Compression and Efficient Processing of Graph Data: From Signal Processing to Deep Learning

▶ **Research theme:** Graph signal processing, graph machine learning, geometric deep learning

▶ **Keywords:** Graph signals, graph summarization, graph neural networks

▶ **Research groups:** Télécom Paris, LTCI, Télécom SudParis, SAMOVAR, Institut Polytechnique de Paris, University of Southern California, STAC Lab

▶ **Advisors:** Jhony H. Giraldo (jhony.giraldo@telecom-paris.fr), Aref Einizade (aref.einizade@telecom-paris.fr), Antonio Ortega (aortega@usc.edu)

▶ **Starting date of PhD and duration**: 3 years starting in Summer/Fall 2026 (September/October)

Context

Graph signal processing (GSP) [1–3] and graph machine learning (GML) [4–6] are research fields that aim to generalize classical concepts from signal processing and machine learning to data defined on graphs. In GSP, classical operations such as filtering, sampling, and reconstruction are extended to graph signals, *i.e.*, functions defined over the vertices of a graph [7,8]. In GML, neural network architectures are adapted to graph-structured data, resulting in models such as graph neural networks (GNNs) [9] and message-passing neural networks (MPNNs) [10]. GSP and GML are crucial because graph-structured data appear in numerous applications, including social networks [3], the web [11], recommender systems [12], biological networks [13], and knowledge graphs [14], among others. A common challenge across these domains is the massive scale of the underlying graphs, which often contain millions or even billions of nodes and edges [15]. Therefore, it becomes essential to obtain compact representations from the original data, such as graphs with fewer nodes and sampled or aggregated graph signals, enabling more efficient learning and processing.

In the GSP literature, a large body of work has addressed the problems of sampling and reconstruction of graph signals [7, 16–18]. These works provide theoretical guarantees for reconstructing graph signals from samples based on the spectral properties of the graph, guiding the design of optimal or near-optimal sampling sets [8, 19, 20]. However, these methods assume the original graph is not modified and consider only the problem of sampling graph signals. In parallel, the GML community has developed principled graph coarsening methods [21], including recent approaches with message-passing guarantees [22, 23]. In graph coarsening, also called graph summarization in some contexts [24], the general objective is to compress a large graph into a smaller, more manageable representation while preserving the information that matters for a downstream task (*e.g.*, visualization, learning, storage). Typically, these approaches only consider topological summarization, whereas we consider both graph and signal summarization.

Establishing a *principled framework for jointly summarizing graph signals and topology* with (*i*) recovery guarantees for the signal and (*ii*) task-level guarantees on the compressed topology remains an open challenge. We address this challenge by estimating *higher-order structures* (*e.g.*, cliques) and using them to induce an aggregation operator and a *coarsened* graph as shown in Figure 1, where maximal cliques are used as the higher-order units. These higher-level structures will enable us (*i*) to summarize signals and reconstruct them with guarantees, and (*ii*) to run GNNs directly on the summarized representation while preserving downstream performance.

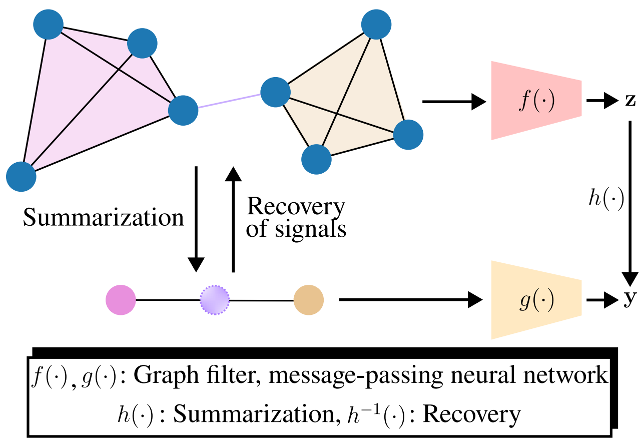

Figure 1: Graph data summarization.

A key question we will study in this PhD thesis is: *When are such higher-order structures informative and practically useful for graph summarization and learning?* We target regimes where cliques are common (*e.g.*, community-like substructures), so that aggregation captures smooth variations in the clique and preserves task performance. However, summarization should be structure-aware; other structures like cycles or paths should be considered with different mapping functions. This problem has broad applications, ranging from graph data compression and storage to efficient processing of relational datasets in domains such as social networks, the web, recommender systems, and biological systems.

Candidate profile

We are looking for candidates:

- Currently holding or finishing an M2 in engineering, data science, computer science, applied mathematics, signal processing, statistics, or equivalent, with a strong background in signal processing and machine learning. The student should have a genuine interest in working in graph signal processing and geometric deep learning.
- Have strong programming skills in Python (including PyTorch).
- Have a genuine interest in understanding the mathematics behind graph signal processing and geometric deep learning (this is a strong requirement).
- Have good communication skills.

Team and location

Télécom Paris and Télécom SudParis are premier engineering schools in France and constituent members of Institut Polytechnique de Paris. The University of Southern California is a leading research university in Los Angeles, California, United States. Both universities are consistently ranked among the best universities worldwide (Shanghai Ranking, QS Ranking). Télécom Paris and Télécom SudParis are located on the outskirts of Paris (around 45 minutes by train from the center of Paris) at the center of the Paris-Saclay cluster—a fast-growing research and industrial ecosystem. The PhD position will be funded by the IMT Futur, Ruptures & Impacts 2026 programme `https://phd.imt.fr`. The student will be integrated within the MM Team at LTCI lab and the ARMEDIA Team at SAMOVAR.

How to apply

Please apply directly through this link including the following:

- A full CV.
- A motivation letter explaining your interest in the position (max 1 page).
- Transcript of records (grades).
- At least one recommendation letter.

The deadline for applying is February 15, 2026.

References

[1] D. I. Shuman, S. K. Narang, P. Frossard, A. Ortega, and P. Vandergheynst, "The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains," *IEEE Signal Processing Magazine*, vol. 30, no. 3, pp. 83–98, 2013.

[2] A. Ortega, P. Frossard, J. Kovačević, J. M. Moura, and P. Vandergheynst, "Graph signal processing: Overview, challenges, and applications," *Proceedings of the IEEE*, vol. 106, no. 5, pp. 808–828, 2018.

[3] A. Ortega, *Introduction to Graph Signal Processing*. Cambridge University Press, 2022.

[4] M. M. Bronstein *et al.*, "Geometric deep learning: Going beyond Euclidean data," *IEEE Signal Processing Magazine*, vol. 34, no. 4, pp. 18–42, 2017.

[5] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu, "A comprehensive survey on graph neural networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 1, pp. 4–24, 2021.

[6] C. Morris, F. Frasca, N. Dym, H. Maron, I. I. Ceylan, R. Levie, D. Lim, M. M. Bronstein, M. Grohe, and S. Jegelka, "Position: Future directions in the theory of graph machine learning," in *ICML*, 2024.

[7] S. Chen, R. Varma, A. Sandryhaila, and J. Kovačević, "Discrete signal processing on graphs: Sampling theory," *IEEE Transactions on Signal Processing*, vol. 63, no. 24, pp. 6510–6523, 2015.

[8] A. Anis, A. Gadde, and A. Ortega, "Efficient sampling set selection for bandlimited graph signals using graph spectral proxies," *IEEE Transactions on Signal Processing*, vol. 64, no. 14, pp. 3775–3789, 2016.

[9] T. N. Kipf and M. Welling, "Semi-supervised classification with graph convolutional networks," in *ICLR*, 2017.

[10] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl, "Neural message passing for quantum chemistry," in *ICML*, 2017.

[11] G. Wang, J. Yu, M. Nguyen, Y. Zhang, S. Yongchareon, and Y. Han, “Motif-based graph attentional neural network for web service recommendation,” *Knowledge-Based Systems*, vol. 269, p. 110512, 2023.

[12] X. Li, L. Sun, M. Ling, and Y. Peng, “A survey of graph neural network based recommendation in social networks,” *Neurocomputing*, vol. 549, p. 126441, 2023.

[13] R. Li, X. Yuan, M. Radfar, P. Marendy, W. Ni, T. J. O’Brien, and P. M. Casillas-Espinosa, “Graph signal processing, graph neural network and graph learning on biological data: A systematic review,” *IEEE Reviews in Biomedical Engineering*, vol. 16, pp. 109–135, 2021.

[14] Z. Ye, Y. J. Kumar, G. O. Sing, F. Song, and J. Wang, “A comprehensive survey of graph neural networks for knowledge graphs,” *IEEE Access*, vol. 10, pp. 75 729–75 741, 2022.

[15] J. Leskovec and A. Krevl, “SNAP Datasets: Stanford large network dataset collection,” http://snap.stanford.edu/data, June 2014.

[16] I. Pesenson, “Sampling in Paley–Wiener spaces on combinatorial graphs,” *Transactions of the American Mathematical Society*, vol. 360, no. 10, pp. 5603–5627, 2008.

[17] M. Tsitsvero, S. Barbarossa, and P. D. Lorenzo, “Signals on graphs: Uncertainty principle and sampling,” *IEEE Transactions on Signal Processing*, vol. 64, no. 18, pp. 4845–4860, 2016.

[18] A. G. Marques, S. Segarra, G. Leus, and A. Ribeiro, “Sampling of graph signals with successive local aggregations,” *IEEE Transactions on Signal Processing*, vol. 64, no. 7, pp. 1832–1843, 2016.

[19] G. Puy, N. Tremblay, R. Gribonval, and P. Vandergheynst, “Random sampling of bandlimited signals on graphs,” *Applied and Computational Harmonic Analysis*, vol. 44, no. 2, pp. 446–475, 2018.

[20] Y. Tanaka, Y. C. Eldar, A. Ortega, and G. Cheung, “Sampling signals on graphs: From theory to applications,” *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 14–30, 2020.

[21] A. Loukas, “Graph reduction with spectral and cut guarantees,” *Journal of Machine Learning Research*, vol. 20, no. 116, pp. 1–42, 2019.

[22] A. Joly and N. Keriven, “Graph coarsening with message-passing guarantees,” *NeurIPS*, 2024.

[23] A. Joly, N. Keriven, and A. Roumy, “Taxonomy of reduction matrices for graph coarsening,” in *NeurIPS*, 2025.

[24] Y. Liu, T. Safavi, A. Dighe, and D. Koutra, “Graph summarization methods and applications: A survey,” *ACM Computing Surveys*, vol. 51, no. 3, 2018.