

A Survey on Distributed Graph Pattern Matching in Massive Graphs

SARRA BOUHENNI, Ecole nationale Supérieure d'Informatique, Algérie and Université de Lyon, Université Lyon 1, LIRIS CNRS, France

SAÏD YAHIAOUI and NADIA NOUALI-TABOUDJEMAT, CERIST, Centre de Recherche sur l'Information Scientifique et Technique, Algérie

HAMAMACHE KHEDDOUCI, Université de Lyon, Université Lyon 1, LIRIS CNRS, France

Besides its NP-completeness, the strict constraints of subgraph isomorphism are making it impractical for graph pattern matching (GPM) in the context of big data. As a result, relaxed GPM models have emerged as they yield interesting results in a polynomial time. However, massive graphs generated by mostly social networks require a distributed storing and processing of the data over multiple machines, thus, requiring GPM to be revised by adopting new paradigms of big graphs processing, e.g., Think-Like-A-Vertex and its derivatives. This article discusses and proposes a classification of distributed GPM approaches with a narrow focus on the relaxed models.

CCS Concepts: • **Theory of computation** → **Distributed algorithms; Graph algorithms analysis; Computing methodologies** → *Distributed algorithms*;

Additional Key Words and Phrases: Graph pattern matching, distributed graphs, graph simulation, subgraph isomorphism

ACM Reference format:

Sarra Bouhenni, Saïd Yahiaoui, Nadia Nouali-Taboudjemat, and Hamamache Kheddouci. 2021. A Survey on Distributed Graph Pattern Matching in Massive Graphs. *ACM Comput. Surv.* 54, 2, Article 36 (February 2021), 35 pages.

<https://doi.org/10.1145/3439724>

1 INTRODUCTION

With the rapid growth of the Internet, huge amounts of data are being generated with every passing minute. The challenges brought about by big data have led to the research community revising traditional techniques and algorithms that were designed for collecting, storing, analyzing, and

This work is supported by the Franco-Algerian collaborative program PHC Tassili BiGreen N°18 MDU 111 and by the DGRSDT grant FNRSDT N°253.

Authors' addresses: S. Bouhenni, Ecole nationale Supérieure d'Informatique, BP M68, Oued Smar, 16309, Algérie, Université de Lyon, Université Lyon 1, LIRIS CNRS, Villeurbanne, 69622, France; email: cs_bouhenni@esi.dz; S. Yahiaoui and N. Nouali-Taboudjemat, CERIST, Centre de Recherche sur l'Information Scientifique et Technique, Ben Aknoun, 16030, Algérie; emails: {syahiaoui, nnouali}@cerist.dz; H. Kheddouci, Université de Lyon, Université Lyon 1, LIRIS CNRS, Villeurbanne, 69622, France; email: hamamache.kheddouci@univ-lyon1.fr.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

© 2021 Association for Computing Machinery.

0360-0300/2021/02-ART36 \$15.00

<https://doi.org/10.1145/3439724>