On Compressing Social Networks
Flavio Chierichetti, et al.

This paper examines the properties of social networks that make them well-suited for compression and then explores how techniques developed for web graph compression can be adapted for social networks. The authors' goal was to store large social networks efficiently while still allowing fast adjacency queries. The authors analyze the web graph compression method introduced in the Permuting Web and Social Graphs paper which achieved a storage efficiency of three bits per edge. The authors also highlight the importance of node ordering in improving compression results, focusing on properties like locality and similarity in social networks, specifically the concept of backlinks compression allowing them to leverage mutual links found in social networks to improve compression. Then, the authors conduct an analysis of experiments, including LiveJournal and Flickr, comparing the performance of different node orderings. The authors assume that social networks have high reciprocity (if one user follows another, typically they are followed back by that user). They also assume that node ordering can yield large savings after compression, particularly due to optimized adjacency queries. One limitation of their backlinks compression scheme is that it can slow down adjacency queries due to long prototype chains. One possible extension noted by the authors is using alternative graph traversal strategies or hybrid compression techniques that consider graph dynamics over time. A key takeaway from the paper is that social networks, while similar in sparsity to web graphs, have distinct structural properties that require modifications to current compression techniques, e.g., reciprocity in social networks could offer crucial savings in compression which may not be a big player in web graphs. The approach successfully leverages the unique properties of social networks to improve compression but comes with trade offs in query efficiency. Future work could explore dynamic compression methods that adapt to evolving social networks rather than relying on static graphs.

Permuting Web and Social Graphs
Paolo Boldi, et al.

This paper explores how the ordering of nodes in a graph influences its ability to be compressed and proposes new strategies for both web and social graph ordering. The authors' goal is to identify orderings that improve the rates of compression while maintaining efficient adjacency queries. They analyze different ordering strategies such as lexicographic sorting, Gray code orderings, and some mixed approaches in order to determine which provide the best tradeoff between compression and query efficiency. There is a substantial difference in the properties of web graphs vs. social graphs, namely web graphs are hierarchically meaning that host-based sorting works well, whereas social networks are more reciprocal and therefore require different techniques. The authors propose hybrid orderings such as combining host-based sorting with lexicographic or Gray ordering, achieving significant compression improvements in both types of graphs. The authors assume that graph compression efficiency is highly dependent on node ordering and that different graph types require distinct ordering strategies. Additionally, the authors assume that static ordering is sufficient for compression, which may not be the case for dynamic graphs where nodes and edges change frequently, like what we observe in practice. One limitation of the paper is that it primarily evaluates static graphs, failing to account for real-world social networks. One possible extension to the paper is to develop adaptive ordering strategies that adjust dynamically as the graph structure changes over time, addressing this limitation. The results of the study indicate that ordering is indeed important in graph compression and that different graph structures require different optimization strategies.

Discussion Questions

Compressing social networks and web graphs are similar in that they both involve storing sparse graphs efficiently by exploiting structural properties such as locality and similarity to achieve better compression. Both rely on smart node ordering to improve adjacency list representations and reduce storage space. However, they differ in that web graphs are more hierarchical, meaning host-based ordering techniques are more effective, while social networks tend to exhibit reciprocity, making bidirectional link compression more useful. Social networks also tend to have denser local structures. The role of having an underlying model in determining how to compress these types of structures is critical because it allows compression techniques to exploit predictable patterns in the data. The most important properties of these models include locality, ensuring that nodes that interact frequently are stored near each other, similarity which allows nodes with similar neighborhoods to be grouped together, reciprocity, and hierarchical structure.