Large-Scale Networks

1a-Introduction
What is a Network?
Why you should care about networks

- **Universal language** for describing complex data
  - Networks from science, nature, and technology are more similar than one would expect

- **Shared vocabulary** between fields
  - Computer science, social science, physics, economics, statistics, biology

- **Data availability** (/computational challenges)
  - Web/mobile, bio, health, and medical
Air-traffic

US air-traffic graph
Facebook

Facebook social graph - 4-degrees of separation [Backstrom-Boldi-Rosa-Ugander-Vigna, 2011]
Epidemics

Spread of tuberculosis disease
Terrorist network

9/11 terrorist network [Krebs, 2002]
Blog Web pages

Political blogs prior to the 2004 U.S. Presidential election
Administration
Timetable

- Title: Large-Scale Network
- UOS code: COMP5313
- Credit points: 6
- Lecture:
  - Thursdays 18h-20h, weeks 1-13
- Lab:
  - Thursdays 20h-21h, weeks 2-10 (11-12)
Myself

- Vincent Gramoli
  - Office: 417
  - Phone: (02) 903 69270
  - Vincent.Gramoli@sydney.edu.au
  - Office hour: Thursday 14h-15h

- Background
  - Distributed computing
  - USA, France, Switzerland
eLearning


- All assessments are communicated through the blackboard website: [https://elearning.sydney.edu.au/](https://elearning.sydney.edu.au/) and piazza.

- Do not send email to the lecturer, but post questions online on piazza [https://piazza.com/sydney.edu.au/](https://piazza.com/sydney.edu.au/)

- Send proof (certificate) within the next 5 days (special consideration) and you will either be excused or you will be evaluated through an alternative mark assessment
eLearning

- Slides will be uploaded right before the lecture

- Video recording should be automatically posted on blackboard
Material
Materials

Textbook: Networks, Crowds and Markets. Easley and Kleinberg

- Required textbook
Materials

Related books (optional)

Materials

Other books related to defence (optional)

...and to programming (optional)
Roadmap
Roadmap

Click here:
Objectives
Course description

- The growing connectedness of modern society translates into simplifying global communication and accelerating spread of news, information and epidemics. The focus of this unit is on the key concepts to address the challenges induced by the recent scale shift of complex networks. In particular, the course will present how scalable solutions exploiting graph theory, sociology, game theory and probability tackle the problems of communicating (routing, diffusing, aggregating) in dynamic and social networks.
Goal

- Understanding
  - Network structures
  - Strategic behaviors
  - Epidemic behaviors

- Networks were studied in many fields
  - Computer science
  - Applied mathematics
  - Sociology

- A synthesis is thus necessary
  - Complexity of network structures
  - Information
  - System with interacting agents

› We cover:
  - Graph theory
  - Information networks
  - Network dynamics
  - Basic Python programming

› We do not cover:
  - Game theory
  - Economics
  - Markets
Learning outcomes

- **Information Seeking**
  - 1. Students understand and can quantify accurately the role of network in communication exchanges.
  - 2. Students understand the technical issues that affect the dissemination of information in a network.
  - 3. Students can analyse probabilistically the relations between communicating entities of a network.
  - 4. Students know the key factors that impact the accuracy and speed of information dissemination and aggregation.

- **Maths/Science Methods and Tools**
  - 5. Students understand the asymptotic complexity and accuracy of graph algorithms.
  - 6. Students know the stochastic methods necessary to evaluate the convergence of various algorithms.
  - 7. Students recognise probabilistic solutions to problems that have no deterministic solutions and apply them thoroughly.

- **Engineering/IT Specialisation**
  - 8. Students have skills to compare experimentally and theoretically the adequacy of different probabilistic solutions.
  - 9. Students are familiar with various types of network models in different contexts like computer science, society or markets
  - 10. Students understand the fundamental structures, dynamics and resource distribution in such models.
Assessments

- Assignment 1
  - Solving problems [20%; due week 6].
  - Covers outcomes 2, 3, 4, 5, 6 and 7.

- Midterm online Quiz
  - Answering multiple choice questions in labs [20%; due week 7].
  - Covers outcomes 1, 2, 3, 4, 5, 7 and 10.

- Assignment 2:
  - Either one of these two tasks:

    (1) Writing a short (4-6p) research paper exploring a research topic related to the course in LaTeX and presenting the related work and an analysis of this topic; or

    (2) Programming an algorithm related to the in C/C++, Java or Python and making a demo of it.

  - [20%; due week 10]. Covers outcomes 5, 7, 8, 9 and 10.

- Final exam [40%; due exam period] Covers all outcomes.
Assessments

- Suggestions for assignment 2

- Programming assignment examples (suggest your own as early as possible):
  - Computing the PageRank on a network, analyzing the result, concluding
  - Extracting properties from a network (clustering coefficient, degree, diameter), analyzing the result, concluding
  - Building a new network, based on Massively Open Online Courses (MOOC) and linking them with pre-requisites links, analyzing, concluding

- Literature assignment examples (suggest your own by as early as possible):
  - Gossip-Based Computation of Aggregate Information. Kempe, Dobra, Gherke.
Graph theory

- The communication can be balanced between staying within small organizational units and cutting across organizational boundaries.

- **Strong ties**, representing close and frequent social contacts, tend to be embedded in tightly-linked regions of the network.

- **Weak ties**, representing more casual and distinct social contacts, tend to cross between these regions.

- At a global scale, it suggests some of the ways in which weak ties can act as *short-cuts* that link together distant parts of the world, resulting in the phenomenon colloquially known as the *six degrees of separation*. 
Graph theory

E-mail communication among 436 employees of HP Lab
Structural balance

- **Social networks** can also capture the **sources of conflict** within a group.

- People may tend to be connected to **many others without being connected together**.

- **Non-interacting clusters** may be symptomatic of a **conflicts** between people.

- **Structural balance** can be used to reason about how **fissures** in a network may arise from the dynamics of conflict and antagonism at a **local level**.
Structural balance

Two rival karate clubs represented by different colors
Network cascades

- A *cascading behavior* spreads from one person to another
  - Like a *biological epidemics*
  - Called “*social contagion*”

- *Cascading effects* may arise from individuals with incentive to adopt the behavior of their neighbors in the network
  - A new behavior starts with a *small set of initial adopters*
  - Then spreads radially outward *through the network*
Network cascades

Ebola 2014 Crisis

Sources: New England Journal of Medicine

The University of Sydney
Network cascades

Ebola 2014 Crisis

Source: Washington Post
Network cascades

E-mail recommendations for a particular Japanese graphic novel spread outward from four initial purchasers [Leskovec et al. The dynamics of viral marketing. ACM Trans. on the Web ‘07]
Applications

- When understood, the properties of networks can be applied elsewhere

- **News spreading:** *Six degrees of separation* is useful to disseminate information rapidly

- **File lookup:** Selecting *shortcuts* adequately can help navigating rapidly to a destination
Applications

A gossip-based protocols allows to rapidly spread data to a large number of computers of the same network.
Applications

The overlay networks use shortcuts to effectively retrieve files in a peer-to-peer file sharing system.
Why Taking this Course?
Why taking this course?

- **Networks are everywhere**
  - Public transportation
  - Social media
  - Protein interactions

- **Give a better understanding of networks**
  - Property identification (low diameter, high clustering coefficient…)
  - Characterization (scale-free, power-law distribution…)

- **Large impact**
  - Marketing: what individuals to target to maximize product adoption?
  - Defense: how to locate a terrorist?
  - Health: how to identify patient 0 and understand epidemics propagation?
Why taking this course?

- Age and size of social networks

Source: CS224W Stanford University - Jure Leskovec.