

PhD in Electrical, Electronics and Communications Engineering

Research Title: Fraudulent reviews identification in online reputation systems and their impact on economy

Funded by	Interdepartmental Centre for SmartData (SmartData@PoliTo) Politecnico di Torino
------------------	---

Supervisor	Marco Mellia – mellia@polito.it Luca Dall'Asta - luca.dallasta@polito.it Paolo Neirotti – paolo.neirotti@polito.it
-------------------	---

Contact	http://bigdata.polito.it
----------------	---

Context of the research activity	<p>The web has become the main means of acquiring information. We use it to look for a restaurant, to find a hotel, to keep us informed, to buy goods, to find which movie to watch, etc. Systems such as TripAdvisor.com, Booking.com, Amazon.com or Netflix.com offer the end-users rating information and suggestions, and let people themselves to contribute to increase the information in a crowdsourced and voluntarily means. Technically, these systems collect user-generated-content, that is leveraged to provide recommendations to users. They are referred as recommendation systems, a subclass of information filtering systems that seeks to predict the "rating" or "preference" that a user would give to an item. Among the difficulties in building such systems, the verification of the actual correctness of the information the users enter is among the major problem. Fraudulent reviews have caused controversies and affect the rating policies at base of the recommendation system itself. This clearly affect the actual economic result of a company, so that identifying and filtering anomalous/suspicious reviews is becoming more and more important. How then the reputation of a business entity affects its economic performance in the real world is of major importance, and how fraudulent reviews may change this is even more critical. Considering news, the problem is linked to the so called "fake-news", i.e., unverified news that quickly spread through the web and Social Networks with the intent to mislead to gain financially or politically, often with sensationalist, exaggerated, or patently false headlines that grab attention. Big data and machine learning have emerged as means to extract valuable information from raw data itself. Graph mining includes</p>
---	--

	<p>techniques to model complex relationship between objects, and highlight common or anomalous patterns in graphs. Exploratory algorithms allow one to detect anomalous patterns, which for instance may represent cliques of users that are misusing of a recommendation system.</p> <p>In this context, the research activities of the candidate will focus on the collection of data, its modeling by means of graphs, and usage of big data techniques to highlight anomalies in the data itself. The project also aims at developing an understanding of the transformations at the value chain level of how new big data-based approach produce in the generation and the appropriation of economic value in the tier of intermediation related to news distribution, recommendation systems about travels, etc.</p> <p><i>The research activity fits in the SmartData@PoliTo interdepartmental center, that brings together competences from different fields, ranging from modeling to computer programming, from communications to statistics. The candidate will join this interdisciplinary team of experts and collaborate with them.</i></p>
--	---

Objectives	<p>The goal of the research activities is threefold:</p> <ol style="list-style-type: none"> 1) Design smart data collection crawlers that can automatically collect data from recommendation systems, i.e., designing smart policies to sample the humongous amount of data they potentially expose. 2) Model the data using graph based modeling, and design and test algorithms to automatically identify possible anomalies which reflect eventual pollution in the recommendation system. 3) Correlate the reputation of a business entity with its economic performance in the real world, and building models to predict how the performance will eventually improve based on a change in the reputation on the online recommendation systems. <p>At the end, the candidate is expected to maturate competences in the data acquisition, data modeling, and data analysis, forming a data scientist who can manage big data approaches to extract useful information from the raw data, including the definition of distributed algorithms for inference and learning. The candidate will develop also an understanding of the economic traits of the digital transformations ignited by big data in the value chains of information-based services.</p>
-------------------	---

Skills and competencies for the development of the activity	<p>The candidate is required to have good programming skills, excellent knowledge of how web platforms works, possibly also considering Big Data frameworks. Knowledge of data modeling techniques and methodologies based on graphs and economic background is appreciated but not strictly required.</p>
--	--