Chapter 15
An Investigation for CA-Based PageRank Validation in View of Power-Law Distribution of Web Data to Enhance Trustworthiness and Safety for Green Cloud

Arnab Mitra
https://orcid.org/0000-0002-7382-089X
Siksha 'O' Anusandhan (Deemed), India

ABSTRACT

An investigation on cellular automata (CA)-based validation of PageRank with reference to the power-law distribution is presented in this chapter to enhance the trustworthiness and safety of Clouds. Web traffic data for several Clouds were analyzed in view of power-law distribution to explore whether they are natural or manmade. Results obtained for CA-based PageRank validation were compared with Alexa®, which further supported the power-law distribution with some cut-offs, which ensured effectiveness and accuracy for CA-based validation of PageRank. Hence, this exploration helps to enhance the trustworthiness for any computed PageRank, and thus, it helps for the automated decision-making process towards an enhanced trust and safety of Clouds.

1 INTRODUCTION

Automated Decision making has a major impact in today's technology-enhanced life: starting from finance sector (e.g., online loan approval) to assessments in online recruitment scenarios (in "https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/automated-decision-making-and-profiling/what-is-automated-individual-decision-making-and-

DOI: 10.4018/978-1-7998-3246-1.ch015

profiling/", accessed on September 10, 2019). Thus, one can easily recognize that the impact of WWW (World Wide Web). It is not difficult to realize that on-line resources are playing a crucial role in the development of our knowledge as they offer easy and mobility access to resources. The resources available at several web servers are often considered as Clouds (Mell et al., 2011). Hence, "health information on the World Wide Web is of variable quality, methods are needed to assist consumers to identify health websites" (Griffiths et al., 2005).

In an online knowledge acquisition scenario, a user intended to collect information over the Clouds often requires the help of a search engine to scroll down the available relevant Clouds (web servers). The search engine provides an index of sorted Clouds based on information relevancy with respect to the user's query. The search engine generated index of Clouds is commonly known as PageRank-ed list. By convention, low the index number refers to high the PageRank. Several dynamic and static PageRank computing algorithms have been offered by researchers and practitioners to improve the trustworthiness of computed PageRank (Pasquinelli, 2009; Kundu et al., 2006; Guha et al., 2015). Though, the major emphasis is found to be present towards the computations of PageRank (Pasquinelli, 2009; Kundu et al., 2006; Guha et al., 2015 and 2016), validations of computed PageRank are also found to be present towards the enhanced reliability and safety (Mitra et al., 2015 and 2017; Mitra, 2019). We found that PageRank validation may result to enhance the reliability of computed PageRank and further may enhance web security (Mitra, 2019).

In this regard, we feel that it is not out of scope to mention that advancements in today's Cloud Computing technologies have further transformed the earlier days' Clouds into the Green Clouds (Baliga et al., 2011; Berl et al., 2010; and Patel et al., 2015). Thus, with synchronization with emerging Green Clouds, an emphasis towards energy-based PageRank calculation (Guha et al., 2015) and modeling of low energy consuming PageRank validation (Mitra et al., 2017 and Mitra, 2019) have also been presented by researchers in recent times.

On the other hand, Power-Law is known as a functional relationship among two measures, where a change in one measure consequences in a proportional adjustment in the other, and it is independent of the first size of those measures (i.e., one measure differs as a power of another). Power-Law distributions may be observed in several natural systems; it shows the capability of a system towards self-organizing when it is continually pushed to a disordered state by any "natural perturbations (Barabási 2003)" (Logan et al., 2013). Further, it may be noted that finest diversity, which is acute for resilience, may also be found from the Power-Law patterns (Logan et al., 2013).

For this reason, Power-Law has appealed for certain interest over years, which often leads to the exploration of amazing physical consequences, "and for its appearance in a diverse range of natural and man-made phenomenon" (Clauset et al., 2009). As described in (Clauset et al., 2009), an amount x conforms a Power-Law if it is drawn from a probability distribution $p(x) \propto x^{-\alpha}$, where $\pm$ is known as exponent / scaling constant. The exponent / scaling parameter (α) typically follows the range $2 < \alpha < 3$, although there might be occasionally exceptions. In run-through, only a few empirical phenomena follow Power-Laws for all values of x. Most frequently, the Power-Law is applicable for values greater than some maximum of x_{min}. In those scenarios, it is considered that the conclusion of the distribution confirms a Power-Law (Clauset et al., 2009).

Several cases for natural and man-made scenarios were analyzed and presented with reference to Power-Law distribution in (Clauset et al., 2009). Besides different uses of Power-Law distribution targeting Cloud and or IoT were presented by several researchers over time (Girau, 2017; Darwish et al., 2018;