Before Graph attention networks (GATs), convolutional neural networks (CNNs) and recursive neural networks (RNNs) were commonly used for tasks involving grid-like data structures. However, these architectures were not well-suited for data represented in irregular domains such as 3D meshes, social networks, or biological networks. Early attempts to extend neural networks to handle graph-structured data focused on recursive approaches, but they didn’t have the ability to handle cyclic, directed, and undirected graphs efficiently.

The introduction of GNNs marked a significant advancement, allowing for the direct processing of various types of graphs. Spectral approaches, such as those utilizing the graph laplacian, initially showed promise but faced challenges with computational complexity and dependence on graph structure. Non-spectral approaches emerged as an alternative, operating directly on the graph and enabling weight sharing similar to CNNs. GATs represent a developed version of GNNs, which has better computational efficiency, flexibility in neighborhood aggregation, and independence from upfront knowledge of the entire graph structure. By employing masked self-attention, GATs can assign varying levels of importance to nodes within a neighborhood, enhancing model capacity and performance.

Graph Attention Networks (GATs) utilize masked self-attention layers to efficiently process graph-structured data. Central to GATs is the attention layer, which calculates attention coefficients to measure the relevance of neighboring nodes features for each node. This attention mechanism allows nodes to selectively aggregate information from their neighbors, assigning varying weights to different nodes within the neighborhood. GATs excel in computational efficiency due to their ability to run in parallel manner, enabling simultaneous self-attention and feature computation across all nodes and edges. This efficiency, combined with the ability to handle diverse neighborhood sizes and assign distinct importances to nodes, makes GATs different and better from spectral-based approaches. Leveraging these mechanisms, GATs achieve state-of-the-art performance in node classification tasks on both transductive and inductive benchmarks, demonstrating their effectiveness in analyzing and modeling graph-structured data.

Experimental evaluations have demonstrated the effectiveness of GATs across a range of benchmarks, including both transductive and inductive learning tasks. State-of-the-art performance has been achieved on well-established datasets, highlighting the potential of GATs for real-world applications.