Catching Numeric Inconsistencies in Graphs

WENFEI FAN, University of Edinburgh & BDBC, Beihang University & SICS, Shenzhen University
XUELI LIU, College of Intelligence and Computing, Tianjin University
PING LU, BDBC, Beihang University
CHAO TIAN, Alibaba Group

Numeric inconsistencies are common in real-life knowledge bases and social networks. To catch such errors, we extend graph functional dependencies with linear arithmetic expressions and built-in comparison predicates, referred to as numeric graph dependencies (NGDs). We study fundamental problems for NGDs. We show that their satisfiability, implication, and validation problems are Σ_2^p-complete, Π_2^p-complete, and coNP-complete, respectively. However, if we allow non-linear arithmetic expressions, even of degree at most 2, the satisfiability and implication problems become undecidable. In other words, NGDs strike a balance between expressivity and complexity. To make practical use of NGDs, we develop an incremental algorithm IncDect to detect errors in a graph G using NGDs in response to updates ΔG to G. We show that the incremental validation problem is coNP-complete. Nonetheless, algorithm IncDect is localizable, i.e., its cost is determined by small neighbors of nodes in ΔG instead of the entire G. Moreover, we parallelize IncDect such that it guarantees to reduce running time with the increase of processors. In addition, to strike a balance between the efficiency and accuracy, we also develop polynomial-time parallel algorithms for detection and incremental detection of top-ranked inconsistencies. Using real-life and synthetic graphs, we experimentally verify the scalability and efficiency of the algorithms.

CCS Concepts: • **Information systems → Inconsistent data**; **Data cleaning**;

Additional Key Words and Phrases: Numeric errors, graph dependencies, incremental validation

ACM Reference format:
Wenfei Fan, Xueli Liu, Ping Lu, and Chao Tian. 2020. Catching Numeric Inconsistencies in Graphs. *ACM Trans. Database Syst.* 45, 2, Article 9 (June 2020), 47 pages.
https://doi.org/10.1145/3385031

Fan is supported in part by ERC 652976, Royal Society Wolfson Research Merit Award WRM/R1/180014, EPSRC EP/M025268/1, Shenzhen Institute of Computing Sciences, and Beijing Advanced Innovation Center for Big Data and Brain Computing. Lu is supported in part by NSFC 61602023. Liu is supported in part by NSFC 61902274.
Authors' addresses: W. Fan, University of Edinburgh & BDBC, Beihang University & SICS, Shenzhen University, 10 Crichton Street, Edinburgh, UK, EH8 9AB; email: wenfei@inf.ed.ac.uk; X. Liu (corresponding author), College of Intelligence and Computing, Tianjin University, 135 Yaguan Road, Jinnan District, Tianjin, China, 300350; email: xueli@tju.edu.cn; P. Lu, BDBC, Beihang University, 37 Xue Yuan Road, Haidian District, Beijing, China, 100191; email: luping@buaa.edu.cn; C. Tian, Alibaba Group, 969 West Wen Yi Road, Yu Hang District, Hangzhou, China, 311121; email: tianchao.tc@alibaba-inc.com.

0362-5915/2020/06-ART9 $15.00
https://doi.org/10.1145/3385031