Graphs Counterfactual Explainability: A Comprehensive Landscape

Tutorial @ AAAI 2024

M.A. Prado-Romero - B. Prenkaj - G. Stilo

Centro Nazionale di Ricerca in HPC,
Big Data and Quantum Computing

aiim

Mario A. Prado-Romero
PhD. Candidate
Gran Sasso Science Institute

Bardh Prenkaj
PostDoc
Sapienza University of Rome

Giovanni Stilo
Associate Prof.
University of L'Aquila

We are with the *Artificial Intelligence & Information Mining* (aiim - pronounced as i'm /aɪm/, and aim /eɪm/)
a collective of *Individuals* (/aɪm/) who share a common *Interest* (/eɪm/) in
Artificial Intelligence, Data Mining, and Machine Learning

ROADMAP

- **Part I: Introduction and Background (20 mins)** [STILO]
  - Wide-spread adoption of GNNs in graph prediction problems
  - How do GNNs work?
  - Applications of different types of GNNs

- **Part II: XAI on Graphs (30 mins)** [PRENKAJ]
  - Issues of black-box models and the importance of interpretability
  - What is a factual explanation?
  - (Briefly) Revisiting GNNExplaner and GraphLIME

- **Part III: Counterfactual Explainations in Graphs (55 mins)** [PRADO]
  - What is a graph counterfactual explanation (GCE)?
  - GCE taxonomy description and method classification
    - Instance-level explainers
      - Search-based
      - Heuristic-based
      - Learning-based
    - Model-level explainers

  - Benchmarking datasets and evaluation metrics (pro et contra)

Coming Soon

Microsite