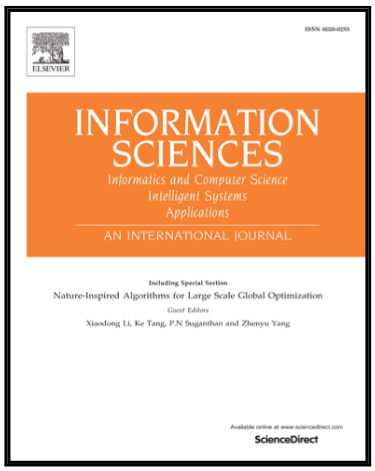

Black Hole Metric: Overcoming the PageRank Normalization Problem

M. Buzzanca, V. Carchiolo, A. Longheu, M. Malgeri, G. Mangioni

PII: S0020-0255(18)30044-6
DOI: 10.1016/j.ins.2018.01.033
Reference: INS 13387

To appear in: *Information Sciences*

Received date: 28 February 2017
Revised date: 10 January 2018
Accepted date: 14 January 2018

Please cite this article as: M. Buzzanca, V. Carchiolo, A. Longheu, M. Malgeri, G. Mangioni, Black Hole Metric: Overcoming the PageRank Normalization Problem, *Information Sciences* (2018), doi: 10.1016/j.ins.2018.01.033

Black Hole Metric: Overcoming the PageRank Normalization Problem

M. Buzzanca[a], V. Carchiolo[a], A. Longheu[a], M. Malgeri[a], G. Mangioni[a]

[a]*Dipartimento di Ingegneria Elettrica Elettronica e Informatica (DIEEI)*
Università degli Studi di Catania, Italy

Abstract

In network science, there is often the need to sort the graph nodes. While the sorting strategy may be different, in general sorting is performed by exploiting the network structure. In particular, the metric PageRank has been used in the past decade in different ways to produce a ranking based on how many neighbors point to a specific node. PageRank is simple, easy to compute and effective in many applications, however it comes with a price: as PageRank is an application of the random walker, the arc weights need to be normalized. This normalization, while necessary, introduces a series of unwanted side-effects. In this paper, we propose a generalization of PageRank named Black Hole Metric which mitigates the problem. We devise a scenario in which the side-effects are particularily impactful on the ranking, test the new metric in both real and synthetic networks, and show the results.

Keywords: Pagerank metric, social networks, trust

1. Introduction

In the vast amount of digital data, humans have the need to discriminate those relevant for their purposes to effectively transform them into useful information, which usefulness depends on the scenario being considered. For instance, in web searching we aim at finding significant pages with respect to an issued query [36], in an E-learning context we look for useful resources within a given topic [9, 40, 41], or in a recommendation network we search for most reliable entities to interact with [15, 14, 21, 4]. All these situations fall under the umbrella of *ranking*, a challenge addressed in these years through different solutions. The most well-known technique is probably the PageRank algorithm [8, 32], originally designed to be the core of the Google (*www.google.com*) web search engine. Since it was published it has been analyzed [34, 6, 24], modified or extended for use in other contexts [46, 19], to overcome some of its limitations, and to address computational issues [25, 35].

PageRank has been widely adopted in several different application scenarios. In this paper, we propose a generalization of PageRank whose motivation stems from the concept of trust in virtual social networks. In this context trust is generally intended as a measure of the assured reliance on a specific feature of someone [28, 16, 1], and it is exploited to rank participants in order to discover the *best* entities that is "safe" to interact with. This