

BAB I

PENDAHULUAN

1.1. Latar Belakang

Jaringan Saraf Graf atau *Graph Neural Networks* (GNN) telah menjadi teknologi utama dalam aplikasi pembelajaran mesin dengan data berstruktur graf. GNN dapat digunakan untuk klasifikasi titik, prediksi tautan, dan sistem rekomendasi [1]. Kemampuan GNN untuk memproses dan mempelajari pola relasional yang kompleks menjadikannya model yang sangat kuat untuk menganalisis berbagai sistem di dunia nyata seperti jaringan sosial, sistem biologi, dan jaringan interaksi lainnya [2]. Meskipun GNN menunjukkan potensi besar, skalabilitasnya masih menjadi tantangan signifikan seiring dengan bertambahnya ukuran graf. Representasi graf di dunia nyata seringkali melibatkan miliaran titik dan triliunan sisi, sehingga menyebabkan beban komputasi yang berat selama pelatihan dan inferensi GNN [3]. Mekanisme *message passing* yang menjadi inti GNN, terutama pada operasi *Sparse Matrix Multiplication* (SpMM) menjadi sangat mahal secara komputasi pada graf yang padat atau berukuran besar [4]. *Overhead* pada komputasi ini menghambat adopsi GNN dalam aplikasi yang membutuhkan pemrosesan cepat atau memiliki keterbatasan sumber daya.

Salah satu pendekatan utama untuk mengatasi tantangan ini adalah sparsifikasi graf [5]. Sparsifikasi bertujuan untuk mengurangi jumlah sisi dalam graf dengan tetap mempertahankan properti struktural atau fungsional yang penting [6]. Dengan mengurangi kepadatan graf, sparsifikasi dapat secara signifikan mempercepat waktu komputasi GNN dan mengurangi penggunaan memori. Selain itu, penghapusan sisi yang redundan atau *noisy* juga berpotensi meningkatkan kinerja generalisasi model GNN [7]. Berbagai teknik sparsifikasi telah dikembangkan, namun mayoritas bersifat *edge-centric* yang mengevaluasi dan memangkas sisi secara individual berdasarkan kriteria tertentu. Contoh dari teknik sparsifikasi saat ini termasuk sampling acak [8], metode berbasis konektivitas lokal [9], [10], sparsifier berbasis cut [11], dan sparsifier spektral yang menjaga spektrum Laplacian graf [12], [13]. Meskipun efektif dalam beberapa skenario, metode *edge-*

centric ini mungkin tidak secara eksplisit mempertahankan struktur penting yang berpusat pada titik-titik paling berpengaruh dalam graf.

Observasi ini memotivasi pergeseran paradigma ke arah pendekatan sparsifikasi yang *vertex-centric*. Alih-alih menilai setiap sisi, pendekatan ini mengidentifikasi sekumpulan titik penting yang membentuk basis atau *backbone* struktural graf, dan kemudian mempertahankan konektivitas lokal di sekitar backbone tersebut. Konsep himpunan dominasi (*dominating set*) dalam teori graf [14] menyediakan kerangka kerja matematis yang alami untuk mengidentifikasi backbone ini. Variasi seperti *Connected Dominating Set* (CDS) telah digunakan dalam membentuk virtual backbones untuk *routing* di jaringan komunikasi [15]. Penelitian ini mengusulkan metode sparsifikasi baru yang memanfaatkan konsep dominasi, yaitu *k*-Weighted Dominating Set (*k*-WDS) [16]. Pemanfaatan *k*-WDS memungkinkan penggabungan bobot sisi dan penggunaan nilai *k* sebagai *threshold* yang dapat disesuaikan. Konsep ini menawarkan fleksibilitas lebih dalam mendefinisikan tingkat "kepentingan" titik dan tingkat dominasi yang diperlukan. Metode yang diusulkan dalam penelitian ini bernama Sparsifikasi *k*-WDS. Metode ini bekerja dengan dua tahap, yaitu menemukan himpunan *k*-WDS dengan melakukan aproksimasi menggunakan algoritma heuristik *Iterated Greedy* (IG), dan mempertahankan semua sisi yang insiden ke titik di *k*-WDS. Pendekatan ini bertujuan untuk mengurangi kompleksitas komputasi GNN dengan tetap menjaga informasi struktural penting yang terkonsentrasi di sekitar titik-titik dominan. Fokus utama penelitian ini adalah menganalisis jaminan konektivitas yang diberikan oleh metode sparsifikasi ini dan mengevaluasi dampaknya terhadap efisiensi pelatihan serta akurasi GNN pada tugas klasifikasi titik.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana rancangan dan implementasi algoritma sparsifikasi graf menggunakan konsep *k*-WDS?

2. Bagaimana jaminan teoretis yang diberikan oleh metode sparsifikasi k -WDS terhadap properti konektivitas graf hasil sparsifikasi?
3. Bagaimana perbandingan waktu pelatihan dan akurasi model GNN pada tugas klasifikasi titik terhadap penerapan sparsifikasi k -WDS sebagai langkah pra-pemrosesan?

1.3. Batasan Masalah

Penelitian ini dibatasi pada aspek-aspek berikut:

1. Analisis teoretis difokuskan pada pembuktian properti konektivitas graf hasil sparsifikasi. Analisis mendalam terhadap properti spektral tidak termasuk dalam cakupan utama.
2. Implementasi dalam menentukan k -WDS menggunakan algoritma heuristik *Iterated Greedy* (IG) seperti yang diusulkan oleh [16].
3. Evaluasi eksperimental terbatas pada dataset *citation network* standar (Cora dan PubMed) dan arsitektur GNN tipe GCN (*Graph Convolutional Network*).

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Merancang dan mengimplementasikan algoritma Sparsifikasi k -WDS sebagai metode sparsifikasi graf yang bersifat *vertex-centric*.
2. Menganalisis dan membuktikan secara teoretis bagaimana metode sparsifikasi k -WDS dapat mempertahankan konektivitas graf.
3. Mengetahui hasil perbandingan waktu pelatihan dan akurasi GNN setelah mengintegrasikan sparsifikasi k -WDS pada langkah pra-pemrosesan.

1.5. Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Manfaat Teoretis:
 - a) Memberikan kontribusi pada bidang teori graf terapan dengan menjembatani konsep optimasi kombinatorial (k -WDS) dengan masalah praktis dalam pembelajaran mesin (sparsifikasi graf).

- b) Menyediakan analisis formal pertama mengenai jaminan properti konektivitas dari konsep sparsifikasi *vertex-centric* berbasis himpunan dominasi.
2. Manfaat Praktis:
- a) Menawarkan metode sparsifikasi baru yang dapat disesuaikan (*tunable* via parameter k) sebagai langkah pra-pemrosesan untuk GNN.
 - b) Memberikan wawasan empiris mengenai trade-off antara tingkat sparsifikasi, efisiensi komputasi, dan akurasi model GNN pada graf dunia nyata.

1.6. Sistematika Penulisan

Tesis ini disusun dalam empat bab utama dengan rincian sebagai berikut:

1. Bab I Pendahuluan
Menguraikan latar belakang masalah, urgensi penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan tesis secara keseluruhan.
2. Bab II Tinjauan Pustaka dan Landasan Teori
Menyajikan kajian literatur terkait GNN, sparsifikasi graf, dan teori dominasi, serta memaparkan dasar-dasar teori graf, GNN, k -WDS, dan algoritma IG yang relevan.
3. Bab III Hasil dan Pembahasan
Mendeskripsikan metode penelitian yang meliputi rancangan algoritma Sparsifikasi k -WDS, analisis teoretis (properti k -WDS, jaminan konektivitas, kompleksitas), dan menyajikan hasil evaluasi eksperimental pada dataset Cora dan PubMed beserta pembahasannya.
4. Bab IV Penutup
Merangkum kesimpulan utama dari penelitian dan memberikan saran untuk pengembangan di masa mendatang, serta menyertakan pernyataan publikasi hasil penelitian.