NATIONAL TECHNICAL UNIVERSITY OF ATHENS

Dimitrios Tsoumakos, PhD
Assoc. Professor
Ph: +30-210-7722515
E-mail: dtsouma@mail.ntua.gr
url: http://www.cslab.ece.ntua.gr/~dtsouma

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING
DIVISION OF COMPUTER SCIENCE
Iroon Politechniou 9, Zografou Campus, 157 80, Athens

Marie Skłodowska-Curie Doctoral Network RELAX: 1 PhD Position

Supervisor: Assoc. Prof. Dimitrios Tsoumakos Contact details: dtsouma@mail.ntua.gr
Proposed PhD Project Titles: RELAX Doctoral Network: **Analytics modelling over uncertain and variable data inputs**
Project Introduction: Many companies, across all industry sectors, are increasingly becoming data companies as they collect, curate, and analyse massive amounts of data to increase productivity and cost-effectiveness, or to develop new data-driven products and services. Similarly, governments and administrations world-wide are increasingly reliant upon data science, for instance, in response to the Covid-19 pandemic. Within each application domain, the volume and rate of producing data increases over time, which has a knock-on impact on the power consumed by the data centres, devices and communication networks that drive data analytics. Consequently, the efficiency of data analytics is increasingly important: to scale analytics to increasingly larger and more complex data sets while maintaining low response times, but also to manage the computational requirements of analytics. Data analytics, however, operate in a complex software ecosystem combining a multitude of components to handle computation, storage, resource management, etc. Efficiency cannot be isolated in a single component, nor can it be delivered as a service. On the contrary, efficiency must permeate the system design. By consequence, data analytics systems need to be built as bespoke software systems that are optimised based on a thorough understanding of the full software stack. This requires developing an understanding of the domains where these systems would be used and embedding such understandings within the design of the software systems developed. The RELAX Doctoral Network brings together 5 cross-disciplinary research groups working across data science, data management, distributed computing and computing systems to pursue a fundamentally new approach to this problem by leveraging the semantics or correctness conditions of applications, with the goal of enhancing scalability, response times, and availability. The Doctoral Network provides a bespoke technical and non-technical training programme and fosters cross-disciplinary and third-party collaborations. Further information on the project including opportunities at the partner universities is available on http://www.relax-dn.eu/.
PhD Positions Description: Data volume has been the initial challenge behind data-intensive computational technologies, leading to the deployment of frameworks such as Hadoop, Spark, Flink, Hive, etc. However,

volume is only one dimension to describe what we refer to as "extreme data": An often-overlooked dimension of data volume relates to the increasing abundance of datasets and the requirement to fuse them. High variability both in types (diversity) and values (high speed), uncertainty, location distribution as well as content-related aspects present challenges in converting data to actionable insights. A crucial factor overlooked in previous approaches relates to how "fit for purpose" utilized data is in the context of business operations and analytics – what is referred to as **data quality**. Analysts direly need to focus on high-impact data**,** i.e., intelligence that has the best potential of driving strategic decisions.

We plan on setting a principled path towards linking data quality and analytics performance via: a) creating models that measurably map dataset interrelations based on structural, semantic and operational data aspects for all data types (structured, unstructured and semi-structured inputs) and b) novel analytics modeling that maps performance (in multiple qualitative terms, rather than simply execution time) to the modeled data features.

PhD project: Analytics modelling over uncertain and variable data inputs
Data sources may regularly exhibit varying levels of uncertainty such as noise and missing attributes. The relationship between data uncertainty and its impact on analytics performance is still cryptic. Similarly, data velocity is a regular source of complexity in analytics. In this Thesis, we plan on addressing the challenging aspects of data uncertainty and data churn in view of a content-centric approach; namely, we wish to define and implement appropriate methods and tools that model, maintain in real time and help analysts select multiple data inputs that change in different aspects and/or contain varying degrees of uncertainty.

Expected Results: Develop algorithms and tools that create performance models for analytics tasks over multiple possible inputs that exhibit varying degrees of i) churn and ii) uncertainty.

Project Key Words (Max 100 characters): Big Data Analytics, data quality, machine-learning, uncertainty, optimization, modelling, streaming data

Full-Time X	Part-Time ☐	**Start Date: March-May 2024** **Application Deadline: 14 Feb 2024**
Funding Body: EU MSCA Doctoral Network "RELAX"		Project Funding Type: funded

Funding Information:

To be eligible for consideration for a RELAX Doctoral Candidate position (including a monthly living allowance of EUR 3,270 plus mobility and family allowances scaled by a country-specific correction coefficient), a candidate must satisfy all the eligibility criteria set by both the EU[1] (funding body) and the School of Electrical and Computer Engineering of the National Technical University of Athens. The Studentship is open to all nationalities under the predefined conditions (see below).

Applicants must be **doctoral candidates,** i.e., not already in possession of a doctoral degree at the date of the recruitment (understood as the recruitment call deadlines) and undertake transnational mobility (see mobility rule below). Researchers who have successfully defended their doctoral thesis but who have not yet formally been awarded the doctoral degree will not be considered eligible.

Mobility Rule: Researchers must not have resided or carried out their main activity (work, studies, etc.) in Greece for more than 12 months in the 36 months immediately before their date of recruitment. Compulsory national service, short stays such as holidays, and time spent as part of a procedure for obtaining refugee status under the Geneva Convention are not taken into account.

The positions are strictly full-time and provide RELAX-related funding for a maximum of 3 years (latest date of hiring: March 2024).

Academic Requirements:

The academic requirements/documents for admission as a doctoral candidate at the School of Electrical and Computer Engineering of NTUA are listed here:
https://www.ece.ntua.gr/uploads/announcements/juD4g3DW/%CE%A0%CE%A1%CE%9F%CE%9A%CE%97%CE%A1%CE%A5%CE%9E%CE%97%20%CE%9D%CE%95%CE%A9%CE%9D%20%CE%A5.%CE%94.%20%CE%91%CE%9A%CE%91%CE%94.%20%CE%95%CE%A4%CE%9F%CE%A5%CE%A3%202023-2024.pdf
An overview (but not the complete list!) is detailed here:
(a) All undergraduate and graduate (Master's) degrees/diplomas of the applicant. A Master's degree or equivalent is necessary.
(b) Detailed transcripts for all degrees submitted in step (a). The degree grade must be greater or equal to 7/10 (or equivalent to the School of ECE 7/10 GPA). Alternatively, the candidate must be officially ranked either in the top 33% of their class relative to the undergraduate degree or in the top 20% of their class for their graduate degree (if any).
(c) Curriculum vitae.
(d) A document describing their academic interests.
(e) Certification of knowledge of the English or other foreign language(s).
(f) Copies of published papers, prizes, awards, etc.
(g) Photocopy of their ID/Passport.
(h) A draft PhD thesis plan.

To Apply:
    Contact Prof. Dimitrios Tsoumakos (dtsouma@mail.ntua.gr)