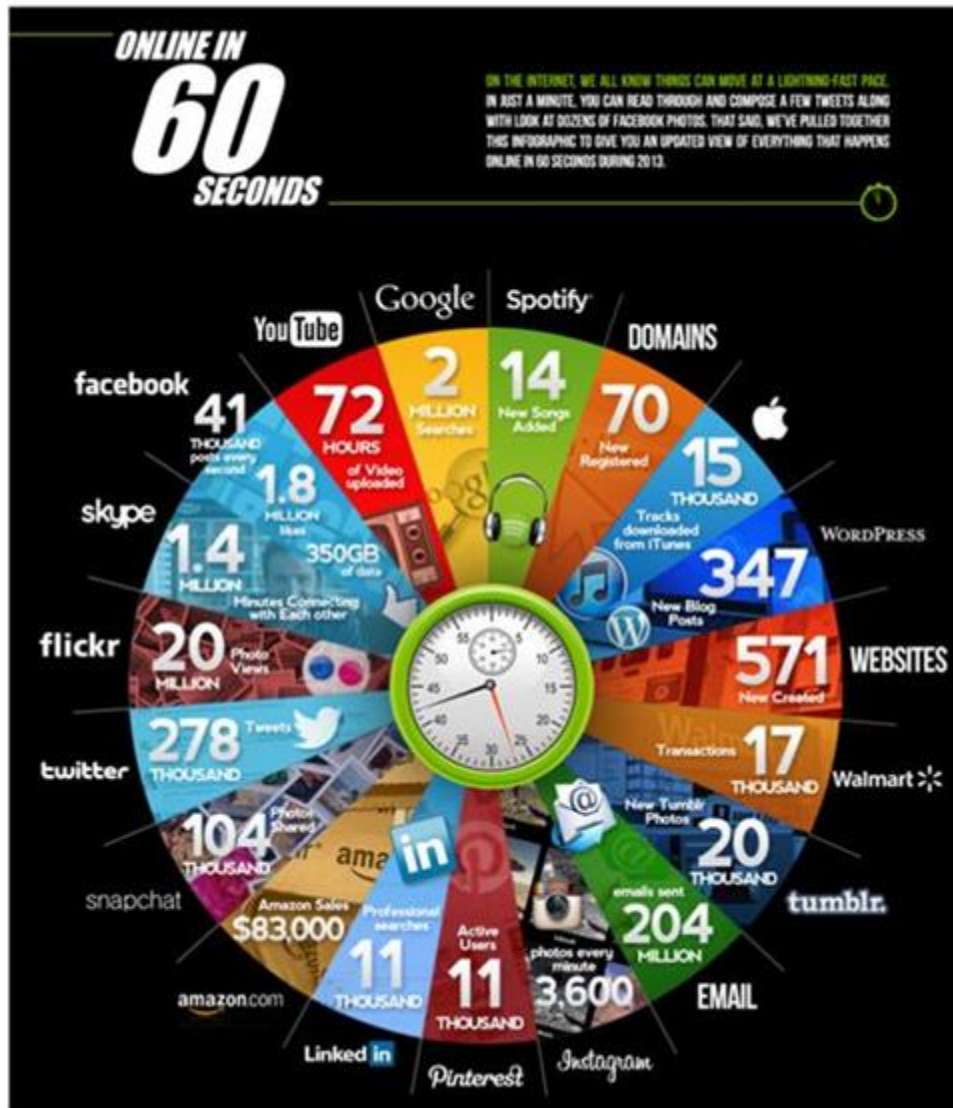


Social Media



Many-to-Many

Rich UGCs in Web 2.0



As of August 2014

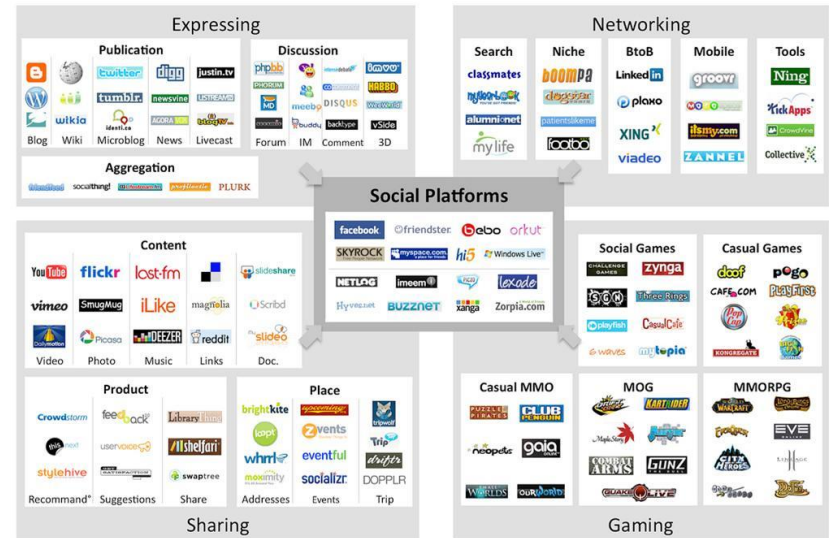
Characteristics of Social Media

- Everyone can be a media outlet
- Rich User Interactions
- User-Generated Contents
- User Enriched Contents
- Collaborative environment
- Collective Wisdom

Characteristics of Social Media



Broadcast Media
Filter->Publish



Social Media
Publish->Filter

- How would these characteristics affect the ways we look into social media

Outline

- **Age of Social Media**
- **Analysis of Social Media**
- **Challenges in Social Media**
- **Course Arrangements**

Social Media Platforms

- Social **Network platforms**

- Three major platforms: Private, Professional, Public



★ The recent apps are all image/video based..

- Media **Sharing platforms**

- YouTube, Instagram, Flickr, ... Vine



- Social **Messaging platforms**

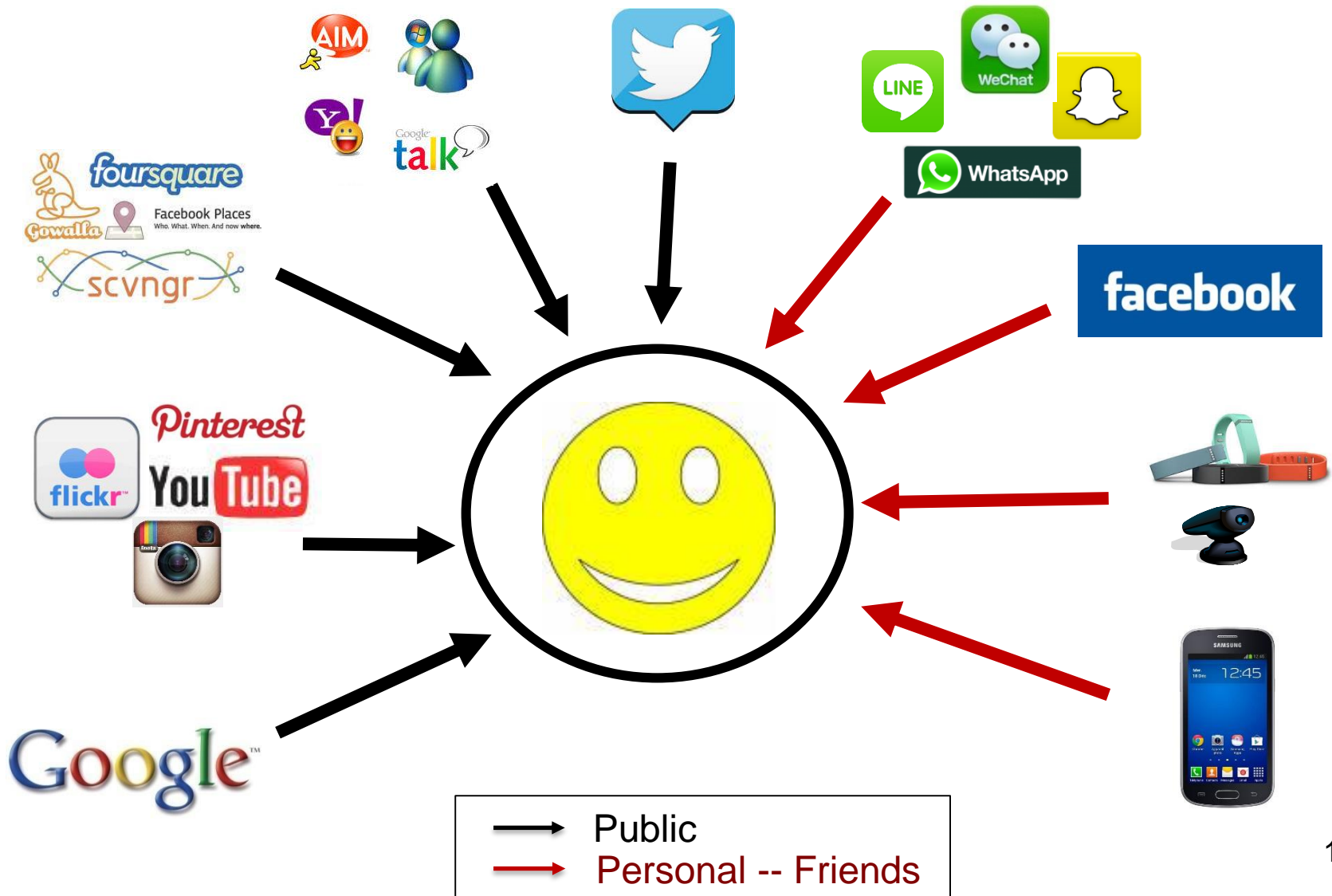
- Whatsapp, LINE, Wechat, SnapChat, ...



- Social **Curation platforms:** Pinterest



Users at Center of Social Media Environment



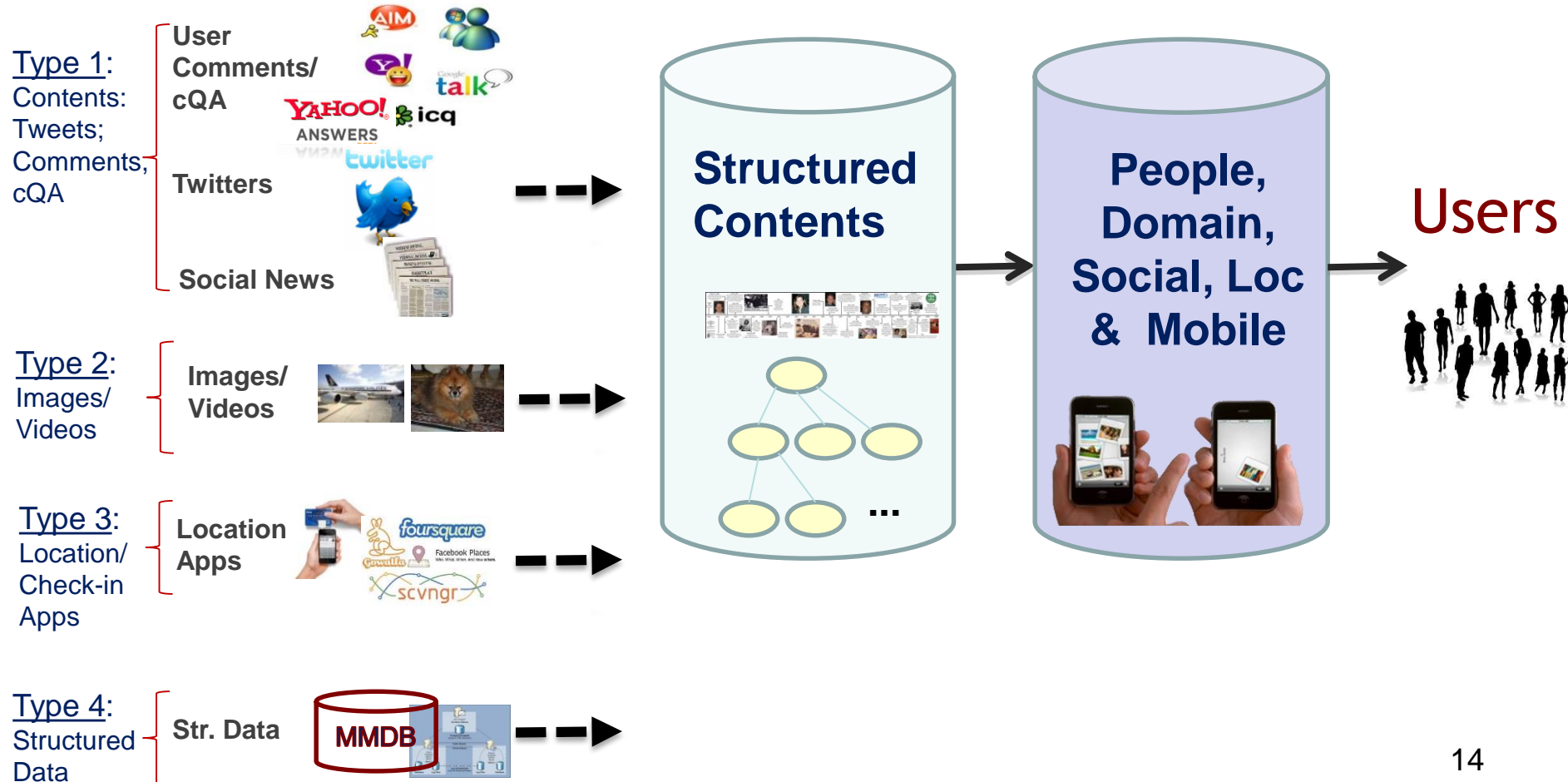
Functions of Social Media Platforms

- 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

NExT Social Observatory at NUS

- Gathering of multi-source data for integrated analysis

Types of UGC's Gathered



Platform as a Service: Indian Elections

Sub-Events Around an Entity (Modi)

<http://live.nextcenter.org>

The screenshot displays the LiveSenze web application interface. The top navigation bar includes the 'LiveSenze' logo, a dropdown menu for 'Narendra Modi' (with a profile picture), and a 'Week:201411' filter. A user profile icon and the name 'chuats' are visible on the right. A left sidebar contains navigation links: Home, Timeline, Sub Topics (selected), Entity Graph, Trending Words, Sentiment, Influence Users, Hot Zone, and System Management. The main content area is titled 'SUB TOPICS' and features a grid of topic cards. Each card has a title, a list of associated terms, and a brief description or sample text. A central grid of 25 small images shows various people, likely related to the election topics.

LiveSenze | Narendra Modi | Week:201411 | chuats

Home / SUB TOPICS

Home
Timeline
Sub Topics
Entity Graph
Trending Words
Sentiment
Influence Users
Hot Zone
System Management

Bjp,Seat,Give,Ticket,Sir
sir please also give ticket to BJP candidate from sitamarhi (Bihar) the birth place of maa janaki I don't give this seat 2 ally

Bjp,Follow,Tweet,Twitter,Rss
there is not even a single symbol of BJP (lotus) in your twitter cover . being from RSS we should not follow "Vyakti pooja"

Ur,Take,Sir,Modus,Frm
thanx sir fr ur great word for the solders plz take the action against naxals

Sir,India,Pm,Wish,Hope
sir,I hope u r the next PM of India sir.

Bjp,Varanasi,Contest,Poll,Seat
BJP's Narendra Modi to contest from MM Joshi's Varanasi seat in Uttar...

Kejriwal,Arvind,Gujarat,Question,Development
A Gujarati Replies to the 16 Questions asked by Arvind Kejriwal to Narendra Modi ...

Media,Aap,Congress,Say,Kejriwal
Who exposed Narendra Modi? Media - No Congress - No AAP - YES

Win,Election,Wish,Sir,Gujarat
happy holi to you sir and wish you all the best for upcoming lok sabha elections. You...

Bjp,Seat,Varanasi,Contest,Candidate
BJP's prime ministerial candidate Narendra Modi to contest from...

India,Pm,Candidate,Party,Election

Speech,Know,Support,Bjp,Thackeray
Raj Thackeray supports Narendra Modi, but his Sena may hurt BJP

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,Odisha
Today WATCH LIVE: Shri Narendra...

Rahul,Gandhi,Gujarat,Turf,Education

Platform as a Service: Indian Elections

Details of a Sub-Event (Modi)

<http://live.nextcenter.org>

The screenshot displays the LiveSenze web application interface. The top navigation bar includes the 'LiveSenze' logo, a user profile for 'Narendra Modi', and a 'Week:201411' dropdown. A left sidebar contains navigation links: Home, Timeline, Sub Topics, Entity Graph, Trending Words, Sentiment, Influence Users, Hot Zone, and System Management. The main content area is titled 'TOPICS DETAIL' and shows the topic 'govt.gujarat.state.need.india' with a count of 2707. Below this, a 'Topic Detail' section lists several tweets related to land allocation and political issues. On the right, an 'Images' section displays a grid of images, including a large image of Narendra Modi speaking at a podium, and a bar chart titled 'GUJARAT DESTINATION NO.1'.

LiveSenze Narendra Modi Week:201411 chuats

Home /

TOPICS DETAIL

govt.gujarat.state.need.india 2707

Topic Detail

- NSUI UP West:** @ narendramodi allotted 8 lakh sq. m land to L&T @ Re1 per sq. m whose actual price is Rs950 per sq. m. State loss 800000xRs949. Why?
- Shrishail s k:** @ narendramodi farmer depend n agricultu it depend n tree,trees r in forest.if forest ll burn, farmer ll burn.Then india?
- 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
- MBS:** @ narendramodi allotted 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 allotted 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?

Images

BREAKING NEWS
2014 POLITICS
MODI TAKES AIM AT KHURSHID
LIVE SAMBALPUR

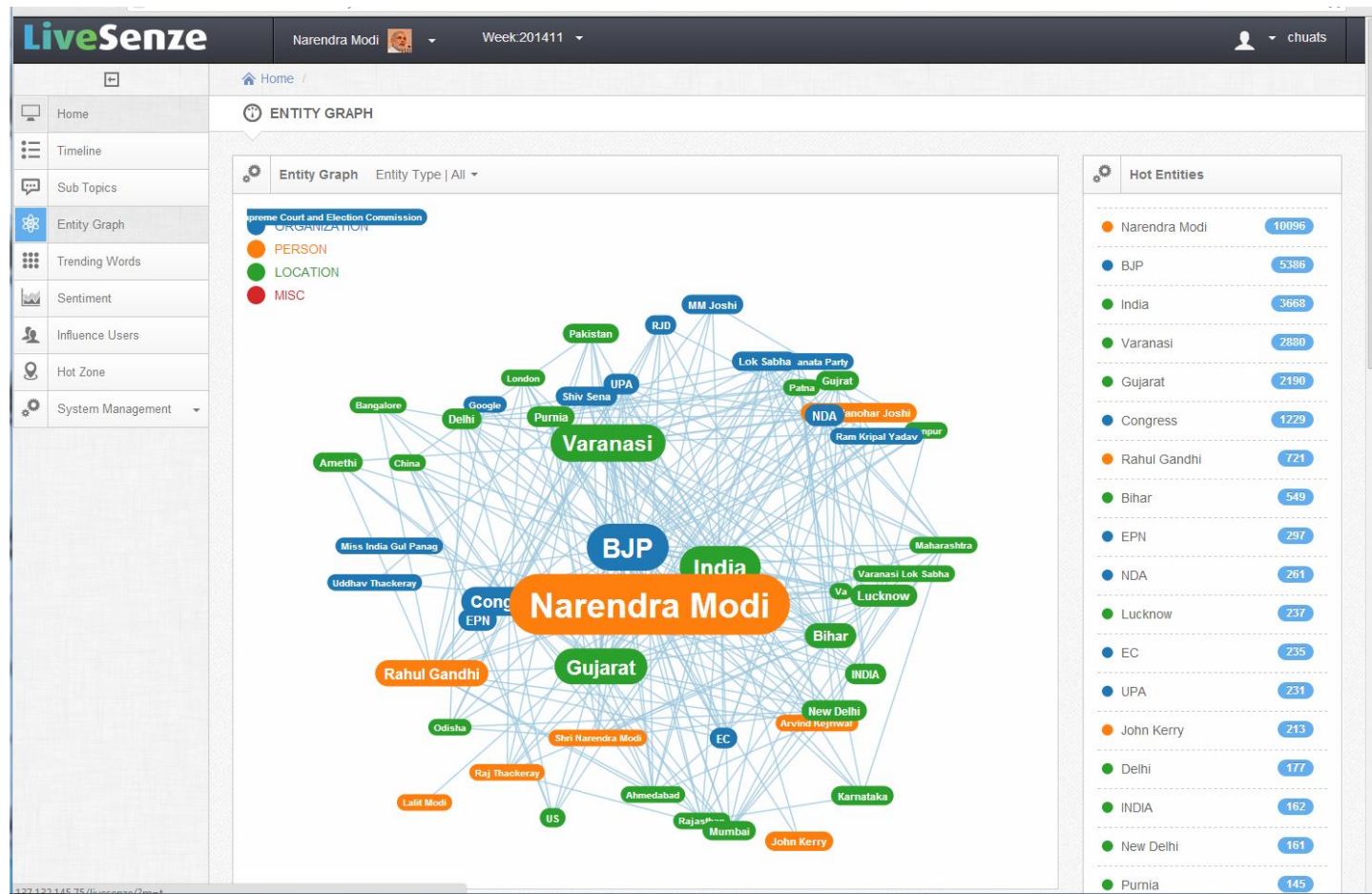
GUJARAT DESTINATION NO.1

16

Platform as a Service: Indian Elections

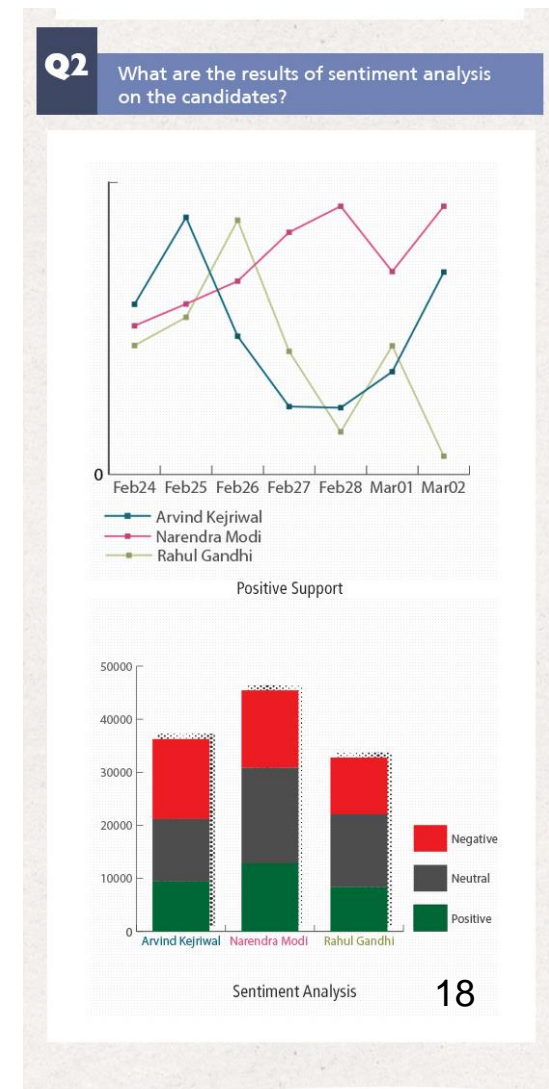
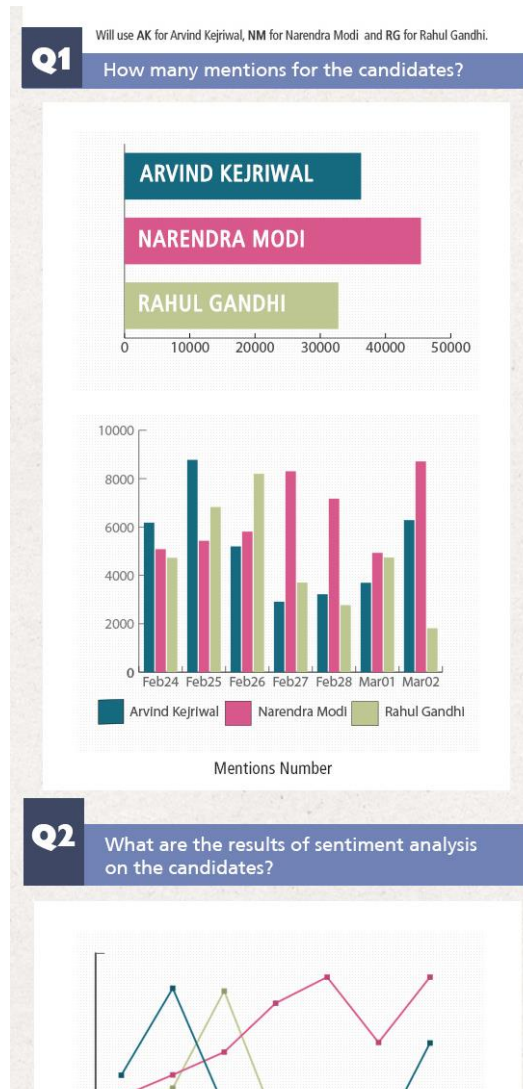
Relations between Entities (Modi)

<http://live.nextcenter.org>



Platform as a Service: Indian Elections

From Data to Event to Reports



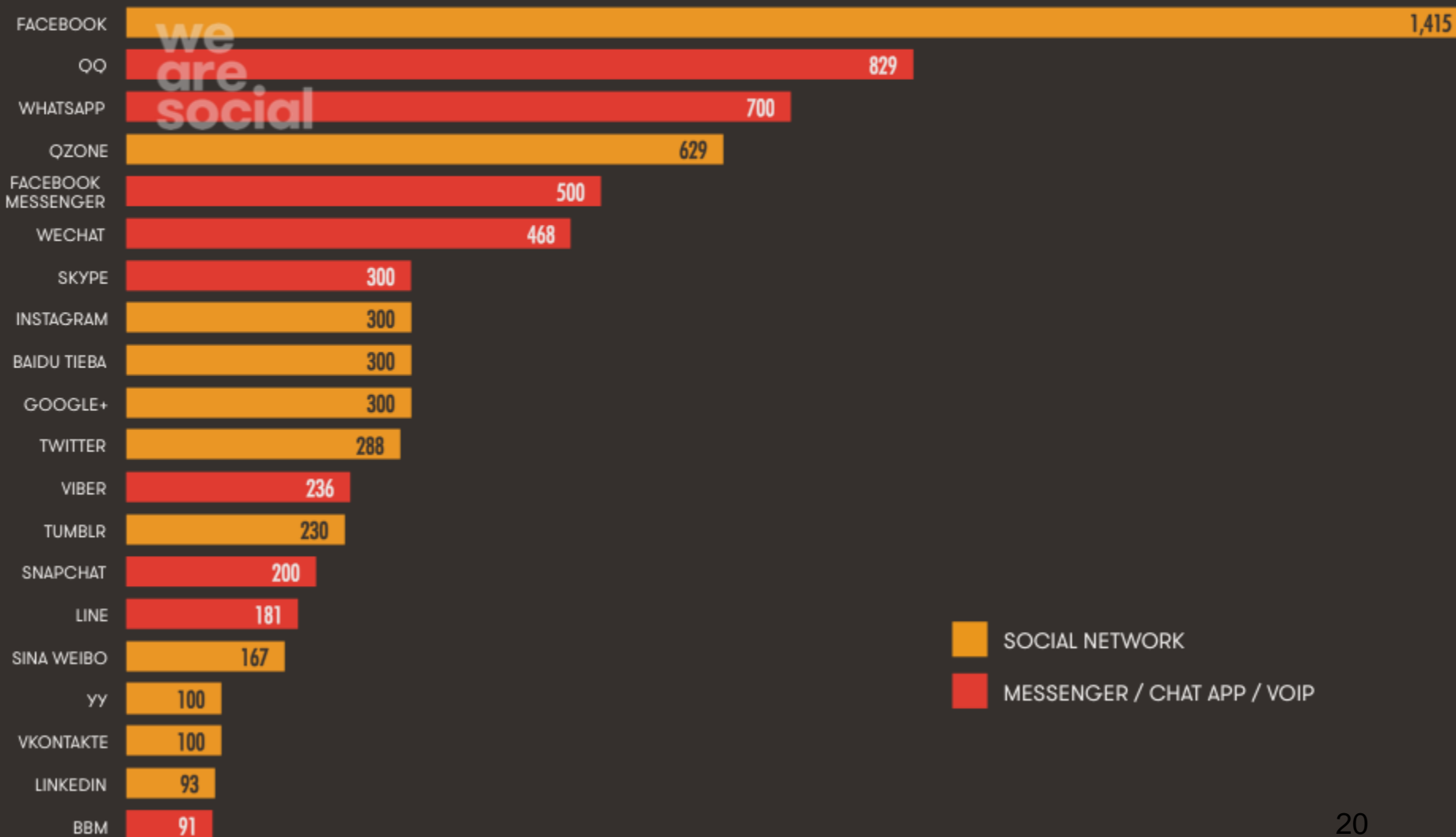
Which Social Platforms are the most influential?

MAR
2015

ACTIVE USERS BY SOCIAL PLATFORM

MOST RECENTLY PUBLISHED MONTHLY ACTIVE USER ACCOUNTS BY PLATFORM, IN MILLIONS

09 MAR 2015

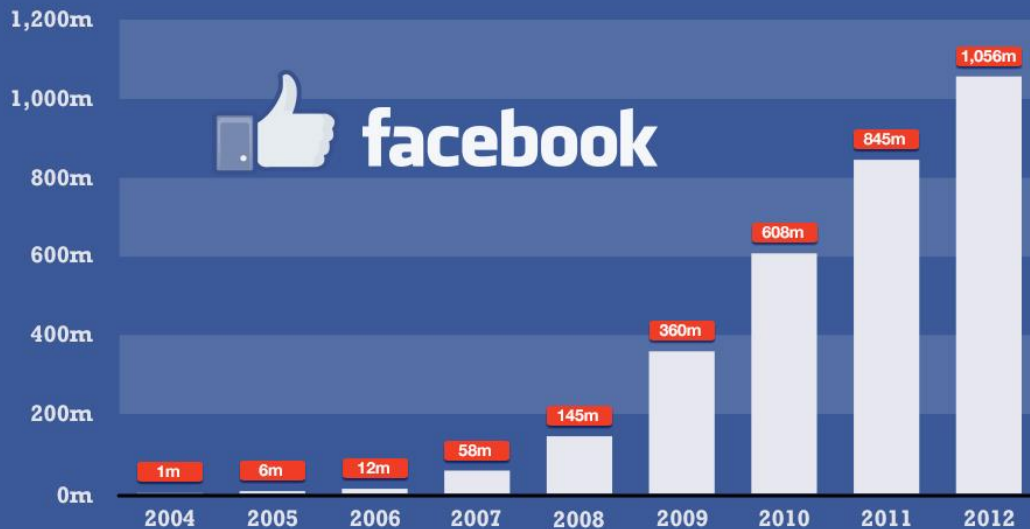


FaceBook

- Facebook is by **far the largest social network** site in the world, overtaking Google in 2007 in terms of user traffic
- It is also the **largest photo sharing site**, with more than 250 million uploads each month (2012)

Facebook's Remarkable User Growth

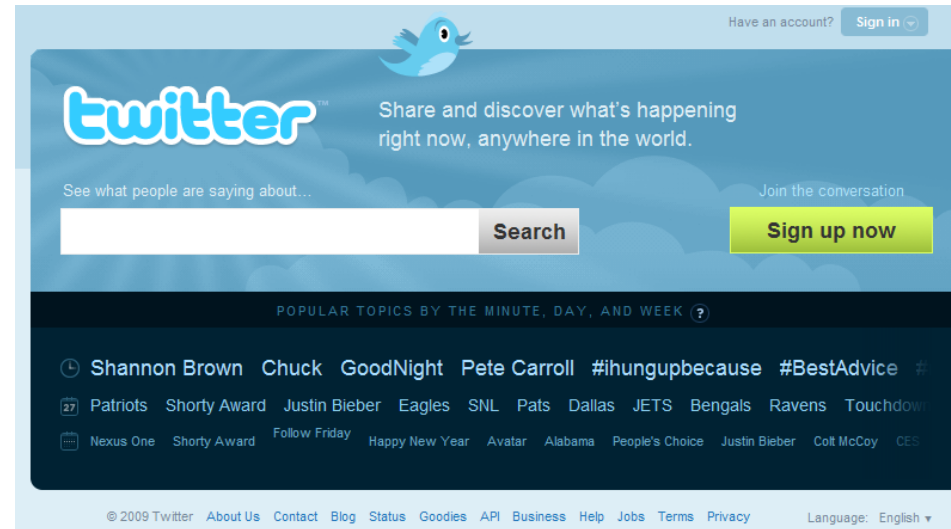
Monthly active users from 2004 to 2012 (at year-end)



	Country	Users ▼
	World	971,426,940
1.	United States	164,958,520
2.	Brazil	65,657,820
3.	India	61,697,760
4.	Indonesia	48,807,580
5.	Mexico	39,810,220
6.	United Kingdom	32,597,460
7.	Turkey	32,260,920
8.	Philippines	30,214,140
9.	France	25,502,360
10.	Germany	25,284,240

Twitter

- Twitter, a Microblogging site, is **the second most popular with over 650 million users**
- **First popular English-speaking microblog**
- Twitter **messages are mostly public**, whereas FB messages are mostly private, they serve different purposes
 - **Twitter** serving **public forum**
 - **FB** targets **private** groups



Characteristics of Twitter

- Its messages are restricted to **140 characters in length**:
 - Encourages users to post **short and frequent messages**
 - However, **messages are informal** and contains abbreviations
Eg: u must be talkin bout the paper but I was thinkin movies
 - **Hard to analyze by formal NLP tools**
- It supports **social functions**, like follow, re-tweet & reply:
 - **“Follow” permits someone to follow anyone**; it is very popular and copied by most social network services
 - **Re-tweets** are widely used too
- It is conversational in nature:
 - **Messages tend not to repeat terms used earlier**, and
 - **Vocabulary used change dynamically**

User's Tweeting Habits Over Time

- A concrete example



@Crushedsj 2013-10-17 13:15:24

detail

not sure yet, i m trynna find a good medical school so probably nus



@Crushedsj 2013-10-17 18:21:41

detail

@AsiaPacNews: NUS professors to ride from Singapore to Sweden for breast cancer research
<http://t.co/cNGTwedSSq> #indonesia



@Crushedsj 2013-10-17 20:41:48

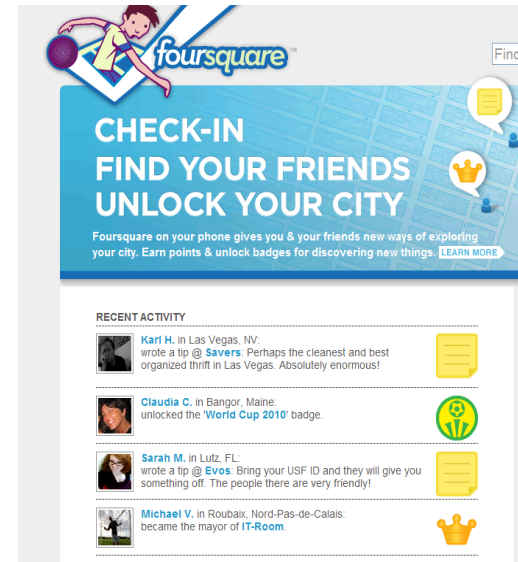
detail

My future school !! I swear I will take all the effort to get in here !!!

Location-Based Social Networks



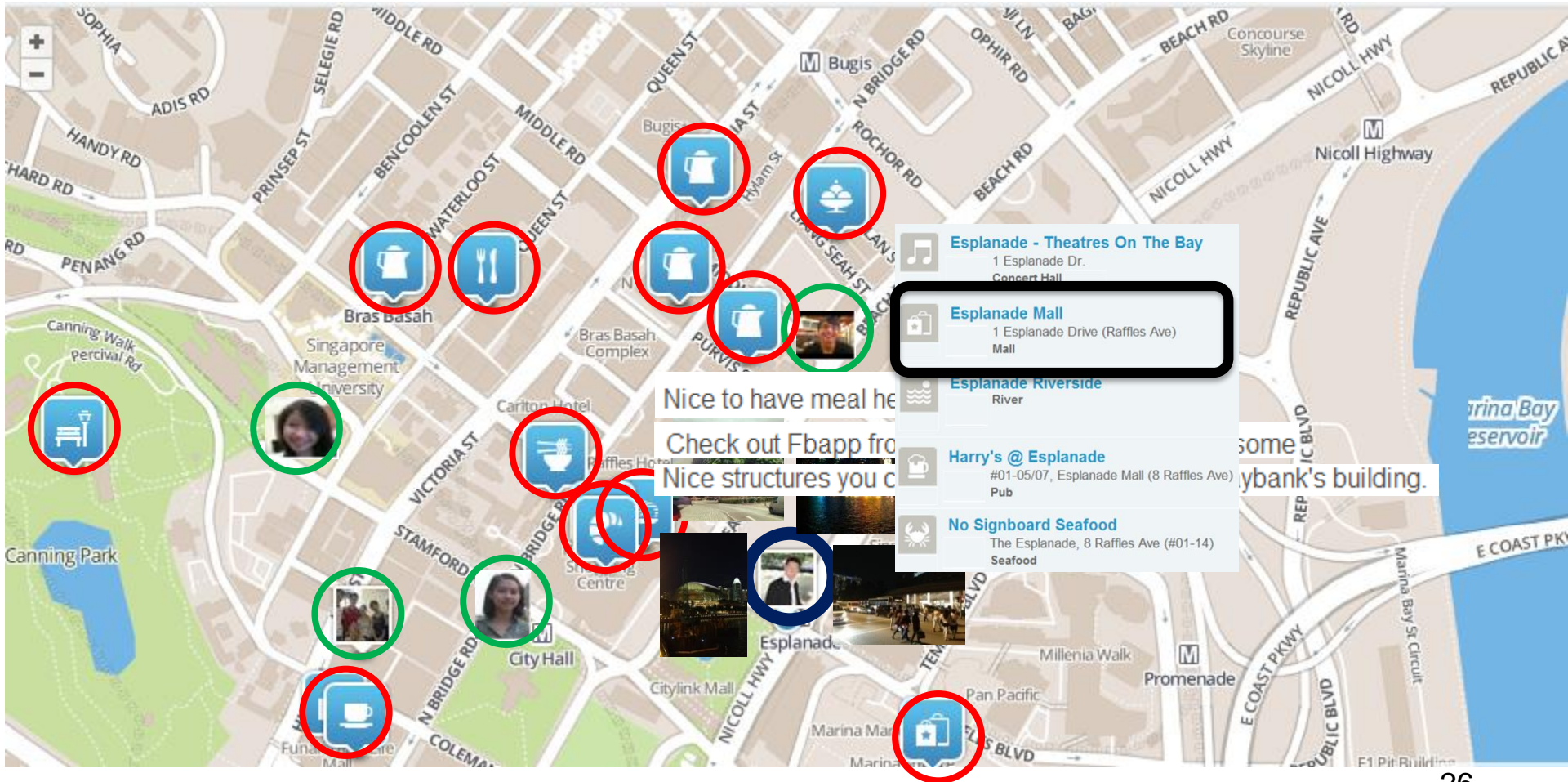
- Encourages users to "check-in" at various local venues, and enter comments and/or upload photos
- Allows many local places to be tagged
- Search support: find list of nearby places; whereabouts of friends, etc



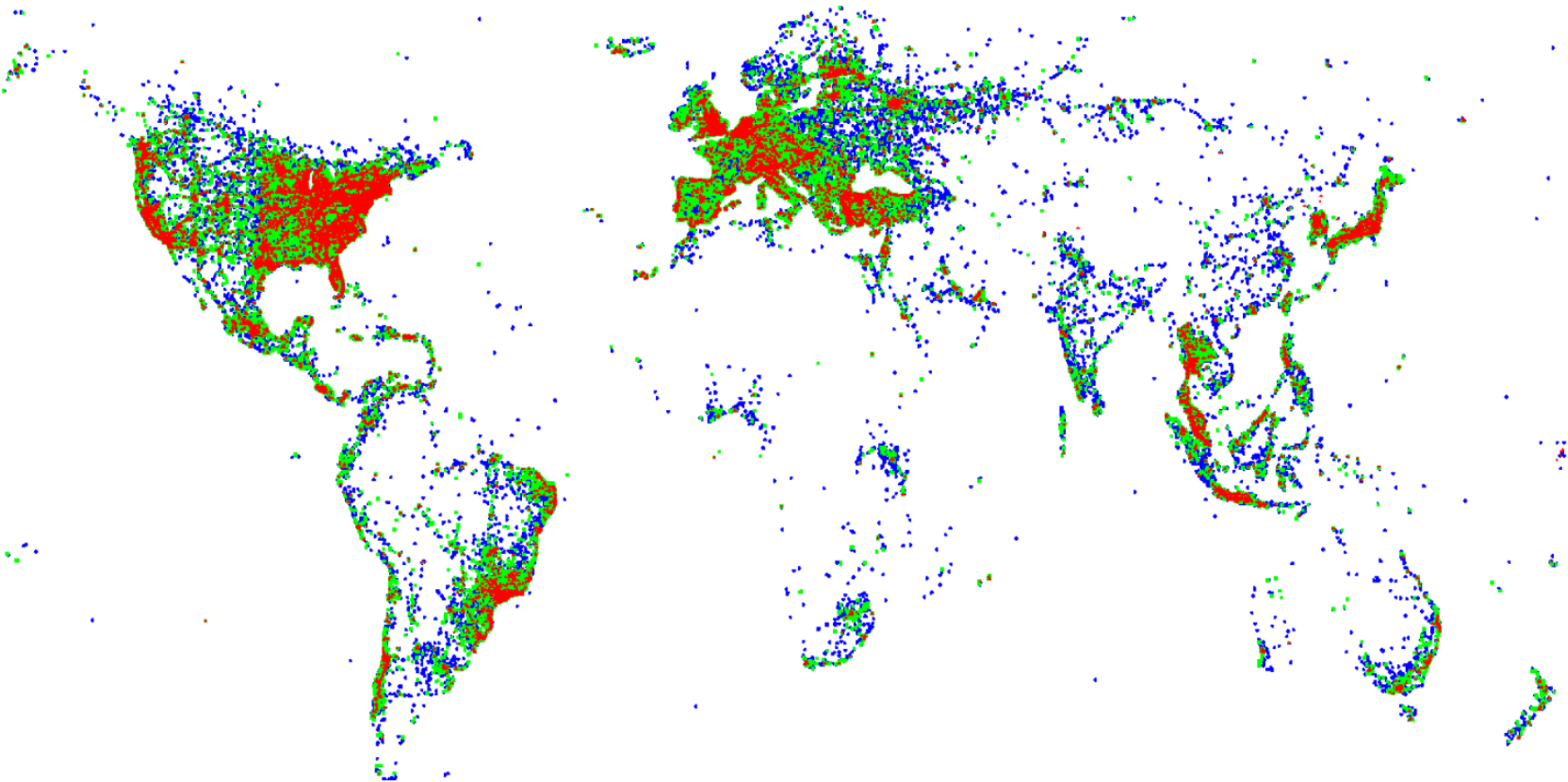
- Many other similar LBSNs, like Gowalla, Yelp etc..



Typical Activities on 4Square



Distributions of 4Square Venues



33 million users, 3.5 billion check-ins , as at April 2013

40+ million users, 4.5 billion check-ins, as at September 2013

A Venue Page in 4Square



Gardens By The Bay

Park and Garden in Singapore

[Report a problem](#)

Address 18 Marina Gardens Dr.
Singapore, Singapore 018953 →

Phone 6420 6848

Website gardensbythebay.org.sg

Hours Likely open (See when people check in)

Gardens by the Bay brings to life NParks' vision of creating a City in a Garden. The Gardens captures the essence of Singapore as the premier tropical Garden City with the perfect environment in which to live and work - making Singapore a leading global city of the 21st century.

9.3 /10

Lots of people like this place

Here Now
3

Total Visitors
13,960

Total Check-ins
19,341



SAVE

<http://4sq.com/9faDil>

SHARE

People talk about:

"... flower dome. nice [super tree](#) walk and there will b bistro right on d..." (3 tips)

"... Forest only, the other [Flower Dome](#) is not worth it" (2 tips)

"You have to visit the [Cloud Forest](#)" in this place. It's unbelievable how..." (2 tips)



Don't come here in the afternoon/evening it's really very hot - ___-

Cheryl S. · July 7, 2012

Save Like - 11 likes



Must see in Singapore. The super tree was so awesome but not as awesome in the photos uploaded by the Tourism of Singapore

William Khoo W. · August 25, 2012

Save Like - 10 likes

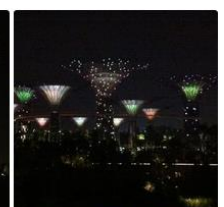
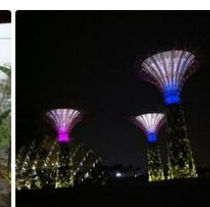


One can best appreciate the garden at night for its flowers are not yet fully grown. It's like a normal spot at day. Make sure to wear your most comfortable shoes because the place is huge.

Candy V. · July 17, 2012

Save Like - 7 likes

Tips



28

Mobile Photo Sharing Sites- Instagram



Instagram



- Fun app: give users capability to transform picture to more professional looking ones
- Bought by Facebook for US\$1 Billion

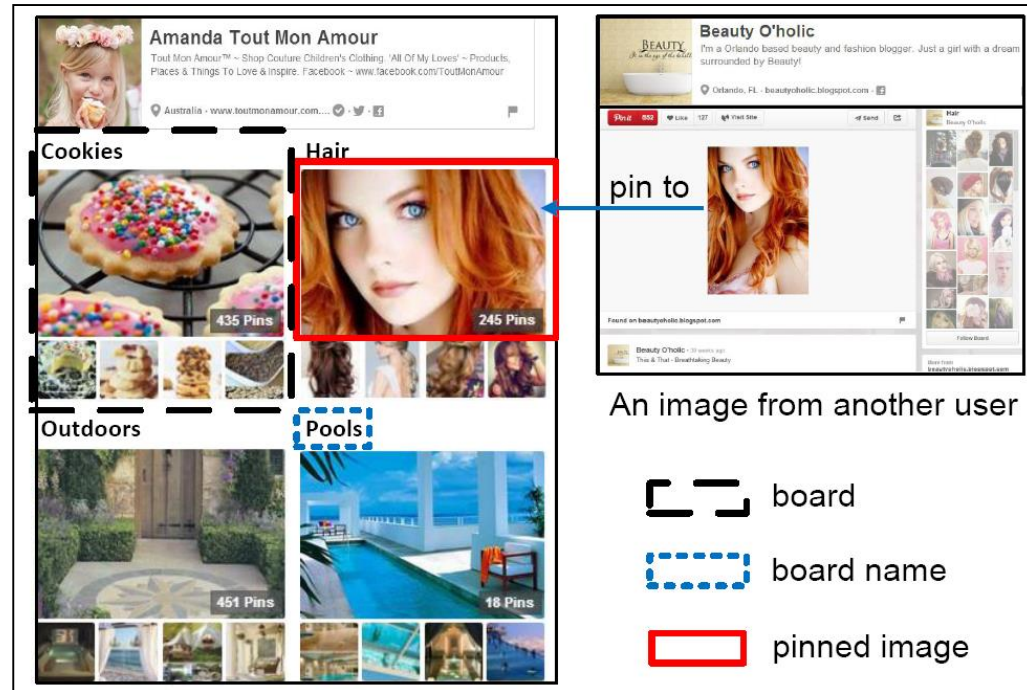


Visual-based Social Media Platforms

- Many emerging popular applications are visual based

- **Pinterest**: founded in 2010; popular with women and is very targeted (~70M users)

- **Vine**: founded in 2012; sharing of 6-sec loop-videos; popular with young people (~40M users)

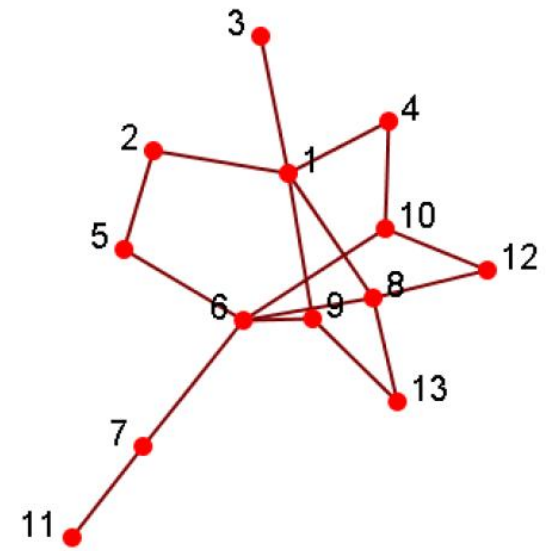


- **Snapchat**: founded in 2011; Sharing 10-sec video moments that erases itself after certain period; 70% of users <24 yrs old (~100M active users); user base will be bigger than Twitter soon



Social Networks between People

- One key aspect of social network systems is the network relationships between users
- Social Network: A social structure made of vertices (individual users) that are related to each other implicitly (similar behavior) or explicitly (friendship)
- Graphical Representation:
 - Vertices => members
 - Edges => relationships
 - Adjacency Matrix/List



Outline

- **Age of Social Media**
- **Analysis of Social Media**
- **Challenges in Social Media**

A Recap: Key Characteristics of Current Generation of Social Media Systems

- **Live, Big-Data and 3M** (Multi-source, Multimedia & Multilingual)
- Support users' need for **communication**, sharing and interaction
- Support **co-viewing and co-creation of contents** (by users and systems)
- **Develop social analytics** aim to understand contents, events and users - targeting at recommendation

What Has Changed in Last 5 Years?

1. Image/Video handling

- Top 3 recent **social media platforms** are all image/video centric



- **Live video** from GoPro

2. Live Location Analytics

- **Sensor devices are everywhere** – capable of multi-form of sensing
- **Multi-sources info**: Location traces, POI and Audio
- Towards better location estimation and mobility analytics.

What Has Changed in Last 5 Years

3. Online and Live are Central

- **Live**: comes the ability for continuous sharing, interactions and feedback..
- Users want to get **instant feedback** – from friends and systems

4. Quality and Structures of Data

- Deteriorating quality of data, with about **70% of UGCs belongs to noise/ spam/ rumors** category
- Key part of making data usable is to structure them: at both knowledge and data level

5. Co-Creation & Co-Invention

- With live instant feedback, comes the possibility to **co-create and co-invent**
- Not just the contents, but systems and design

Summary

- We live in new era of Social Media. It is really Big and full of heterogeneous relations of type “Many – Many”
- The data is not just Big, but multi-source and of different modality:
 - Text from Microblogs and other Social Forums
 - Location From Location-Based Social Networks
 - Images from Image Sharing Services
 - Video from Video Sharing Services
- We live in new era of Social Media. It is really Big and full of heterogeneous relations of type “Many – Many”
- Data become more noisy, New types of data emerge:
 - Live data
 - Sensor data
 - Etc.

Course Schedule

Wk	Date	Lecture/Tutorial Topics	Remarks
1.	9 Nov	L1: Introduction to Social Networks & Challenges	Details of Assgn 1
2.	9 Nov	L2: Text Processing	T1: Text representation
3.	9 Nov	L3: Location and Image Data Processing, Clarification on Assign. 1	T2: Location and Image data representation
4.	10 Nov	L4: Introduction to Retrieval and Classification	T3: Introduction to Information Retrieval and Classification
5.	10 Nov	L5: Source Fusion and Evaluation	T4: Data Source Fusion
6.	10 Nov	L6: Recent Study	T5: Additional Topics in Social Media
7.	12 Nov	L7: Wrap up	Summary of the above
8.	12 Nov	Group presentations of the Assignment 1	Assign 1 Due

Assessment:

- 50%: Lecture Participation
- 50%: 1 Programming Assignment

Assignment -1

- In addition to theory and knowledge, one key aspect of this course is hands-on exercises
- One assignment (groups of 3):

- **Prediction of User Demography**

Given info from **multiple social media platforms from three geographical regions**, determine the demography of users such as:

- **Age**
- **Gender**
- **Education Level**
- **Occupation Industry**
- **Relationship Status.**



- **10 min. group presentations** of projects after the last lecture on **Thursday Nov. 12, 2015**

Assignment -2

- Presentation should be 15 min long include:
 - Team members and their roles introduction
 - Demography prediction Approach Architecture and the Motivation behind.
 - Evaluation Results:
 - In terms of MACRO Precision, MACRO Recall, MACRO F-Measure
 - Based on those users, who have mentioned their real age (In ground truth files, the column “Real Age” is not empty) and for every city.

Assignment -3

- DATASET:
- <http://lms.comp.nus.edu.sg/research/NUS-MULTISOURCE.htm>
- DESCRIPTION OF THE DATA IS IN PAPER*.
- Please, ask **any questions** during the conference and after: **farseev@u.nus.edu**



*Aleksandr Farseev, Liqiang Nie, Mohammad Akbari, and Tat-Seng Chua. 2015.
Harvesting Multiple Sources for User Profile Learning: a Big Data Study
In Proceedings of the 5th ACM on International Conference on Multimedia
Retrieval (ICMR '15).

Assignment -4

- All slides and will be here:
 - <http://farseev.com/ainlfruct.html>
- Recommended software to use:
 - **KNIME** (No programming required)
<https://www.knime.org/>
 - **Python** and it's Machine Learning Support
 - **Any other language** you like. Just make it work ;)

<http://next.comp.nus.edu.sg/opportunities>

- **RESEARCH INTERN**
- **ARCHITECT / JAVA DEVELOPER**



Next Lesson

- **Text Processing**