- **Social Media Analytics**

1. Introduction
2. Social media & social networks
3. Social media data
4. Applications of social media
5. Challenges, biases and limitations
6. Text and reference books
7. Technical issues related social data analysis

Social media

- Social Media: Web and mobile based Internet applications that allow the creation, access and exchange of user generated contents that is ubiquitously available
  - Fun video ☺
- Social networking sites: Facebook, Twitter, Wikipedia..
- Blogs, wikis, news, online forums,
- Majority of them yield unstructured data

Social media cont….

- Social media is affecting various aspects of our life
- Mobile/tablets based apps

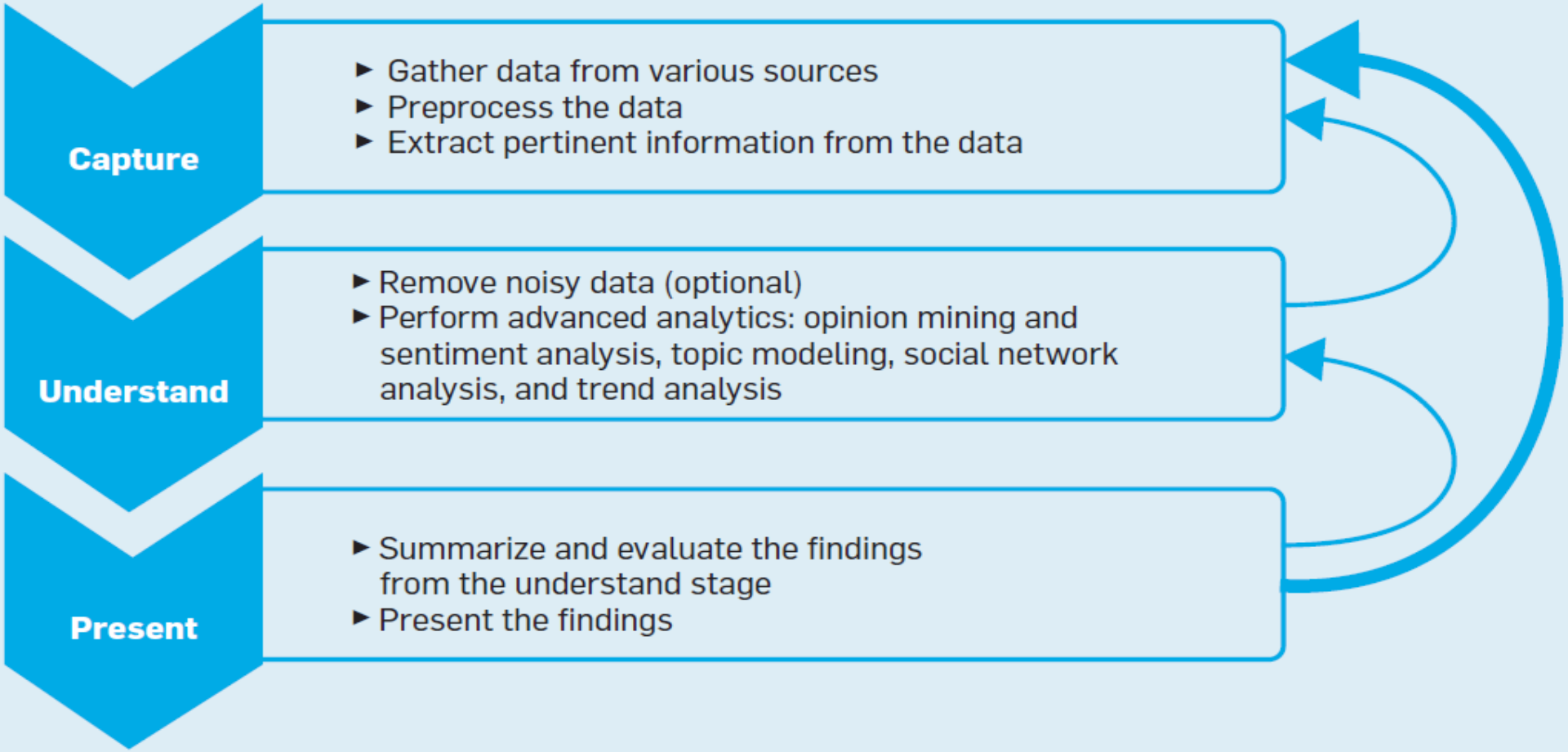
- Twitter feeds for sentiment analysis

- Computational Social Science

Social media analytics

- Social Media Analytics deals with development and evaluation of tools and frameworks to collect, monitor, analyze, summarize, and visualize social media data
  - Facilitate conversation and interaction between online users
  - Extracts useful patterns and information

Figure 1. Social media analytics process.



- **Network analysis** + machine learning + natural language processing (NLP) + statistics

Social network analysis (SNA)

- SNA provides a set of concept and metrics for systematic study of social network graphs
  - Used to understand underlying structure, connections and theoretical properties
  - A social network graph consists of nodes (users) and associated relationships (edges)
  - Direct: friendship, Indirect: voting, tagging and commenting

- To identify the relative importance of different nodes (edges) within the network
  - To identify key influencers in viral marketing

- Used to model network dynamics and growth
- Personalized recommendations and to detect sub communities
- Nicholas Christakis on Social Networks

Unstructured data

- Scraping: collecting online data from social media in the form of unstructured data, e.g. metadata, image tags, messages.
    - Social media data available through APIs
    - Due to commercial values, websites often impose various restrictions
    - DataSift, Gnip; Thomson Reuters for News data
- Opinion Mining: Automatic systems to determine human opinion, e.g. sentiment analysis, relevance etc.

1. Introduction
2. Social media & social networks
3. Social media data
4. Applications of social media
5. Challenges, biases and limitations
6. Text and reference books
7. Technical issues with social data analysis

Applications in business and management

- **Retail** companies - to harness their brand awareness, service improvement, advertising/marketing strategies, identifying influencers

- **Finance**: to determine market sentiment, news data for trading
  - Sentiment of random sample of twitter were correlated with Dow Jones Industrial Average prices
  - Twitter data to forecast individual NASDAQ stock prices

Public health and sociology

- Given that two people have been in approximately the same geographic locale at same time, on multiple occasions, how likely they know each other ? –social ties (PNAS, 2013)
  - Geotagged Flickr photos

- Forecasting the Influenza season using Wikipedia (MIT Tech. Review, Nov 3, 2014)
  - Wikipedia access logs + Center for Disease Control & Prevention (CDC) influenza-like illness reports

- Monitoring diseases: HealthMap

Government and public officials

- Monitoring public perception on political candidates, election campaigns and announcements
- Prediction at national level of happiness, unemployment etc.
  - Use of social media metrics to improve the share-ability and reach of articles
  - Social media job loss index: econprediction.eecs.umich.edu

- An article on real world applications
- Crime Patrol to identify "potential lone wolves"
  - Sudden change in behavior
  - www.orgnet.com

- DeitY & DST (India) DARPA and NSF (USA), IPTS(EU), and Social Computing laboratory by the Chinese Academy of Sciences

Social media startups

- Health: healthMap
- Third party data providers: SEMRUSH, Gnip
- To measure market sentiment: Social Market Analytics
- Text Mining of social networks: Ayasdi
- Social networks for crime control: orgnet
- Twitter for unemployment prediction

Some ideas..

- Encourage voluntary participation
- Social networks and latrine adoption ?
  - Track toilet use using smartphones and tablets
- Cell phone data to develop disaster response systems/management
  - Stampede, floods ?
- Tools for better visualization and understanding

Challenges, biases and limitations

- Introduction
- Often contains data and metadata - not readily treated using traditional analysis tools
  - e.g. tags, implicit and explicit social networks

- Holistic data sources: combining data from different sources to get meaningful insights (microblogs, blogs, real-time markets, customer data, reviews)

- Quality v/s quantity, Garbage in and garbage out
- Google Flu Trends

Challenges cont.

- Restrictions imposed by websites on data collection

- How social media providers change the sampling and filtering of data streams ?
  - Platform specific sampling problems: streaming APIs of twitter, are not an accurate representation of the overall platform data

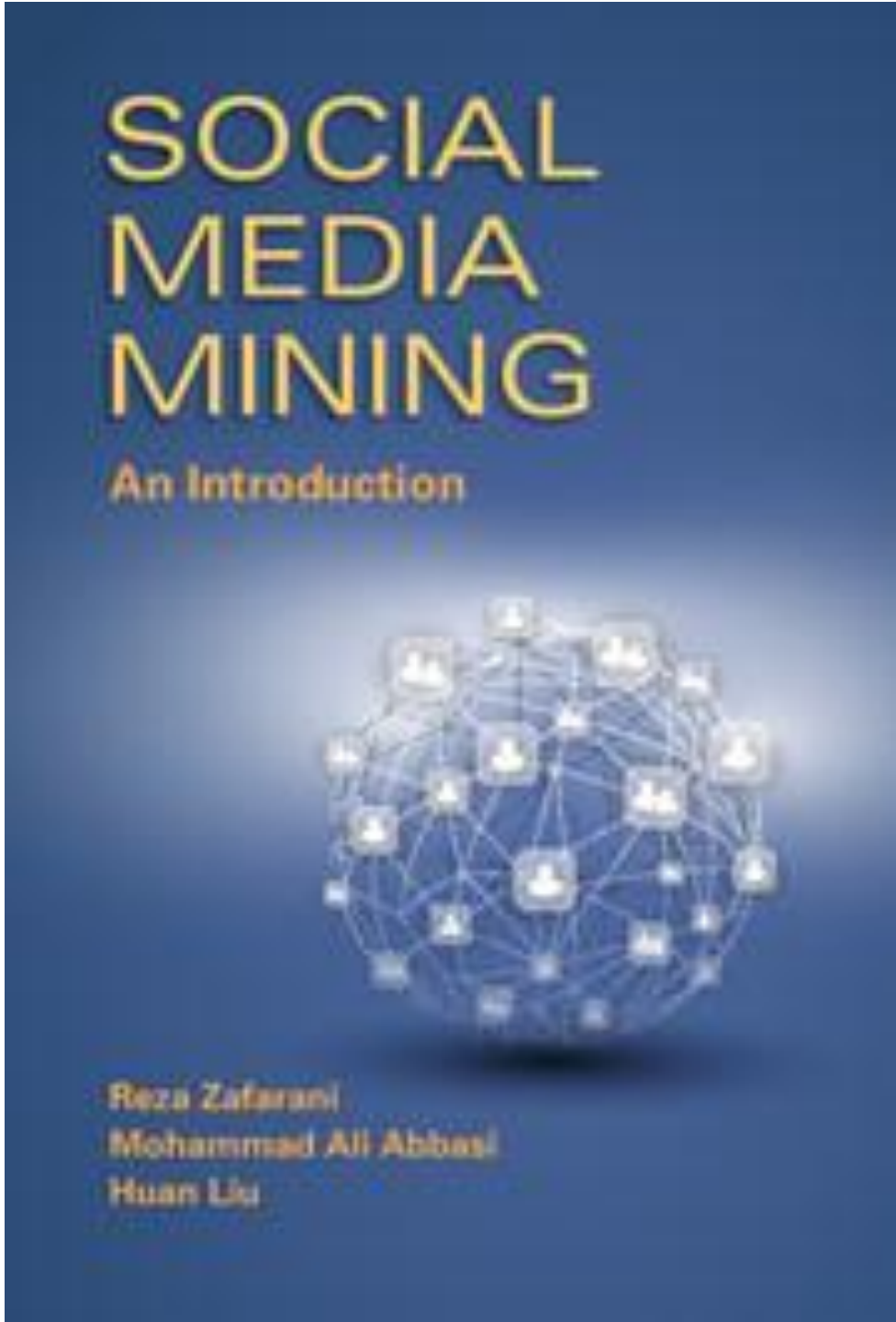
- Analysis may misrepresent the real world
  - Proxy population bias: very relevant in the Indian context
- Spread of unsubstantiated rumors
  - Rumors about Ebola
  - Fake news on social media

Challenges cont.

- Distortion of human behavior: social platforms are build to serve specific, practical purpose- not necessarily to represent social behavior
  - Alter ego: Professionally managed accounts of prominent individuals
- Nonhumans: social bots and spammers
- Replication of results
  - social media platforms forbid the retention or sharing of data sets
- Over fitting
  - Performance of a technique should take into account the number of feature being used; Feature hunting
- Social data is dynamic in nature and their sheer size pose significant computing challenges

Text book

- Social Media Mining: An Introduction
  Zafarani et al. 2014



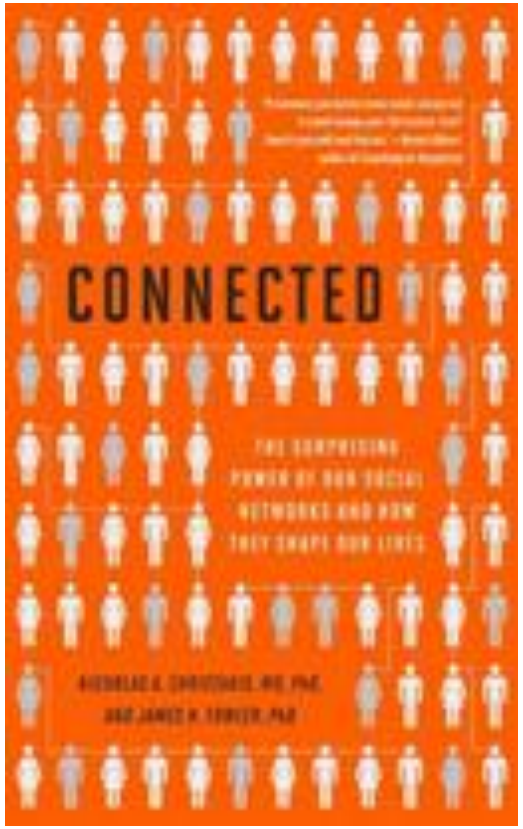
Getting started with social media..

- Gephi
- igraph

- Connected….

  by Nicholas Christakis

- Marketing Analytics



Different data sources

- Open publicly accessible academic alliance: DERP
- GitHub
- Wikipedia
  - HTTP based APIs that allows programmable access and scraping
  - Open source toolkit MediaWiki
- Facebook and Twitter data can also be accessed with some restrictions
  - JavaScript-based APIs, and return tagged data in XML, CSV or JSON.
- World Bank Databank ; data.gov.in

- News feeds
- Location and time sensitive feeds

Technical issues

- Social media data: XML, JSON, real-time financial data, spatial data
- Social media programmatic access:
  - Protect the raw data, but provide simple metrics
  - Google Trends, Google Analytics
- Data cleaning
  - DataWrangler
- Data Analysis Tools
  - Transformation tools: transforms textual inputs into tables, maps, charts etc.
    - Zoho
  - Analysis Tools: Gephi, Twitter Data Analytics

Application areas

- Economics and Finance
- Sociology/Psychology
- Marketing, management and organization science
- Geospatial: civil and environmental sciences
- Healthcare and public health
- Mathematics and Statistics
- Computer science


Thanks!