

An Efficient PageRank Approach for Urban Traffic Optimization

Florin POP¹, Ciprian DOBRE¹

¹Faculty of Automatic Control and Computers, University Politehnica of Bucharest, Splaiul Independentei 313, Bucharest 060042, Romania

Email: F. Pop - florin.pop@cs.pub.ro, C. Dobre* - ciprian.dobre@cs.pub.ro;

*Corresponding author

Abstract

The cities are not static environments. They change constantly. When we talk about traffic in the city, the evolution of traffic lights is a journey from mindless automation to increasingly intelligent, fluid traffic management. In our approach, presented in this paper, reinforcement learning mechanism based on cost function is introduced to determine optimal decisions for each traffic light, based on the solution given by Larry Page for page ranking in Web environment [1]. Our approach is similar with work presented in [2] and [3]. We consider that the traffic lights are controlled by servers and a score for each road are computed based on efficient PageRank approach and is used in cost function to determine optimal decisions. We demonstrate that the cumulative contribution of each car in the traffic respects the main constrain of PageRank approach, preserving all the properties of M -matrix considering in our model.

Keywords: Optimization, PageRank, Intelligent Traffic Lights, Learning Mechanisms, Traffic Control

1 Introduction

The PageRank model [1] uses the structure of the Web to build a Markov chain considering a nonnegative and irreducible matrix M for transition probability (called also a *primitive* matrix). This property guarantees that, according to Frobenius test for primitivity [4], a stationary vector exists as a solution for the PageRank problem.

In our approach, the main role of a computer server that will serve a traffic light is to compute the congestion cost of different monitored road segments. The congestion is a real effect of dynamic and unpredictable behavior of drivers and the traffic optimization (in terms of cost reduction) will contribute to solve partially the congestion problem. We consider a directed graph having streets as nodes and introduce the *centrality* of a road as a measure of popularity of the road on different path (a more central road will be part of more driving routes). The two roads are connected if they are part in the same route. This is similar to how Web pages are connected. In the Web model, in general, the pages are connected, so we could introduce the density in order to measure number of connection related to number of resources. In the traffic model, the road density (considered as number of roads in a city region) is consequently higher and hence more jams. The road density is translated in our model to graph density which has an effect in graph representation (the graph matrix is not sparse).

The urban growth patterns are considered in [5] as development units correlated and able to reproduce the observed morphology of cities and the area distribution of sub-clusters in an urban system. Another model try to map out fractal models of urban growth [6, 7]. Peterson concluded that *“such a model emphasizes how uncoordinated local decision making can generate global patterns that give cities their particular sizes and shapes. It suggests that a city’s overall structure can emerge naturally out of individual actions without any coordinating authority”*.

In this paper we prove that an adapted PageRank approach over the proposed graph offers a stationary solution for the computation of centrality road-segments as a new mathematical applied problem in engineering. Also, we prove that the cumulative effect of each car that offers information about traffic preserves the properties of the M -matrix.

2 The PageRank Model for the Urban Traffic

The transition probability matrix M is build from the structure of the city, to be stochastic and primitive. The Markov model represents the graph with a square matrix P whose element p_{ij} is the probability of moving from state i (road i) to state j (road j , directed connected with road i) in a time-step [8]. In the M matrix some rows can contain all zeros, which is the natural assumption for cities (a specific route is composed of some road-segments). We will replace all zeros rows with $\frac{1}{n}e^T$, where e^T is the row vector of all ones and n is the order of the matrix (the number of roads). Therefore, we will compute a new nonnegative and irreducible matrix $M \in R^{n \times n}$ as follows:

$$M(i, j) = \begin{cases} \frac{p_{ij}}{\sum_{k=1}^n p_{ik}}, & \sum_{k=1}^n p_{ik} \neq 0 \\ \frac{1}{n}, & \sum_{k=1}^n p_{ik} = 0 \end{cases}$$

Now, considering $g \in R^n$, a column vector that contains the centrality scores of all roads, the PageRank mathematical problem can be written as $g^T (\alpha M + \frac{1-\alpha}{n} e e^T) = g^T$, where $0 \leq \alpha \leq 1$. For the urban traffic model we can change the uniform vector $\frac{1}{n} e^T$ with the more general vector v^T (*teleportation vector*), $v(i)$ representing the probability of road i to be considered part of a route. We will define $\bar{M} = \alpha M + (1-\alpha) e v^T$. It is clear that this change makes \bar{M} irreducible and aperiodic: all nodes are connected (all roads are connected in cities where the graph of roads is always connected). The additional normalization equation $g^T e = 1$ ensures that g^T is a probability vector, so it can be used in order to compute the congestion cost. The PageRank problem is now: $g^T \bar{M} = g^T$. The row vector g^T can be found by solving the eigenvector problem $g^T \bar{M} = g^T$, or by solving the homogeneous linear system $g^T (\bar{M} - I_n) = 0^T$.

Theorem 1 *Let's consider the general problem $g^T \bar{M} = g^T$, where $\bar{M} = \alpha M + (1-\alpha) e v^T$ for a specific teleportation vector v and a specific α ($0 \leq \alpha \leq 1$). Then, the solution of the problem $g^T \bar{M} = g^T$ can be written as:*

$$g^T = (1-\alpha) v^T \sum_{k=0}^{\infty} \alpha^k \cdot M^k.$$

Proof. We will start with the initial problem:

$$g^T = g^T \bar{M} = g^T (\alpha M + (1-\alpha) e v^T) = \alpha g^T M + (1-\alpha) g^T e v^T = \alpha g^T M + (1-\alpha) v^T$$

where we have used $g^T e = 1$ since $\|g\|_1 = 1$. Hence, $g^T (I_n - \alpha M) = (1-\alpha) v^T$ and so

$$g^T = (1-\alpha) v^T (I_n - \alpha M)^{-1} = (1-\alpha) v^T \sum_{k=0}^{\infty} \alpha^k \cdot M^k.$$

□

Theorem 2 (*Convex Combinations of Teleportation Vectors*): *For any convex combination of q teleportation vectors, $v^T = \sum_{i=1}^q a_i v_i^T$, with $a_i \geq 0$ and $\sum_{i=1}^q a_i = 1$, we have the convex combination:*

$$g_v^T = \sum_{i=1}^q a_i g_{v_i}^T.$$

Proof.

$$\sum_{i=1}^q a_i g_{v_i}^T = \sum_{i=1}^q \left((1-\alpha) a_i v_i^T \sum_{k=0}^{\infty} \alpha^k \cdot M^k \right) = (1-\alpha) \left(\sum_{i=1}^q a_i v_i^T \right) \left(\sum_{k=0}^{\infty} \alpha^k \cdot M^k \right) = (1-\alpha) v^T \sum_{k=0}^{\infty} \alpha^k \cdot M^k = g_v^T.$$

□

So, a convex combination ($a_i \geq 0$) of teleportation vectors will lead to the same convex combination of PageRank vectors. This also demonstrate that the convex cumulative contribution of each car and traffic light in the traffic follows the main constrain of PageRank approach, preserving all the properties of the M -matrix. For our system we make the hypothesis that cars are equipped with wireless devices that run a mobile navigation application. Such a mobile application uses the device's sensors (e.g., GPS sensor, accelerometer) and constantly sends the sensed traffic conditions (travelling speed, position) to a computer server (the car's contribution). For the driver, the incentive for using such an application might consist in the dynamic adjustment of his/her traveling path, based on the returned feedback.

An important result was obtained related to representation of a stochastic traffic bound. The main result prove that the burstiness bound and the bound of long-term average rate are separately connected to the fractal dimension that is the measure of the local self-similarity together with the small-scale factor and the Hurst parameter that is the measure of the long-range dependence together with the large-scale factor a of traffic [9]. Having this bound, the basis to reliably distinguish flood traffic from aggregated one is reliable detection of signs of DDOS flood attacks. Reliably distinguishing DDOS flood traffic from aggregated traffic becomes a tough task mainly due to the effects of flash-crowd traffic [10].

2.1 Convergence

Considering the power method to solve the PageRank problem, we have the following:

$$g_1^T = g_0^T \overline{M}, g_2^T = g_1^T \overline{M}, \dots, g_N^T = g_{N-1}^T \overline{M}$$

Applying this gives $g_N^T = g_0^T \overline{M}^N$. Because \overline{M} is irreducible and aperiodic, $\lim_{N \rightarrow \infty} g_0^T \overline{M}^N = g^T$, for all initial approximation of centrality vectors g_0 .

Theorem 3 (*Fast PageRank*): *For any approximation of centrality vectors g_0^T , the power method, $g_N^T = g_0^T \overline{M}^N$ will converge.*

Proof. If N denotes the number of steps then we can approximate g^T and considering ϵ the precision threshold, we have:

$$\begin{aligned} g_1^T &= g_0^T (\alpha M + (1 - \alpha) e v^T) = \alpha g_0^T M + (1 - \alpha) v^T \\ &\dots \\ g_N^T &= g_{N-1}^T (\alpha M + (1 - \alpha) e v^T) = \alpha g_{N-1}^T M + (1 - \alpha) v^T. \end{aligned}$$

We can approximate $g^T = g_N^T$ with the precision ϵ , thus we have:

$$\epsilon = (1 - \alpha) \sum_{K=N+1}^{\infty} \alpha^k = \alpha^{N+1}.$$

So, having $0 \leq \alpha \leq 1$ we conclude that the method converges. \square

Another important result that sustains the convergence of this method it represented by the following theorem:

Theorem 4 *Given the spectrum of the stochastic and irreducible matrix M is $\{1, \lambda_2, \dots, \lambda_n\}$, the spectrum of the primitive stochastic matrix $\bar{M} = \alpha M + (1 - \alpha)ev^T$ is $\{1, \alpha\lambda_2, \dots, \alpha\lambda_n\}$.*

Proof. The proof is give in [8], Theorem 5.1 \square

This theorem shows that if all λ_i values are lower than 1, than the new values $\alpha\lambda_i$ are also lower than 1, so the convergence of the method is preserved for the \bar{M} matrix. Therefore, it is important to have a near-optimal model, and convergence plays an important role here because it leads to fewer iterations to be executed until the algorithm reaches a solution for a centrality vector (g) in PageRank problem.

2.2 Stability

We will discuss the stability of the proposed method as an analysis of centrality of the vector g^T , considering the dependency $\bar{M} = \bar{M}(\alpha)$ and $g^T = g^T(\alpha)$. We have the following theorem:

Theorem 5 *If $g^T(\alpha) = [g_1(\alpha), g_2(\alpha), \dots, g_n(\alpha)]$ is a centrality vector, then*

$$\left| \frac{dg_i(\alpha)}{d\alpha} \right| \leq \frac{1}{1 - \alpha}$$

and

$$\left\| \frac{dg^T(\alpha)}{d\alpha} \right\| \leq \frac{2}{1 - \alpha}.$$

Proof. The proof is give in [8], Theorem 7.2 \square

This theorem makes it apparent that the sensitivity of the centrality vector as a function of α is primarily governed by the size of $\frac{1}{1-\alpha}$. If α is close to 1, then centrality is sensitive to small changes in α . As α becomes smaller, the effects of the probability given by v^T are increased. In the urban traffic model this is not an accurate situation, because the teleportation vector is computed based on car behavior and depends of its context. So, it is more desirable (at least in this respect) to choose α close to 1 (between [0.75, 0.90], or 0.85 like in the Web model [1]). However, if α is too close to 1, then the method will be unstable, and the convergence rate slows.

Theorem 6 We have: $K_\infty(I_n - \alpha\bar{M}) = \frac{1+\alpha}{1-\alpha}$.

Proof.

$$K_\infty(I_n - \alpha\bar{M}) = \|(I_n - \alpha\bar{M})\|_\infty \cdot \|(I_n - \alpha\bar{M})^{-1}\|_\infty.$$

The ∞ -matrix norm is the maximum absolute row sum. If the node in the traffic model has a positive number of links (this is true for all cities), then the corresponding diagonal element is $I_n - \alpha\bar{M}$ is 1. Then, the sum to α is in absolute value, so $\|(I_n - \alpha\bar{M})\|_\infty = 1 + \alpha$. Now, we will prove that the row sums of $(I_n - \alpha\bar{M})^{-1}$ are $\frac{1}{1-\alpha}$. Because of $(I_n - \alpha\bar{M})e = (1 - \alpha)e$ the row sums of $(I_n - \alpha\bar{M})$ are $1 - \alpha$. Based on [4] and [11] we have that the $(I_n - \alpha\bar{M})$ is a positive and irreducible matrix, so $\|(I_n - \alpha\bar{M})^{-1}\|_\infty = \frac{1}{1-\alpha}$. So, $K_\infty(I_n - \alpha\bar{M}) = \frac{1+\alpha}{1-\alpha}$. \square

In the ideal case K must be 1, but, in this case α will be 0. If α is too close to 1, then the method will be unstable because $\lim_{\alpha \rightarrow 1} K_\infty(I_n - \alpha\bar{M}) = \infty$. For the proposed interval $\alpha \in [0.75, 0.90]$ we have $K_\infty(I_n - \alpha\bar{M}) \in [7, 19]$, so the matrix $I_n - \alpha\bar{M}$ is hard conditioned. For a Web value, $\alpha = 0.85$, we have $K_\infty = \frac{37}{3} \approx 12.33$.

3 Decomposition of Traffic Model

Random paths can be observed in different routes in the cities. It is possible to decompose the path in different smaller parts, so $g = \sum_{i=1}^r \frac{\omega_i}{c_i} g_i$ where ω_i is a timed weight for the part i and c_i is the number of outbound links of the part i . On the other hand, the matrix \bar{M} can be written as $\bar{M} = \sum_{i=1}^r b_i M_i$, where r is the number of different smaller parts (or a range) of a route. The decomposition of the routes for the driver has the following significance: more routes means more possibilities to choose from, so the driver is constantly facing the challenge of choosing the route that best fits its own preferences, which is not always easy to do, especially considering the ever-changing dynamic traffic conditions in large cities.

In the decomposition model there are two suppositions: (i) g is the PageRank solution for each M_i matrix, so $g^T = g^T M_i$ which means that the centrality of the roads is kept in the decomposition model or (ii) g_i is the PageRank solution for each M_i matrix, so $g_i^T M_i = g_i^T$ which means that the centrality of the roads is not kept in the decomposition model and is computed as a temporal dependency by the formula $g = \sum_{i=1}^r \frac{\omega_i}{c_i} g_i$.

Theorem 7 The decomposition model with $\bar{M} = \sum_{i=1}^r b_i M_i$ and conservative centrality $g^T = g^T M_i, \forall i$, of the PageRank problem is stable.

Proof.

$$\begin{aligned} g^T &= g^T \left(\alpha \sum_{i=1}^r b_i M_i + (1 - \alpha) e v^T \right) = \alpha \sum_{i=1}^r b_i g^T M_i + (1 - \alpha) v^T = \\ &= \alpha \sum_{i=1}^r b_i g^T + (1 - \alpha) v^T = g^T \left(\alpha \sum_{i=1}^r b_i I_n + (1 - \alpha) e v^T \right). \end{aligned}$$

So, the $(I_n - \alpha \bar{M})$ matrix becomes $(I_n - \alpha \sum_{i=1}^r b_i I_n) = (1 - \alpha \sum_{i=1}^r b_i) I_n$ which is a diagonal matrix with non-zero diagonal elements. So, the stability of the method is proven. \square

For the second case (ii), the global teleportation vector can be deduced as:

$$v^T = \frac{1}{1 - \alpha} \sum_{i=1}^r \left(\frac{\omega_i}{c_i} - \alpha b_i \right) g_i^T,$$

which highlights the possibility for decomposition of a global teleportation vector. This result makes it easy to generate new centrality vector (g) from existing and computed centrality vectors. This could be applied for router nodes placement problem in Wireless Mesh Networks designed for traffic light monitoring [12].

Also, the problem of the stabilization of descriptor systems in continuous-time via static output-feedback is important for monitoring phase in order to collect the p_{ij} elements of matrix P used for the computation of matrix M . Sufficient conditions are derived for the closed-loop system to be admissible (i.e., stable, regular, and impulse-free) and are proposed in [13].

4 Discussions

The new area of research is oriented on next generation data technologies in support of collective and computational intelligence [14]. Next generation of data technologies are used together to capture, integrate, analyze, mine, annotate and visualize distributed data made available from various community users in a meaningful and collaborative for the organization manner. Also, in this context the processing and command synthesis for memory-limited complex systems are very important, considering the limitation of traffic lights [15].

The European Commission's report "*European transport policy for 2010: time to decide*" recognizes the important role played by the transport infrastructure in the modern European economies and identifies a number of major European policy goals that are expected to help streamline existing transport systems [16]. In particular, the report shows that the development of new infrastructures is not anymore a reliable solution as well as investing in technologies based on intelligent transport systems (ITSs), as predictions show that in the future traffic congestions can be solved mainly by technology. It is estimated that only in Europe

traffic congestion affects approximately 10% of the existing transport network, with significant financial implications. According to the conclusion of the analysis given by Gilles Duranton and Matthew Turner [17] “*increased provision of interstate highways and major urban roads is unlikely to relieve congestion of these roads*”. Thus, it becomes apparent that the way to alleviate congestion and utilize the existing infrastructure more efficiently are ITSs.

An intelligent transport systems can be seen as a cyber-physical networking systems which are made up of various physical systems that are heterogeneous in nature. The data processing and analysis in these systems is the most important aspect for solving difficult problems like traffic congestion. Li and Zhao demonstrate that the power laws may yet be a universality of data in these systems, a useful point of view for data modeling [18,19]. If we consider the periodic trends, like traffic evolution patterns, the long-range power-law cross-correlation can be applied. Diverse methods can be generalized to eliminate the trend effects, such as moving average, singularity value decomposition and can be used for detecting the cross-correlation of traffic signals, like in metropolis where the traffic jam is becoming a puzzle for citizens [20].

We envision a society with IT systems that do respond to the actual people’s needs. In particular, the theoretical model can inspire the development of solutions that will impact the domain of transportation. Today phones can be used to collect information (e.g., evidence about waste, noise), and mobilize people together for a common goal. Systems can further use the sensed information to construct high-level models that enable people better understand various social and environmental phenomena. The stage for these phone-centric sensing campaigns was recently set up by various projects targeting various domains. For traffic, congestion alone can severely impact both the environment and human productivity (because of wasted hours waiting in congestions). Traffic jams cost Europe alone €135 billion a year, and drivers lose five days per year sitting in traffic. This is why mobile phone systems such as MIT VTrack or Mobile Millennium are already used to provide fine-grained traffic information on a large scale, using mobile phones to collect traffic data, pinpoint specific problems, and facilitate services such as accurate travel time estimation for improved commute planning. We believe that our theoretical proposed model can, in fact, sustain the development of such intelligent transportation systems in the future, to deliver better results with and for the drivers in large urban environments.

5 Conclusions

In this paper we describe a new efficient approach of a classical mathematical problem formulated by Larry Page a solution for web search. Our approach adapts this problem to the urban traffic and demonstrates that

the usage of this method to compute the centrality of each segment of road can improve the traffic and reduce congestions. The presented solution can lead to support for traffic applications designed to support traffic in large urban transportation systems. The approach can be implemented into Intelligent Transportation Systems that eventually will lead to solving problems related to congestions, finding better routes for drivers, meaning less time spent in traffic by individual cars, less pollution and fuel saving. As future study, some the decomposition approach of this method will be studied.

Author's contributions

All authors carried out the proofs. All authors conceived of the study and participated in its design and coordination. All authors read and approved the final manuscript.

Acknowledgments

This work was partially supported by project "ERRIC - Empowering Romanian Research on Intelligent Information Technologies/FP7-REGPOT-2010-1", ID: 264207. The research presented in this paper is supported by Romanian projects: "TRANSYS - Models and Techniques for Traffic Optimizing in Urban Environments" (Contract No. 4/28.07.2010, Project CNCSIS-PN-II-RU-PD ID: 238) and "*SORMSYS - Resource Management Optimization in Self-Organizing Large Scale Distributes Systems*" (Contract No. 5/28.07.2010, Project CNCSIS-PN-II-RU-PD ID: 201). The work has been co-funded by the Sectorial Operational Program Human Resources Development 2007-2013 of the Romanian Ministry of Labor, Family and Social Protection through the Financial Agreement "*POSDRU/89/1.5/S/62557*".

References

1. Page L, Brin S, Motwani R, Winograd T: **The PageRank Citation Ranking: Bringing Order to the Web.** Technical Report 1999-66, Stanford InfoLab 1999, [<http://ilpubs.stanford.edu:8090/422/>]. [Previous number = SIDL-WP-1999-0120].
2. Sheng-Chun L, Liang-Jiang K, Mu-Ren L, Rong-Sen Z: **The effects of intelligent traffic light on the crossing traffic flow.** *ACTA PHYSICA SINICA* 2009, **58**(4):2266–2270.
3. Yousef KM, Al-Karaki JN, Shatnawi AM: **Intelligent Traffic Light Flow Control System Using Wireless Sensors Networks.** *JOURNAL OF INFORMATION SCIENCE AND ENGINEERING* 2010, **26**(3, SI):753–768.
4. Meyer CD (Ed): *Matrix analysis and applied linear algebra.* Philadelphia, PA, USA: Society for Industrial and Applied Mathematics 2000.
5. Makse HA, Havlin S, Stanley HE: **Modeling Urban-Growth Patterns.** *Nature* 1995, **377**(6550):608–612.
6. Peterson I: **The shapes of cities: mapping out fractal models of urban growth.** *Science News* 1996, **149**:8–9.
7. Cattani C: **Harmonic wavelet approximation of random, fractal and high frequency signals.** *Telecommunication Systems* 2010, **43**:207–217, [<http://dx.doi.org/10.1007/s11235-009-9208-3>].

8. Langville AN, Meyer CD: **Deeper inside pagerank**. *Internet Mathematics* 2004, **1**:2004.
9. Ming L, Wei Z: **Representation of a Stochastic Traffic Bound**. *IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS* 2010, **21**(9):1368–1372.
10. Ming L, Wei Z: **A Model to Partly but Reliably Distinguish DDOS Flood Traffic from Aggregated One**. *Mathematical Problems in Engineering* 2012, **2012**(2012):12 pages.
11. Berman A, Plemmons RJ: *Nonnegative matrices in the mathematical sciences*. Computer science and applied mathematics, New York, NY: Academic Press 1979.
12. Khafa F, Barolli A, Sanchez C, Barolli L: **A simulated annealing algorithm for router nodes placement problem in Wireless Mesh Networks**. *Simulation Modelling Practice and Theory* 2011, **19**(10):2276 – 2284, [<http://www.sciencedirect.com/science/article/pii/S1569190X10001905>].
13. Chaabane M, Tadeo F, Mehdi D, Souissi M: **Robust Admissibilization of Descriptor Systems by Static Output-Feedback: An LMI Approach**. *MATHEMATICAL PROBLEMS IN ENGINEERING* 2011.
14. Bessis N, Khafa F: *Next Generation Data Technologies for Collective Computational Intelligence*. Springer Publishing Company, Incorporated, 1st edition 2011.
15. Toma C: **Advanced signal processing and command synthesis for memory-limited complex systems**. *Mathematical Problems in Engineering* 2012, **2012**, [<http://www.scopus.com/inward/record.url?eid=2-s2.0-84855274291&partnerID=40&md5=a5a49953965abf2cea5597bab7674b31>].
16. ***: **Intelligent Transport Systems - Thematic Research Summary**. Tech. rep., Transport Research Knowledge Center, Directorate-General for Energy and Transport 2009.
17. Duranton G, Turner M: **The fundamental law of road congestion: Evidence from US cities**. Tech. rep., National Bureau of Economic Research 2009.
18. Ming L, Wei Z: **Visiting Power Laws in Cyber-Physical Networking Systems**. *Mathematical Problems in Engineering* 2012, **2012**(2012):13 pages.
19. Stanley H: **Phase transitions: power laws and universality**. *Nature* 1995, **378**:554.
20. Zhao X, Shang P, Lin A, Chen G: **Multifractal Fourier detrended cross-correlation analysis of traffic signals**. *Physica A: Statistical Mechanics and its Applications* 2011, **390**(21-22):3670 – 3678, [<http://www.sciencedirect.com/science/article/pii/S0378437111004651>].