# RoadRank: Traffic Diffusion and Influence Estimation in Dynamic Urban Road Networks

Tarique Anwar[*†] Chengfei Liu[*] Hai L. Vu[*] Md. Saiful Islam[*]

[*]Swinburne University of Technology, Melbourne, Australia
[†]NICTA, Melbourne, Australia
{tanwar,cliu,hvu,mdsaifulislam}@swin.edu.au

## ABSTRACT

With the rapidly growing population in urban areas, these days the urban road networks are expanding at a faster rate. The frequent movement of people on them leads to traffic congestions. These congestions originate from some crowded road segments, and diffuse towards other parts of the urban road networks creating further congestions. This behavior of road networks motivates the need to understand the influence of individual road segments on others in terms of congestion. In this work, we propose `RoadRank`, an algorithm to compute the influence scores of each road segment in an urban road network, and rank them based on their overall influence. It is an incremental algorithm that keeps on updating the influence scores with time, by feeding with the latest traffic data at each time point. The method starts with constructing a directed graph called *influence graph*, which is then used to iteratively compute the influence scores using probabilistic diffusion theory. We show promising preliminary experimental results on real SCATS traffic data of Melbourne.

## Categories and Subject Descriptors

H.2.8 [**Database Management**]: Database Applications—*Spatial databases and GIS, Data mining*

## General Terms

Algorithms, Experimentation

## Keywords

Influential roads; Road networks; Traffic diffusion

## 1. INTRODUCTION

Traffic congestion remains a big challenge in the 21st century due to the rapid growth of population and their mobility demand within the urban areas [1]. Congestion often starts in few confined places within the network and propagates through various connected road segments. The propagation of congestion is due to the diffusion (or movement) of traffic from one road segment to another where the amount of traffic on any particular road segment is *influenced by* and *influences* that of others. The level of influence one road segment can have on others in terms of congestion depends on their spatial and temporal attributes and is not the same across the network. Clearly, high influential road segments play a more important role in the building up of congestion and will need to be better managed in order to alleviate or delay the congestion onset.

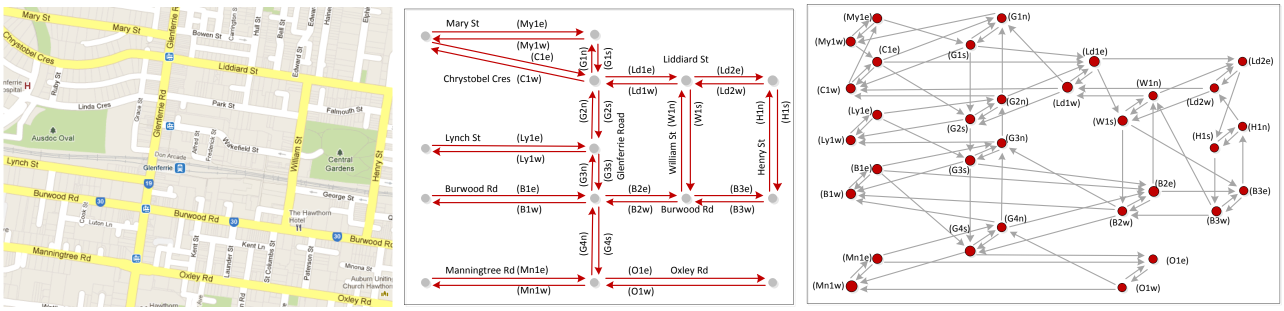
Identifying important nodes in a network is a problem common to many different kinds of information networks [3]. In the recent years, information propagation and diffusion has been a hot area of research in social networks [6, 8, 4]. Several existing works focus on estimating the user influences [5, 7] and analyzing the influence propagation [2] in different types or platforms of social networks. Though there exists many works on such varied problems, we could not find any research on traffic diffusion in road networks and identifying influential road segments.

In this paper, we propose our novel algorithm named `RoadRank` that computes the influence scores (called *roadrank scores*) of each road segment in an urban road network. To deal with the dynamic nature of traffic on a road network, it also updates the scores incrementally with time, by feeding in the latest traffic measures at each timestamp. The method starts with constructing a directed graph called *influence graph*. It is used to compute the traffic diffusion from one road segment to another, based on the collected traffic measures. Finally the *roadrank* scores are iteratively computed using probabilistic diffusion theory for the starting point of time. At each successive timestamps, they are iteratively updated based on the new traffic measures. We also show preliminary experimental results on real SCATS traffic data of Melbourne.

## 2. PROBLEM STATEMENT

**Traffic data**: The developed urban areas these days are well equipped with urban traffic control systems (UTCS) like Sydney coordinated adaptive traffic system (SCATS) and split cycle offset optimisation technique (SCOOT). These traffic management systems log the aggregated traffic movement data on each of the road segments in real-time. This data do not contain individual movement trajectories, rather they are different traffic measures indicating the count of vehicles crossing a road segment during the green time called


*CIKM'15,* October 19-23, 2015, Melbourne, VIC, Australia
© 2015 ACM. ISBN 978-1-4503-3794-6/15/10...$15.00
DOI: http://dx.doi.org/10.1145/2806416.2806588.




(a) Actual road map (b) Road network (c) Influence graph

**Figure 1: Influence graph of road networks**

the *traffic volume* and the ratio of the effectively used green time to the total available green time called the *degree of saturation*, at each signal cycle. The proposed `RoadRank` algorithm uses these traffic measures to compute the influence scores of all road segments in the given road network.

**Road networks**: Urban roads exist in the form of a physical network spatially spread over a large urban area. A *road* is a generally used term to mean a publicly accessible way for transportation. It has no standard range for its length and can vary from a very short one to a very long. A *road segment* $r_i$ is defined as the smallest unit of a road having its traffic flow in a single direction. Thus a road is composed of one or more road segments, where the two opposite directions of traffic flow form different road segments. A *road intersection point* (or intersection point in short) $\iota_i$ is defined as the point connecting two or more road segments.

Definition 1: (**Road Network**) A road network is defined as $\mathcal{N} = (\mathcal{I}, \mathcal{R})$ comprising a set of road intersection points $\mathcal{I} = \{\iota_1, \iota_2, \ldots, \iota_{n_\iota}\}$ as nodes that are connected among themselves by the set of directed road segments $\mathcal{R} = \{r_1, r_2, \ldots, r_{n_r}\}$ as its links, where each road segment $r_i$ associates a set of feature values with itself, including the different traffic measures $r_i.f$. ■

**Problem definition**: Let us suppose we have a given urban road network $\mathcal{N}$ having all recorded historical traffic data $\mathcal{D}$ corresponding to each of the road segments in the network. Let $\mathcal{T} = \langle t_0, t_1, \ldots, t_{l-1}, t_l \rangle$ be the timestamps of the recorded data, where $t_0$ is the very first record, and $t_l$ is the latest record that keeps on updating with time. Thus the dataset $\mathcal{D} = \langle d_0, d_1, \ldots, d_{l-1}, d_l \rangle$ becomes an incrementally updating vector, where each $d_i$ is the traffic data recorded at time $t_i$. The problem of estimating the influence of road segments is to incrementally compute an *influence score* for each road segment $r_i$ in $\mathcal{N}$ corresponding to each $d_i$ (at $t_i$) and rank them based on this measure. The influence score (also called roadrank score to be defined in Section 3.3) of a road segment $r_i$ gives a measure of how much the traffic on $r_i$ influences that on the global network $\mathcal{N}$ because of the traffic diffusion going on via its linked road segments.

# 3. ROADRANK ALGORITHM

## 3.1 Road Influence Graph Construction

Most of the roads exist as two-way roads which are partitioned into two parts from the middle for traffic of the two opposite directions. Each of the two parts undergo different kinds of traffic flow patterns. To accommodate this property, the road network considers the two traffic directions as separate road segments that share common road intersection points in the road network. Thus the road network $\mathcal{N}$ is a directed graph with intersection points as nodes and directed road segments as links, where for the two-way roads there are two different oppositely directed links between the same pair of intersection points.

The method starts with constructing a *road influence graph* $\mathcal{G}^{inf}$ from $\mathcal{N}$, having the road segments as nodes and the directed edges between them represent the *influences* relationship.

Definition 2: (**Road Influence Graph**) Given a road network $\mathcal{N} = (\mathcal{I}, \mathcal{R})$, the corresponding road influence graph $\mathcal{G}^{inf} = (\mathcal{V}, \mathcal{E})$ is constructed by adding each road segment $r_i \in \mathcal{R}$ as a node $v_i \in \mathcal{V}$, and establishing a directed link $e_i \in \mathcal{E}$ from $v_j$ to $v_k$ if these two conditions hold- *i*) there exist at least one intersection point $\iota_l$ that is a common intersection for the roads $r_j$ and $r_k$, and *ii*) the direction of traffic flow leads the movement from $r_j$ to $r_k$. The directed link $e_i$ shows that the traffic on road segment $r_j$ (or node $v_j$) influences the traffic on road segment $r_k$ (or node $v_k$). ■

We assume that the traffic rule allows a moving vehicle to take all the turns including left-, right-, and U-turns at intersection points on all the roads in the network (though in real there exist restrictions). Figure 1 shows an example of the influence graph construction from a real road network segment, showing the map in Figure 1(a). Considering only the yellow major roads to keep the example simple (all the roads including small ones are considered in real), Figure 1(b) shows the equivalent road network, which is then used to construct the influence graph in Figure 1(c). A traversal from one node to another in the influence graph gives the path of the traffic flow as well as the traffic diffusion process. One such example is $(My1e) \xrightarrow{inf} (G1s) \xrightarrow{inf} (G2s) \xrightarrow{inf} (G3s) \xrightarrow{inf} (B2e) \xrightarrow{inf} (B3e) \xrightarrow{inf} (B3w)$.

## 3.2 Traffic Diffusion Computation

Information diffusion is a widely studied research area in social media, where the information propagates from one sources to another. In road traffic networks, there is a physical movement of the traffic from one road segment to another. We model this movement as a traffic diffusion process, in which the traffic diffuses among differently concentrated road segments through their connectivities. The directed links in the influence graph constructed in Section 3.1 show the direction and order of movement of this traffic. To find the exact amount of traffic that diffuse from one road segment to another in a directed link, we compute a measure based on their real time traffic measures.

Let $Vo = \langle vo_1, vo_2, \ldots, vo_{n_r} \rangle$ and $DS = \langle ds_1, ds_2, \ldots, ds_{n_r} \rangle$ be the vectors having the traffic volume and degree of saturation measures of all the road segments $r_i \in \mathcal{R}$. We compute two similarity matrices $S^{vo}$ and $S^{ds}$ for the $Vo$ and $DS$ vectors respectively such that $S^{vo}_{ij} = GSim(vo_i, vo_j)$ and $S^{ds}_{ij} = GSim(ds_i, ds_j)$. Equation 1 defines the Gaussian similarity measure $GSim(x_i, x_j)$, where $\sigma^2 = \frac{1}{n} \times \sum_{i=1}^{n} (x_i - \mu)^2$ is the variance of all values in the vector.

$$GSim(x_i, x_j) = \exp\left(\frac{-(x_i - x_j)^2}{2 \times \sigma^2}\right) \tag{1}$$

Definition 3: (**Traffic Diffusion**) The *traffic diffusion* $td(r_i \rightarrow r_j)$ is defined as the actual amount of traffic diffused from $r_i$ to $r_j$. We compute this measure by using Equation 2, where $\alpha_{ij}$ is a balance factor called *traffic condition factor*, whose value is based on the real traffic conditions. ∎

$$td(r_i \rightarrow r_j) = \begin{cases} \alpha_{ij} \times S^{vo}_{ij} + (1 - \alpha_{ij}) \times S^{ds}_{ij}, & \text{if } (r_i \rightarrow r_j) \in \mathcal{E} \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

The *traffic condition factor* $\alpha_{ij} \in [0, 1]$ is a measure that quantifies the actual traffic condition on the roads. For example, let $r_i$ and $r_j$ be a narrow and a wide road segment respectively. When a certain number of vehicles move from $r_i$ to $r_j$, the $ds_i$ will be greater than $ds_j$. Thus even though all the vehicles from $r_i$ move on to $r_j$, their similarity in the $DS$ measure becomes low, because of the difference in their width. The traffic condition factor captures this aspect of road segments to balance the $Vo$ and $DS$ similarity measures and aid in accurately computing the traffic diffusion. Setting this value to 1 makes Definition 3 behave as volume diffusion, where as 0 makes it behave as degree of saturation diffusion, and setting an appropriate value[1] in between them makes it behave as the traffic congestion diffusion.

Definition 4: (**Traffic Diffusion Probability**) The *traffic diffusion probability* $tdp(r_i \rightarrow r_j)$ is defined as probability of the traffic that diffuse from $r_i$ to $r_j$. We compute this measure by using Equation 3, where $r_i \xrightarrow{inf} r_k$ means $r_i$ *influences* $r_j$ in $\mathcal{G}^{inf}$ (in other words, $r_i \rightarrow r_k \in \mathcal{E}$). ∎

$$tdp(r_i \rightarrow r_j) = \frac{td(r_i \rightarrow r_j)}{\sum_{\forall k : r_i \xrightarrow{inf} r_k} td(r_i \rightarrow r_k)} \tag{3}$$

## 3.3 Ranking

In the road traffic network scenario, the top-k influential road segments can be identified by ranking and selecting the top-k of them. Moreover, as the road networks are dynamic in nature because of the frequently changing traffic, the ranking is time-sensitive, and needs to be updated with time based on the real-time traffic measures. Therefore we start with computing the ranking at timestamp $t_0$, and keep on updating the ranking at each new timestamps $t_1, t_2, \ldots t_l$ to always have the ranking obtained from the latest collected traffic measures.

[1]It can be done based on experimental results and inputs from the experienced traffic management people.

**Ranking at $t_0$:** Our ranking algorithm borrows some concepts of the well known PageRank algorithm [3]. The PageRank algorithm computes a ranking of webpages to estimate their importance to Web navigators. It initializes each of the pages with a small value as their page rank score (PR($p_i$)), and iteratively uses the linkages ($L$) among them to compute their new page rank score (PR($p_j$)) using equation 4, where $d \in [0, 1]$ is the damping factor typically set to 0.85 [3], $\text{prob}(p_j|p_i) = \frac{1}{out-degree(p_i)}$ is the transition probability from webpage $p_i$ to webpage $p_j$, and $l_{ij} \in L$ is the hyperlink from page $p_i$ to $p_j$.

$$\text{PR}(p_j) = (1 - d) + d \times \sum_{\forall p_i : l_{ij} \in L} \text{prob}(p_j|p_i) \times \text{PR}(p_i) \tag{4}$$

The traffic on road networks consists of vehicles moving on its road segments. In the proposed `RoadRank` algorithm, we consider that the traffic moving from one road segment $r_i$ to another road segment $r_j$ as the recommendation made by $r_i$ to $r_j$. Over a period of time, also new vehicles start and existing vehicles stop randomly on different road segments. Let $vpn$ be the probability that a new random vehicle starts on any road segment in the given road network $\mathcal{N}$ at timestamp $t_r$, such that $(1 - vpn)$ gives the probability that the vehicle was already there in $\mathcal{N}$. For the new random vehicle that started in $\mathcal{N}$, let $vp(r_j)$ be the probability that the vehicle started on $r_j$, such that $(1 - vp(r_j))$ gives the probability that the vehicle started on any other road segment $r_k \in \{\mathcal{V} - r_j\}$. Thus $vpn \times vp(r_j)$ gives the probability that a new random vehicle started on $r_j$, and $(1 - vpn \times vp(r_j))$ gives the probability that the vehicles currently on $r_j$ were there on $\mathcal{N}$ since before. We compute the roadrank score of a road segment $r_j$ by considering and aggregating these two scenarios using Equation 5, where $RR(x)$ denotes the updated score of $x$ and $RR'(x)$ denotes the score of $x$ in the previous iteration. For the new traffic, it simply counts the probability $vpn \times vp(r_j)$, and for the existing traffic it counts the summation of all roadrank scores diffused from the neighboring road segments, multiplied by the probability $(1 - vpn \times vp(r_j))$. The diffused roadrank scores are computed as multiplication of the previous roadrank scores $(RR'(r_i))$ of the roads influencing $r_j$ with their traffic diffusion probability.

$$\begin{aligned} RR(r_j) &= f(\text{new traffic}) + f(\text{existing traffic}) \\ &= vpn \times vp(r_j) + (1 - vpn \times vp(r_j)) \\ &\quad \times \sum_{\forall i : r_i \xrightarrow{inf} r_j} (tdp(r_i \rightarrow r_j) \times RR'(r_i)) \end{aligned} \tag{5}$$

To start with, the roadrank scores $RR(r_i)$ for all the road segments are set to 1. Equation 5 is then used to iteratively update the $RR(r_j)$ scores from the previously available $RR'(r_i)$ scores. Each iteration updates $RR(r_j)$ for all $r_j \in \mathcal{R}$ using $RR'(r_i)$, and the iterations are repeated until convergence is achieved.

**Incremental ranking at each new timestamp $t_l$:** If we closely monitor the traffic on road networks in very short intervals of time, we will notice that the change in traffic in successive timestamps are very smooth. Therefore we can expect that in these short intervals of time there will be no big change also in the roadrank scores. Let $RR^0(r_j), RR^1(r_j), \ldots, RR^l(r_j)$ denote the roadrank scores at timestamps $t_0, t_1, \ldots, t_l$ respectively, based on the traffic datasets $d_0, d_1, \ldots, d_l$ respectively. Similarly $vpn^l$, $vp^l(r_j)$, and $tdp^l(r_i \rightarrow r_j)$ denote the values at timestamp $t_l$ based on the dataset $d_l$. To compute the score at the latest timestamp $t_l$, the roadrank scores of all $r_i$ are initialized as those computed from the previous timestamp $t_{l-1}$, i.e., $RR'^l(r_i) = RR^{l-1}(r_i)$. Equation 6 is then used to iteratively compute latest roadrank scores. The iterations are continued until it reaches the convergence.

$$RR^l(r_j) = vpn^l \times vp^l(r_j) + (1 - vpn^l \times vp^l(r_j)) \times \sum_{\forall i: r_i \xrightarrow{inf} r_j} \left( tdp^l(r_i \to r_j) \times RR'^l(r_i) \right) \quad (6)$$

Thus at each new timestamp, instead of recomputing the roadrank scores by initializing them to 1, we incrementally update them. It significantly improves the performance by requiring comparatively fewer iterations and making the task much faster. In this way, the algorithm is able to maintain the roadrank scores with the real time data.

# 4. EXPERIMENTAL RESULTS

To show the performance of `RoadRank`, we conducted experiments on real traffic data collected from the urban Melbourne road network. The considered Melbourne subnetwork consists of 1444 road segments and 581 intersection points, where the traffic measures are logged by 493 SCATS sites. In the dataset, many road segments have their traffic measures missing because of faulty SCATS sensors. We regain those missing data by applying a data repairing technique. The traffic measures are recorded for each signal cycle that differs for the different SCATS sites. To make them consistent, we made slots of 5 minutes each and aggregate the traffic measures during that time slot for all the road segments.

**Table 1: Top-5 influential road segments**

| Rank | Road segment | RR Score |
|---|---|---|
| | *03-02-2012 08:05 AM* | |
| 1. | Hoddle St (Victoria Parade to Elizabeth St) | 02.57095 |
| 2. | Hoddle St (Elizabeth St to Albert St) | 02.00645 |
| 3. | Mills St (Canterbury Rd to Danks St) | 01.89356 |
| 4. | Heidelberg Rd (Hoddle St to The Esplanade) | 01.83253 |
| 5. | Heidelberg Rd (The Esplanade to Hoddle St) | 01.81797 |
| | *03-02-2012 10:05 AM* | |
| 1. | Hoddle St (Victoria Parade to Elizabeth St) | 03.49890 |
| 2. | Hoddle St (Elizabeth St to Albert St) | 02.50019 |
| 3. | Heidelberg Rd (Hoddle St to The Esplanade) | 02.38343 |
| 4. | Heidelberg Rd (The Esplanade to Hoddle St) | 02.35427 |
| 5. | Hoddle St (Elizabeth St to Victoria Parade) | 02.31045 |

Table 1 shows the 5 top-ranking influential road segments in the considered network on 03-12-2012 (Monday) at 08:05 AM and 10:05 AM. These results are obtained by setting the traffic condition factor $\alpha_{ij}$ to 0.5, and getting the values of $vpn$ and $vp(r_j)$ from the dataset. A few road segments of Hoddle St and Heidelberg Rd are there in the list which are well known as a congestion hotspot for both the network operators and the commuters during the aforementioned time period. The ranking varies greatly for different values of $\alpha$. We obtained the rankings at $\alpha = 0.00, 0.25, 0.50, 0.75, 1.00$, and compared the top-100 of them using the Kendal's Tau measure. The measures we obtained are $\text{KT}(RR_{\alpha=0}, RR_{\alpha=0.25}) = 2409$, $\text{KT}(RR_{\alpha=0}, RR_{\alpha=0.5}) = 4516$, $\text{KT}(RR_{\alpha=0}, RR_{\alpha=0.75}) = 6428$, and $\text{KT}(RR_{\alpha=0}, RR_{\alpha=1}) = 7590$. These values show the difference in behavior of Vo diffusion and DS diffusion. A good balance between these two diffusion can be obtained through $\alpha_{ij}$ by incorporating the real traffic conditions to improve the ranking accuracy.

The ranking also keeps on changing with time based on the new traffic measures. Comparing the previous ranking to that obtained at 10:05 AM, we found that while few road segments in the top order remain same, those in the lower order change greatly with $\text{KT}(RR_{08:05AM}, RR_{10:05AM}) = 3794$ for the top 100 segments. Also some small road segments sometimes become influential for a short period of time but quickly goes down in the next time point as the congestion scenario changes. The convergences for $t_0 =$ 08:00 AM and 10:05 AM are achieved[2] in 922 (taking 36 seconds) and 1193 (taking 42 seconds) iterations respectively, which take significantly lesser iterations and time in incrementally computing the ranking at subsequent timestamps $t_l$. Normally a traffic signal takes one to three minutes to complete a cycle. With this much interval of available time for processing the data, our algorithm is efficient enough to perform the computations in real time.

# 5. CONCLUSION

In this paper, we devised a probabilistic traffic diffusion model to identify the most influential road segments in an urban road network. We presented `RoadRank` as a novel algorithm to compute the influence scores (or roadrank scores) of each road segment, which are updated incrementally with time to deal with the dynamic nature of traffic. The method starts with constructing a directed *influence graph* that is used to compute the traffic diffusion from one road segment to another at each timestamp. The *roadrank* scores are iteratively computed using probabilistic diffusion theory. We conducted experiments on historical traffic data collected from the 493 SCATS sites in Melbourne (Australia), and presented insights to the real traffic congestion scenario.

# Acknowledgment

This research was supported by the NICTA, ARC Discovery Projects DP120102627 and DP140103499, and Future Fellowship grant FT120100723. NICTA is funded by the Australian Government through the Department of Communications and the Australian Research Council through the ICT Centre of Excellence Program.

# 6. REFERENCES

[1] T. Anwar, C. Liu, H. L. Vu, and C. Leckie. Spatial partitioning of large urban road networks. In *Proc. EDBT*, 2014.

[2] B. Bi, Y. Tian, Y. Sismanis, A. Balmin, and J. Cho. Scalable topic-specific influence analysis on microblogs. In *Proc. WSDM*, 2014.

[3] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1-7), Apr. 1998.

[4] M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *Proc. WWW*, 2009.

[5] A. Goyal, F. Bonchi, and L. V. Lakshmanan. Learning influence probabilities in social networks. In *Proc. WSDM*, 2010.

[6] D. Gruhl, R. Guha, D. Liben-Nowell, and A. Tomkins. Information diffusion through blogspace. In *Proc. WWW*, 2004.

[7] J. Herzig, Y. Mass, and H. Roitman. An author-reader influence model for detecting topic-based influencers in social media. In *Proc. HT*, 2014.

[8] D. Kempe, J. Kleinberg, and E. Tardos. Influential nodes in a diffusion model for social networks. In *Proc. ICALP*, 2005.

[2] Performed on a Core i5 computer with 8GB RAM