The Lattice Structure of Flow in Planar Graphs

Samir Khuller *
Institute for Advanced Computer Studies
University of Maryland
College Park, MD 20742

Joseph Naor †
Computer Science Department
Technion
Haifa 32000, Israel

Philip Klein ‡
Computer Science Department
Brown University
Providence, RI 02912

Abstract

Flow in planar graphs has been extensively studied, and very efficient algorithms have been developed to compute max-flows, min-cuts and circulations. We show intimate connections between solutions to the planar circulation problem and with "consistent" potential functions in the dual graph. We show that the set of integral circulations in a planar graph very naturally form a distributive lattice whose maximum corresponds to the shortest path tree in the dual graph. We further characterize the lattice in terms of unidirectional cycles with respect to a particular face called the root face. We show how to compactly encode the entire lattice, and also show that the set of solutions to the min-cost flow problem forms a sub-lattice in our lattice.

*Partially supported by NSF grant CCR-8906949, CCR-9103135 and CCR-9111348. Part of this research was done while this author was supported by an IBM Graduate Fellowship at Cornell University.

†Part of this work was done while this author was at Stanford University, and supported by contract ONR N00014-88-K-0166.

‡Research supported by NSF grant CCR-9012357.

1. Introduction

Maximum flow has been one of the most well studied problems in the area of algorithms (both in the fields of Computer Science and Operations Research) over the last thirty years. It has applications in solving efficiently a large set of problems, e.g., many VLSI problems, transportation problems and communication networks.

Flow in planar graphs has been extensively studied, and very efficient algorithms have been developed to compute max-flows, min-cuts and circulations. There is a wealth of ideas in solving these problems efficiently for this class of graphs [FF, Ha, HJ, IS, Jo, JV, MN, Re, KN]. (The algorithms vary for different versions of the same basic flow problem.) Very efficient parallel and sequential algorithms can also be developed by exploiting the planar structure of the graph.

Recently, Miller and Naor [MN] have pointed out that the most general formulation of the maximum flow problem in planar graphs must allow for the existence of many sources and sinks. Unlike the case of arbitrary graphs, where sources and sinks can be merged, for planar networks there is no obvious reduction of the multiple source/sink problem to a single source/sink problem. Miller and Naor [MN] further showed that the case where the demands of the sources and sinks are fixed is equivalent to a *circulation problem* (with lower bounds on edge capacities). They also gave an efficient algorithm for computing a circulation.

Our objective is to study the *structure* of the set of *integral solutions* to the circulation problem for planar graphs. We review the relation between solutions to the planar circulation problem and consistent potential functions in the dual graph and show that there is a 1-1 correspondence between them. We then show that the set of circulations in a planar graph very naturally form a distributive lattice (under an appropriate definition of meet and join operations for the lattice). We further characterize the lattice in terms of unidirectional cycles with respect to a particular face called the root face. We show that the top (bottom) element of the lattice is the circulation in which there are no clockwise (counterclockwise) residual cycles around the root face. It turns out that the flow functions computed by [Ha], [HJ], [Jo] and [MN] correspond to the top element of the circulation lattice. (This is essentially a matter of notation; if we reverse the direction of the dual edges, their algorithms will be computing the bottom element in the lattice.) It is interesting to note that if a planar graph also contains vertex capacities, i.e., there is a limit on the amount of flow that is allowed to go through a vertex, then the set of feasible circulations does *not* form a lattice (see Subsection 3.4).

The lattice representing all feasible circulations is clearly of *exponential* size, since there are exponentially many solutions to the circulation problem. We provide a compact encoding of the entire lattice by providing a directed acyclic graph such that the predecessor-closed subsets of this partial order correspond to elements in the lattice. Although this dag may be large, its size depends on the maximum edge-capacity – we can represent it succinctly, in polynomial size. This compact encoding of the partial order provides in turn a compact encoding of the lattice elements.

The *minimum cost circulation* problem is that of obtaining a circulation of minimum cost in a network whose edges have both capacities and costs per unit of flow. The problem is equivalent to the transshipment problem and has wide applicability to a variety of optimization problems [AMO]. One of the motivations for the research reported here is to investigate new approaches to solving the minimum cost circulation problem in planar networks. For such networks, we interpret the cost function as a function on the lattice, and we show that it is a modular function. It follows that the set of solutions to the minimum cost circulation problem form a sublattice. We show that this sublattice consists of the feasible circulations for a network derived from the original network. This allows us to provide a succinct representation for the set of minimum cost circulations as well. Minimizing a modular function over a lattice is a well known problem in operations research and can be solved by computing the minimum cut (max flow) in a network obtained from the dag corresponding to the circulation lattice. However, this approach does not directly yield a new polynomial-time min-cost circulation algorithm, since the dag is so large. At the end of this paper, we briefly discuss two possible approaches to obtaining such an algorithm.

Other examples of problems whose solution set has similar structure, are the stable marriage problem and the minimum cut problem. Picard and Queyranne [PQ] have shown that the set of all minimum cuts forms a distributive lattice where the join and meet operations are defined as intersection and union respectively. The structure of the solution set of the stable marriage problem has been extensively investigated in the book by Gusfield and Irving [GI]. They show that the set of all stable marriages also forms a distributive lattice, and show that a compact encoding of the entire solution set is possible. They provide many applications of the lattice structure, e.g., computing an egalitarian stable marriage solution.

2. Preliminaries and Terminology

We begin by defining the *circulation* problem. Consider a directed graph G, with each edge e having an integral *lower* and *upper* bound on its capacity, denoted respectively by $\ell(e)$ and $u(e)$ ($[\ell, u]$). When we speak of the capacity of an edge without specifying whether it is a lower or upper capacity, we mean its upper capacity.

We are required to find a flow function $f : E \rightarrow Z$ that is feasible in that the following two conditions are satisfied:

capacity constraints: $\forall e \in E : \ell(e) \leq f(e) \leq u(e)$ (The flow on each edge is between the lower and upper bounds on its capacity.)

conservation constraints: $\forall v \in V : \sum_{e \in in(v)} f(e) = \sum_{e \in out(v)} f(e)$ (The flow into each node equals the flow out of the node.)

The circulation problem is that of finding a feasible flow function (such a flow function may not even exist). In the *maximum flow* problem, two distinguished vertices are added to the graph, a source and a sink, and the aim is to maximize the amount of flow entering the sink. Note that a flow problem in which the flow value is prespecified can be reduced to a circulation problem.

From now on we will be restricting our attention to planar graphs. Let $G = (V, E)$ be a directed embedded planar graph. The graph partitions the plane into connected regions called *faces*. For each edge $e \in E$, let $D(e)$ be the corresponding *dual edge* connecting the two faces bordering e. Let $D(G) = (F, D(E))$ be the *dual graph* of G where F is the set of faces of G and $D(E) = \{D(e) | e \in E\}$. The dual graph is planar too, but may contain self loops and multiple edges. We refer to graph G as the *primal graph*.

$-\ell$

$[\ell, u]$

u

Nodes of the dual graph

Nodes of the primal graph

Figure 1: Construction of directed dual graph

There is a 1-1 correspondence between primal and dual edges; the direction of a primal edge e induces a direction on $D(e)$. We use a right-hand rule: if the right hand's thumb points in the direction of e, then the index finger points in the direction of $D(e)$ (with the palm facing downwards). We refer to dual edges as capacitated as well, where the capacity of edge $D(e)$ is equal to that of edge e (see Fig. 1).

We have the following equivalence rules that relate the orientation of an edge $e = (v \rightarrow w)$, the sign of its flow $f(e)$, and its lower and upper capacity bounds.

1. The edge $v \rightarrow w$ with flow $f(e)$ is equivalent to the edge $w \rightarrow v$ with flow $-f(e)$.

2. The edge $v \to w$ with capacities $[\ell, u]$ is equivalent to the edge $w \to v$ with capacities $[-u, -\ell]$.

3. The edge $v \to w$ with capacities $[\ell, u]$ is equivalent to two antiparallel edges: $v \to w$ of capacities $[0, u]$ and $w \to v$ with capacities $[0, -\ell]$.

4. Let e_1 and e_2 be two parallel edges that are oriented in the same direction with capacities $[\ell_1, u_1]$ and $[\ell_2, u_2]$ respectively. The two edges can be replaced by one edge with capacity $[\ell_1 + \ell_2, u_1 + u_2]$ and flow $f(e_1) + f(e_2)$.

The *residual graph* is defined with respect to a given circulation. Let $e = (v \to w)$ be an edge with capacities $[\ell, u]$ and flow f. In the residual graph, e is replaced by two darts, $v \to w$ with capacities $[0, u - f]$ and $w \to v$ with capacities $[0, f - \ell]$. A directed cycle is said to be *residual* with respect to the given circulation if every edge in the cycle has positive upper capacity in the residual graph.

Miller and Naor [MN] have shown that for planar graphs, the maximum flow problem should be formulated with respect to many sources and sinks. They show how to reduce this problem to a circulation problem with lower bounds on the edges if the demands of the sources and sinks are given. This is done by returning the flow back from the sinks to the sources via a spanning tree.

An important tool for computing flow functions in planar graphs is the notion of a potential function $p : F \to Z$ in the dual graph. This function was first introduced by Hassin [Ha] and its usage was later elaborated by [HJ],[Jo] and [MN]. Let e be an edge in the graph G, and let $D(e) = (g, h)$ be its corresponding edge in the dual graph such that $D(e)$ is directed from g to h. The potential difference over e is defined to be $p(h) - p(g)$. The following proposition, proved in [Ha] and [Jo], can be easily verified.

Proposition 2.1: *Let $C = c_1, \ldots, c_k$ be a cycle in the dual graph and let $f_1, \ldots, f_k$ be the potential differences over the cycle edges. Then* $\sum_{i=1}^{k} f_i(e) = 0$.

It follows from the proposition that the sum of the potential differences over all the edges adjacent to a primal vertex is zero.

A potential function is defined to be *consistent* if the potential difference over each edge is between the upper and lower bounds on the capacity (i.e., $\ell \leq p(h) - p(g) \leq u$). Such a potential function induces a circulation in the graph by defining the flow on an edge as the potential difference over it. Clearly the flow on the edge satisfies the capacity constraints; by using the previous proposition it is easy to see that the flow conservation constraints are also satisfied. Once we fix the potential of some particular face as zero, all the other potentials can be normalized with respect to this face. We will assume that all consistent potential functions are normalized and we call the face whose potential is set to zero the *root face*. We will assume that the planar embedding is such that the root face is the infinite face.

How is a consistent potential function computed ? Let us consider the dual graph where each edge $D(e)$ with capacity $[\ell, u]$ is split by Rule 3 into two antiparallel edges, where one edge has capacity u, and the other capacity $-\ell$. Miller and Naor [MN] show that if a solution to a circulation problem exists, then there cannot be negative cycles in the dual graph. Hence, a natural way for computing a consistent potential function would be the following: choose an arbitrary face as the root face; the potential of face h is defined to be the length of the shortest path in the dual graph from the root face to h. It follows from properties of shortest paths that the resulting potential function is indeed consistent. We will refer to this potential function as the *shortest path* potential function.

It is not hard to see that there is a one-to-one correspondence between consistent (and normalized) potential functions and circulations. Given a legal circulation C, a corresponding potential function can be constructed. To do that, the capacity of every edge is replaced by its actual flow in C. That is, if the flow on edge e is of value f_e, then edge $D(e)$ is replaced by two parallel edges in opposite directions, where one edge has capacity f_e, and the other has capacity 0. The potential of face h is the length of the shortest path in the dual graph from the root face to h. It is not hard to see that this potential function induces circulation C.

We shall henceforth view a potential function as a vector where the entries correspond to the potentials of the faces, and the potential of the *root face* (an arbitrary but fixed face) is always equal to zero. Since there is a one-one correspondence between circulations and potential vectors, we will use both terms to refer to a circulation. We also assume that all potential values are integral, as we are interested only in integral solutions to the circulation problem.

3. The lattice structure

Our aim in this section is to investigate the structure of the set of legal circulations in G. We will show that this set forms a distributive lattice and also explore its structure. Given two consistent vectors, P_1 and P_2, we say that $P_1 \geq P_2$ if for all components i, $P_1(i) \geq P_2(i)$. We say that circulation C_1 *dominates* C_2, if for their corresponding potential vectors, P_1 and P_2, $P_1 \geq P_2$. We use the term $\mathcal{P}$ to refer to set of all consistent potential vectors. It is easy to see that $\mathcal{P}$ is a partial order under the dominance relation (also written as $(\mathcal{P}, \preceq)$).

We now show that the partial order $(\mathcal{P}, \preceq)$ is, in fact, a *distributive lattice*.

A distributive lattice is a partial order in which

1. Each pair of elements has a greatest lower bound, or *meet*, denoted by $a \wedge b$, so that $a \wedge b \preceq a, a \wedge b \preceq b$, and there is no element c such that $c \preceq a, c \preceq b$ and $a \wedge b \prec c$.

2. Each pair of elements has a least upper bound, or *join*, denoted by $a \vee b$, so that $a \preceq a \vee b, b \preceq a \vee b$, and there is no element c such that $a \preceq c, b \preceq c$ and $c \prec a \vee b$.

3. The *distributive* laws hold, namely $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$ and $a \wedge (b \vee c) = (a \wedge b) \vee (a \wedge c)$.

We show that $(\mathcal{P}, \preceq)$ is a distributive lattice by presenting appropriate definitions for meet and join.

Given two circulations C_1 and C_2 (represented as P_1 and P_2), we define the meet as the circulation induced by the potential vector $P_m = \min(P_1, P_2)$. Clearly, the face at zero potential in both circulations stays at zero potential. Every face g is assigned a potential equal to $\min(P_1(g), P_2(g))$ where $P_i(g)$ is the potential of g in C_i. Similarly, the join is defined as $P_j = \max(P_1, P_2)$.

The following theorem shows that P_m and P_j are consistent potential vectors assuming that P_1 and P_2 are consistent.

Theorem 3.1: *The partial order $(\mathcal{P}, \preceq)$ is a distributive lattice, with the meet and join defined appropriately.*

Proof: We first show that the meet and join are consistent potential assignments. Let g and h be faces in the dual graph bordering primal edge e. The potential across e is $p_1(h) - p_1(g)$ and $p_2(h) - p_2(g)$ respectively in each circulation. If $p_1(h) \leq p_2(h)$ and $p_1(g) \leq p_2(g)$, then the meet is clearly consistent. (Similarly, if p_2 is the smaller potential for both g and h.) If $p_1(h) \leq p_2(h)$ and $p_2(g) \leq p_1(g)$, then it follows that $\ell \leq p_1(h) - p_2(g) \leq u$ (since $p_1(h) - p_1(g) \geq \ell$, and $p_2(h) - p_2(g) \leq u$). The last case is when $p_2(h) \leq p_1(h)$ and $p_1(g) \leq p_2(g)$, and it follows that $\ell \leq p_2(h) - p_1(g) \leq u$ (since $p_2(h) - p_2(g) \geq \ell$ and $p_1(h) - p_1(g) \leq u$). Hence P_m is a consistent potential assignment.

The proof that P_j is a consistent potential assignment is almost identical. It is also easy to see that they are the g.l.b and l.u.b respectively. This establishes that $(\mathcal{P}, \preceq)$ is a lattice.

Let a, b and c be any integers. Then, $\min(a, \max(b, c)) = \max(\min(a, b), \min(a, c))$ and $\max(a, \min(b, c)) = \min(\max(a, b), \max(a, c))$. Hence, the distributive laws hold for the lattice $\mathcal{P}$. $\square$

It is easy to see that a lattice has a unique minimum and maximum, P_b and P_t, referred to as *bottom* and *top* respectively. We now provide a simple characterization for them. Let us denote by P the shortest path potential vector, in which the potential of a face is exactly its distance from the root face in the dual graph. The following lemma shows that P corresponds precisely to the top of the lattice.

Lemma 3.2: *The potential vector P_t is equal to P.*

Proof: The shortest path problem can be cast as a linear program, with a variable x_h for each face h, in which the objective is to maximize $\sum_h x_h$, subject to $x_r = 0$, where r is the root face, and subject to inequalities $x_g \leq x_h + u_e$ for each edge $e = h \rightarrow g$ of the dual graph. Any vector x satisfying these constraints is a circulation. The top element of the lattice clearly maximizes the objective function over all lattice elements. □

The following lemma, which characterizes the bottom of the lattice, follows by symmetry.

Lemma 3.3: *Let a potential vector P be computed as follows: The potential of face h in P is the length of the shortest path from h to r, the root face, multiplied by -1. Then, the vector P is equal to P_b.*

3.1. Eliminating lower bounds

The existence of the lattice provides us with a simple way of getting rid of lower bounds on edges. To do that, we define a new lattice $\mathcal{P}'$ by normalizing the vectors in $\mathcal{P}$ with respect to P_b, the bottom element of the lattice. Each vector $P \in \mathcal{P}$ is replaced by a new vector P', where $P' = P - P_b$ and subtraction is performed componentwise. This is the same as computing the residual graph with respect to P_b.

Lemma 3.4: *Let G' be the residual graph of G with respect to the circulation P_b. Then*
(1) the lower bounds on the capacity of the edges in G' are zero, and
(2) the lattice of feasible circulations in G' is isomorphic to that in G.

Proof: By the additivity property of flow, each circulation $P \in \mathcal{P}$ can be written as the sum of two circulations, P_b and some other circulation Q. Hence, the lemma follows. □

3.2. Unidirectional cycles and the lattice

In this section we establish a connection between the lattice and unidirectional cycles. Recall that we assumed the planar embedding was such that the infinite face is the root face. Each simple cycle divides the sphere into two nonempty disjoint sets of faces, called regions. The region containing the root face is designated the *exterior* region; the other region is *interior.* In a traversal of a directed cycle, all faces that border the cycle on its right are in the same region, the cycle's *right-hand region.*

Definition 1: *A directed cycle is* clockwise *if the cycle's right-hand region is interior. Otherwise, the cycle is* counterclockwise.

Let us adopt the following convention that follows from the right-hand rule defined in Section 2. Pushing positive flow through a directed cycle C is equivalent to increasing the potentials of the faces in C's right-hand region.

A circulation is said to be *maximal* in the clockwise direction ("clockwise maximal," for short) if there are no clockwise residual cycles with respect to the circulation. "Maximal in the counterclockwise direction" is defined similarly.

We begin by characterizing the top and bottom of the lattice.

Theorem 3.5: *A circulation is clockwise maximal if and only if it corresponds to P_t. A circulation is counterclockwise maximal if and only if it corresponds to P_b.*

Proof: We consider the first statement; the second follows by symmetry. First we show that P_t is clockwise maximal. Let Γ be any clockwise cycle; we shall show that Γ is not residual with respect to P_t because some edge of Γ has zero residual capacity.

By Lemma 3.2, P_t is the shortest path vector. Let T be the shortest path tree in the dual graph, rooted at the root face. Since T spans all faces, there must be some face h in the interior of Γ whose parent g in T is in the exterior of Γ. Then $P(h) = P(g) + b$, where b is the capacity of the edge $D(e) = g \to h$ in the dual graph. Thus b is also the capacity of the edge $e \in \Gamma$ in the primal graph. But the flow $f(e)$ defined by P is $P(h) - P(g) = b$, so the residual capacity is zero.

Conversely, suppose P is a circulation with respect to which there exists a clockwise residual cycle Γ. Since Γ is clockwise, the interior does not contain the root face. Since Γ is residual, we can therefore increase the potentials of all faces in the interior by some positive amount without violating the constraints. Hence P is not the top element of the lattice.

$\square$

It is tempting to believe that the dominance relation in the lattice can be stated in terms of saturating clockwise cycles. That is, if $P_1 < P_2$, then circulation P_2 can be obtained from P_1 by saturating clockwise cycles. Unfortunately, the following counterexample shows that this is not true. Let c_1 and c_2 be clockwise cycles such that c_1 is contained in the interior of c_2. We construct two circulations, P_1 and P_2, such that $P_1 < P_2$. To construct P_1, take P_b and push one unit of flow in the cycle c_1 in the clockwise direction. To construct P_2, take P_b and push one unit of flow in the cycle c_2 in the clockwise direction. Obviously, $P_1 < P_2$, but the only way to obtain circulation P_2 from P_1 is to push a unit of flow in the cycle c_2 in the clockwise direction and in the cycle c_1 in the counterclockwise direction.

However, in the next section we will show that every circulation can be obtained from P_b by pushing flow through a set of clockwise cycles.

3.3. The region growing algorithm

In this section we will give a generic algorithm for obtaining any circulation P, from P_b the bottom circulation of the lattice, by saturating only clockwise cycles. We assume that circulation P is given as input to the algorithm. This algorithm will be used to prove that the difference between P_b (or P_t) and any other circulation is a unidirectional set of cycles. The algorithm can also be used to obtain the top or bottom elements of the lattice from any given circulation. In Section 5.2, we briefly discuss a possible approach to computing a minimum-cost circulation based on the idea of the region growing algorithm.

The Region-Growing Algorithm:

1. Let $Q \leftarrow P_b$
2. Let R be the set of faces on which Q and P agree; $\Delta \leftarrow \min_i\{P(i) - Q(i) | i \in F - R\}$
3. For all faces $f \in F - R : Q(f) \leftarrow Q(f) + \Delta$
4. If $Q \neq P$, then Goto Step 2.

The correctness of the algorithm is trivial, and clearly the algorithm can be modified to use the top of the lattice as the initial value of Q. We now prove that the algorithm has an interesting property.

Lemma 3.6: *The region R remains connected during all stages of the algorithm.*

Proof: By the construction of Lemma 3.4, we can assume without loss of generality that all capacity lower bounds in G are zero, and that P_b is the all-zeroes circulation. Let f be the assignment of flows to edges defined by the circulation P. Let G' be the graph obtained from G by setting upper bounds as follows:

$$u'(e) = \begin{cases} f(e) & \text{if } f(e) > 0 \\ 0 & \text{otherwise} \end{cases}$$

It is easy to check that P is the shortest path potential vector for G'. Thus the potential P assigns to a face f is the distance of f from the root face in a graph with nonnegative edge-lengths. Now consider the region-growing algorithm. Because we started with the zero circulation, at any point t in the algorithm, R consists of all faces whose potentials in P are no more than some value, say v_t. That is, R consists of faces whose distance from the root is no more than v_t. Clearly R is therefore connected. □

The last lemma provides us with the following view of the region-growing algorithm. Initially, R only contains the root face. At each step: push a certain amount of flow (Δ) on the boundary of R in the clockwise direction; annex to R the faces bordering it whose potential has reached the desired value.

For example, to obtain P_t from from any circulation P, we will run the algorithm "backwards". Initially, R only contains the root face. At each step, the boundary of R is saturated, and the faces bordering saturated edges are annexed to R. (A saturated edge e is an edge, whose flow has either reached its upper bound or its lower bound, and no more flow can be added to it in the clockwise direction.)

It is not hard to see that for an efficient implementation of this algorithm, all we need is a shortest-path tree in the residual dual graph. This tree can be computed in $O(n\sqrt{\log n})$ by Fredrickson's algorithm [Fr], where n is the number of faces in the graph.

3.4. Vertex capacities

An interesting version of planar flow is the case where vertices as well as edges have capacity constraints [KN]. Vertex capacities may arise in various contexts such as computing vertex disjoint paths in graphs [KS], and in various network situations when the vertices denote switches and have an upper bound on their capacities. For the case of general graphs this problem can be reduced to the version with only edges having capacity constraints by a simple idea of "splitting" vertices into two and forcing all the flow to pass through a "bottleneck" edge in-between. In planar graphs, this reduction may *destroy* the planarity of the graph and thus cannot be used.

The following example shows that the set of feasible circulations with vertex capacity constraints does not form a lattice when circulations are reperesented by potential vectors. This may explain in part why it is harder to design efficient algorithms for this case.

Let $G = (V, E)$ be a planar graph with 5 vertices where v_1, v_2, v_3, v_4 form a directed anticlockwise cycle and v_5 is connected to all other vertices as follows: the edges from v_1 and v_3 are directed towards v_5 and the edges from v_5 to v_2 and v_4 are directed away from v_5. The capacity of vertex v_5 is c, and the capacity of each edge is $4c$.

We choose the following two feasible circulations. In circulation C_1, the flow on edges v_1, v_5 and v_5, v_2 is c and the flow on edges v_3, v_5 and v_5, v_4 is 0. In circulation C_2, the flow on edges v_1, v_5 and v_5, v_2 is 0 and the the flow on edges v_3, v_5 and v_5, v_4 is c. (The flow on the edges on the cycle v_1, v_2, v_3, v_4 is not important in both C_1 and C_2).

It is not hard to see that either the meet or the join of C_1 and C_2 will generate the circulation in which the flow on edges v_1, v_5 and v_5, v_2 is c and the flow on edges v_3, v_5 and v_5, v_4 is also c. Clearly, this circulation is infeasible (due to the capacity at v_5 being violated).

4. The partial order

There is a partial order associated with every distributive lattice. We will investigate the partial order that is associated with the lattice $\mathcal{P}$. Our exposition will follow Gusfield and Irving [GI, Ch. 2] and Grätzer [Gr, Ch. 2].

Let $\mathcal{P}[f = i]$ denote the set of all circulations such that the potential of face f is equal to i. Obviously, $\mathcal{P}[f = i]$ induces a sublattice of $\mathcal{P}$. We call a lattice element *irreducible* if for some face f and potential value i, it is the bottom element of $\mathcal{P}[f = i]$. Let $I(\mathcal{P})$ denote the set of all irreducible elements of the lattice. We define the partial order $(I(\mathcal{P}), \preceq)$ as the partial order on $I(\mathcal{P})$ where the dominance relation is inherited from $\mathcal{P}$. Clearly, $I(\mathcal{P})$ has a unique minimum and maximum since it contains both P_b and P_t.

For a partial order R, a subset S is said to be *closed* in R if for every $s \in S$, the predecessors of s are also in S. The following theorem is proved in [Gr, p. 72, Thm. 9] and [GI, Thm. 2.2.1].

Theorem 4.1: *Define a mapping from the closed subsets of $I(\mathcal{P})$ into $\mathcal{P}$ by $S \mapsto \vee S$. Then this mapping is one-to-one and onto. Moreover, if closed subsets S and S' of $I(\mathcal{P})$ correspond to circulations P and P', respectively, then P dominates P' if and only if $S \subseteq S'$.*

It is clear that $I(\mathcal{P})$ may have exponential size if the capacities are exponential. However, we will see that a different partial order can be constructed such that Theorem 4.2 still holds, yet this partial order has a more regular structure which enables us to represent it *succinctly.*

The elements P_1 and P_2 are called *consecutive* elements in the lattice $\mathcal{P}$, if P_2 covers P_1, i.e. there is no element Q such that $P_1 < Q < P_2$. Suppose elements P_1 and P_2 are consecutive and $P_1 < P_2$. The *minimal difference* between P_1 and P_2 is defined to be the pair (f, i), where f is the face on which P_1 and P_2 differ and i is the potential of f in P_1. (Obviously, the potential of face f in P_2 is $i + 1$). We denote by $\mathcal{D}$ the set of all minimal differences in $\mathcal{P}$.

Notice that we can assume without loss of generality that consecutive elements differ in only one face since we can assume that there are no edges in the graph whose lower bound on the capacity is equal to the upper bound. One way of seeing this follows from Section 3.1 where lower bounds are eliminated, and then such edges have zero capacity and can be removed from the graph.

Lemma 4.2: *Suppose that the potential of face f in P_b and P_t is p and q, respectively. Then, for all i, $p \leq i \leq q$, there exist consistent potential vectors in which face f has potential i.*

Proof: The proof follows from the Region Growing Algorithm. Run the algorithm so as to obtain the top element of the lattice. For some k, at the end of Step $k-1$, $Q(f) \leq i$, yet at the end of Step k, $Q(f) \geq i$. By decreasing Δ appropriately at Step k, $Q(f) = i$, and the potential vector obtained is consistent. $\square$

The last lemma implies that for consecutive elements P_1 and P_2 where $P_1 < P_2$, the potential of face f in P_2 is bigger than its potential in P_1 by precisely one unit. Hence, we can denote a minimal difference by (f, i), i.e., the potential of face f is increased from i to $i+1$.

A *maximal chain* in a lattice is a chain of consecutive elements that starts at P_b and ends at P_t. An interesting property of distributive lattices is that each maximal chain contains *all* the minimal differences. The minimal differences appear on each maximal chain in some order and each minimal difference appears exactly once.

We can now define the partial order $(K(\mathcal{P}), \preceq)$. Let $D_1, D_2 \in \mathcal{D}$; then $D_1 < D_2$ if and only if D_1 precedes D_2 on every maximal chain in $\mathcal{P}$. The motivation for defining this partial order follows from Theorem 19 of [Gr, p. 75], which states that every distributive lattice is isomorphic to a ring of sets. A ring of sets is a distributive lattice where the elements are subsets defined over a base set and the join and meet operations are respectively defined as intersection and union. Let $\mathcal{R}$ denote the ring of sets isomorphic to $\mathcal{P}$. Then, the closed subsets of the partial order $(K(\mathcal{P}))$ are in fact in one-one correspondence with the elements of $\mathcal{R}$. This leads us to the next theorem, whose proof follows from [GI, Thm. 2.4.4] and which relates the partial orders $I(\mathcal{P})$ and $K(\mathcal{P})$.

Theorem 4.3: *There is a one-one correspondence between the closed subsets of $I(\mathcal{P})$ and $K(\mathcal{P})$.*

We are now ready to simplify the partial order $K(\mathcal{P})$ and define a dag $T(\mathcal{P}) = (\mathcal{D}, E)$ which has a succinct description. The vertex set of $T(\mathcal{P})$ is again $\mathcal{D}$, the set of minimal differences. The edge set of $T(\mathcal{P})$ is defined as follows:

- For face f that takes potential values between p and q, there is a directed chain $(f,p) \rightarrow (f, p+1), \rightarrow \ldots \rightarrow (f, q-1)$. (Such a chain is called an f-chain.)
- For adjacent faces f and g that take potential values between p_1 and q_1, and p_2 and q_2, respectively, and the edge from f to g has capacity b. There is a "ladder" between the f-chain and the g-chain: $(f, p_1) \rightarrow (g, p_1 + b), (f, p_1 + 1) \rightarrow (g, p_1 + b + 1), \ldots, (f, x) \rightarrow (g, x + b)$, where $x = \min\{q_1 - 1, q_2 - b - 1\}$.

The next theorem relates the closed subsets of the dag $T(\mathcal{P})$ and the elements of $\mathcal{P}$. The intuitive reason for its correctness follows from the fact that the shortest path information can be completely recovered from the constraints on adjacent faces.

Theorem 4.4: *There is a one-one correspondence between the closed subsets of $T(\mathcal{P})$ and the elements of $\mathcal{P}$.*

Proof: Let S be any closed subset of $T(\mathcal{P})$. Let f be any face and denote the lower and upper bounds on its potential values by p and q. Since S is a closed subset, the intersection between the f-chain and S is a subchain starting at (f, p) and ending at (f, x_f), where $x_f \leq q - 1$. The potential vector corresponding to S is defined as follows: for each face f, assign its potential to be x_f. To see that this is a consistent potential vector, let f and g be any two adjacent faces where the capacity of the edge from f to g is b. If $x_g - x_f > b$, then S cannot be a closed subset since $(g, x_g) \in S$ where as $(f, x_g - b) \notin S$.

The correspondence in the other direction is proven very similarly. Given a consistent potential vector where face f has potential x_f, the closed subset S is constructed as follows: for each f-chain, the subchain from (f, p) to $(f, x_f - 1)$ belongs to S. Again, for any adjacent faces f and g, since $x_g - x_f \leq b$, S is a closed subset. $\square$

We now consider the simple example in Fig. 2 (a). The face f_1 is chosen to be the root face. The bottom element of the lattice corresponds to the smallest possible potential vector which is $(0, 1, -4)$ (these are the potentials of the faces f_1, f_2, f_3 respectively. We can now modify the graph, by eliminating lower bounds on the edge capacities (by constructing the residual graph with respect to P_b). We now get the graph in Fig. 2 (b). We construct its dual graph in Fig. 2 (c). This is the graph for which we would like to encode all feasible potential vectors. The range of potentials for both f_2 and f_3 can easily be seen to be $0 \ldots 2$. We thus construct the two chains $(f_i, 0), (f_i, 1)$ $(i = 2, 3)$ (see Fig 3). The ladder edges are added from (f_3, j) to (f_2, j) $(j = 0, 1)$. This gives us $T(\mathcal{P})$ whose closed subsets encode all the feasible solutions. Clearly, there are six closed subsets of this DAG, i.e., $\Phi, \{A\}, \{A, B\}, \{A, C\}, \{A, B, C\}, \{A, B, C, D\}$. Each closed subset corresponds to a set of minimal differences, that we can add to P_b to generate an integer circulation. These closed subsets are in 1-1 correspondence with the set of circulations of the graph, namely, $(0, 1, -4), (0, 1, -3), (0, 2, -3), (0, 1, -2), (0, 2, -2), (0, 3, -2)$. These circulations are obtained by adding the minimal differences to the potential vector P_b (bottom of the lattice). It is easy to see that the shortest path potential vector corresponds to the top of the lattice.

In the stable marriage problem, it was shown that every partial order can be associated with some instance of the problem [GI]. An interesting question is whether there exists a subset of the set of dags that has some "nice" characterization such that there exists a planar circulation instance that can be associated with each dag in the subset. Unfortunately, it seems that dags

f_1

-1 5 -2 5

5

-1

f_2 f_3

$[1,5]$ $[1,5]$ $[2,5]$

(a) Graph and its dual graph

f_1

4 1

2

f_2 f_3

4

2

(b) Residual graph w.r.t P_b

4 0 1 2

4

0

figure=f1.ps (c) Dual graph from residual graph

Figure 2: Figure to illustrate example

$(f_2, 1)$ D C $(f_3, 1)$

figure=f2.ps $(f_2, 0)$ B A $(f_3, 0)$

Figure 3: Figure to illustrate DAG $T(\mathcal{P})$

corresponding to planar circulation instances have very specialized structure: (i) There is a 1-1 correspondence between faces in the planar graph and subsets of vertices in the dag that induce an acyclic tournament, either directly, or by implication. (The f-chains). (ii) There is a special structure connecting these subsets (the "ladders") that has to correspond to capacities in the circulation instance.

5. Minimum cost flow

In the minimum cost circulation problem, each edge has in addition to its capacity, an associated cost $c(e)$ (sometimes written as c when the edge is clear from the context). The costs on the edges are assumed to be antisymmetric and may be positive as well as negative. The aim is to compute a feasible circulation such that the *cost* is minimized, where the cost is defined as

$$\sum_{e \in E} f(e)c(e)$$

In a planar graph, the cost of a circulation can be expressed as a function of the potentials of the faces and *face costs*. The cost of a face g is defined as follows. Traverse the boundary of the face clockwise, since the graph is directed some edges are traversed in the forward direction and some in the reverse direction.

$$c(g) = \sum_{e \in forward(g)} c(e) - \sum_{e \in reverse(g)} c(e)$$

The cost of the circulation is $\sum_{e \in E} f(e)c(e)$, and is the same as $\sum_{g \in F} p(g)c(g)$.

Let f be a function defined on a lattice $\mathcal{L}$, and let $a, b \in \mathcal{L}$. The function f is called *modular* if

$$f(a) + f(b) = f(a \vee b) + f(a \wedge b)$$

The next proposition is immediate.

Proposition 5.1: *The cost function of a planar circulation is modular.*

We now show that the solutions to the minimum cost circulation problem form a sublattice. Let f denote the cost function in a circulation and let $P_1, P_2 \in \mathcal{P}$ be any two minimum cost circulations. Since $f(P_1)+f(P_2) = f(P_1 \vee P_2)+f(P_1 \wedge P_2)$, the cost of $f(P_1 \vee P_2)$ and $f(P_1 \wedge P_2)$ has to be minimum as well.

5.1. Representing the minimum cost solutions

Having shown that the minimum-cost circulations of G form a sublattice, we now describe how to construct a network G' whose feasible circulations are exactly the minimum-cost circulations of G. Hence, it follows that the machinery discussed in Section 4 for representing the set of feasible circulations, can also be applied to represent the set of minimum-cost circulations, i.e., a partial order whose closed subsets correspond to the minimum-cost circulations can be constructed.

For any network G (not just for a planar network), once we have a single minimum-cost circulation C, we can represent all minimum-cost circulations as $C + \{C' : C' \text{ is a circulation in } G'\}$, where G' is a network derived from G and C. This representation is analogous to the representation of all solutions to a linear system or differential equation as a single solution, plus the set of solutions to the homogeneous equations.

To compute G', we first derive reduced edge-costs $\hat{c}$ from the original costs c in the residual graph of G with respect to C. Reduced edge-costs are all nonnegative, and have the property that the cost of any cycle in G_C is the same whether we use original or reduced costs. In particular, any cycle in G_C that has zero-cost with respect to the original costs also has zero-cost with respect to the reduced costs, and hence contains only edges that have zero reduced cost. Let G' be the subgraph of G_C consisting of edges with zero reduced cost.

Lemma 5.2: *The set of min-cost circulations in G is $\{C + C' : C' \text{ a circulation in } G'\}$.*

Proof: Let C' be any circulation in G'. Since G' is a subgraph of G_C, C' consists of a collection of cycles of flow in G_C such that $C + C'$ is a circulation in G. Since G' contains only edges with zero reduced cost, every cycle in C' has zero cost. Hence the cost of $C + C'$ is the same as that of C, and is hence minimum.

Conversely, let C_1 be any min-cost circulation in G. Then $C_1 - C$, being the difference between two circulations, is itself a circulation in G_C, and is hence composed of cycles of flow.

If any such cycle of flow had negative cost, it could be added to C to reduce C's cost, so there are no negative cycles. Similarly, there are no positive-cost cycles, else they could be subtracted from C_1 to reduce its cost. Thus the difference $C_1 - C$ consists of a collection of zero-cost cycles of flow. By the remarks above, each such cycle consists of edges with zero reduced cost, so $C_1 - C$ is a circulation in G'. $\square$

For completeness, we describe one standard construction for computing reduced costs. Since C is minimum-cost, every cycle in G_C has nonnegative cost [FF]. Obtain an auxiliary graph from G_C by adding a node s and zero-cost edges from s to every original node. Next, in the auxiliary graph compute shortest path distances $d(v)$ of each node v from the added node s, using an algorithm, e.g. Floyd-Warshall, that depends only on the nonexistence of negative cycles.

Now for each edge $e = uv$ in G_C, we define the *reduced cost* $\hat{c}(e) = c(e) + d(u) - d(v)$. By Bellman's equations, $d(v) \leq c(e) + d(u)$, so each reduced edge-cost is nonnegative. Furthermore, it is easy to check that the cost of any cycle in G_C is the same whether we use original edge-costs or reduced edge-costs, since the $d(v)$'s cancel out as we traverse the cycle.

5.2. Directions for future research

An outstanding open question is whether a better algorithm for computing minimum cost circulations in planar graphs can be found. In what follows, we will outline two possible approaches for this problem.

Minimizing a modular function defined on a lattice is a well known problem in operations research. We briefly review its solution. (See e.g. [GI, pp. 130-133, Ir, To] for more details and proofs.) Let the cost of every vertex in the dag $T(\mathcal{P})$ be equal to the cost of the corresponding face in the original graph. It is not hard to see that the minimum cost circulation problem can be restated as the problem of finding the predecessor-closed set of minimum cost in $T(\mathcal{P})$, where the cost of a closed set is defined to be the sum of the costs of its members. The problem of computing the minimum cost closed set can be reduced to computing the minimum cut in the following graph, denoted by T:

- Connect all positive cost vertices to a source and all negative cost vertices to a sink.
- The capacity assigned to edges adjacent to the source or sink is equal to the absolute value of the cost of the vertices to which they are adjacent.
- All other edges have infinite capacity.

Solving this problem directly, by computing a maximum flow, would take too long, because the graph T is too big. However, some algorithm based on maximum flow would be interesting for the following reason.

Most algorithms for computing the minimum cost circulation have the following form: they start from an initial circulation and generate circulations of smaller cost until a minimum cost circulation is obtained. In the dag $T(\mathcal{P})$, a one-one correspondence can be established between its closed sets (or the circulations in $\mathcal{P}$) and the cuts separating the source from the sink in the graph T. Thus, all these algorithms can be viewed as algorithms that implicitly compute the minimum cut in T. On the contrary, an algorithm that computes the minimum cut in T via a maximum flow can be considered a dual algorithm to all other minimum cost circulation algorithms. The question therefore arises: can the special structure of T be exploited to find the maximum flow much more quickly, say by considering only intermediate solutions (feasible flows) of a certain form ?

A different approach to computing a minimum cost circulation follows from the Region Growing Algorithm. What happens when this algorithm is applied to the minimum cost circulation problem ? At each step of the algorithm we have to first decide whether to push some amount of flow in the clockwise direction, and then decide which faces to annex to R. These decisions will depend on the cost of the boundary of R in the clockwise direction. The easy case is when the cost is negative: then the boundary is saturated and the faces that border on saturated edges are annexed to R. The difficulty arises when the boundary has positive cost. Then, there is no gain in pushing more flow on the boundary in the clockwise direction. From the existence of the Region Growing Algorithm, we know that there is at least *one* face f that borders R and can be annexed to it, i.e., the potential of f has reached its value in some optimal solution. Does there exist a simple criterion for determining which face that is ?

Another intriguing research question concerns dynamic computation of feasible or minimum-cost circulations in planar networks. Suppose that we are given a circulation and assume that the capacities are changed on edges of a single face. It follows from the correspondence between circulations and shortest paths that it is easy to derive a new circulation from an old one in $O(n\sqrt{\log n})$ time [Fr] by using the notion of reduced cost. What can be said about the analogous problem for minimum-cost flow ?

Acknowledgement: We would like to thank Edith Cohen, Tomas Feder and Dan Gusfield for helpful discussions.

References

[AMO] R. K. Ahuja, T. L. Magnanti and J. B. Orlin, Network flows: theory, algorithms and applications, Prentice Hall, 1992.

[FF] L. R. Ford and D. R. Fulkerson, *Maximal flow through a network,* Canadian Journal of Mathematics, Vol. 8, pp. 399-404 (1956).

[Fr] G. N. Frederickson, *Fast algorithms for shortest paths in planar graphs, with applications,* Siam Journal on Computing, Vol.16, pp. 1004-1022 (1987).

[Gr] G. Grätzer, Lattice Theory: first concepts and distributive lattices, W. H. Freeman and Company, San Francisco, 1971.

[GI] D. Gusfield and R. W. Irving, The stable marriage problem, The MIT Press, 1990.

[Ha] R. Hassin, *Maximum flows in* (s,t) *planar networks,* Information Processing letters, Vol. 13, page 107 (1981).

[HJ] R. Hassin and D. B. Johnson, *An* $O(n\log^2 n)$ *algorithm for maximum flow in undirected planar networks*, Siam Journal on Computing, Vol. 14, pp. 612-624 (1985).

[II] Hiroshi Imai and Kazuo Iwano, *Efficient sequential and parallel algorithms for planar minimum cost flow*, SIGAL, International Symposium on Algorithms, Tokyo (1990).

[Ir] M. Iri, *Structural theory for the combinatorial systems characterized by submodular functions*, In Progress in Combinatorial Optimization, pp. 197-219, Academic Press, 1984.

[IS] A. Itai and Y. Shiloach, *Maximum flow in planar networks,* Siam Journal of Computing, Vol. 8, pp. 135-150 (1979).

[Jo] D. B. Johnson, *Parallel algorithms for minimum cuts and maximum flows in planar networks,* Journal of the ACM, Vol. 34, (4), pp. 950-967 (1987).

[JV] D. B. Johnson and S. Venkatesan, *Using divide and conquer to find flows in directed planar networks in* $O(n^{1.5}\log n)$ *time,* Proceedings of the 20th Annual Allerton Conference on Communication, Control and Computing, University of Illinois, Urbana-Champaign, Ill., pp. 898-905 (1982).

[KN] S. Khuller and J. Naor, *Flow in Planar Graphs with Vertex Capacities,* Proceedings of Integer Programming and Combinatorial Optimization Conference, (1990), pp. 367-383. Also: to appear in Algorithmica, Special Issue on Network Flow.

[KS] S. Khuller and B. Schieber, *Efficient parallel algorithms for testing k-connectivity and finding disjoint s-t paths in graphs,* Siam Journal on Computing, Vol. 20, No. 2, pp. 352-375 (1991).

[MN] G. L. Miller and J. Naor, *Flow in planar graphs with multiple sources and sinks*, Proceedings of the 30th Annual Symposium on Foundations of Computer Science (1989), pp. 112-117.

[PQ] J. Picard and M. Queyranne, *On the structure of all minimum cuts in a network and its applications*, Mathematical Programming Study, Vol. 13, pp. 8-16, (1980).

[Re] J. H. Reif, *Minimum $s - t$ cut of a planar undirected network in $O(n \log^2 n)$ time,* Siam Journal on Computing, Vol. 12, No. 1, pp. 71-81 (1983).

[To] D. Topkis, *Minimizing a submodular function on a lattice*, Operations Research, Vol. 26, pp. 305-321 (1978).