

# Social Media Analytics

## SMA Basics

### — SL01 —

Philippe Cudré-Mauroux, Mourad Khayati & Dingqi Yang

`{firstname.lastname}@unifr.ch`

# TABLE OF CONTENTS — SL01

## 1. Lecturers

### Lecturers

## 2. Course Information

## 3. Social Media

## 4. Social Media Analytics (SMA)

# PROF. DR. PHILIPPE CUDRÉ-MAUROUX



- ▶ Academic background:
  - ▶ M.Sc. from EPFL + Eurécom Institute
  - ▶ Ph.D. from EPFL (EPFL best thesis award, EPFL Press Mention)
  - ▶ Worked for Hewlett-Packard (CH), IBM Watson (USA), U.C. Berkeley (USA), Microsoft Research Asia (China), M.I.T. (USA)
  - ▶ Full Prof @ UNIFR since 09/2016
- ▶ Research interests:
  - ▶ Big data
  - ▶ Scientific data
  - ▶ Linked data management

# DR. MOURAD KHAYATI

- ▶ Academic background:
  - ▶ M.Sc. from INSA Lyon
  - ▶ Ph.D. from University of Zürich
  - ▶ Senior researcher @ UNIFR since 06/2015
- ▶ Research interests:
  - ▶ Time Series
  - ▶ Matrix decomposition techniques
  - ▶ Recovery of missing values



# DR. DINGQI YANG

- ▶ Academic background:
  - ▶ M.Sc. from Télécom SudParis (France)
  - ▶ Ph.D. Université Pierre et Marie Curie Paris VI (France)
  - ▶ Senior researcher @ UNIFR since 04/2015
- ▶ Research interests:
  - ▶ Social media data mining
  - ▶ Computational social science
  - ▶ Pervasive big data analytics



# TABLE OF CONTENTS — SL01

1. Lecturers

2. Course Information  
Introduction

3. Social Media

4. Social Media Analytics (SMA)

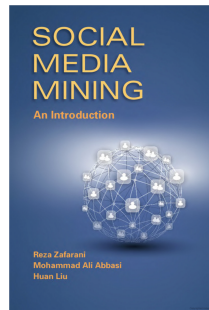
# LITERATURE

## ► Textbook:

- Reza Zafarani, Mohammad A. Abbasi and Huan Liu, *Social Media Mining: an Introduction*, Cambridge university Press, 2014.

- Available at the DIUF library

- Available as ebook from:  
<http://dmml.asu.edu/smm/book/>



## ► References (available at the DIUF library):

- Albert-László Barabási, *Network Science*, Cambridge university press, 2016.
- Charu C. Aggrawal, *Social Network Data Analytics*, Springer, 2011.

# GOALS OF THIS COURSE

The main things we will learn in this course:

- ▶ To learn how to collect, represent and mine social media data.
- ▶ Hands-on knowledge of a **toolbox** of social graph algorithms.
- ▶ To find and analyze communities in social media networks.
- ▶ To understand relationships in real-world social media networks.



# SYLLABUS

1. **Social Media Analytics Basics**
2. **Networks & Graphs**
3. **Network measures**
4. **Community Structure Analysis**
5. **Social Recommendation**
6. **Geo-Social Media Analytics**
7. **Network models**
8. **Diffusion, Assortativity & Influence**
9. **Crowdsourcing**
10. **Link Analysis**
11. **Recap**

# ADMINISTRATION/1

## TAs AND COURSE ACCESS

### Teaching Assistants (TA)

- ▶ Names: Rana Hussein, Natalia Ostapuk, Laura Rettig and Akansha Bhardwaj
- ▶ Contact: `firstname.lastname@unifr.ch`

### Course Access

- ▶ Enroll to the course on ILIAS (if not done yet)
- ▶ Check ILIAS frequently

# ADMINISTRATION/2

## COURSE CONTENT

### Course text book

- ▶ *Social Data Mining: an Introduction*, 1st edition, Cambridge University press, Reza Zafarani, Mohammad Ali Abbasi and Huan Liu, 2014.

### Slides

- ▶ Slides will be available on ILIAS shortly after the lecture.
- ▶ Many slides are adapted from various sources including research papers, tutorials, other courses, Web pages, etc.

# ADMINISTRATION/3

## ORGANIZATION AND LOGISTICS

### Lectures

- ▶ C 230, Thursday 09:15 - 12:00
- ▶ One lecture/lab per week
- ▶ Exception:
  - ▶ April 25, 2019 (Easter break)
  - ▶ May 23, 2019 (Project presentations)

### Office hours (Starting March 7, 2019)

- ▶ Discuss projects' related problems only
- ▶ Send an email to the TA to set an appointment

# ADMINISTRATION / 4

## PROJECTS

- ▶ The list of projects will be introduced on February 28, 2019
- ▶ The projects cover different topics in the course
- ▶ Each project consists on implementing a set of algorithms discussed in the course.
- ▶ Each project should be done in a group of **two** people.

# ADMINISTRATION/5

## EXERCISES

- ▶ In order to take part of the final exam, **five** homework assignments have to be passed. An assignment is passed if about half of the tasks are correctly/seriously solved.
- ▶ The assignments are published one week before the dedicated lab.
- ▶ Each assignment is an individual work. Group work is not allowed.
- ▶ The assignment work should be uploaded to ILIAS by the latest the day before the lab day.

# ADMINISTRATION/6

## EVALUATION

**Exam participation:** Only students who passed the exercises will be allowed to take part of the final exam.

**Exam grade:** As part of the course assessment the written final exam (grade between 1 and 6) must be completed with a passing grade.

### Evaluation:

- ▶ Project presentation (May 23, 2019): 30%
- ▶ Final exam (June 2019): 70 %

# GENERAL REMARKS/1

- ▶ **Hands-on exercises** are an important part of this course: an abstract understanding of the concepts is not good enough.
- ▶ You must be able to **apply** your knowledge to **new** examples. Use exercises to practice this during the semester.
- ▶ Often, it is most effective to first solve exercises/algorithms on paper and later key them in on the computer.
- ▶ Theoretical and programming tasks.



## GENERAL REMARKS/2

- ▶ During lectures/labs:
  - ▶ Interaction is welcome; ask questions.
  - ▶ Speed up/slow down the progress.
  - ▶ Additional explanations if desired.
- ▶ During the lectures we solve many examples. Please participate and take notes.

# PREREQUISITES

## Data structures and algorithms

- ▶ Data types, procedures and functions
- ▶ Time/space complexity

## Programming skills

- ▶ Ability to edit, compile, and execute a program
- ▶ Python/C and basic knowledge in Matlab

## Basic knowledge of statistics, probability, linear algebra and math

- ▶ Variance, correlation and density
- ▶ Matrix operations and optimization functions

# ACKNOWLEDGMENTS

- ▶ The slides are based on the textbook Social Media Mining: An introduction from R. Zafarani M.L. Abbasi and H. Liu.
- ▶ Kurt Ranaltar created the initial Latex template of the slides.
- ▶ Michael H. Böhlen from UZH created a refined latex template of the slides.

# TABLE OF CONTENTS — SL01

1. Lecturers

2. Course Information

3. Social Media  
Overview

4. Social Media Analytics (SMA)

# QUICK QUIZ

Examples of:

- ▶ Social media networks
- ▶ Data analysis
- ▶ Recommender systems
- ▶ Privacy issues

# DEFINITION

- ▶ Social media is the collection of online communication channels dedicated to community-based input, interaction, content-sharing and collaboration.
- ▶ Websites and applications used to deploy forums, microblogging, social networking, social bookmarking, social curation, and wikis are considered as social media platforms.
- ▶ Social Media is a media designed to be disseminated through social interaction.

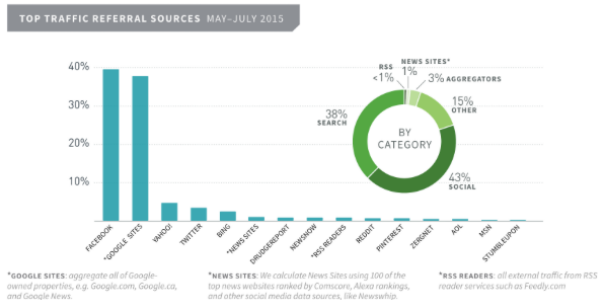
# SOCIAL MEDIA: BIG CHANGE/1

- ▶ Web is no longer a static library that people passively browse.
- ▶ Web became a place where people:
  - ▶ Consume and create content
  - ▶ Interact with other people



# SOCIAL MEDIA: BIG CHANGE/2

- Facebook surpassed Google in referral traffic for large publications for the first time in 2015: <http://fortune.com/2015/08/18/facebook-google/>.





# SOCIAL MEDIA: BIG CHANGE/3

- ▶ Big Data
  - ▶ Billions of users and contents
  - ▶ Textual, Multimedia (image, videos, etc.)
  - ▶ Billions of connections
- ▶ Data access
  - ▶ Data is easily accessible on Social Media
  - ▶ Datasets
  - ▶ Crawling the web

# SOCIAL MEDIA DATA

- ▶ Social tagging: CiteULike, Bibsonomy, MovieLens, Delicious, Flickr, etc.
- ▶ Twitter Firehose/API
- ▶ MySpace data (real-time data, multimedia content, etc.)
- ▶ Stanford Large Network Dataset Collection (SNAP): Social networks, Citation networks, Collaboration networks, location-based online social networks, etc.  
<https://snap.stanford.edu/data/>

# SOCIAL MEDIA CATEGORIES

- ▶ *Communication*: direct interaction:
  - ▶ Social Networks (Facebook, MySpace, LinkedIn)
  - ▶ Blogs (Blogger, Wordpress)
  - ▶ Location-based Social Networks (Foursquare, Facebook places): “Check in” to visited places.
- ▶ *Collaboration*: Social Bookmarking (StumbleUpon, Delicious) and Social News (Digg, Reddit)
- ▶ *Multimedia*: photography and video sharing
  - ▶ Photography (Flickr, Photobucket, Picasa)
  - ▶ Video Sharing (YouTube, Dailymotion, Vimeo)
  - ▶ Livecasting (Skype, Ustream)

# LOCATION-BASED SOCIAL NETWORKS

## FOURSQUARE



- ▶ Large-scale POI data sources  $\Rightarrow$  Local search and recommendations
- ▶ Real-world application (crime detection):  
<http://prediction.heaney.ch>
- Location privacy



# SOCIAL MEDIA TOOLS

## HootSuite, TweetDeck, Seesmic

- ▶ Used to update multiple social profiles at once
- ▶ Can choose which ones to send to

## Metrics

- ▶ Google Analytics, Insight: Track fans/friends/subscribers, posts, visits, links clicks, page views, etc.
- ▶ Google Reader (track RSS feeds), Google Alerts (track keywords across the web)

# TABLE OF CONTENTS — SL01

1. Lecturers
2. Course Information
3. Social Media
4. Social Media Analytics (SMA)  
Social Media Analytics

# DEFINITION

- ▶ Social media analytics is the process of collecting and analyzing data from social media platforms to extract meaningful information.
- ▶ SMA introduces basic concepts and principal algorithms suitable for analyzing massive social media data.
- ▶ SMA is an interdisciplinary field (computer science, social sciences, etc.).

# APPLICATION OF SMA

- ▶ SMA is useful to apply when you want to:
  - ▶ Study the properties of a network to improve its effectiveness.
  - ▶ Analyze a network (offline or online) to perform a quantitative or qualitative research.
  - ▶ Visualize your data to discover patterns in relationships (or interactions).
  - ▶ Follow the paths that information follows in social networks.



# CHALLENGES

- ▶ Size of data
  - ▶ The social media data is tremendous which makes it difficult to collect, store and analyze
  - ▶ Sampling data might be biased and not representative of the full data
- ▶ Noise removal/reduction
  - ▶ The noise detection is a complicated task
  - ▶ Removing too much noise yield to too sparse networks that can render data useless
- ▶ Empirical evaluation
  - ▶ Ground truth is often not existing
  - ▶ Creating synthetic data requires domain knowledge expertise