

Local clustering on graphs

Supervisor(s): Harry Sevi and Argyris Kalogeratos

Contact: <fname.sname@ens-paris-saclay.fr>

Project type: Research

Theme: Graph Machine Learning

Group: Machine Learning and Massive Data Analysis (MLMDA)

Lab: Centre Borelli, ENS Paris-Saclay, Gif-sur-Yvette

Funding: under the auspices of the Industrial Data Analytics & Machine Learning Chair (IdAML)

Description

The increasing size of available datasets makes centralized computation too expensive. Instead of analyzing and inferring the whole structure of a very large graph, local graph clustering methods aim to only find small meaningful clusters. These methods take as input a graph and a seed node, and they return as output a cluster including the seed in a running time that depends on the size of the output cluster, but that is independent of the size of the input graph. Local graph clustering has become an important learning technique for analyzing several large-scale graphs and has been applied to solve other learning and combinatorial optimization problems.

Probably the first to propose a local clustering algorithm was the work in [1], which employed random walks on graphs to explore the local structure close to the seed node. Today, state-of-the-art local clustering algorithms are based on heat kernel PageRank [2]. Similar to the standard [1, 3] heat kernel, the heat kernel PageRank relies also on random walks.

The internship aims to investigate the state-of-the-art local clustering algorithms and potentially propose a novel local graph clustering.

- See references :[1, 2, 3, 4, 5]

Keywords: Clustering network nodes, local methods, random walks, graph embeddings, unsupervised and semi-supervised learning.

Indicative references

- [1] D. A. Spielman and S.-H. Teng, “A local clustering algorithm for massive graphs and its application to nearly linear time graph partitioning,” *SIAM Journal on Computing*, vol. 42, no. 1, pp. 1–26, 2013.
- [2] F. Chung and O. Simpson, “Computing heat kernel pagerank and a local clustering algorithm,” *European Journal of Combinatorics*, vol. 68, pp. 96–119, 2018.
- [3] R. Andersen, F. Chung, and K. Lang, “Local graph partitioning using pagerank vectors,” in *Annual IEEE Symposium on Foundations of Computer Science (FOCS’06)*. IEEE, 2006, pp. 475–486.
- [4] K. Fountoulakis, D. Wang, and S. Yang, “p-norm flow diffusion for local graph clustering,” in *Inter. Conf. on Machine Learning*. PMLR, 2020, pp. 3222–3232.
- [5] K. Scaman, A. Kalogeratos, and N. Vayatis, “A greedy approach for dynamic control of diffusion processes in networks,” in *IEEE Inter. Conf. on Tools with Artificial Intelligence (ICTAI)*. IEEE, 2015, pp. 652–659.

Doors open afternoon – 14 December 15h00-18h00: The members of the Graph Machine Learning team will be happy to welcome the interested students at our lab. We can talk about this subject and the research opportunities at Centre Borelli. Please contact Argyris Kalogeratos to express your interest and allocate a time slot.