Wiley Online Library



Original Article

Summarizing and understanding large graphs

Danai Koutra, U Kang, Jilles Vreeken, Christos Faloutsos

First published: 18 May 2015
https://doi.org/10.1002/sam.11267
Cited by: 7

About
Access

Abstract

How can we succinctly describe a million - node graph with a few simple sentences? Given a large graph, how can we find its most “important” structures, so that we can summarize it and easily visualize it? How can we measure the “importance” of a set of discovered subgraphs in a large graph? Starting with the observation that real graphs often consist of stars, bipartite cores, cliques, and chains, our main idea is to find the most succinct description of a graph in these “vocabulary” terms. To this end, we first mine candidate subgraphs using one or more graph partitioning algorithms. Next, we identify the optimal summarization using the minimum description length (MDL) principle, picking only those subgraphs from the candidates that together yield the best lossless compression of the graph—or, equivalently, that most succinctly describe its adjacency matrix.

Our contributions are threefold: (i) : we provide a principled encoding scheme to identify the vocabulary type of a given subgraph for six structure types prevalent in real - world graphs, (ii) : we develop VoG, an efficient method to approximate the MDL - optimal summary of a given graph in terms of local graph structures, and (iii) : we report an extensive empirical evaluation on multimillion - edge real graphs, including Flickr and the Notre Dame web graph.

Citing Literature

About Wiley Online Library
Help & Support

Copyright © 1999-2018 John Wiley & Sons, Inc. All rights reserved

WILEY