Missing data completion in wastewater network databases: the added-value of Graph Convolutional Neural Networks.

Yassine Belghaddar, Carole Delenne, Nanée Chahinian, Ahlame Begdouri, Abderrahmane Seriai

HAL Id: hal-03264958
https://hal.science/hal-03264958

Submitted on 18 Jun 2021

Missing data completion in wastewater network databases: the added-value of Graph Convolutional Neural Networks.

Yassine Belghaddar[1,2,3], Carole Delenne[1], Nanée Chahinian[1], Ahlame Begdouri[2], and Abderrahmane Seriai[3]

[1]HSM, Univ Montpellier, CNRS, IRD, Montpellier, France (yassine.belghaddar@usmba.ac.ma)

[2]LSIA, Univ. Sidi Mohamed Ben Abdellah, Fez, Morocco

[3]Berger-Levrault, Pérols, France

Wastewater networks are mandatory for urbanization. Their management, which includes reparation and expansion operations, requires precise information about their underground components, mainly pipes. For hydraulic modelling purposes, the characteristics of the nodes and pipes in the model must be fully known via specific, complete and consistent attribute tables. However, due to years of service and interventions by different actors, information about the attributes and characteristics associated with the various objects constituting a network are not properly tracked and reported. Therefore, databases related to wastewater networks, when available, still suffer from a large amount of missing data.

A wastewater network constitutes a graph composed of nodes and edges. Nodes represent manholes, equipment, repairs, etc. while edges represent pipes. Each of the nodes and edges has a set of properties in the form of attributes such as diameters of the pipes. In this work, we seek to complete the missing attributes of wastewater networks using machine learning techniques. The main goal is to make use of the graph structures in the learning process, taking into consideration the topology and the relationships between their components (nodes and edges) to predict missing attribute values.

Graph Convolutional Network models (GCN) have gained a lot of attention in recent years and achieved state of the art in many applications such as chemistry. These models are applied directly on graphs to perform diverse machine learning tasks. We present here the use of GCN models such as ChebConv to complete the missing attribute values of two datasets (1239 and 754 elements) extracted from the wastewater networks of Montpellier and Angers Metropolis in France. To emphasize the importance of the graph structure in the learning process and thus on the quality of the predictions, GCNs' results are benchmarked against non-topological neural networks. The application on diameter value completion, indicates that using the structure of the wastewater network in the learning process has a significant impact on the prediction results especially for minority classes. Indeed, the diameter classes are very heterogeneous in terms of number of elements with a highly majority class and several classes with few elements. Non-topological neural networks always fail to predict these classes and affect the majority class value

to every missing diameter, yielding a perfect precision for this class but a null one for all the others. On the contrary, the ChebConv model precision is slightly lower (0.93) for the majority class but much higher (increases from 0.3 to 0.81) for other classes, using only the structure of the graphs. The use of other available information in the learning process may enhance these results.