A Network Analysis of NIPS Conference Papers

Group №26: Yueran Liang, Peilin Kang, Yawen Hou, Zhechen Su

November 2019

1 Purpose

Machine learning has been a very popular subject since the last decade. The quantity of papers related with various subjects in the field of machine learning has been increasing exponentially over years. In the current project, we aim to explore the network of NIPS papers. Neural Information Processing Systems (NIPS) is one of the top machine learning conferences in the world. We will study the evolution in the relationships between the topics and citations, and the connections between individual researchers.

2 Dataset

The NIPS paper dataset is taken from Kaggle [1]. It covers topics ranging from deep learning and computer vision to cognitive science and reinforcement learning. In terms of structure, the dataset includes the title, authors, abstracts, and extracted text for all NIPS papers from the first 1987 conference to the 2018 conference. This dataset contains 7241 papers in total, involving more than 10 000 authors. In addition, we will also scrape papers from 2018 and 2019.

3 Network

We will convert the dataset into the following hypothesized networks (might change as we explore the dataset):

- Citation Network: Each paper will represents a node in the network. There will be a directed edge from paper A and paper B if paper A cites paper B. We can measure the similarities between the references of each paper, and also pinpoint the super-hubs (very popular papers cited by numerous others).

- Collaboration Network: Each author will represent a node in the network. There will be a undirected edge linking author A and author B if they have collaborated in the redaction of a paper. We will be able to pinpoint the super authors (the ones that produced a phenomenal quantity of papers).

- Topic Network: Extracting the majors topics from a paper is a highly valuable functionality, since when a researcher is looking for papers to support his/her opinion, it is hard to read through each paper and extract the ones related to the topics he/she is interested in. Using LDA and its variants, we will extract topics from the papers, and we will define each paper as a distribution of topics. Each topic will be a node, and we will add a weighted edge between 2 topics that appear in the same paper. The weight will be determined by the distribution of the topics in the paper.

4 Research Questions

- Trends in regards to cited papers and the evolution of their prevalence from 1987 until 2019
- Collaboration network patterns with respect to authors
- First insight into changes over the years with respect to collaborations
- Clusters of the cited documents
- Clusters of topics and their relationship

---

[1] https://www.kaggle.com/benhamner/nips-papers