Secteur des Sciences
et Technologies

Invitation à la soutenance publique de thèse de

Monsieur Ilkka KIVIMÄKI

Master of Science

Pour l’obtention du grade de Docteur en sciences de l’ingénieur et technologie

« Distances, centralities and model estimation methods based on randomized shortest paths for network data analysis»

qui se déroulera

le lundi 26 novembre 2018 à 15h00
Salle Vaes (A.50)
Place des Doyens, 1
1348 Louvain-la-Neuve

Membres du jury :

Prof. Marco Saerens (UCLouvain), supervisor
Prof. Charles Pecheur (UCLouvain), chairperson
Prof. François Fouss (UCLouvain), secretary
Prof. Michel Verleysen (UCLouvain)
Prof. Gianluca Bontempi (ULB, Belgium)
Dr. Bram Van Moorter (NINA, Norway)

The emergence of networks and network data in different forms in the near past has given rise to development of new data analysis methods with a shift in focus from vector spaces to graph topologies. While some of these methods have been designed specifically for analyzing networks, others have been transformed from more traditional methods, for instance from the machine learning and data mining literature, to function also with networks or graph-based data. The main focus in the thesis is on the recently developed randomized shortest paths (RSP) framework, from which multiple interesting graph-based measures can be derived. These measures can be used either for analyzing networks directly or as tools augmenting existing machine learning or data mining methods. Essentially, the RSP framework defines a Gibbs-Boltzmann probability distribution on the set of paths connecting two nodes on a graph. The distribution focuses mostly on shortest paths but also assigns non-zero probability to longer connections between the two nodes. The extent of this spread of the probability mass is controlled by a temperature parameter, inspired by a thermodynamical analogy. A low temperature enforces the distribution to focus more on shortest paths, whereas a high temperature causes a more spread-out distribution. From this distribution several meaningful quantities can be derived and computed conveniently, which have interesting connections with more traditional network measures that have been defined considering either shortest paths or random walks on the network. As one highlight, the thesis presents the free energy distance, which is a novel metric for measuring the distance between two nodes on a graph, not only based on the shortest path length, but by taking into account all possible paths between the nodes and their lengths or costs. The free energy distance is appealing theoretically in many ways, and it has been shown to provide desirable results in different network analysis tasks, some of which are presented in the thesis. The thesis also studies other graph node distance measures, partly in comparison to the free energy distance. In addition, different measures for quantifying the centrality of a node in a network or the overall functionality of a network are proposed based on the RSP framework. The theory of RSPs is also extended with model estimation methods for fitting the RSP model to a dataset of trajectories on a network. In order to demonstrate and evaluate the derived methods the thesis deals with multiple application domains, ranging from the evaluation of ecological landscapes and quantifying centrality in an urban street network to clustering of text document collections. On the other hand, the methods are developed in a general theoretical setting and also discussed from different viewpoints in order to enable their use in other possible domains.