An Overview of CS512
@Spring 2016

JIAWEI HAN
COMPUTER SCIENCE
UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN

JANUARY 21, 2016
Data and Information Systems (DAIS) Course Structures at CS/UIUC

- Three main streams: Database, data mining and text information systems
  - Seminar: Yahoo!-DAIS Seminar: (Not CS591 seminar, no credit given)
- Database Systems:
  - Database management systems (CS411: Fall+Spring)
  - Advanced database systems (CS511: Fall)
  - Human-in-the-loop Data Management (CS 598: Aditya Parameswaran)
- Data mining
  - Intro. to data mining (CS412: Fall+Spring)
  - Data mining: Principles and algorithms (CS512: Spring (Han))
- Text information systems
  - Introduction to Text Information Systems (CS410: Spring (Zhai))
  - Advance Topics on Information Retrieval (CS 598: Fall (Zhai))
  - Social & Economic Networks (CS 598: Hari Sundaram)
Coursera Data Mining Specialization

- Data Visualization: John Hart
- Pattern Discovery in Data Mining: Jiawei Han
- Text Retrieval and Search Engines: ChengXiang Zhai
- Cluster Analysis in Data Mining: Jiawei Han
- Text Mining and Analytics: ChengXiang Zhai
- Data Integration and Data Warehousing: Kevin Chang
- Capstone Data Mining Capstone (6 weeks)
Topic Coverage: CS512 @ 2016

- Class introduction (0.5 wk)
- Introduction to networks (1.5 week)
- Mining heterogeneous information networks (2 weeks)
- Text mining (2 weeks)
  - 1st midterm exam (0.5 week) — 1st Lect. of 7th week
- Truth finding (1 week)
- Stream data mining (1 week)
  - Spring break (1 week)
- Mining spatiotemporal and social media data (1 week)
- Advances on classification (1 week)
- Advances on outlier analysis (1 week)
- Selected class survey presentation (1.5 week)
  - 2nd midterm exams (0.5 week) — 2nd Lect. of 15th week
- Class research project presentation (final week + exam week)
Class Information

- **Instructor:** Jiawei Han ([www.cs.uiuc.edu/~hanj](http://www.cs.uiuc.edu/~hanj))
  - Lectures: Tues/Thurs 9:30-10:45am (0216 Siebel Center)
  - Office hours: Tues/Thurs. 10:45-11:30am (2132 SC)

- **Teach Assistants:**
  - Fangbo Tao (ftao2), Shi Zhi (shizhi2)

- **Prerequisites** (course preparation)
  - CS412 (offered every semester) or consent of instructor
  - General background: Knowledge on statistics, machine learning, and data and information systems will help understand the course materials

- **Course website** (bookmark it since it will be used frequently!)
  - [https://wiki.cites.illinois.edu/wiki/display/cs512/Lectures](https://wiki.cites.illinois.edu/wiki/display/cs512/Lectures)

- **Textbook:**
  - Yizhou Sun and Jiawei Han, *Mining Heterogeneous Information Networks: Principles and Methodologies*, Morgan & Claypool, 2012
  - J. Han, M. Kamber, J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2011
  - A set of recent published research papers (see course syllabus)
Textbook & Recommended Reference Books

- **Textbook**
  - Yizhou Sun and Jiawei Han, *Mining Heterogeneous Information Networks: Principles and Methodologies*, Morgan & Claypool, 2012
  - Jiawei Han, Micheline Kamber, Jian Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Morgan Kaufmann, 2011

- **Recommended reference books**
  - P. S. Yu, J. Han, and C. Faloutsos (eds.), *Link Mining: Models, Algorithms, and Applications*, Springer, 2010
Assignments: (2 assignments) 10% total

Two midterm exams: (20% each) 40% in total

Research project proposal (one-page): 0% (due at the end of 4th week)

Class attendance (3%): Max-allowed missing: 3 classes, then -1% for each 3 misses

Class presentation and/or research survey (12% total)

- Survey report: [expect to be comprehensive and in high quality, ≈ 20 pages]
  - Encourage to align up with your research project topic domain
  - Hand-in together with companion presentation slides [due at the end of 12th week]

Class presentation: May use 10 min. class survey presentation to replace the survey report (consent of instructor)—contents must closely aligned with the class content and in very high technical quality

Final course project: 35% (due at the end of semester)

Evaluated by class (50%) and TA + instructor (50%) collectively!
Final course project: 35% (due at the end of semester)

- The final project will be evaluated based on (1) technical innovation, (2) thoroughness of the work, and (3) clarity of presentation.

- The final project will need to hand in: (1) project report (length will be similar to a typical 8-12 page double-column conference paper), and (2) project presentation slides (which is required for both online and on-campus students).

- Each course project for every on-campus student will be evaluated collectively by instructor (plus TA) and other on-campus students in the same class.

- The course project for online students will be evaluated by instructors and TA only.

- Group projects (both survey and research): Single-person project is OK, also encouraged to have two as a group, and team up with other senior graduate students, and will be judged by them.
Where to Find Reference Papers?

- Course research papers: Check reading list and list of papers at the end of each set of chapter slides
- Major conference proceedings that will be used
  - DM conferences: ACM SIGKDD (KDD), ICDM (IEEE, Int. Conf. Data Mining), SDM (SIAM Data Mining), ECMLPKDD (Principles KDD), PAKDD (Pacific-Asia)
  - DB conferences: ACM SIGMOD, VLDB, ICDE
  - ML conferences: NIPS, ICML
  - IR and Web conferences: SIGIR, CIKM, WWW, WSDM
  - Social network conferences: ASONAM
- Other related conferences and journals
  - IEEE TKDE, ACM TKDD, DMKD, ML
- Use course Web page, DBLP, Google Scholar, Citeseer
From Data to Networks to Knowledge: An Evolutionary Path!

Han, Kamber and Pei, Data Mining, 3rd ed. 2011
Yu, Han and Faloutsos (eds.), Link Mining: Models, Algorithms and Applications, 2010
Sun and Han, Mining Heterogeneous Information Networks, 2012
Wang and Han, Mining Latent Entity Structures, 2015

Y. Sun: SIGKDD’13 Dissertation Award
C. Wang: SIGKDD’15 Dissertation Award