

Early Adopter Proposal: NSF/TCPP Core Curriculum Initiative

PDC technology and the analysis of large networks

Niloy Ganguly, Bivas Mitra and Animesh Mukherjee
Department of Computer Science and Engineering,
Indian Institute of Technology, Kharagpur, India -- 721302

Introduction

The last few years have seen an unprecedented activity around supporting and analyzing large-scale social networks. There has been meteoric growth of industries like Facebook, Twitter etc. Parallel and distributed computing technology are being used in their full fledge in the design of novel analysis methods for such large-scale graphs in order to gain both speedup and accuracy. It has been felt that, today, these industries offer the best and the most lucrative jobs and absorb a sheer majority of the undergraduate students in the Computer Sciences. Hence, there is an urgent need to foster an appropriate training programme that would make these students suitable for such vacancies. Therefore, motivated by the job market scenario for the undergraduate students, we present an in-depth curricular structure in the following that would systematically equip and educate them with all the necessary knowledge for a position in these industries. The course structure is in principle categorized into four different modules with the first one being appropriate for the pre-final year students while all the other three qualifying as final year courses. The course structure assumes knowledge of algorithm analysis as a pre-requisite, which also already runs as a core course in the Department. The courses are ordered in a way in which they should be imparted for having the best and the most productive outcomes.

Distributed Algorithms

Course Objective: Objective of this course is to provide a comprehensive understanding about basic distributed algorithms. Moreover, this course will also provide concepts of its usage across different core subjects of computer science.

Course Structure:

General Distributed Algorithms – Logical clock and causal ordering, Global State Collection, Wave and Traversal algorithm, Leader Election Algorithm, Minimal Spanning Tree algorithm, Agreement Protocol

Distributed Operating System – Distributed Mutual Exclusion, Distributed Deadlock Detection

Distributed Networking algorithms – Balanced Sliding Window Protocol, Distributed Routing

Distributed Database Management – Distributed transaction, Commit protocol

Peer-to-Peer Networks and Mobile Computing

Course Objective: The course will deal with various protocols based upon which fixed as well as mobile P2P network functions:

Course Structure:

Real-World P2P Network – Structured, Unstructured and semi-structured p2p networks

Routing in P2P Network -- CAN, Chord, DHT, Incentives for Peer-to-Peer Routing
Search & Replication Strategies -- Distributed Replication – Design and Issues
BitTorrent – Session based distributed file-sharing.
P2P Video-on-Demand: Design motivation and issues.
Mobile P2P – Basics of wireless and mobile technology, issues and challenges, mobility models, JXTA Mobile P2P Architecture
Mobile Phones -- The New Platform for People-Centric Sensing, Social and Personal Applications, Localization, Exploiting Mobility, Location Privacy, Location Dependency/Awareness, Location and Context Awareness

Social Networks

Course Objective: The objective of this course would be to develop a thorough understanding of the concepts of social network analysis, social dynamics, data mining for social networks, parallel programming for analyzing large social networks and human centric computation.

Course Structure:

Social Network Analysis (SNA) -- Types of networks, measurement and collection of network data, structural and locational properties of social network - centrality, prestige, prominence and related concepts, cohesive subgroups

Data Mining and Parallel programming -- Scope of data mining in SNA and the aligned applications. Parallel infrastructure for crawling network data, data mining processes, pattern discovery, semantic web and recommender systems. Parallel implementation of graph analysis (centrality and community analysis) and graph traversal techniques (BFS, DFS).

Human Based Computation -- Classes of human-based computation, methods of human-based computation, incentives for participation, crowdsourcing as a model for problem solving, design of ESP Game.

Social Dynamics -- Opinion dynamics, cultural and language dynamics, crowd behaviour, social spreading phenomena.

Complex Networks (Elective Course)

Course Objective: This course would study the various models and concepts put forward to explain the emergence of large networks all around us.

Course Structure:

Types of network -- Social networks, Information networks, Technological networks, Biological networks.

Properties of network -- Small world effect, transitivity and clustering, degree distribution, scale free networks, maximum degree; network resilience; mixing patterns; degree correlations; community structures; network navigation.

Random Graphs -- Poisson random graphs, generalized random graphs, the configuration model, power-law degree distribution, directed graph, bipartite graph. *Models of network growth* -- Price's model, Barabasi & Albert's model, other growth models, vertex copying models.

Processes taking place on networks -- Percolation theory and network resilience, Epidemiological processes.

Applications -- Search on networks, exhaustive network search, guided network search, network navigation; network visualization, semantic zooming.

Evaluation Plan

The evaluation plan would consist of two parts (a₁). Lab Demonstration – each student has to perform certain lab experiments or (a₂). Term Project – each student perform a mini-research project (this would be done in the elective course of complex networks) and (b). Regular mid and end semester evaluation. The sample Lab demonstration is noted below:

(A₁)Lab Demonstrations

Distributed Algorithms

Create a LAN with 2 systems. Implement balanced sliding widow protocol.

Create a LAN with 10 systems (Arbitrary topology). Implement distributed version of Floyd-Warshall algorithm to build routing tables.

Implement termination detection algorithm for a tree topology LAN with 10 systems.

Peer-to-peer networks

Create a simple peer-to-peer network between two PCs, Identify the proper cable to connect the two PCs, Configure workstation IP address information, Test connectivity using the ping command.

P2p simulators: Narses, 3LS, NeuroGrid, PeerSim , P2PSim, Omnet++

PlanetLab Deployment of a p2p network

Mobile phone programming: Extract data from light sensor, gyroscope, accelerometer, cameras, microphones, GPS, barometer of a smart phone

Basic communication models of smartphones, e.g., WiFi-direct, Bluetooth.

Android programming -- Key Android Concepts, choosing IDE, Basic Widgets.

Social Networks

Developing parallel infrastructure for crawling large network data sets like Twitter, Facebook, DBLP etc.

Sequential and parallel implementation of centrality measures – e.g., PageRank, hub-authority.

Sequential and parallel implementation of clustering methods – e.g., Louvain, Infomap.

Parallelizing multi-agent based simulation methods and distributed AI concepts.

(A₂) Term projects in Complex Networks

A number of mini projects either exploring the dynamical properties of networks or building applications based such properties shall be floated. All these projects shall represent novel research problems each of which would have the potential to qualify as a full-fledged research article. This would range from analyzing (i) social networks -- Twitter, Facebook, movie-actor, (ii) transportation networks – Railways, (iii) scientific citation/collaboration network, (iv) diffusion processes – linguistic diffusion, information diffusion, popularity and viral diffusion, epidemic spreading.

(B) Regular Evaluation:

There will be a mid-term evaluation (a 2-hour written examination) and an end-term evaluation (a 3-hour written examination) and would comprise 50-60% of the total marks.

Showcasing

Department Symposia: A mini symposia will be held in the department once a year where best lab demos will be displayed, the best term projects will be presented in a poster session.

Competition: The students would be especially encouraged to participate in various competitions e.g., Google summer of code etc.

Outreach

Adjunct faculty: For each course, special initiative would be taken by the instructors to have adjunct faculty from industry and research labs who will discuss specialized architectures, protocols used in industry to tackle large graphs and networks.

Web Page: A detailed webpage maintaining all the information as well as lecture notes will be publicly available.

Video Lecture: Government of India has taken special initiative under the aegis of NPTEL to make video lectures of undergraduate/postgraduate courses and disseminate it to the lower-ranked engineering colleges in India. The four courses would accordingly be videoed – IIT Kharagpur has complete infrastructure to support that.

Qualification of the instructors

The three instructors writing this proposal are part of the Complex Network Research Group cse.iitkgp.ac.in/resgrp/cnerg/ which is group specialized in research of large networks. The group has published in top conferences and journals over the years. The three instructors are highly qualified maintaining a lot of collaboration with academia as well as industry around the world.

Budget

Head	Expenses
Equipment and lab setup	\$900
Teacher's compensation	\$900
Adjunct faculty	\$400
Workshop organization	\$300
Total	\$2500