On a Traffic Sensing Problem

Weizhen Gu
wg04@txstate.edu

Xingde Jia
jia@txstate.edu

Department of Mathematics
Texas State University, San Marcos, Texas 78666

Abstract

Let $D = (V, E)$ be a digraph. A *flow* of D is a function $f : E(D) \to \mathbb{R}^+ \cup \{0\}$ with the property that, for every vertex v of D,

$$\sum_{e \in E_D^+(v)} f(e) = \sum_{e \in E_D^-(v)} f(e),$$

where $E_D^+(v)$ denotes the set of arcs with tail at vertex v, and $E_D^-(v)$ the set of arcs with head at vertex v. A subset F of the arc set $E(D)$ is called an *edge control set* of D if there is an arc e of D such that $f_1(e) \neq f_2(e)$ for any different flows f_1 and f_2. An edge control set F is said to be *minimal* if every proper subset of F is not an edge control set of D. Let D be a connected digraph and let A be the subset of $E(D)$ consisting of all arcs of D that do not lie on any directed cycle of D. It is proved in this paper that a subset F of $E(D)$ is a minimal edge control set of D if and only if the underlying graph of $D - (F \cup A)$ is a spanning forest of the underlying graph of D. This result has applications in traffic sensing networks.

1 Introduction

Every year many miles of new roads and highways are built and put in use in almost every urban area to accommodate the growing number of vehicles. However, the expansion of roads and highways still cannot keep up with the growth rate of the number of vehicles. More travel by more vehicles aggravates traffic congestion and causes multiple delays for drivers. According to a recent study by the Texas Transportation Institute [2], Americans spent 3.7 billion hours in travel delays in 2003, with a total cost of more than $63 billion. On average, Americans spent 47 hours stuck in traffic in 2003, while ten years earlier they were spending 40 hours a year. Traffic congestion becomes one of the major obstacles for the further development of many urban areas, affecting millions of people's daily life. Constructing more roads and highways may improve the situation, but it is very costly, and in many cases it is impossible due to the existing structures. We have to use the current road network more efficiently. To that end, we need to know the current state of traffic flows in the region. Live traffic data for a region can be used to direct traffic flow to improve traffic throughput without adding new roads. Accurate traffic data can also be used to support new road construction decisions. In order to collect accurate traffic data, sensors have to be placed on the streets and roads to measure the flow of traffic. One of the key questions in the collection of traffic data is where to place the traffic sensors. Of course one would like to use the least possible number of sensors to reduce cost and shorten the traffic data processing time.

To study the mathematical model for this traffic data collection problem, we use a directed graph, or a digraph to model a road network of a city or a region. At any given moment, we consider the traffic flow (measured by the number of cars per minute, for instance) as weights for the arcs of the digraph. A subset F of arc set of a digraph is called an edge control set if the traffic flow can be completely determined by the flow on the arcs in F. In this paper we establish a relation between minimal edge control sets of a digraph and spanning trees of underlying graph of the digraph. This relation determines not only the minimal number of arcs (sensors) to be monitored (used) in a traffic network but also the locations of these arcs (sensors).

2 Flows

Let D be a digraph. For each vertex v of D, let $E^+(v)$ (or $E^-(v)$) be the set of arcs with tail (or head) at v, i.e.,

$$\begin{aligned} E^+(v) &= \{\, uv \in E(D) \;:\; u \in V(D) \,\}, \\ E^-(v) &= \{\, vu \in E(D) \;:\; u \in V(D) \,\}. \end{aligned}$$

A function $f : E(D) \to \mathbb{R}^+ \cup \{0\}$ is called a *flow* of D if

$$\sum_{e\in E^+(v)} f(e) = \sum_{e\in E^-(v)} f(e) \quad \text{for every} \quad v \in V(D) \tag{1}$$

Flows and cycles are discussed extensively in C.-Q. Zhang's book: *Integer flows and cycle covers of graphs*,[4] which is a major reference and a source of topics and problems about flows and cycles. Let X and Y be disjoint subsets of $V(D)$. We use $[X,Y]_D$ to denote the set of arcs with tails in X and heads in Y. In case of no confusion, we simply write $[X,Y]$ instead. $[X,Y]_D$ is called an *edge cut set* of D if X and Y form a partition of $V(D)$ such that $[Y,X]_D = \emptyset$. The following useful lemma is proved in [4].

Lemma 1. *Let D be a digraph. If f is a flow of D, then for each partition (X,Y) of $V(D)$,*

$$\sum_{e\in[X,Y]} f(e) = \sum_{e\in[Y,X]} f(e).$$

Lemma 2. *Let D be a digraph. If every arc of D lies on a directed cycle of D, then there exists a flow f of D such that $f(e) \neq 0$ for every $e \in E(D)$.*

Proof. Suppose that there is a flow f of D such that $f(e) = 0$ for some arc $e \in E(D)$. Since D is a finite digraph, we can pick a flow f of D such that the number of arcs e with $f(e) = 0$ is minimal. Let

$$E_0 = \{\, e \in E(D) \;:\; f(e) = 0 \,\} \qquad \text{and} \qquad m = |E_0|.$$

Pick $e_0 \in E_0$. Note that e_0 lies on a directed cycle, say,

$$v_0, v_1, v_2, \cdots, v_t, v_0,$$

where

$$e_i = v_i v_{i+1} \quad \text{for } i = 0, \ldots, t-1, \qquad \text{and} \qquad e_t = v_t v_0.$$

We now define a function $\hat{f} : E(D) \to \mathbb{R}^+ \cup \{0\}$ by

$$\hat{f}(e) = \begin{cases} f(e) + 1 & \text{if } e = e_i,\ i = 0, 1, \ldots, \text{ or } t, \\ f(e) & \text{otherwise.} \end{cases}$$

First note that the number of arcs e with $\hat{f}(e) = 0$ is at most $m - 1$. Therefore, we only need to prove that $\hat{f}$ is a flow of D, i.e.,

$$\sum_{e\in E^+(v)} \hat{f}(e) = \sum_{e\in E^-(v)} \hat{f}(e) \qquad \text{for every} \quad v \in V(D).$$

It follows from the definition of $\hat{f}$ that we only need to verify the above equation for v_i, $i = 0, 1, \ldots, t$.

$$\begin{aligned} \sum_{e\in E^+(v_i)} \hat{f}(e) &= \sum_{\substack{e\in E^+(v_i)\\ e\neq e_i}} \hat{f}(e) + \hat{f}(e_i) = \sum_{\substack{e\in E^+(v_i)\\ e\neq e_i}} f(e) + (f(e_i) + 1) \\ &= \sum_{e\in E^+(v_i)} f(e) + 1 = \sum_{e\in E^-(v_i)} f(e) + 1 \\ &= \sum_{\substack{e\in E^-(v_i)\\ e\neq e_{i-1}}} f(e) + (f(e_{i-1}) + 1) = \sum_{\substack{e\in E^-(v_i)\\ e\neq e_{i-1}}} \hat{f}(e) + \hat{f}(e_{i-1}) \\ &= \sum_{e\in E^-(v_i)} \hat{f}(e). \end{aligned}$$

Therefore, $\hat{f}$ is also a flow. The proof is complete. □

Theorem 3. *Let D be a digraph with $e' \in E(D)$. Then $f(e') = 0$ for every flow f of D if and only if there exists an edge cut set containing e'.*

Proof. (⇐) Let $[X, Y]$ be an edge cut set that contains e'. Since X and Y form a partition of $V(D)$, by Lemma 1,

$$\sum_{e \in [X,Y]} f(e) = \sum_{e \in [Y,X]} f(e).$$

Since $[Y, X] = \emptyset$, $\sum_{e \in [Y,X]} f(e) = 0$ which forces $\sum_{e \in [X,Y]} f(e) = 0$. Noting that $f(e) \geq 0$ for all $e \in E(D)$, we have that $f(e') = 0$.

(⇒) Suppose that $f(e') = 0$ for every flow f of D, but e' does not belong to any edge cut set. Let V_1 and V_2 be subsets of $V(D)$ that form a partition of $V(D)$ with the following properties:

$$e' \in [V_1, V_2] \qquad \text{and} \qquad |[V_2, V_1]| \text{ is the least.}$$

Note that $[V_2, V_1] \neq \emptyset$. Pick an arc $e_1 = xy \in [V_2, V_1]$. Since $f(e') = 0$ for every flow f of D, by Lemma 2, e' does not lie on any directed cycle. It follows that e' and e_1 are not parallel arcs and don't lie on any directed cycle. Let $e' = uv$. Without loss of generality, we may assume that $v \neq x$ and there does not exist a directed path from v to x.

Let W_x be the subset of V_2 defined by

$$W_x = \{z \in V_2 \ : \ D \text{ contains a directed path from } z \text{ to } x\}.$$

Define

$$X = V_1 \cup W_x \cup \{x\} \qquad \text{and} \qquad Y = V_2 - (W_x \cup \{x\}).$$

Then $u \in X$ and $v \in Y$. Moreover, X and Y form a partition of $V(D)$ with $|[Y, X]| \leq |[V_2, V_1]| - 1$. This contradicts the choice of V_1 and V_2. The proof is complete. □

Corolary 4. *Let D be a digraph with $e \in E(D)$. Then e lies on a directed cycle of D if and only if there is a flow of D such that $f(e) \neq 0$.*

Proof. An arc is not contained in an edge cut set if and only if it is on a directed cycle. □

3 Minimal Edge Control Sets

Definition 5. *Let D be a digraph. A subset F of $E(D)$ is called an* edge control set *if, for any two different flows f_1 and f_2 of D, there exists an arc $e \in F$ such that $f_1(e) \neq f_2(e)$.*

In other words, a set F of arcs of D is an edge control set if and only if

$$f_1(e) = f_2(e) \text{ for all } e \in F \qquad \Rightarrow \qquad f_1(e) = f_2(e) \text{ for all } e \in E(D).$$

For any digraph D, let $G(D)$ denote the underlying graph of D.

Lemma 6. *Let D be a digraph such that every arc lies on a directed cycle. If F is an edge control set of a digraph D, then $G(D - F)$ contains no cycles.*

Proof. Suppose that $G(D - F)$ contains the following cycle:

$$C: \quad v_1,\ v_2,\ \ldots,\ v_k. \tag{2}$$

It follows from Lemma 2 that there exists a flow f of D such that $f(e) > 0$ for all $e \in E(D)$. Now we divide our proof according to whether or not C is a directed cycle in D.

Case I: C is a directed cycle. Define $f_1 : E \to \mathbb{R}^+ \cup \{0\}$ by

$$f_1(e) = \begin{cases} f(e) + 1, & \text{if } e \in E(C), \\ f(e), & \text{otherwise.} \end{cases}$$

Since C does not contain any arcs in F, we have that $f(e) = f_1(e)$ for all $e \in F$. We now show that f_1 is also a flow of D. We only need to verify, for every $i : 1 \le i \le k$,

$$\sum_{e \in E_D^+(v_i)} f_1(e) = \sum_{e \in E_D^-(v_i)} f_1(e).$$

Denote $e' = v_{i-1}v_i$ and $e'' = v_iv_{i+1}$. Then

$$\begin{aligned}\sum_{e \in E_D^+(v_i)} f_1(e) &= f_1(e') + \sum_{\substack{e \in E_D^+(v_i)\\ e \neq e'}} f(e) = 1 + \sum_{e \in E_D^+(v)} f(e)\\ &= 1 + \sum_{e \in E_D^-(v)} f(e) = f_1(e'') + \sum_{\substack{e \in E_D^-(v_i)\\ e \neq e''}} f(e) = \sum_{e \in E_D^-(v_i)} f_1(e).\end{aligned}$$

Therefore, f_1 is a flow of D. Recall that f and f_1 are equal on F but $f \neq f_1$, which contradicts the assumption that F is an edge control set of D.

Case II: C is not a directed cycle of D. Assume that

$$v_{i_1},\ v_{i_2},\ \ldots,\ v_{i_h}, \tag{3}$$

are the vertices of C that have degree in $G(D)$ larger than two, which are arranged in the same order as they are on the cycle (2). Since every edge is on a directed cycle, then the path of C between any two consecutive vertices in the above sequence (3) is directed. For the simplification of notation, we assume that the pathes change directions at *every* vertex v_{i_j}, $j = 1, 2, \ldots, h$. Note that h must be an even integer. Assume, without loss of generality, that the pathes of C have the following indicated directions (see Figure 1).

$$P_1 : v_{i_1} \to v_{i_2},\quad P_2 : v_{i_3} \to v_{i_2},\quad \ldots,\quad P_{h-1} : v_{i_{h-1}} \to v_{i_h},\quad P_h : v_{i_1} \to v_{i_h}.$$

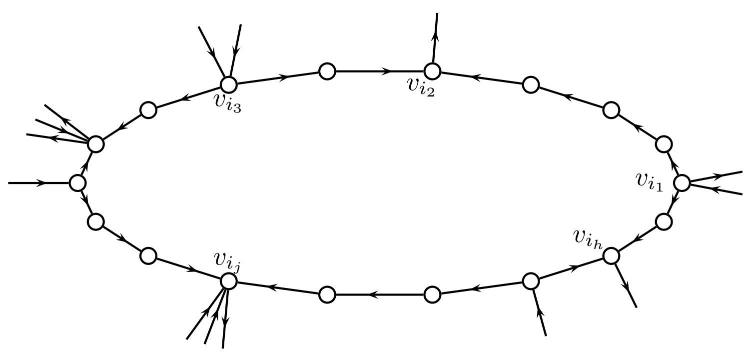

Figure 1: C is not a directed cycle.

Define $f_1 : E(D) \to \mathbb{R}^+ \cup \{0\}$ by

$$f_1(e) = \begin{cases} f(e) + (-1)^j, & \text{if } e \in E(P_j),\ j = 1, 2, \ldots, h,\\ f(e), & \text{otherwise}\end{cases}$$

Clearly, $f_1 \neq f$. Suppose that v_i is on the path P_j. If $\deg_{G(D)}(v_i) = 2$, then v_i is an interior vertex of P_j. Let e' and e'' be the edges incident with v_i. Since f is a flow of D, we have $f(e') = f(e'')$. Hence

$$f_1(e') = f(e') + (-1)^j = f(e'') + (-1)^j = f_1(e'').$$

If $\deg_D(v_i) > 2$, then we may assume that $v_i = v_{i_j}$ is the end vertex of directed pathes P_{j-1} and P_j. Suppose that $e' \in E(P_{j-1})$ and $e'' \in E(P_j)$ are the edges incident with v_{i_j}. We may assume, without loss of generality, that j is even. Then both e' and e'' are in $E_D^+(v_{i_j})$. Therefore,

$$\begin{aligned}\sum_{e\in E_D^+(v_{i_j})} f_1(e) &= f_1(e') + f_1(e'') + \sum_{\substack{e\in E_D^+(v_{i_j})\\ e\neq e',e''}} f_1(e)\\ &= f(e') + (-1)^{j-1} + f(e'') + (-1)^j + \sum_{\substack{e\in E_D^+(v_{i_j})\\ e\neq e',e''}} f(e) = \sum_{e\in E_D^-(v_{i_j})} f_1(e).\end{aligned}$$

Hence f_1 is also a flow of D, which is a contradiction. Therefore, $G(D-F)$ contains no cycles. The proof is complete. □

An edge control set F of D is said to be *minimal* if any proper subset of F is not an edge control set of D. The following lemma provides a necessary condition for an edge subset of a digraph to be a minimal edge control set.

Lemma 7. *Let D be a digraph with $F \subset E(D)$. If F is a minimal edge control set of D, then every arc in F lies on a directed cycle of D.*

Proof. Suppose that F contains an arc e that does not belong to any directed cycle of D. Let f_1 and f_2 be flows such that $f_1(e) = f_2(e)$ for all $e \in F - \{e^*\}$. By Corollary 4, $f_1(e^*) = f_2(e^*) = 0$. Since F is a minimal edge control set, $f_1(e) = f_2(e)$ for all $e \in E(D)$. Hence $F - \{e^*\}$ is also an edge control set of D, contradicting the minimality of F. □

Theorem 8. *Let D be a connected digraph with $F \subseteq E(D)$. Suppose that each arc of D lies on a directed cycle. Then F is a minimal edge control set of D if and only if $G(D-F)$ is a spanning tree of $G(D)$.*

Proof. Suppose that F is a minimal edge control set of D. By Lemma 6, $G(D-F)$ has no cycle. To prove $G(D-F)$ is a spanning tree of G, it is sufficient to prove that $G(D-F)$ is connected. If $G(D-F)$ is disconnected, then F contains an edge cut set, namely F^*. Pick an edge e^* from F^*. By Theorem 3, $f(e^*) = 0$ for any flow f. It follows from Corollary 4 and Lemma 7 that e^* does not lie on any directed cycle and hence $e^* \notin F$. This contradicts to the choice of e^*. Therefore, $G(D-F)$ is a spanning tree of G.

Conversely, we assume that $G(D-F)$ is a spanning tree of D. Let f_1 and f_2 be the flows such that $f_1(e) = f_2(e)$ for all $e \in F$. Let T be the graph obtained from $D-F$ by deleting all edges e so that $f_1(e) = f_2(e)$. Since $G(D-F)$ is a tree, $G(T)$ is a forest. We assume that T contains an arc. Pick a leaf e' from T. Without loss of generality, we may assume that $e' = uv$, an arc from u to v. Note that both $E_D^+(u)$ and $E_D^- - \{e'\}$ are subsets of F. So

$$\begin{aligned}\sum_{e\in E_D^+(u)} f_1(e) &= \sum_{e\in E_D^+(u)} f_2(e) = \sum_{e\in E_D^-(u)} f_2(e)\\ &= f_2(e') + \sum_{\substack{e\in E_D^-(u)\\ e\neq e'}} f_2(e) = f_2(e') + \sum_{\substack{e\in E_D^-(u)\\ e\neq e'}} f_1(e)\\ &= f_2(e') - f_1(e') + \sum_{e\in E_D^-(u)} f_1(e) = f_2(e') - f_1(e') + \sum_{e\in E_D^+(u)} f_1(e).\end{aligned}$$

Thus $f_2(e') - f_1(e') = 0$. So $f_1(e') = f_2(e')$. This fact contradicts to the definition of T. Therefore, $E(T) = \emptyset$. In other word, T consists of all isolated vertices. It follows that $f_1(e) = f_2(e)$ for all $e \in E(D)$, which implies that F is an edge control set of D.

To complete the proof, we need to prove that F is a minimal edge control set. Let $A \subset F$ be a minimal edge control set of G. By the first part of the theorem, $G(D-A)$ is a spanning tree. Since $G(D-F)$ is also a spanning tree of $G(D)$ and all spanning trees of $G(D)$ have the same size, $|E(G(D-A))| = |E(G(D-F))|$ implies that $|F| = |A|$. Therefore, F is a minimal edge control set of D. □

Corolary 9. *Let D be a connected digraph such that each arc lies on a directed cycle. If F is a minimal edge control set of D, then $|F| = |E(D)| - |V(D)| + 1$.*

Proof. This follows from Theorem 8 and the fact that every spanning tree of $G(D)$ contains $|V((G(D))|-1 = |V(D)|-1$ edges. $\square$

Corolary 10. *Let D be a connected planar digraph with r regions. If each arc lies on a directed cycle, then every minimal edge control set contains $r-1$ arcs.*

Proof. This follows immediately from Corollary 9 and Euler's formula $|V(G(D))| - |E(G(D))| + r = 2$. $\square$

The following theorem is a generalization of Theorem 8 and Corollary 9.

Theorem 11. *Let D be a connected digraph. Let A be the subset of $E(D)$ consisting of all arcs that do not lie on any directed cycles. Then subset F of $E(D)$ is a minimal edge control set of D if and only if $G(D-(F\cup A))$ is a spanning forest of $G(D)$. Furthermore,*

$$|F| = |E(D)| - |V(D)| - |A| + k(A),$$

where $k(A)$ is the number of components of $G(D-A)$.

Proof. By the definition of A and Lemma 7, $F\cap A = \emptyset$. Let $k = k(A)$ and let $D_1, D_2, \ldots, D_k$ be k components of $D-A$. For any i with $1 \le i \le k$, let $F_i = F \cap E(D_i)$. Note that if an arc of D lies on a directed cycle C of D, then any arc of C is not in A. Thus, C is also a directed cycle of $D-A$. Furthermore, C is also a directed cycle of D_i, if $e \in E(D_i)$. Thus for any integer i with $1 \le i \le k$, D_i is a connected digraph with the property that each arc lies on a directed cycle of D_i. By Theorem 8, F_i is a minimal edge control set of D_i if and only if $G(D_i - F_i)$ is a spanning tree of $G(D_i)$. It is clear that $F = \cup_{i=1}^{k} F_i$ is a minimal edge control set if and only if $G(D-(F\cup A))$ is a spanning forest of $G(D)$. Moreover by Corollary 9,

$$|F_i| = |E(D_i)| - |V(D_i)| + 1 \qquad \text{for} \quad 1 \le i \le k.$$

Therefore,

$$|F| = \sum_{i=1}^{k} |F_i| = \sum_{i=1}^{k} (|E(D_i)| - |V(D_i)| + 1) = |E(D)| - |V(D)| - |A| + k(A).$$ $\square$

4 An Example

To find a minimal edge control set for a digraph, we only need to first delete the arcs that are not on any directed cycles and then find a spanning forest for its underlying graph (not necessarily connected) of the resulting digraph. There are many algorithms to construct a spanning tree for a connected graph. The following example gives a direct construction for a minimal edge control set, which is basically the break-cycle method for constructing a spanning tree.

Let D be a digraph such that each arc is on a directed cycle. A subset F of arcs is to be constructed by the following steps.

Step 1. Let $F := \emptyset$ and $H := D$.

Step 2. While $E(H) \neq \emptyset$, pick any edge $e \in E(H)$, let

$$F := F \cup \{e\}, \qquad H := H - C_H(e),$$

where $C_H(e)$ is the set of cut edges of $G(H-e)$.

It is fairly easy to show that $G-F$ is a spanning tree of $G(D)$. Therefore, F is a minimal edge control set of D. This direct construction is explained in the following example.

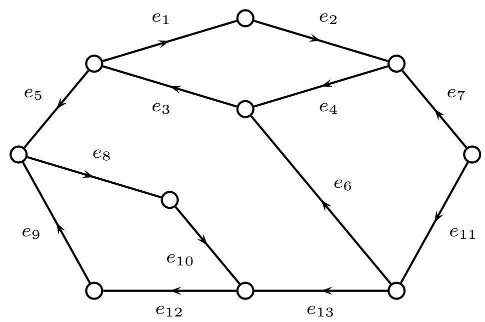

Figure 2: A digraph D with each arc being in a directed cycle.

We begin with $F = \emptyset$ and $H := D$ and then choose an edge, say e_1. Then $F = \{e_1\}$, $C_H(e_1) = \{e_1, e_2\}$, and $H := H - C_H(e_1)$ shown in Figure 3.

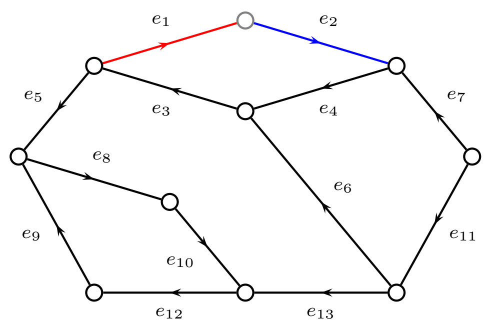

Figure 3: The first digraph H and the corresponding F.

Since $E(H) \neq \emptyset$, we select e_3 to be added to F. So $F = \{e_1, e_3\}$, $C_H(e_3) = \{e_3, e_5, e_{13}\}$, and $H := H - C_H(e_3)$ displayed in Figure 4, where the arcs in F are in red, while other removed arcs are in blue.

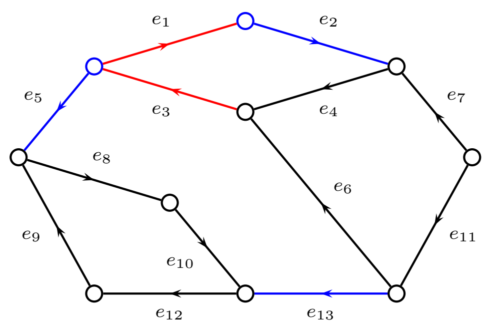

Figure 4: The second digraph H and the corresponding F.

We now add e_4 to F and then $C_H(e_4) = \{e_4, e_6, e_7, e_{11}\}$. The remaining subgraph $H := H - C_H(e_4)$ is shown in Figure 5.

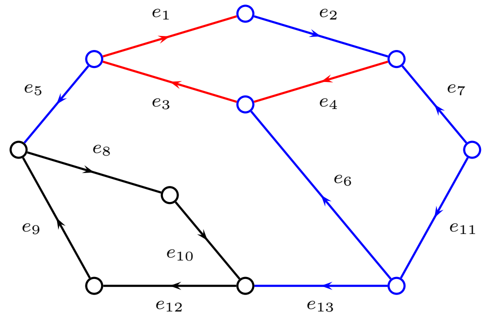

Figure 5: The third digraph H and corresponding F.

Finally, we add e_8 into F. Therefore, $F = \{e_1, e_3, e_4, e_8\}$, $C_H(e_8) = \{e_8, e_9, e_{10}, e_{12}\}$, and the new H consists of all isolated vertices, i.e , $E(H) = \emptyset$. Therefore, $F = \{e_1, e_3, e_4, e_8\}$ is a minimal edge control set of G.

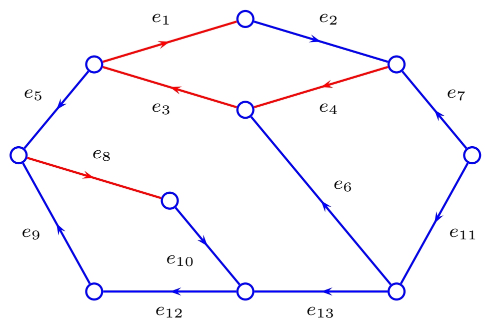

Figure 6: The constructed minimal edge control set F (red) and the spanning tree (blue).

If we take this digraph as a model of a road network, the possible locations for placing sensors are on the red arcs and the minimum number of sensors needed to collect the traffic flow information for the network is four. Note that the weights (flows) on arcs of a digraph are not considered when constructing a minimal edge control set in the way described above. Therefore the sum of the weights of arcs in a minimal edge control set could be relatively large or small. However, to collect traffic flow information, it is reasonable to place sensors on streets that have a constantly large traffic flow. In order to do so, we should find a minimum spanning tree of the underlying graph of a digraph, instead of finding an arbitrary spanning tree.

The red arcs on each of the following digraphs form a minimal edge control set of D. But one edge control set has the largest sum of the weights on the red arcs, while the other has the smaller sum of the weights on the red arcs.

Acknowledgement. The authors would like to thank Professor Vadim Ponomarenko for helpful suggestions which led a major revision of the paper. The authors would also like to thank Professor Jian Shen for several helpful discussions.

References

[1] G. Chartrand and P. Zhang, *Introduction to Graph Theory*, The Walter Rudin Student Series in Advanced Mathematics, McGraw Hill, Inc., 2005.

[2] Texas Transportation Institute, *The 2005 Urban Mobility Report*, May 2005.

[3] D. B. West, *Introduction to Graph Theory*, 2nd Ed., Prentice Hall, Inc., New Jersey, 2001.

[4] C.-Q. Zhang, *Integer Flows and Cycle Covers of Graphs*, Monographs and Textbooks in Pure and Applied Mathematics **205**, Marcel Dekker, Inc., 1997.