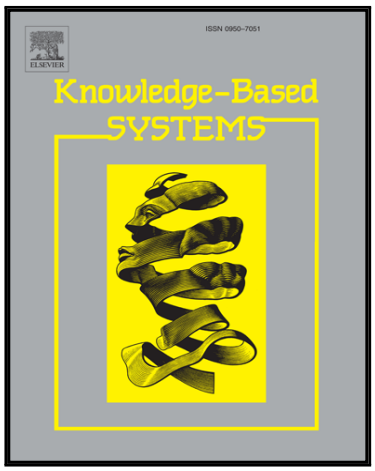

Forward Backward Similarity Search in Knowledge Networks

Baoxu Shi, Lin Yang, Tim Weninger

PII: S0950-7051(16)30484-1
DOI: 10.1016/j.knosys.2016.11.025
Reference: KNOSYS 3749

To appear in: *Knowledge-Based Systems*

Received date: 9 August 2016
Revised date: 28 November 2016
Accepted date: 30 November 2016

Please cite this article as: Baoxu Shi, Lin Yang, Tim Weninger, Forward Backward Similarity Search in Knowledge Networks, *Knowledge-Based Systems* (2016), doi: 10.1016/j.knosys.2016.11.025

Forward Backward Similarity Search in Knowledge Networks

Baoxu Shi

Department of Computer Science and Engineering, University of Notre Dame, 384E Nieuwland Science Hall, Notre Dame, IN 46556 USA

Lin Yang

Department of Computer Science and Engineering, University of Notre Dame, 326E Cushing Hall Hall, Notre Dame, IN 46556 USA

Tim Weninger*

Department of Computer Science and Engineering, University of Notre Dame, 353 Fitzpatrick Hall, Notre Dame, IN 46556 USA

Abstract

Similarity search is a fundamental problem in social and knowledge networks like GitHub, DBLP, Wikipedia, etc. Existing network similarity measures are limited because they only consider similarity from the perspective of the query node. However, due to the complicated topology of real-world networks, ignoring the preferences of target nodes often results in odd or unintuitive performance. In this work, we propose a dual perspective similarity metric called Forward Backward Similarity (FBS) that efficiently computes topological similarity from the perspective of both the query node and the perspective of candidate nodes. The effectiveness of our method is evaluated by traditional quantitative ranking metrics and large-scale human judgement on four large real world networks. The proposed method matches human preference and outperforms other similarity search algorithms on community overlap and link prediction. Finally, we demonstrate top-5 rankings for five famous researchers on an academic collaboration network to illustrate how our approach captures semantics more intuitively than other approaches.

Keywords: knowledge graph, similarity measures, graph search

1. Introduction

Computing the similarity of two or more objects in an information network is the main focus of a large amount of scientific research and technological development. Friendship recommendation in social networks is one example, but web search, community detection, general link prediction, list augmentation, and dozens of other application areas are all singularly dependent upon some notion of similarly in the underlying networks.

Similarity is multi-faceted; various traits can be used to determine similarity depending on the specific problem domain. Entire fields of research are dedicated to the development of algorithms that effectively and efficiently retrieve objects similar to some query-object, *e.g.*, information retrieval, computer vision, and databases (broadly speaking). Researchers and practitioners understand that network topology plays a critical role in the identification of object

*Corresponding author
Email addresses: bshi@nd.edu (Baoxu Shi), lyang5@nd.edu (Lin Yang), tweninge@nd.edu (Tim Weninger)