

Thesis
B.Sc.

Thesis
M.Sc.

IDP,
Guided
Research

Automatic Syn- thesis of High Quality Network- and Traf- fic Datasets based on Real Data

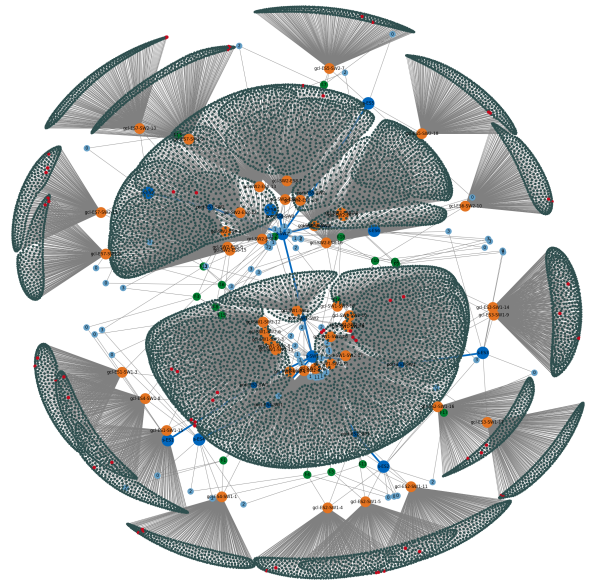
Motivation

Applications such as machine learning require high quality datasets for training to perform in a meaningful way. Machine learning in the context of computer networks therefore requires high quality descriptions of networks. In this context, a network contains information such as topology, node capabilities, traffic characteristics, and more. High quality means that the dataset contains realistic networks with realistic parameters as well a balanced mix of different scenarios. Therefore, basing the synthesis on real world networks [1] is a first starting point.

Based on the usecase these requirements might change. Therefore, the synthesis process should be parameterized to enable generation of data for different usecases.

Topic can be extended to a Master's thesis.

[1] <http://www.topology-zoo.org/>



Your Task

- Identify and classify real world dataset sources
- Generalize datasets (e.g. as probability distributions)
- Automate the process of new dataset synthesis
- Analyze the quality of synthetic datasets based on different metrics

Requirements

- Experience working with Linux and bash (or similar)
- Experience in Python or Go

Contact

Max Helm helm@net.in.tum.de
Benedikt Jaeger jaeger@net.in.tum.de

