

Constraint-based Explanation for Incomplete Graphs

1st Qi Song

Amazon.com

qison@amazon.com

2nd Peng Lin

Washington State University

peng.lin@wsu.edu

3rd Hanchao Ma

Case Western Reserve University

hxm382@case.edu

4th Yinghui Wu

Case Western Reserve University

yxw1650@case.edu

Abstract—This paper proposes *constraint-based explanation* for graphs, a novel approach that capitalizes on a set of graph data constraints Σ to explain missing data and query answers for incomplete graphs. An explanation is a sequence of operational enforcement of Σ that consistently leads to informative instantiations towards the recovery of missing elements (*e.g.*, attribute values, edges). We show that constraint-based explanations allow us to understand why data is missing, explore how to recover missing elements, and diagnose failure of data constraints. We provide a formal representation for the semantics of graph incompleteness under constraints, and formulate the problem of Σ -explanation, which computes optimal explanations with guarantees on the informativeness. We show the decision problem of Σ -explanation is a complement of corresponding validation problem, and showcase of relativizing upper bounds (*e.g.*, in Δ_2^P for graph keys). We develop feasible algorithms for Σ -explanation in support of online scenarios and missing query answers. Using real-world graphs and benchmarks, we experimentally verify the effectiveness and efficiency of our algorithms.

Index Terms—Graphs, Data Constraints, Data Provenance

REFERENCES