Augmented Betweenness Centrality for Environmentally Aware Traffic Monitoring in Transportation Networks

RAMI PUZIS,1 YANIV ALTSHULER,2 YUVAL ELOVICI,3 SHLOMO BEKHOR,4 YORAM SHIFTAN,5 and ALEX (SANDY) PENTLAND2
1University of Maryland Institute for advanced Computer Studies (UMIACS), College Park, Maryland, USA
2MIT Media Lab, Cambridge, Massachusetts, USA
3Telekom Innovation Laboratories and Department of Information Systems Eng., Ben-Gurion University
4Transportation Research Institute, Technion–Israel Institute of Technology, Haifa, Israel
5Civil and Environmental Engineering, Technion, and the Goldstein UAV and Satellite Center

Network planning and traffic flow optimization require the acquisition and analysis of large quantities of data such as the network topology, its traffic flow data, vehicle fleet composition, emission measurements and so on. Data acquisition is an expensive process that involves household surveys and automatic as well as semiautomatic measurements performed all over the network. For example, in order to accurately estimate the effect of a certain network change on the total emissions produced by vehicles in the network, assessment of the vehicle fleet composition for each origin–destination pair is required. As a result, problems that optimize nonlocal merit functions become highly difficult to solve. One such problem is finding the optimal deployment of traffic monitoring units. In this article we suggest a new traffic assignment model that is based on the concept of shortest path betweenness centrality measure, borrowed from the domain of complex network analysis. We show how betweenness can be augmented in order to solve the traffic assignment problem given an arbitrary travel cost definition. The proposed traffic assignment model is evaluated using a high-resolution Israeli transportation data set derived from the analysis of cellular phones data. The group variant of the augmented betweenness centrality is then used to optimize the locations of traffic monitoring units, hence reducing the cost and increasing the effectiveness of traffic monitoring.

Keywords Betweenness; Monitoring; Traffic Assignment

INTRODUCTION

Intelligent transportation systems (ITS) serve an important role in reducing vehicle emissions and improving air quality, with its main contribution being threefold: first, the introduction of new technologies that reduce vehicle emissions; second, the provision of information regarding both driving behavior that encourage more “eco driving styles” and travel behavior that assist in choosing better modes and routes to reduce emissions; and third, the replacement of activities that require travel, for example, by trading meetings for communication mediated activities such as video conferences.

There are many technical opportunities to improve fuel efficiency and economy of motor vehicles to reduce emissions. A review of various possibilities of reducing vehicle energy consumption can be found in Atabani et al. (2011), presenting advanced engine and transmission technology as well as alternative power sources.

Improved driving behavior (such as using less acceleration and deceleration while driving) is shown to significantly reduce fuel consumption and emission (see e.g., Barkenbus, 2010;
Barth & Boriboonsomsin, 2009a). The work of Toledo (2010) has shown that drivers who received feedback on their driving from in-vehicle data recorders were able to reduce their fuel consumption by up to 14%. Appropriate driving feedback is also shown to save travel time, as can be seen in Toledo & Bein- haker (2006). Various technologies, such as in-vehicle monitor assistants, associated with training programs, can help drivers learn more about their driving patterns, advising them regarding certain aspects of their driving that should be improved in order to better serve environmental policies.

In a broader perspective, travel behavior, like many other aspects of daily life, is being transformed by information technology (IT). Accessibility can no longer be measured only in terms of travel time, distance, or generalized travel cost. IT provides users with virtual access to a rapidly growing range of activities. E-commerce has become a catalyst for structural changes in the freight transportation industry and is changing where freight moves, the size of typical shipments and the time within which goods must be delivered. Some of the potential effects of IT on transportation both personal and freight, were explored in Golob and Regan (2001), and similar results were later demonstrated in Larsson and Ericsson (2009). More general attempts to model human mobility patterns were discussed in Gonzalez et al. (2008), Song et al. (2010), and Simini et al. (2012). This was even further expanded in Câmara Pereira et al. (2012), suggesting using the internet itself as a transportation-oriented sensor .

Another important aspect where IT can significantly affect driving patterns is the growing penetration of global positioning system (GPS) based navigation systems. The number of drivers using these systems, from small dedicated devices to navigation applications in smartphones, grows constantly. There is no doubt that routing algorithms embedded in navigation systems may affect global driving patterns we witness in global, national, and urban transportation networks. Most routing algorithms available nowadays consider total travel time, distance, or ease of driving as their optimization goals. However, a significant impact could be generated if personal navigation systems would also include environment-oriented considerations into their route planning engines and policies, while analyzing data gathered from the transportation network state in real time. Such techniques can be used either for reducing emission, fuel consumption, or the risk of accidents involving carriers of hazardous materials (Kim et al., 2011).

The development of such systems, however, requires the placement of monitoring units—units that will provide the input used for routes planning. Finding the optimal deployment for those units is therefore an important challenge. Ideally, each transportation network (global, national, and urban) should have been built while incorporating a monitoring station alongside each of its routes and intersections. This, however, is not feasible due to financial and operational considerations. One way to reduce, for example, the cost of air quality monitoring is estimating emissions based on various traffic properties such as flow, speed, and vehicle fleet composition (see Kean et al. (2003) or Barth and Boriboonsomsin (2009b)). Furthermore, the cost of traffic monitoring can be significantly reduced by deployment strategies that would maximize the utility of traffic-related monitoring, given a predefined number of monitoring units (such as by employing correlations with upstream or downstream traffic; Chandra & Al-Deek, 2009).

In order to efficiently cope with this challenge, the traffic pattern of the users must be thoroughly studied. The analysis of mobility trends and demand forecasting in transportation networks relies nowadays heavily on household survey data that provide the required input for calibrating the mathematical models that represent decisions people make related to travel (Stopher et al., 2006). However, those models generally involve solving complex problems, which in turn require significant computational resources. In this article we propose to employ computational methods borrowed from the analysis of complex networks as an alternative to classical traffic assignment models.

Shortest path betweenness centrality (BC) stands for the ability of an individual node to control the communication flow in the networks (see Freeman, 1977; also see Anthonisse, 1971). Formally, for a node $v$ it denotes the total portion of shortest paths between every pair of nodes in the network that pass through $v$ (see more details later in this article). In recent years betweenness was extensively applied for the analysis of various complex networks (Barthélemcy, 2004; Strogatz, 2001) including among others social networks (Scott, 2000; Wasserman & Faust, 1994), computer communication networks (Faloutsos et al., 1999; Yook et al., 2002), and protein interaction networks (Bork et al., 2004). In Holme (2003) it is shown that betweenness is highly correlated with congestion in particle hopping systems. Extensions of the original definition of BC (Brandes, 2008; White & Borgatti, 1994) are applicable for directed and weighted networks, as well as for multilayer networks where the underlying infrastructure and origin–destination overlay are explicitly defined (Puzis et al., 2008).

The main contribution of this work is showing the applicability of betweenness centrality and certain augmented types of it for obtaining traffic flows through links of a transportation network. We show how betweenness centrality can be augmented using certain transportation specific measurements, thereby achieving a strong positive correlation with the traffic flows in transportation networks. One of the advantages of betweenness centrality is that it simultaneously considers all shortest paths between an origin and a destination. Another advantage of betweenness centrality employed in this paper is its natural extension to group betweenness centrality which was effectively applied in communication networks for optimizing the deployment of collaborative traffic monitoring systems (Dolev et al., 2009; Puzis et al., 2009). In this article we use the proposed betweenness-driven traffic assignment model in order to optimize the locations of monitoring units in transportation networks.

The rest of the article is organized as follows: Relevant literature regarding emission monitoring is given in article the
second section. The third section describes the transportation data that was used in this study. The fourth section provide short overview of traffic assignment models and betweenness centrality respectively. In the fifth section we discusses the correlation between betweenness centrality and traffic flow and provide methods to further increase this correlation. Our proposed approach for generating efficient deployment schemes is discussed in the sixth section. Two methods are presented and their performance in terms of solution efficiency and computation time is discussed. Concluding remarks can be found in the final section.

RELATED WORK

Vehicle air pollutant emissions are considered one of the major environmental issues. The problem stems from a constant growth in motorization rate in relation to the technological means for reduction of pollutant emissions (De Nevers, 2000).

Air pollutant concentrations in the Tel Aviv Metropolitan area have been on the rise during the last couple of decades. The metropolitan area has the highest air pollutant concentrations from vehicles (Israeli Ministry of Environmental Protection, 2011). High concentrations of primary pollutants (NOx and hydrocarbons) are monitored at early morning hours, when traffic volumes are high.

In Israel, gasoline is sold using octane numbers determined by Israeli standards that are less rigorous compared to those of the European Community. The gasoline sold in Israel contains large amounts of unsaturated hydrocarbons (benzene and toluene), which contribute to the formation of ozone. Low-quality gasoline creates conditions that lead to high gasoline consumption and high pollutant emissions, because of wear of engine parts and the catalytic converter. New regulations adapted during the last decade have improved gasoline quality in Israel, but older model vehicles that have a catalytic converter in bad shape emit large amounts of air pollutants (Israeli Ministry of Environmental Protection, 2011).

Although technologies are introduced periodically for pollutant emission reduction, vehicle traffic volume is on the rise throughout the years, and consequently so are pollutant emissions. This trend strengthens the presumption that it is necessary to deal with metropolitan air pollution issues using a combination of actions including air pollutant reduction per vehicle, and vehicle traffic volume reduction policies in order to reduce total air pollutant emissions (Israeli Ministry of Environmental Protection, 2011).

The basic emission model includes the product of two variables: the emission factor of the pollutant and the level of activity of a certain type of vehicle. The pollutant emissions from a vehicle depends on both variables and the emission factor depends on vehicle speed (Parra et al., 2006).

Transport emissions include both global emissions of greenhouse gases (GHGs) and local emissions of oxides of nitrogen (NOx), volatile organic compounds (VOC), carbon monoxide (CO), and particular matter (PM). The main factors affecting vehicle emissions are summarized in Litman (2011), including vehicle type and model, load, fuel type, and operating conditions including speed and acceleration, road type, weather, and engine temperature. Various researches have attempted to improve the modeling of emission from transportation (Cortés et al., 2008; Rentziou et al., 2012) and emphasize the need for accurate data to achieve reliable emission estimates (Armstrong & Khan, 2004). Emission modeling can be done either by a bottom-up approach (Colvice et al., 2001) which needs to tackle several problems: to collect real local data regarding traffic conditions at different hours and days, to accurately estimate the emissions generated by the actual fleet according to these conditions, to estimate the composition of the fleet, and to estimate the mileage driven by the fleet, spatially and temporally resolved. Various researchers have attempted to improve the data use such as Reynolds and Broderick (2000), who used real-time data obtained from induction loops and closed-circuit televisions (CCTV) as well as statistical data to calculate emissions and to study the impact of various policies on emissions (Ross Morrow et al., 2010; Yan & Crookes, 2010).

Recent research has demonstrated that average speed, and perhaps even simple estimates of the amount of delay and the number of vehicle stops on a roadway, are insufficient to fully capture the environmental impacts of intelligent transportation system (ITS) strategies such as adaptive traffic signal control. Specifically, for the same average speed, one can observe widely different instantaneous speed and acceleration profiles, each resulting in very different fuel consumption and emission levels. One application for such quantification of the environmental impacts of ITS alternatives is given in Rakha and Ahn (2004).

In Li & Ouyang (2011) it is observed that several transportation researchers have discussed the trade-off between accuracy and coverage, for given limited resources of sensor devices. The work of Lam & Lo (1990) proposed a heuristic approach to select locations for traffic volume count sensors in a roadway network. Yang & Zhou (1998) proposed a sensor deployment framework to maximize such utilities. This framework has been extended to accommodate turning traffic information (Bianco et al., 2001), existing installations and O-D information content (Ehler et al., 2006), the screen line problem (Yang et al., 2006), time-varying network flows (Fei et al., 2007) and unobserved link flow estimation (Hu et al., 2009). The work of Bianco et al. (2006) considered the problem of locating the minimum number of counting sensors on the network nodes in order to determine arc flow volumes of the entire network. They showed that the problem is NP-complete and analyzed selected networks in an attempt to find an approximation algorithm. A similar work can be found in Castillo et al. (2011), where link flow is estimated using past observations. In Li & Ouyang (2011) a reliable facility location model is proposed to optimize traffic surveillance benefit from synthesized sensor pairs (e.g., for travel time estimation) in addition to
individual sensor flow coverage (e.g., for traffic volume statistics), while considering probabilistic sensor failures.

There are several path-based algorithms proposed in the literature designed to solve the traffic assignment problem. The Goldstein–Levitin–Polyak Gradient Projection (GP) method was first applied in Bertsekas & Gafni (1982), and further adapted to solve the problem for large networks in Jayakrishnan et al. (1994). The basic notion behind the GP algorithm is to enlarge the pathset by adding the current shortest path to the pathset. As iterations proceed, the algorithm checks which path in the pathset is the shortest, and deviates flows from the non-shortest paths in the pathset to the shortest path. In theory, the number of routes increase with increasing number of iterations, but to reach convergence a relatively small number of iterations (and consequently routes) is needed (typically 5–10 routes per origin-destination pair; Bekhor & Toledo, 2005).

**TRANSPORTATION NETWORK DATA SET**

The widespread use of cellular phones in Israel enables the collection of accurate transportation data. Given the small size of the country, all cellular companies provide nationwide coverage. As shown in Bekhor et al. (2011), the penetration of cellular phones in the Israeli market is very high, even to lower income households, and especially among individuals in the ages of 10 to 70 years (the main focus of travel behavior studies). Such penetration enables a comprehensive study of travel behavior that is based on the mobility patterns of randomly selected mobile phones in the Israeli transportation system. This data was shown in Bekhor et al. (2011) and (Gur et al., 2009) to provide a high quality coverage of the network, tracking 94% of the trips (defined as at least 2 km in urban areas, and at least 10 km in rural areas). The resulting data contained a wealth of traffic properties for a network of over 6,000 nodes, and 15,000 directed links. In addition, the network was accompanied with an origin–destination (OD) matrix, specifying start and end points of trips.

The network was created for the National Israeli Transportation Planning Model (Gur et al., 2009). In urban areas the network contains arterial streets that connect the interurban roads. For each link there is information about the length (km), hierarchical type, free-flow travel time (min), capacity (vehicles per hour), hourly flow (vehicles per hour), and congested travel time (min). The hourly flows and congested travel times were obtained from a traffic assignment model that loads the OD matrix on the network links, as explained in the next section.

Based on the data set described earlier we have created a network structure, assigning running indices from 1 to 6716 to the nodes (junctions). We have examined the directed variant of the network where each road segment between two junctions was represented as either one or two directed links between the respective nodes.

In order to get a basic understanding of the network we first extracted and studied several of its structural properties. We have partitioned the network into structural equivalence classes of the nodes and biconnected components and computed the betweenness centrality indices of the nodes (see Lorrain & White, 1971; and more in Freeman, 1977; Lerner 2005). Structurally equivalent vertices have exactly the same neighbors and the set of these vertices is called a structural equivalence class. As can be seen in Table 1, the number of structural equivalence classes in the network equals roughly the number of vertices, where the size of the largest class is 3. This means that there are no "star-like" structures in the network and alternative paths between any two vertices are either longer than two hops or have other links emanating from the intermediate vertices. The number of biconnected components, however, is low compared to the number of nodes, meaning that there are significant regions of the network that can be cut out by merely disconnecting a single node.

Table 1  Structural properties (Israeli transportation network).

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nodes</td>
<td>6,716</td>
</tr>
<tr>
<td>Edges (undirected representation)</td>
<td>8,374</td>
</tr>
<tr>
<td>Edges (directed representation)</td>
<td>15,823</td>
</tr>
<tr>
<td>Number of structural equivalence classes</td>
<td>6,655</td>
</tr>
<tr>
<td>Largest equivalence class</td>
<td>3</td>
</tr>
<tr>
<td>Number of bi–connected components (BCC)</td>
<td>931</td>
</tr>
<tr>
<td>Average BCC size</td>
<td>8.2</td>
</tr>
<tr>
<td>Largest BCC</td>
<td>5,778</td>
</tr>
</tbody>
</table>

BETWEENNESS, GROUP BETWEENNESS, AND TRAFFIC ASSIGNMENT METHODS

Traffic Assignment

Traffic assignment is defined as the process that allocates a given set of trip demands to a specific transportation network. A transportation network is a directed graph network, composed of nodes and links, which respectively represent junctions and roads. Special nodes (centroids) represent the traffic zones, and dummy links represent the centroid connectors to the network. The demand is represented by a matrix, in which the origins and destinations correspond to the centroids in the network. In this article, it is assumed that the demand from each origin–destination pair is fixed for a given period of time (AM peak hour).

There are a great number of possible allocations of the traffic onto the network. Many factors influence drivers’ route choices. Generally, it is assumed that travel time is the main factor in choosing a route. Other factors, such as travel cost, are assumed to be correlated to travel times.

When drivers are assumed to know all travel costs exactly, and choose the best route accordingly, the behavioral model associated with these two assumptions is known as the
Figure 1  A map of the Israeli transportation network that was used for this article (color figure available online).

Figure 2  A map of the urban area of Tel Aviv, a subset of the Israeli transportation network (color figure available online).
**deterministic user equilibrium** problem. According to this model, each driver selects the route that minimizes journey costs. The equilibrium is achieved when no driver can unilaterally reduce his or her travel cost. This assumption is used in most transportation software packages, such as the TransCAD software used in the National Transportation Planning model.

The traffic assignment algorithm in most software packages performs well for the first few iterations. However, as the iterations proceed, the convergence rate becomes increasingly slow for two main reasons: First, the descent direction is computed by solving the linear approximation of the objective function. Close to the equilibrium point, this direction tends to be orthogonal to the solution (Bertsekas & Gallager, 1987). Second, the step-size determination is a unique value for all origin–destination flows, since it is performed at the link level.

The algorithm in Figure 3 outlines the primary steps of most traffic assignment methods. The most computationally intensive part of the algorithm is line 4 where shortest paths are computed. After the shortest path is computed for and the flows are assigned on the network’s links, there is no need to keep the path flows. Therefore, most algorithms have the advantage of small computer storage demands, since they work with link flows only.

Unfortunately, this optimization also results in a loss of information that may be required for efficient monitoring of the network. For example, after the traffic assignment process is complete, it is difficult to reproduce the portion of the traffic flow shared by two arbitrary links. This problem can be efficiently tackled by employing the data structure maintained for efficient calculation of group betweenness centrality (Puzis et al., 2007a).

**Betweenness and Group Betweenness Centrality**

Shortest path betweenness centrality (BC) is defined as the total fraction of shortest paths between each pair of vertices that pass through a given vertex (Freeman, 1977). Let $G = (V, E)$ be a directed transportation network where $V$ is the set of junctions and $E$ is the set of directed links as described in the previous section. Let $\sigma_{s,t}$ be the number of shortest paths between the origin vertex $s \in V$ and the destination vertex $t \in V$. Some variants of Betweenness relieve the shortest path constraint allowing deviations from the minimal distance between the two vertices (Dolev et al., 2010) or even equally considering all paths or random walks (Freeman et al., 1991; Newman, 2005).

In the rest of this article we will refer to the shortest or “almost” shortest paths between two vertices as *routes*. Let $\sigma_{s,t}(v)$ be the number of routes from $s$ to $t$ that pass through the vertex $v$. BC can hence be expressed by the following equation:

$$BC(v) = \sum_{s,t \in V} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}.$$  \hspace{1cm} (1)

Note that in this definition we include the end vertices ($s$ and $t$) in the computation of betweenness since we assume that vehicles can be inspected also at their origin and at the point of their destination.

Betweenness centrality (BC) can be naturally extended to group betweenness centrality (GBC) (Everett & Borgatti, 1999). Given a group of vertices ($M \subseteq V$) GBC accounts for all routes that pass through at least one member of the group. This way GBC estimates the utility of monitoring the traffic in several locations better than the sum of BC values (respectively traffic flow) of these locations. Let $\sigma_{s,t}(M)$ be the number of routes from $s$ to $t$ passing through at least one vertex in $M$:

$$GBC(M) = \sum_{s,t \in V} \frac{\sigma_{s,t}(M)}{\sigma_{s,t}}.$$  \hspace{1cm} (2)

GBC can be efficiently computed using the algorithm presented in (Puzis et al., 2007a). Note that GBC computation requires a preprocessing stage, whose time complexity can be either $O(|V| \cdot |E|)$ or $O(|V|^3)$ depending on the number of groups that should be evaluated. After the $O(|V|^3)$ preprocessing, the
time complexity of computing GBC of a single group \((M)\) is \(O(|M|^3)\). Note that the time required to compute GBC of a single group does not depend on the size of the network. This fact enables using combinatorial optimization methods for finding a group of given size with the maximal GBC (Puzis et al., 2007b). In the past, GBC optimization methods were successfully applied in communication networks for optimizing network traffic monitoring (Dolev et al., 2009; Puzis et al., 2009).

**AUGMENTED VARIANTS OF BETWEENNESS FOR TRAFFIC ASSIGNMENT**

Betweenness centrality as defined in the preceding section does not meet the requirements for a reasonable traffic assignment model and cannot be used for optimizing the location of monitors as is. Next we demonstrate and discuss this problem, and show how the basic definition of BC can be augmented in order to better fit the traffic assignment and the monitors placement problems in transportation networks.

After computing BC of all the links in the given transportation network, we can easily see that the distribution of betweenness centrality follows a power law (Figure 4, left). Long-tail distributions such as the power law suggest that there is a nonnegligible probability for existence of vertices whose betweenness centrality can be arbitrarily high (in contrast to the exponential flow distribution depicted in Figure 4, right). Exponential distribution of traffic flow through the network links does not have a long right tail suggesting (as expected) that the maximal flow through the network links is bounded. The different nature of these two distributions suggests that BC as already defined will overestimate the actual traffic flow through vertices and links, especially through most central. The primary reason for this overestimation is the fact that BC (as defined in earlier section) considers shortest paths only, without taking into account congestions.

Next we would like to calculate the correlation between BC and traffic flow. Notice that although the correlation is statistically significant the square error is very low \((R^2 = 0.1086)\) as shown in Figure 5a. Every point in this figure represents a vertex with the x-axis corresponding to the traffic flow obtained via traffic assignment model and y-axis corresponding to the computed BC.

At this point we should note that this is the main motivator for the development of the methods presented in this article, namely, betweenness methods that analyze the network while taking into consideration transportation specific features such as free flow time, and behavior in different times of the day (see next sections, as well as Figures 5b, 5c, and 5d).

We next discuss augmented variants of the BC measure that significantly improve the correlation with the traffic flow.

### Origin–Destination-Based Betweenness Centrality

According to The definition of BC (Eq. 1) assumes equal weights of routes between every pair of vertices in the network. In other words, every vertex acts as an origin and as a destination of traffic. We would now like to utilize the measured origin–destination (OD) flow matrix in order to prioritize network regions by their actual use. For this, we use the following altered definition for Betweenness, as suggested in Puzis et al. (2008):

\[
BC_{\text{hop}}(v) = \sum_{s,t \in V} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}} \cdot OD_{s,t}
\]

where \(OD\) is the actual measured origin-destination matrix. This method produces a better correlation \((R^2 = 0.3664)\) between the theoretic (BC) and the measured traffic flow (see Figure 5b).
Betweenness in Weighted Networks

In order to further improve our ability to estimate the predicted network flow using the network’s topology, we note that both BC calculation methods (Eqs. 1 and 3) assume that routes are chosen according to shortest path strategy based on hop counting. In this section, we retain the shortest path assumption but use weighted links for calculating the betweenness score. We recompute BC on the directed transportation network, weighting links using their free-flow travel time.

Let $BC_{ft}(v)$ denote the betweenness of a node $v$ computed with respect to the free-flow travel time. Figure 5c shows significant improvements in the correlation between the measured traffic flow and the theoretical $BC_{ft}$ values computed with respect to the OD matrix and free-flow travel time link weights ($R^2 = 0.5371$).

We can see that there are few links whose flow was significantly overestimated by the $BC_{ft}$ measure. Tracking back those data points into the traffic network, we can see that the seven outlying links in Figure 5c are consecutive road segments, taken from one of the most traffic congested highways in Israel (the Ayalon Highway). The freeflow travel times in these points are very low, resulting in over-utilization of this road and congestions that are not taken into account by $BC_{ft}$. See an example of the points in Figure 6.

Next we further improve this correlation by first applying the traffic assignment model and then using the resulting travel times to define the shortest paths for BC.

Betweenness Centrality Based on Traffic Assignment

Let $BC_{ct}(v)$ denote the betweenness of a node $v$ computed with respect to the travel times obtained using traffic assignment model. Group betweenness centrality, on which we will elaborate in the next section, can be computed based on these times and then effectively utilized for optimizing locations of traffic monitoring units.

Computing betweenness using only the obtained travel times results in $R^2 = 0.6559$. This result can be further improved because not all drivers “hit the road” during peak hours. While traveling off-peak the drivers utilize the free-flow travel time through the main roads and do not diverge sideways to avoid congestion. $BC_{ft}$ would
predict the flow created by those drivers. On the other hand, 
$BC^{ct}$ better predicts flows created by drivers traveling during peak hours.

Next we combine both the BC computed with respect to the free-flow travel time and the obtained time by taking a weighted average, namely:

$$BC^{ct-ft}(v) = (1 - \alpha) \cdot BC^{ct}(v) + \alpha \cdot BC^{ft}(v) \quad (4)$$

The resulting centrality index can achieve higher correlation with the measured average traffic flow. The maximal achievable correlation equals $R^2 = 0.6875$, and is obtained for $\alpha = 0.3$ as shown in the Figure 7 (plot $ct - ft$).

The maximum at $\alpha = 0.3$ suggests that approximately third of the drivers utilize the free-flow travel times. The combinations of the BC variants, and in particular the combinations of $BC^{ct}$ with other measures, are reported here for completeness. These results are inferior to the results reported in the next section (i.e., $BC^{itr}$) mainly because they require using state-of-the-art traffic assignment for obtaining the average travel times.

In a similar way we also examine the combinations of $BC^{ft}$ and $BC^{ct}$ with the classical definition of betweenness ($BC^{hop}$):

$$BC^{hop-ft}(v) = (1 - \alpha) \cdot BC(v) + \alpha \cdot BC^{ft}(v) \quad (5)$$

$$BC^{ct-hop}(v) = (1 - \alpha) \cdot BC^{ct}(v) + \alpha \cdot BC(v) \quad (6)$$
Surprisingly, the combination of regular $BC$ (with respect to the OD matrix) and $BC^{ct}$ computed also with respect to travel times obtained from traffic assignment model produces the best results getting $R^2$ above 0.7. It is possible to examine also various linear combinations of all three betweenness measures discussed already, producing a yet better match to the original traffic assignment model.

**Iterative Traffic Assignment Using Betweenness Centrality**

In previous subsections we have demonstrated that flows obtained using betweenness centrality are highly correlated ($R^2 \approx 0.7$) with flows obtained from a traffic assignment model. Next we incorporate BC calculation within a simple assignment framework (e.g., the algorithm in Figure 3, lines 3–6) to obtain even higher correlation. We split the BC computation into several iterations in order to dynamically update the time to travel through links based on the BC computation itself. We use the same polynomial model for estimating travel times as the function of flow that was used to generate the data set.

Let $k$ be the number of iterations. For simplicity, we split the input OD matrix into $k$ equal shares. The model can be further improved by non-equal step size similar to state-of-the-art traffic assignment models (see Figure 3, line 7).

Mobility oriented BC calculation, as depicted in Figure 8, starts with null flows through all links. During the first iteration BC is calculated for all links based on the $\frac{1}{k}$ of the OD flows. This calculation is similar to $BC^{ft}(v)$ where free-flow travel times were used to compute routes. After the first iteration the traffic flows through all links are set to the resulting BC values. Then, the travel times are updated based on the time model used. The updated travel times are used for the next computation of BC. The Betweenness values obtained during this iteration, as well as the next iterations, are aggregated and added to the links’ traffic flows. Before each iteration takes place, the links’ travel times are updated based on the latest aggregated flows. We denote the BC values calculated using this method as $BC^{itr}(v)$.

Figure 9 demonstrates the efficiency of the proposed method. Note that three iterations are enough in order to significantly improve the correlation of BC and flows obtained from the traffic assignment model.

After only eight iterations the $R^2$ of both the travel times and the flows compared to state-of-the-art traffic assignment are 0.768 and 0.869, respectively. The squared error provides mostly qualitative comparison of the various BC variants with respect to to their correlation to traffic flows calculated using standard traffic assignment methods. We conclude this section by presenting the Pearson correlation between traffic assignment and the various variants of BC in Table 2. All correlations are significant at the level of 0.000.

**OPTIMIZING THE LOCATIONS OF MONITORING STATIONS**

Assume for example that assessment and forecasting of regional air pollution level are needed. It is possible to estimate

<table>
<thead>
<tr>
<th>BC variant</th>
<th>$R^2$</th>
<th>Pearson correlation</th>
<th>Sig. (two-tailed)</th>
<th>$n$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$BC(v)$</td>
<td>0.1086</td>
<td>0.329**</td>
<td>0.000</td>
<td>15,493</td>
</tr>
<tr>
<td>$BC^{hop}(v)$</td>
<td>0.3664</td>
<td>0.617**</td>
<td>0.000</td>
<td>15,493</td>
</tr>
<tr>
<td>$BC^{ff}(v)$</td>
<td>0.5371</td>
<td>0.734**</td>
<td>0.000</td>
<td>15,493</td>
</tr>
<tr>
<td>$BC^{itr}(v)$</td>
<td>0.6559</td>
<td>0.802**</td>
<td>0.000</td>
<td>15,493</td>
</tr>
<tr>
<td>$BC^{itr}(v)$</td>
<td>0.8690</td>
<td>0.896**</td>
<td>0.000</td>
<td>15,493</td>
</tr>
</tbody>
</table>

*Note*. Sig., significance.

**Significance level.
the level of various pollutants using mathematical models (e.g. Barth & Boriboonsomsin, 2009b; Kean et al., 2003). One of the primary inputs to these models is the vehicle fleet composition. It is therefore important to assess the distribution of vehicle types for every origin–destination pair in order to obtain accurate estimation of pollution levels along the road segments of the trip.

If a single monitoring station is to be located in order to estimate the vehicle fleet mix, we would like to place it on the node having the most traffic passing through it. However, locating several monitors on the nodes having the highest flow will result in a suboptimal deployment.

For example, a large number of heavy-duty trucks could be observed along the arterial roads of some region while from a global perspective the number of such trucks would be significantly lower. The reason for such bias is the location of monitors along one commonly used route that results in the same vehicles being accounted for by several monitors during a single trip. We would like to avoid such redundant monitoring for two reasons: reduce double counting and increase the cost effectiveness of monitoring.

This problem is typically solved by a double-layered traffic assignment where the first layer solves the assignment problem to find the net flow passing by a group of monitors while the second layer consists of an optimization algorithm that finds a group with the highest net flow (Li & Ouyang, 2011). Unfortunately, this requires significant computational effort due to the large number of traffic assignment problems being solved during the optimization.

A similar problem in communication networks was solved using group betweenness centrality (GBC) (Dolev et al., 2009; Everett & Borgatti, 1999; Puzis et al., 2009). Since GBC is an immediate natural extension of BC, all variants of BC described in the fifth Section are applicable to GBC. Similar to BC that can be used for estimating the link flow with high correlation, GBC can be used for estimating the net number of distinct vehicles passing by the monitors distributed across the network.

In contrast, to monitor location optimization using traffic assignment, the existing GBC algorithm is able to evaluate a single group ($M$) in time that scales as $O(|M|^3)$ (see Puzis et al., 2007a). It is therefore preferable to use GBC within optimization algorithms that require evaluation of a large number of relatively small groups.

In fact, finding a set of nodes of given size that has the maximal GBC (or net flow) is a (NP-)hard problem\(^1\). Several combinatorial optimization techniques can be used to find a group of nodes of given size that has the largest GBC. In the following discussion we refer to a greedy approximation algorithm for the monitors location optimization problem (Greedy) (Dolev et al., 2009), a classical depth first branch and bound (DFBnB) heuristic search algorithm (Korf & Zhang, 1995), and recently proposed potential search (Stern et al., 2011).

The Greedy approximation algorithm chooses at every stage the node that has the maximal contribution to the GBC of the already chosen group. The approximation factor of the Greedy algorithm as reported in Dolev et al. (2009) is:

$$e - \frac{1}{e} \approx 0.632$$

Both the heuristic search algorithms DFBnB and the potential search provably find the group having the maximal GBC. The Greedy algorithm and DFBnB were previously compared in Puzis et al. (2007b) in the context of monitoring optimization in computer communication networks. The authors have shown that in preferential attachment networks (Barabasi & Albert, 1999) the Greedy algorithm produced results that are 0.3% far from optimal. Given the fact that finding a group of a given size having the maximal GBC is a hard problem, the Greedy algorithm is good enough for any practical purpose.

Moreover, it means that monitors can be deployed/activated gradually as they are needed (or the budget permits). It is not necessary to know ahead of time the total number of monitors that will be deployed in order to find their optimal locations. When additional monitors are needed their locations can be suggested based on up-to-date network data and the current deployment. The effectiveness of such deployment is very close to optimal, in both communication and transportation networks.

Figure 10 illustrates the outcome of the selection of one to 39 inspection locations using the Greedy algorithm. This figure provides guidance for the utility versus cost trade-off of traffic monitoring. As expected, the marginal value of additional monitors gradually decreases as more of them are added, reaching potential traffic coverage of 30% when 39 monitoring stations are deployed. This result is limited though by the imperfect correlation between theoretic betweenness centrality and the measured traffic flow. As betweenness-based mobility models develop, accurate optimization algorithms (especially those that provide optimality guarantees) will be increasingly more relevant for the optimization of deployment locations of traffic measurement and monitoring stations.

\(^1\)It can be proved by a straightforward reduction from the minimal vertex cover problem that the problem of maximizing GBC is NP-complete.
It is interesting to note that both the DFBnB and the potential search algorithms are anytime search algorithms (Zilberstein, 1996). Their execution can be stopped at any point of time, yielding the best solution found so far. Therefore, in the following experiments we limit the search time to one hour, simulating a quasi-real-time optimization constraint. Still, as can be seen in Figure 11, the running time of the Greedy algorithm is by far lower than one hour, for the entire Israeli transportation system.

When DFBnB and potential search algorithms cannot complete the search process within the given time bounds they produce a close to optimal solution and an estimate of its optimality (i.e., certificate). The certificate is computed by dividing the best solution found so far by the upper bound on the optimal solution. The upper bound is computed using admissible heuristic functions and is maintained by the search algorithms for efficient pruning of the search space. Figure 12 shows that potential search produces higher certificates for its solutions within the one-hour time bound for all sizes of the monitors deployment. Note that results produced by the heuristic search algorithms are only relevant if they are above the approximation factor of the Greedy algorithm \( (e - 1/e \approx 0.632) \). Otherwise, 0.632 can be used as the optimality guarantee also for heuristic search algorithms, as results produced by those algorithms are at least as good as the result of the Greedy optimization.

**DISCUSSION AND CONCLUSIONS**

In this article we have discussed the problem of finding optimal locations for traffic monitoring units. For this problem, known to be of high complexity, we demonstrated computationally efficient approximation that is based on shortest path group betweenness centrality (GBC). We showed that there is a high correlation between traffic flow obtained using an augmented betweenness centrality (BC) measure and traffic flow obtained using state of the art traffic assignment model.

Using a comprehensive data set that covers the Israeli transportation network we have first performed a simple analysis of the network and its properties, showing that there was only a small positive correlation between the traffic flow of links and their BC when the original definition of betweenness was used. We then revised the original BC definition, demonstrating that when additional known properties of the links are taken into consideration during the network analysis (e.g., time to travel through links, capacity, etc.), a much stronger correlation can be achieved. Taking into account that routes change as a function of the traffic congestion we have shown that a significantly higher correlation can be achieved when applying BC interactively and updating network properties at each iteration. Using this method we have demonstrated correlation values of \( R^2 = 0.8698 \) and \( R^2 = 0.768 \) for traffic flow and time to travel through links respectively.

We had demonstrated that our proposed method of using BC for optimizing the locations of any (reasonable) number of monitoring units enables us to perform such optimizations using low computational resources. Furthermore, the proposed methods can also be used in order to rapidly estimate the dynamic changes of traffic flow and of optimal deployment of monitors caused by changes in the origin-destination matrix or infrastructure changes. Such a technique can be useful for traffic simulation and prediction systems, such as DynaMIT (Ben-Akiva et al., 2002), resulting in an environmental-awareness module, to be added to these systems.

Another interesting issue revolves around the stochastic nature of traffic patterns, as road conditions may be varying or be time dependent. In addition, demand patterns may also be influenced by dynamic properties such as vehicle fleet composition, variations across days and within days, and so on. In this respect, it should be noted that using our proposed method enables the dynamic addition (or removal) of monitors, as new units can be deployed/activated gradually as they are needed (due to operational requirements or budget constraints).

In addition, we should note that it is not necessary to know ahead of time the total number of monitors that will be deployed in order to find their optimal locations. Upon requirement of
additional monitors units, their locations can be suggested based on up-to-date network data as well as the current deployment (the effectiveness of this method is very close to optimal both in communication and transportation networks).

Another interesting issue to be considered is the trade off between the number of monitoring units and their quality with respect to the number of vehicles each unit can monitor simultaneously (to be denoted as the units’ sampling rate, a number ranging between 0 and 1). Note that higher sampling rates directly implies a higher cost per unit. Therefore, the overall cost of the monitoring system can be modeled as:

\[
\text{Overall cost} = \text{Cost per unit} \times \text{number of units}
\]

whereas the overall monitoring performance of the system can be modeled as:

\[
\text{Overall performance} = \text{System monitoring prediction} \times \text{sampling rate}
\]

\[= f_{BC} \times (\text{number of units}) \times f_{\text{sampling}} \times (\text{cost per unit})\]

For a given budget, the decision whether to deploy a higher number of units or to invest in units of better monitoring capabilities can be directly resolved by studying the functions \(f_{BC}\) and \(f_{\text{sampling}}\). Although the first was thoroughly studied in the previous sections, analyzing the effect of the sampling rate over the performance of the system is a much simpler task (Dolev et al., 2010). With low sampling rates, GBC becomes proportional to the sum of BC values of the group members (as the number of redundant inspections reduces with the sampling rate). We can therefore consider a guideline saying that traffic monitors with very low sampling rates can be deployed on the most central nodes in the network, even if it means deploying several monitors on the same node. However, when the overall sampling rate of monitors deployed on each node is relatively high, then the set of monitored nodes should be chosen wisely using a more rigorous execution of the optimization algorithm.

Notice that BC- and GBC-based deployments have the same utility when selecting a single monitor as expected. However, GBC-based strategy continuously improves the traffic coverage as more monitors are added with the marginal utility of each additional monitor slowly decreasing. It is interesting to note that the effectiveness of GBC-based deployment is much higher and the effectiveness of BC-based deployment is much lower in transportation networks compared to social networks, as reported for example in Tubi et al. (2007).

Our approach can also be used for a variety of other transportation-related problems, such as finding an efficient deployment of emergency response service units, that would achieve maximal coverage for a given number of units (Cheu et al., 2010).

Finally, it should be noted that the problem of finding an optimal (static or dynamic) deployment for monitoring units is also related to other monitoring problems, such as monitoring for evading targets by a flock of unmanned air vehicles (UAVs). In this problem, however, the fact that the paths of the UAVs is unconstrained (as they are flying in the air) makes the calculation of a near-optimal monitoring strategy fairly easy (Altshuler et al., 2008). A more theoretical approach to this problem that studies the complexity of possible strategies can be found in Altshuler and Bruckstein (2011). It is interesting to mentioned that in those variants as well, the topological properties of the network along which the “targets” can move significantly influences the ability of monitoring units to track them, as well as the efficiency of the latter (Altshuler et al., 2005, 2006).

REFERENCES


from the north-eastern Iberian peninsula (Catalonia, Spain). *Environmental Pollution, 140*(2), 200–219.