Evaluating the Impact of Articles with Geographical Distances between Institutions

Xiaomei Bai[1,2], Jie Hou[1], Hongzhuang Du[1], Xiangjie Kong[1] and Feng Xia[1]
[1]School of Software, Dalian University of Technology, Dalian 116620, China
[2]Computing Center, Anshan Normal University, Anshan 114007, China
xjkong@acm.org

ABSTRACT

Evaluating the impact of scholarly papers plays an important role for addressing recruitment decision, funding allocation and promotion, etc. Yet little is known how actual geographic distance influences the impact of scholarly papers. In this paper, we leverage the law of geographic distance and citations between different institutions to weight quantum Pagerank algorithm for objectively measuring the impact of scholarly papers. The results indicate that the weighted quantum PageRank algorithm can better differentiate the impact of scholarly papers compared to PageRank algorithm.

Keywords

Scholarly big data; Article impact; Quantum PageRank

1. INTRODUCTION

Evaluating the impact of scholarly papers is a significant research topic. It plays a crucial role for scholars' promotion and funding allocation.

Measuring the impact of scholarly papers can be divided into three categories: citations, citations-based structured metrics, and Altmetrics [1]. Citations are the most influential metric for evaluating the impact of scholarly papers. Citations-based structured metrics mainly consider the structured features in scholarly networks such as citation network, co-authors network, and paper-author network, and so on. For example, PageRank algorithm is the earliest used to measure the impact of scholarly papers. The importance of each node in citation network is closely related to the importance of its each adjacent node. Altmetrics mainly focus on quantifying the scholarly impact based on the activities in social media platforms such as downloads, mentions, views, etc.

However, little is known how actual geographic distance influences the impact of scholarly papers. In some previous researches, citations were regarded as equally important regardless of the influence of actual citation distance between institutions. It is a challenge for us to find that the amount of citations varies with geographical location. In this paper, we mainly explore the relationship between citations and geographical distances between institutions. According to the citation law, we construct the weighted quantum PageRank algorithm to objectively measure the impact of scholarly papers.

WWW'17 Companion, April 3–7, 2017, Perth, Australia.
ACM 978-1-4503-4914-7/17/04.
http://dx.doi.org/10.1145/3041021.3053061

2. METHOD

2.1 Data description

The American Physical Society (APS) dataset contains all the papers published in Physical Review. It spans across 9 different journals: Physical Review A, B, C, D, E, I, L, ST and Review of Modern Physics, from 1970 to 2013. Each paper includes its title, names, every author's affiliations, date of publication and a list of cited papers. Our experiments are based on the PRC dataset, which is a subset of APS dataset.

2.2 Citation weight

The relationship between citations and actual geographical distances between institutions follows an exponential distribution dependence.

$$z = z_0 + ce^{\frac{-x}{t}} \tag{1}$$

where z represents citations, z_0 represents the initial value, x represents actual geographical distances, c is a constant and t represents time. According to the citation law, we construct a weighed quantum PageRank algorithm. The weight is defined as a relative citation weight, which is a ratio between a geographical distance and the maximum geographical distance.

2.3 Weighted quantum PageRank

Based on the quantum PageRank algorithm [2], we define a weighted quantum PageRank algorithm. First, the initial state of our quantum walk is

$$|\phi_0\rangle := \sum_{j=0}^{N-1} |\phi_j\rangle \tag{2}$$

where ϕ_0 represents the initial state, ϕ_j represents the important degree of each node in the quantum network. The initial state of each node is constructed by the weighted PageRank scores. Second, we construct a general transform matrix

$$U = \pi S \tag{3}$$

where U represents the general transform matrix, S is the swap operator, and π is the unitary matrix. In order to objectively quantify the impact of scholarly papers, we use the average importance of each node to represent its score of paper impact.

$$\langle Q_{i,m} \rangle := \frac{1}{M} \sum_{m=1}^{M} Q_{i,m} \tag{4}$$

where $\langle Q_{i,m} \rangle$ is the final score of each node in citation network. M represents the repeat number of weighted PageRank algorithm.

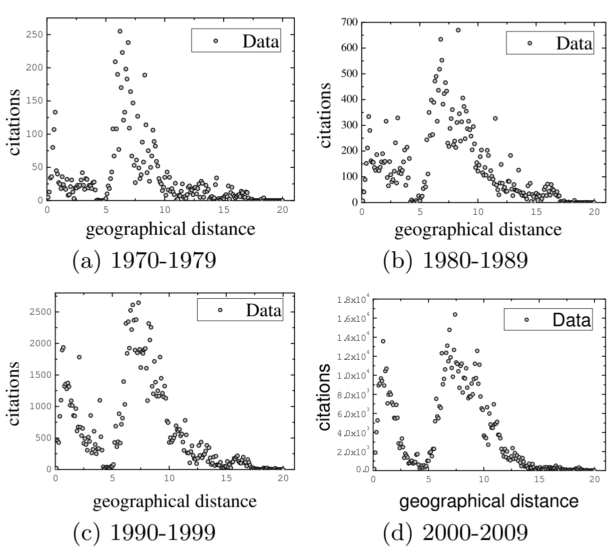
3. RESULTS

3.1 Citation trends

Figure 1 shows the citation trends in different periods. We can observe that citations fast decrease from 0 Mm to 5 Mm, then sustainedly increase and reach the peak at around 7Mm. Following citations rapidly decline from 7 Mm to 20 Mm. The fluctuated citation trend drives us to explore its reason. We find that the uneven geographical distribution between institutions brings about the fluctuated citation trend. Atlantic ocean separates North America from Europe that results into the formation of the peak.

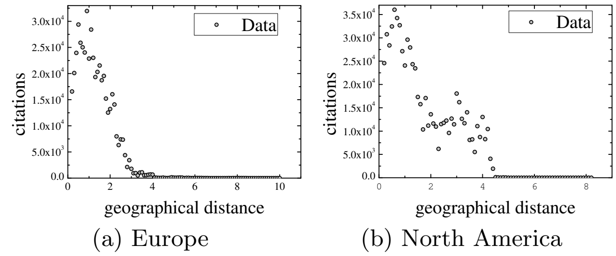
Figure 2 shows the relationship between citations and actual geographical distances. We can observe that there are similar citation trends in Europe and North America. According to Figure 2a and Figure 2b, we can find that more citations are, closer geographical distances between different institutions are.

Figure 3 shows the citation trends without the ocean. We can observe that the changed trend of citations is similar to the citation trends in Europe and North America. It follows an exponential distribution dependence. The citation law is used to weight quantum PageRank to objectively measure the impact of scholarly papers.



(a) 1970-1979 (b) 1980-1989 (c) 1990-1999 (d) 2000-2009

Figure 1: Comparing the citation trend in different periods.



(a) Europe (b) North America

Figure 2: Comparing the citation trend between Europe and North America.

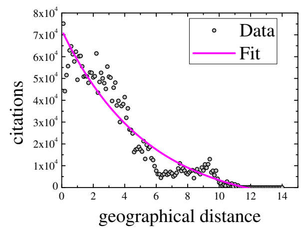


Figure 3: The relationships between citations and actual geographical distances.

3.2 Comparing PageRank and weighted quantum PageRank

Tabel 1 shows the scores of PageRank and weighted quantum PageRank. We have omitted the same part "10.1103/PhysRevC." about the DOI of papers in the Table 1. According to Tabel 1, we can observe that weighted quantum PageRank can better distinguish the impact of scholarly papers compared to PageRank.

Table 1: Comparing between PageRank and weighted quantum PageRank

DOI	PageRank	Weighted quantum PageRank
55.540	0.0007856	0.0001911
16.629	0.0015125	0.0001753
63.024001	0.0003945	0.0001784
53.R1483	0.0003945	0.0001213
53.2809	0.0003945	0.0001023

4. CONCLUSIONS

The weighted quantum PageRank algorithm can better distinguish the impact of papers compared to PageRank algorithm. The weighted quantum PageRank algorithm can be calculated by parallel way that brings a faster speed on conducting the experiments.

5. REFERENCES

[1] M. K. Cheung. Altmetrics: too soon for use in assessment. *Nature*, 494(7436):176–176, 2013.

[2] G. D. Paparo, M. Müller, F. Comellas, and M. A. Martin-Delgado. Quantum google algorithm. *The European Physical Journal Plus*, 129(7):1–16, 2014.