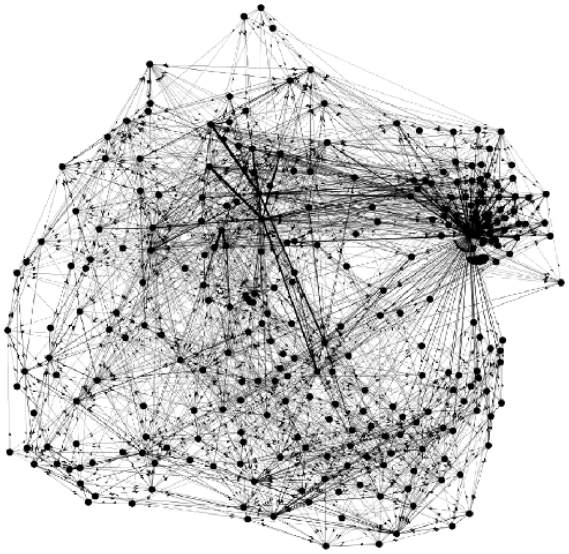
Network Analysis

MAP-I Curricular Unit in Paradigms



Summary

This document presents a proposal of a Ph.D. level course intended as a Curricular Unit in Technology for the MAP-I Doctoral Programme in Informatics. It is offered jointly by Universidade do Porto and Universidade de Aveiro. The course will cover complex networks concepts, theories, visualizations and algorithms. Topics include methods for structural characterization, pattern detection, information extraction, influence detection, and business intelligence in social media.

Instructors

Pedro Ribeiro	DCC, FCUP	pribeiro@dcc.fc.up.pt
José Luis Oliveira	DETI, UA	jlo@ua.pt
Sérgio Nunes	DEI, FEUP	sergio.nunes@fe.up.pt

Course Objectives

A wide variety of real and artificial structures can be intuitively represented by networks. The huge increase in available data in recent years as led to the emergence of network science as a pivotal interdisciplinary academical field dedicated precisely to the study and analysis of complex networks. From an application point of view, this course will mainly focus on biological and social data, but the methodologies given are general and flexible enough to be applied to any kind of network. The first part of the course covers the fundamentals in graph theory, models and common network metrics. The second part follows a data mining approach, with emphasis on subgraph mining, community discovery, information cascades and the challenges of large scale network analysis. The third and final part addresses the fundamentals of web information retrieval and link analysis.

Some similar courses on other universities are:

- "Social and Information Network Analysis", Jure Leskovec, Stanford University
- "Network Science and Models", Patrick Jaillet, MIT
- "Networks", Jon Kleinberg, Cornell University
- "Networks", Mason Porter, Oxford University
- "Network Science", Albert-Lszl Barabsi, Northeastern University
- "Network Theory", Mark Newman, University of Michigan
- "Network Analysis and Modeling", Aaron Clauset, Santa Fe Institute
- "Network Science Analytics", Gonzalo Mateos, University of Rochester

Learning Outcomes

At the end of the this course the students should be able to:

- explain key concepts and algorithms in complex network analysis;
- apply a range of techniques for characterizing network structure;
- define methodologies for analyzing networks of different fields;
- demonstrate knowledge of recent research in the area and exhibit technical writing and presentation skills.

Teaching Methods and Evaluation

The course is composed by theoretical-practical classes, discussions, student assignments and presentations. The student evaluation is based on the following key components:

- 50% Final Exam
- 40% Course Assignments and Presentations
- 10% Participation in Class

Course Content

The course is structured in three parts: (I) Introduction to Network Analysis, (II) Graph Mining and Patterns Discovery, (III) Information Networks and the World Wide Web

Part I - Introduction to Network Analysis

1. **Graph Theory Concepts:** basic definitions; graph types; graph examples; graph representations, classical graph algorithms.
2. **Basic metrics:** diameter, clustering coefficient, shortest paths, centralities.
3. **Network visualization:** graph formats; graph drawing; graph layout methods and algorithms; software experimentation.
4. **Graph models:** Erdös-Rényi random models; small-world and Watts-Strogatz model; scale-free networks, preferential attachment and Barabási-Albert model; other models.

Part II - Graph Mining and Patterns Discovery

1. **Subgraph Mining:** subgraphs as local metrics; subgraph census; network motifs concept and algorithms; graphlet degree distribution; weighted motifs.
2. **Communities:** network community detection; modularity optimization; spectral clustering; overlapping communities; other methods.
3. **Network Characterization and Comparison:** comparative approaches; network dynamics; information cascades; role mining.
4. **Large Scale Network Analysis:** peta-scale networks; graph storage and representation; parallel and distributed graph algorithms.

Part III - Information Networks and the World Wide Web

1. **Information retrieval overview:** Ranking models, Scoring, Term weighting, Evaluation
2. **Web crawling and link analysis:** Web retrieval, Ranking, PageRank.
3. **Information extraction:** Named entity recognition, normalization, relation extraction.
4. **Applications**

Main Bibliography

- Newman, M. "Networks: An Introduction", Oxford University Press, 2010.
- Easley, D., Kleinberg, J. "Networks, Crowds, and Markets: Reasoning About a Highly Connected World". Cambridge University Press, 2010.
- Croft, W. B., Metzler, D., and Strohman T. "Search engines: Information retrieval in practice". Reading: Addison-Wesley, 2010. (available online at `http://ciir.cs.umass.edu/irbook/`)

Additional Bibliography

- Manning, C., Raghaven, P. and Schutze, H. "Introduction to Information Retrieval", Cambridge University Press, 2008.
- Barabsi, A. "Network Science" (available online at `http://barabasi.com/networksciencebook/`)
- Junker, B. and Schreiber, F. "Analysis of Biological Networks". Wiley, 2008.
- West, D. "Introduction to Graph Theory (2nd Edition)". Prentice Hall, 2001.
- Newman, M., Barabási, A. and Watts, D.. "The Structure and Dynamics of Networks". Princeton University Press, 2006.
- Wasserman, S. & Faust, K. "Social network analysis: Methods and applications". New York: Cambridge University Press, 1994.
- Russel, M.A. "Mining the Social Web". O'Reilly Media, Inc. 2011.

Software

- Cytoscape: An Open Source Platform for Complex Network Analysis and Visualization (multiplatform/java based) - `http://www.cytoscape.org/`
- Gephi: an open source graph visualization and manipulation software (multiplatform/java based) - `http://gephi.org/`
- gtrieScanner: Quick Discovery of Network Motifs `http://www.dcc.fc.up.pt/gtries/`
- GraphCrunch 2: Software tool for network modeling, alignment and clustering (Windows and Linux) - `http://bio-nets.doc.ic.ac.uk/graphcrunch2/`
- SNAP: a C++ libary for working with massive network datsets (Windows, Linux, Mac) - `http://snap.stanford.edu/`

Instructors Team

The team of instructors is actively involved in research in the field of complex network analysis and information retrieval. A short biography of the instructors is provided below:

Pedro Ribeiro (`http://www.dcc.fc.up.pt/~pribeiro`) is an invited Assistant Professor at the University of Porto, Computer Science Department and a Senior Researcher at the Center for Research in Advanced Computing Systems. He holds a PhD from the University of Porto in the field of Network Science, with a thesis centered around efficient and scalable detection of network motifs. His main research topics are in the fields of advanced algorithms and data structures, network analysis and parallel and distributed computing.

José Luis Oliveira (`http://www.ieeta.pt/~jlo`) holds a PhD from University of Aveiro where is currently an Associated Professor. He is the coordinator of the bioinformatics group of University of Aveiro (http://bioinformatics.ua.pt/) and his main research interests are in the area of distributed systems, information retrieval and biomedical informatics. He has been involved in more than 20 research projects and he has more than 200 publications in books, book chapters, journals and conferences.

Sérgio Nunes (`http://www.fe.up.pt/~ssn`) is an Assistant Professor at the Faculty of Engineering of the University of Porto (FEUP), in Portugal, and a Senior Researcher at the Information and Computer Graphics Systems Unit at INESC TEC. He has a Masters in Information Management (FEUP, 2004) and a PhD in Informatics Engineering (FEUP, 2010) in the field of Information Retrieval, with a dissertation entitled "Information Retrieval on Time-Dependent Collections". His main research interests are in the fields of Information Retrieval, Information Management, Web Technologies, and Multimedia.