

Research Article

A framework for clustering massive graph streams*

Charu C. Aggarwal , Yuchen Zhao, Philip S. Yu

First published: 22 September 2010

<https://doi.org/10.1002/sam.10090>

Cited by: 3

* Submission to Best of SDM 2010 Issue.



About



Access

ols



ire

Abstract

In this paper, we examine the problem of clustering massive graph streams. Graph clustering poses significant challenges because of the complex structures which may be present in the underlying data. The massive size of the underlying graph makes explicit structural enumeration very difficult. Consequently, most techniques for clustering multidimensional data are difficult to generalize to the case of massive graphs. Recently, methods have been proposed for clustering graph data, though these methods are designed for static data, and are not applicable to the case of graph streams. Furthermore, these techniques are especially not effective for the case of massive graphs, since a huge number of distinct edges may need to be tracked simultaneously. This results in storage and computational challenges during the clustering process. In order to deal with the natural problems arising from the use of massive disk-resident graphs, we propose a technique for creating *hash-compressed microclusters* from graph streams. The compressed microclusters are designed by using a hash-based compression of the edges onto a smaller domain space. We provide theoretical results which show that the hash-based compression continues to maintain bounded accuracy in terms of distance computations. Since clustering is a data summarization technique, it can also be naturally extended to the problem of evolution analysis. We provide experimental results which illustrate the accuracy and efficiency of the underlying method. Copyright © 2010 Wiley Periodicals, Inc. *Statistical Analysis and Data Mining* 3: 399 - 416, 2010

Citing Literature



About Wiley Online Library



Help & Support



Opportunities



Connect with Wiley



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

WILEY