COM5507 Social Media Data Acquisition and Processing
Week 1a. Introduction to the Course

Lectured by: Dr. Xinzhi ZHANG
Research Assistant Professor, Department of Journalism
Hong Kong Baptist University
5 Sep 2018 @ CityU M5064

Agenda

- Who am I?
- What is this course about?
- What is the assessment method?

- **Xinzhi ZHANG [張昕之], Ph.D.**
  - www.drxinzhizhang.com
  - https://scholar.google.com/citations?user=iOFeIDIAAAAJ&hl=en
  - @xin_zhi_zhang
  - https://github.com/xzzhang2

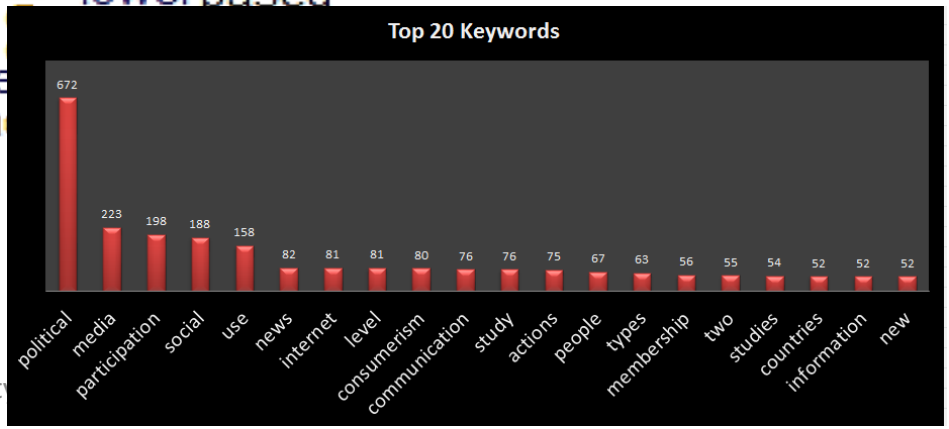
- **Current Position**
  - **Research Assistant Professor,** Department of Journalism, Hong Kong Baptist University
  - **Programme Director,** the Data and Media Communication Concentration (DMC)

- **Professional Experience**
  - **Research Assistant Professor**, Department of Journalism,
    Hong Kong Baptist University (2016 – Present)
  - **Lecturer**, School of Professional Education & Executive Development,
    The Hong Kong Polytechnic University (2014 – 16)
  - **Postdoctoral Fellow**, Department of Media and Communication,
    City University of Hong Kong (2013 – 14)

- **Education**
  - **Ph.D.** in Media and Communication,
    City University of Hong Kong (2009 – 13)
  - **M.A.** with *Distinction*, in Communication and New Media,
    City University of Hong Kong (2008 – 09)
  - **B.A.** in Broadcasting Journalism,
    Guangzhou University, China (2004 – 08)

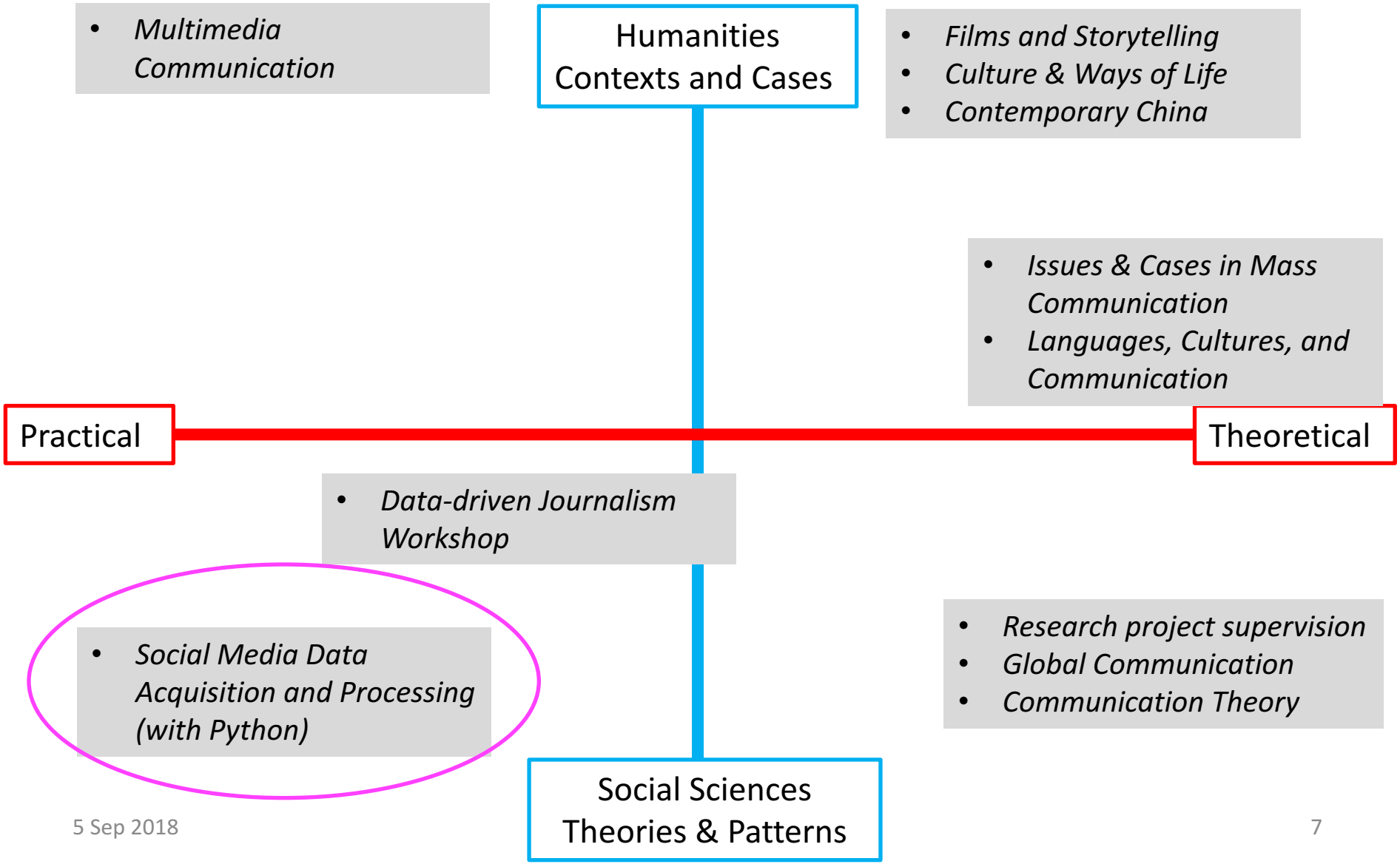
My research areas

- Emerging information and communication technologies and social change
- Comparative political communication
- Digital humanities and digital popular culture
- Data-driven journalism and data visualization
- Computational methods and digital media
- …



Note: The word cloud and the corresponding word frequency are generated from **three recent papers** *with* **R.**

My teaching


Humanities
Contexts and Cases
Practical
Theoretical
Social Sciences
Theories & Patterns
• Multimedia Communication
• Films and Storytelling
• Culture & Ways of Life
• Contemporary China
• Issues & Cases in Mass Communication
• Languages, Cultures, and Communication
• Data-driven Journalism Workshop
• Social Media Data Acquisition and Processing (with Python)
• Research project supervision
• Global Communication
• Communication Theory

Is this course easy?

Is this course happy?

Am I getting lost?

This course works on different types of (online) data – and fetch them

This course works on different types of (online) data – and process them

This course (may) be hard…

This course is interdisciplinary

http://chasingdeer.co.uk/analyst-venn-diagram.html

https://www.r-bloggers.com/data-science-in-businesscomputational-social-science-in-academia/

https://knightlab.northwestern.edu/2013/06/28/want-to-build-a-data-journalism-team-youll-need-these-three-people/

This course is informative and helpful!

Course Overview

- Unit 1: Data science fundamentals and preliminary Python programming (week 1 – 3)
- Unit 2: Automated web data collection (week 4 – 7)
- Unit 3: Data (pre-)processing and data management (week 8 – 9)
- Unit 4: Data exploration (week 10 – 11)

Assessment Components

Type of Assessment	Weighting	CILOs to be addressed	Description of Assessment Tasks
Class participation and tutorial tasks	30%	1, 3, 4	Students will be introduced in lectures and guided readings to the key concepts and methods on data acquisition and processing in the digital age.
Individual exercises	40%	2, 3, 5	Students develop and test customized algorithms individually to collect and process social media data.
Group project and presentation	30%	1, 2, 4, 5	Students work in teams to collect, process, and analyze social media data and present their findings in data product and an oral presentation.

Assessment

- **Class participation and tutorial tasks (30%)**
- Attendance will be taken each week and there will be regular “attendance checkpoints.” “
- Attendance check points” will be announced every 3 weeks (from week 2 – week 13). Full-attendance in each of the check point will be ***awarded*** 1 mark in their final course evaluation;
- (Ir-)regular small quiz and bonuses.
- Students are expected to participate in the lecture sessions, tutorial exercise, and project-based consultation meetings punctually and actively.

Assessment

- **Individual Exercises (40%)**
  - 1 automated online data acquisition challenge (20%)
  - 1 automated data processing (data cleaning) challenge (20%)
  - for each assignment:
    - 500– 800 words in English;
    - defining the problems
    - identifying the (online) data sources
    - presenting the codes and analytical process step by step
    - briefly reporting the results

Assessment

- **Group Project (30%): Presentation (15%) and report (15%)**
  - 3 - 4 students per group;
  - Develop a data-driven investigation and storytelling project based on first-hand automated web data collection and present the research questions, methods and approaches, and findings;
  - Each group, with all the members, should meet and discuss the project with the instructor at least twice throughout the semester;
  - In-class presentation, 12 - 15 minutes;
    1. appropriateness of materials;
    2. proper application of data science skills;
    3. organization – clear and logical flow, coherent and cohesive;
    4. delivery – use of high concepts, clear and focused presentation;
  - Written reports
    1. A4 pages 12 – 15 pages in English (excluding references, codes, and appendix)
    2. focusing on questions, and data analytical steps (the pipelines)
    3. interpreting the results

Assessment

- **Workload distribution declaration for Group Project**
  - **NO "free rider" please.**
  - You can discuss, negotiate and allocate the workload on your own within your team.
  - A **declaration form** will be submitted with your final project report.
  - During the in-class presentation, a table clearly show the **labor distribution** is required.

Weekly Schedule

- Teaching Plan.

Class Rules

- 1. On Class-Participation
  - No attendance requirement, but the participation is counted into the final assessment.
  - The sign-up sheet
  - The regular participation
  - Please be punctual

Class Rules

- 2. In-class rules.
  - "**Silence is golden**" - as long as you keep quite…
  - Eating and drinking?
  - You may bring your own laptop
  - Medium of instruction?