

Query-Oriented Summarization of RDF Graphs

Šejla Čebirić¹, François Goasdoué², and Ioana Manolescu¹

¹ INRIA & U. Paris-Saclay, France ² U. Rennes 1 & INRIA, France

Abstract. The Resource Description Framework (RDF) is the W3C’s graph data model for Semantic Web applications. We study the problem of RDF graph summarization: given an input RDF graph \mathbf{G} , find an RDF graph $\mathbf{S}_\mathbf{G}$ which summarizes \mathbf{G} as accurately as possible, while being possibly orders of magnitude smaller than the original graph. Our approach is *query-oriented*, i.e., querying a summary of a graph should reflect whether the query has some answers against this graph. The summaries are aimed as a help for query formulation and optimization. We introduce two summaries: a *baseline* which is compact and simple and satisfies certain accuracy and representativeness properties, but may oversimplify the RDF graph, and a *refined* one which trades some of these properties for more accuracy in representing the structure.