

## Abstrak

Pemodelan data menggunakan *graph* telah diterapkan oleh banyak aplikasi dan sistem berskala besar dalam berbagai bidang. Data tersebut direpresentasikan sebagai *graph* dengan *node* yang mewakili sebuah objek dan *edge* menandakan hubungan antara dua objek. Untuk memahami karakteristik sebuah *graph*, maka dibutuhkan teknik *graph summarization*.

Pada tugas akhir ini, digunakan metode CANAL (*Categorization of Attributes with Numerical Values based on Attribute Values and Link Structures of Nodes*) untuk meringkas *graph*. Metode ini merupakan pengembangan dari metode *Aggregation-Based Graph summarization* yang melakukan peringkasan dengan mengelompokkan serta menggabungkan *node* kedalam sebuah *super node* dengan menggali pengetahuan dari data untuk menemukan *cutoff* yang digunakan dalam pengelompokan *node* secara otomatis. Metode CANAL memperbaiki metode *graph summarization* SNAP dan k-SNAP yang masih mempunyai kelemahan dalam menangani data dengan atribut numerik. Kedua metode tersebut hanya dapat menangani *categorical node attribute*, sehingga ketika dihadapkan dengan atribut numerik pengguna masih harus melakukan pengelompokan secara manual berdasarkan pengetahuan mereka terhadap data yang digunakan.

Hasil dari sistem yang akan dibangun merupakan sebuah *graph summary* yang merepresentasikan *pattern* hubungan antar kelompok dalam ringkasan. *Pattern* tersebut dapat digunakan untuk membantu memahami informasi yang tersembunyi didalam *graph* asli. Dari ringkasan yang dihasilkan oleh metode CANAL kemudian dinilai kualitasnya dan dibandingkan dengan kualitasnya dengan ringkasan yang berasal dari *cutoff* manual. Perbandingan tersebut menunjukkan bahwa kualitas ringkasan dari CANAL memiliki kualitas baik yang setara dengan kualitas ringkasan dengan *cutoff* manual.

**Kata kunci:** *graph summarization, Aggregation-Based Graph summarization, node attribute, link structure, interestingness measure*