CMSC 451: Lecture 17
Network Flow: Extensions
Thursday, Nov 9, 2017

Reading: Section 7.7 in KT.

Extensions of Network Flow: Network flow is an important problem because it is useful in a wide variety of applications. We will discuss two useful extensions to the network flow problem. We will show that these problems can be reduced to network flow, and thus a single algorithm can be used to solve both of them. Many computational problems that would seem to have little to do with flow of fluids through networks can be expressed as one of these two extended versions.

Circulation with Demands: There are many problems that are similar to network flow in which, rather than transporting flow from a single source to a single sink, we have a collection of *supply nodes* that want to ship flow (or products or goods) and a collection of *demand nodes* that want to receive flow. Each supply node is associated with the amount of product it wishes to ship and each demand node is associated with the amount that it wishes to receive. The question that arises is whether there is some way to get the products from the supply nodes to the demand nodes, subject to the capacity constraints. This is a *decision problem* (or *feasibility problem*), meaning that it has a yes-no answer, as opposed to maximum flow, which is an *optimization problem*.

We can model both supply and demand nodes elegantly by associating a single numeric value with each node, called its *demand*. If $v \in V$ is a demand node, let d_v the amount of this demand. If v is a supply node, we model this by assigning it a negative demand, so that $-d_v$ is its available supply. Intuitively, supplying x units of product is equivalent to demanding receipt of $-x$ units.[1] If v is neither a supply or demand node, we let $d_v = 0$.

Suppose that we are given a directed graph $G = (V, E)$ in which each edge (u, v) is associated with a positive capacity $c(u, v)$ and each vertex v is associated with a supply/demand value d_v. Let S denote the set of *supply nodes* ($d_v < 0$), and let T denote the set of *demand nodes* ($d_v > 0$). Note that vertices of S may have incoming edges and vertices of T may have outgoing edges. (For example, in Fig. 1(a), we show a network in which each node is each labeled with its demand and each edge with its capacity.)

Recall that, given a flow f and a node v, $f^{\text{in}}(v)$ is the sum of flows along incoming edges to v and $f^{\text{out}}(v)$ is the sum of flows along outgoing edges from v. We define a *circulation* in G to be a function f that assigns a nonnegative real number to each edge that satisfies the following two conditions.

Capacity constraints: For each $(u, v) \in E$, $0 \leq f(u, v) \leq c(u, v)$.

Supply/Demand constraints: For vertex $v \in V$, $f^{\text{in}}(v) - f^{\text{out}}(v) = d_v$.

[1] I would not advise applying this in real life. I doubt that the IRS would appreciate it if you paid your \$100 tax bill by demanding that they send you $-\$100$ dollars.

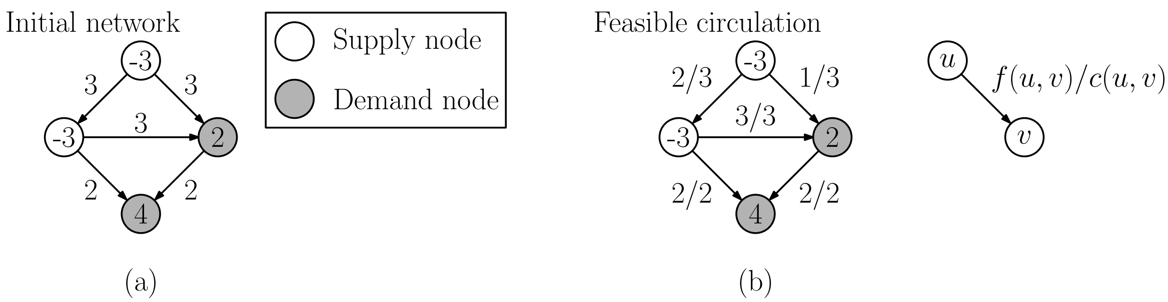

Fig. 1: (a) A circulation network and (b) a feasible circulation.

For example, in Fig. 1(b) we show a valid circulation for the network of part (a). Observe that demand constraints correspond to the flow-balance in the original max flow problem, since if a vertex is not in S or T, then $d_v = 0$ and we have $f^{\text{in}}(v) = f^{\text{out}}(v)$. Also it is easy to see that the total demand must equal the total supply, otherwise we have no chance of finding a feasible circulation. That is, we require that

$$\sum_{v \in V} d_v \;=\; 0 \qquad \text{or equivalently} \qquad -\sum_{v \in S} d_v \;=\; \sum_{v \in T} d_v.$$

Define $D = \sum_{v \in T} d_v$ denote the *total demand*. (Note that this is equal to the negation of the total supply, $\sum_{v \in S} d_v$.)

Reducing Circulation to Max-Flow: Rather than devise an algorithm for the circulation problem, we will show that we can reduce any instance G of the circulation problem into an equivalent network flow problem. The input to our reduction is a network $G = (V, E)$. For each vertex v, let d_v denote the demand value and for each edge (u, v), let $c(u, v)$ denote its capacity. First, observe that we may assume that sum of supplies equals total demand (since if not, we can simply answer "no" immediately.) Otherwise:

- Create a new network $G' = (V', E')$ that has all the same vertices and edges as G (that is, $V' \leftarrow V$ and $E' \leftarrow E$)
- Add to V' a *super-source* s^* and a *super-sink* t^*
- For each supply node $v \in S$, we add a new edge (s^*, v) of capacity $-d_v$
- For each demand node $u \in T$, we add a new adge (u, t^*) of capacity d_v

(The output of the reduction is illustrated in Fig. 2(b).) We then invoke any max-flow algorithm on G'. Recalling that D denotes the total demand, we check whether the value of the maximum flow equals D. If so, we answer "yes," G has a feasible circulation, and otherwise we answer "no," G does not have a feasible circulation. (For example, in Fig. 2(c), there exists a flow of value $D = 6$, implying that the original network has a circulation.)

We prove correctness below, but intuitively, the newly created edges from s^* will be responsible for providing the necessary supply for the supply vertices of S and newly created edges into t^* are responsible for draining off the excess demand from the vertices of T. Suppose that we now compute the maximum flow in G' (by whichever maximum flow algorithm you like).

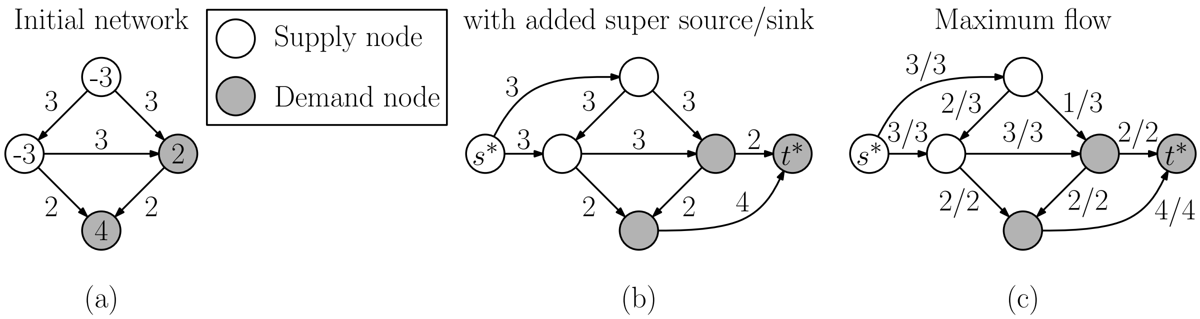

Fig. 2: Reducing the circulation problem to network flow.

Lemma: There is a feasible circulation in G if and only if G' has an s^*-t^* flow of value D.

Proof: ($\Rightarrow$) Suppose that there is a feasible circulation f in G. The value of this circulation (the net flow coming out of all supply nodes) is clearly D. We can create a flow f' of value D in G', by saturating all the edges coming out of s^* and all the edges coming into t^*. We claim that this is a valid flow for G'. Clearly it satisfies all the capacity constraints. To see that it satisfies the flow balance constraints observe that for each vertex $v \in V$, we have one of three cases:

- ($v \in S$) The flow into v from s^* matches the supply coming out of v from the circulation.
- ($v \in T$) The flow out of v to t^* matches the demand coming into v from the circulation.
- ($v \in V \setminus (S \cup T)$) We have $d_v = 0$, which means that it satisfies flow conservation by the supply/demand constraints.

($\Leftarrow$) Conversely, suppose that we have a flow f' of value D in G'. It must be that each edge leaving s^* and each edge entering t^* is saturated. Therefore, by the flow conservation of f', all the supply nodes and all the demand nodes have achieve their desired supply/demand constraints. All the other nodes satisfy their supply/demand constraints because by the flow conservation of f' the incoming flow equals the outgoing flow. Therefore, the resulting flow is a circulation for G.

It is not hard to see to that the reduction can be performed in $O(n + m)$ time by a simply analysis of the network's structure. Thus, the overall running time is dominated by the time to compute the network flow (which is $O(nm)$ according to the current state-of-art).

Circulations with Upper and Lower Capacity Bounds: Sometimes, in addition to having a certain maximum flow value, we would also like to impose minimum capacity constraints. That is, given a network $G = (V, E)$, for each edge $(u, v) \in E$ we would like to specify two constraints $\ell(u, v)$ and $c(u, v)$, where $0 \leq \ell(u, v) \leq c(u, v)$. A circulation function f must satisfy the same demand constraints as before, but must also satisfy both the upper and lower flow bounds:

(New) Capacity Constraints: For each $(u, v) \in E$, $\ell(u, v) \leq f(u, v) \leq c(u, v)$.

Demand Constraints: For vertex $v \in V$, $f^{\text{in}}(v) - f^{\text{out}}(v) = d_v$.

Henceforth, we will use the term *upper flow bound* in place of *capacity* (since it doesn't make sense to talk about a lower bound as a capacity constraint). An example of such a network is shown in Fig. 3(a), and a valid circulation is shown in Fig. 3(b).

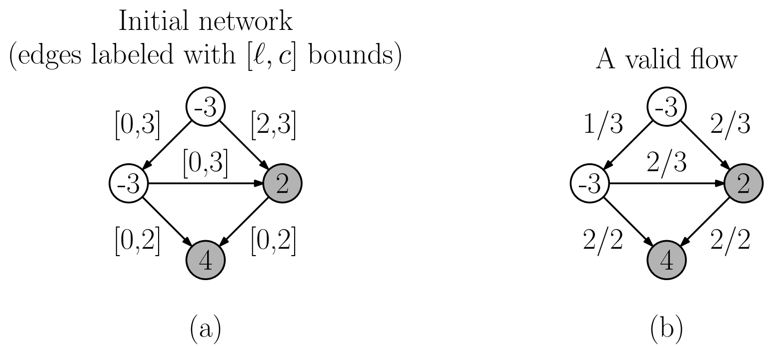

Fig. 3: (a) A network with both upper and lower flow bounds and (b) a valid circulation.

We will reduce this problem to a standard circulation problem (with just the usual upper capacity bounds). To help motivate our reduction, suppose (for conceptual purposes) that we generate an initial (possibly invalid) circulation f_0 that exactly satisfies all the lower flow bounds. In particular, we let $f_0(u, v) = \ell(u, v)$ (see Fig. 4(a)). This circulation may be invalid because f_0 need not satisfy the demand constraints (which, recall, provide for flow balance as well). We will modify the supply/demand values to compensate for this imbalance. Since the lower-bound constraints are all satisfied, it will be possible to apply a standard circulation algorithm (without lower flow bounds) to solve the problem.

To make this formal, for each $v \in V$, let L_v denote the *excess flow* coming into v in f_0, that is

$$L_v \;=\; f_0^{\text{in}}(v) - f_0^{\text{out}}(v) \;=\; \sum_{(u,v)\in E} \ell(u, v) - \sum_{(v,w)\in E} \ell(v, w).$$

(Note that this may be negative, which means that we have a flow deficit.) If we are lucky, then $L_v = d_v$, and v's supply/demand needs are already met. Otherwise, we will adjust the supply and demand values so that by (1) computing any valid circulation f_1 for the adjusted values and (2) combining this with f_0, we will obtain a flow that satisfies all the requirements.

How do we adjust the supply/demand values? We want to generate a net flow of d_v units coming into v and cancel out the excess L_v coming in. That is, we want f_1 to satisfy:

$$f_1^{\text{in}}(v) - f_1^{\text{out}}(v) \;=\; d_v - L_v.$$

(In particular, this means that if we combine f_0 and f_1 by summing their flows for each edge, then the final flow into v will be d_v, as desired.)

How do we determine whether there exists such a circulation f_1? Let's consider how to set the edge capacities. We have already sent $\ell(u, v)$ units of flow through the edge (u, v), which implies that we have $c(u, v) - \ell(u, v)$ capacity remaining. (Note that unlike our definition of residual graphs, we do not want to allow for the possibility of "undoing" flow. Can you see why not?)

We are now ready to put the pieces together. Given the network $G = (V, E)$ as input (with vertex demands d_v and lower and upper flow bounds $\ell(u, v)$ and $c(u, v)$):

1. Create an initial pseudo-circulation f_0 by setting $f(u, v) = \ell(u, v)$ (see Fig. 4(a))
2. Create a new network $G' = (V', E')$ that has all the same vertices and edges as G (that is, $V' \leftarrow V$ and $E' \leftarrow E$)
3. For each $(u, v) \in E'$, set its adjusted capacity to $c'(u, v) \leftarrow c(u, v) - \ell(u, v)$
4. For each $v \in V'$, set its adjusted demand to $d'_v \leftarrow d_v - L_v$, where, $L_v = f_0^{\text{in}}(v) - f_0^{\text{out}}(v)$ (see Fig. 4(b))

Now, invoke any (standard) circulation algorithm on G' (see Fig. 4(c)). Note that there is no need to consider lower flow bounds with G', because f_0 has already taken care of those. If the algorithm reports that there is no valid circulation for G', then we declare that there is no valid circulation for the original network G. If the algorithm returns a valid circulation f_1 for G', then the final circulation for G is the combination $f_0 + f_1$ (see Fig. 4(d)).

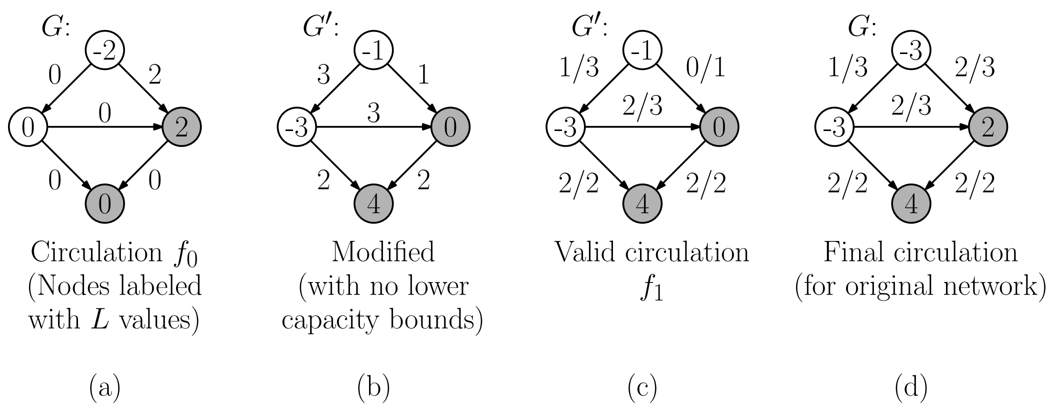

Fig. 4: Reducing the circulation problem with upper and lower flow bounds to a standard circulation problem.

To establish the correctness of our reduction, we prove below that its output $f_0 + f_1$ is a valid circulation for G (with lower flow bounds) if and only if f_1 is a valid circulation circulation for G'.

Lemma: The network G (with both lower and upper flow bounds) has a feasible circulation if and only if G' (with only upper capacity bounds) has a feasible circulation.

Proof: (Sketch. See KL for a formal proof.) Intuitively, if G' has a feasible circulation f' then the circulation $f(u, v) = f'(u, v) + \ell(u, v)$ can be shown to be a valid circulation for G and it satisfies the lower flow bounds. Conversely, if G has a feasible circulation (satisfying both the upper and lower flow bounds), then let $f'(u, v) = f(u, v) - \ell(u, v)$. As above, it can be shown that f' is a valid circulation for G'. (Think of f' as f_1 and f as $f_0 + f_1$.)

Note that (as in the original circulation problem) we have not presented a new algorithm. Instead, we have shown how to *reduce* the current problem (circulation with lower and upper flow bounds) to a problem we have already solved (circulation with only upper bounds). Again, the running time will be the sum of the time to perform the reduction, which is easily seen to be $O(n+m)$ plus the time to compute the circulation, which as we have seen reduces to the time to compute a maximum flow, which according to the current best technology is $O(nm)$ time.

Application: Survey Design: To demonstrate the usefulness of circulations with lower flow bounds, let us consider an application problem that arises in the area of data mining. A company sells k different products, and it maintains a database which stores which customers have bought which products recently. We want to send a survey to a subset of n customers. We will tailor each survey so it is appropriate for the particular customer it is sent to. Here are some guidelines that we want to satisfy:

- The survey sent to a customer will ask questions only about the products this customer has purchased.
- We want to get as much information as possible, but do not want to annoy the customer by asking too many questions. (Otherwise, they will simply not respond.) Based on our knowledge of how many products customer i has purchased, and easily they are annoyed, our marketing people have come up with two bounds $0 \leq c_i \leq c_i'$. We will ask the ith customer about at least c_i products they bought, but (to avoid annoying them) at most c_i' products.
- Again, our marketing people know that we want more information about some products (e.g., new releases) and less about others. To get a balanced amount of information about each product, for the jth product we have two bounds $0 \leq p_j \leq p_j'$, and we will ask at least p_j customers about this product and at most p_j' customers.

We can model this as a bipartite graph G, in which the customers form one of the parts of the network and products form the other part. There is an edge (i, j) if customer i has purchased product j. The flow through each customer node will reflect the number of products this customer is asked about. The flow through each product node will reflect the number of customers that are asked about this product.

This suggests the following network design. Given the bipartite graph G, we create a directed network as follows (see Fig. 5).

- For each customer i who purchased product j we create a directed edge (i, j) with an upper flow bounds of 1, respectively. This models the requirement that customer i will be surveyed at most once about product j, and customers will be asked only about products they purchased.
- We create a source vertex s and connect it to all the customers, where the edge from s to customer i has lower and upper flow bounds of c_i and c_i', respectively. This models the requirement that customer i will be asked about at least c_i products and at most c_i'.

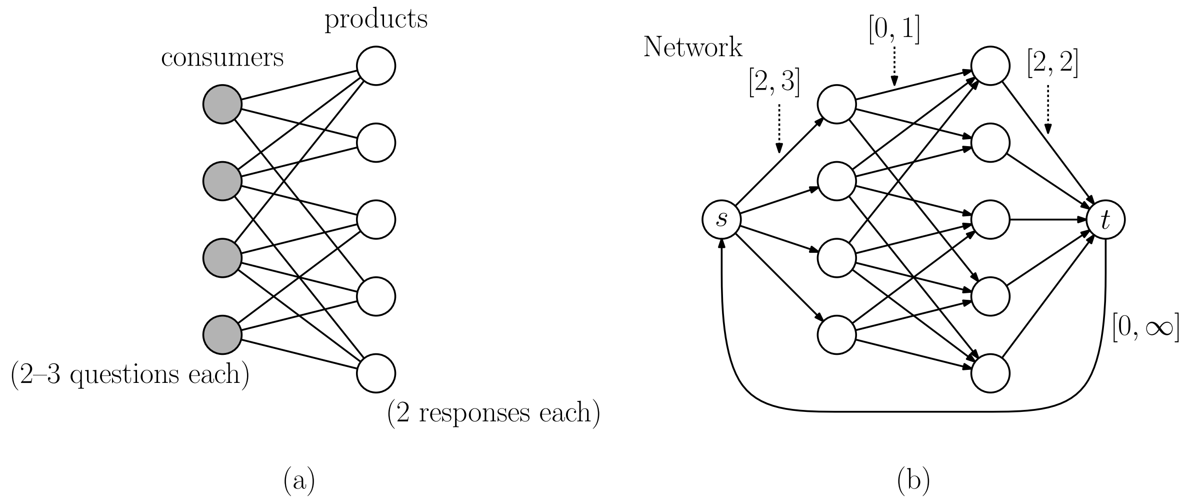

Fig. 5: Reducing the survey design problem to circulation with lower and upper flow bounds. (Assuming $[c_i, c'_i] = [2, 3]$ for all customers and $[p_j, p'_j] = [2, 2]$ for all products.)

- We create a sink vertex t, and create an edge from product j to t with lower and upper flow bounds of p_j and p'_j. This models the requirement that there are at least p_j and at most p'_j customers will be asked about product j.
- We create an edge (s, t). Its lower bound is set to zero and its upper bound can be set to any very large value. This is needed for technical reasons, since we want a circulation.
- All node demands are set to 0.

The correctness of the reduction is established in the following lemma.

Lemma: There exists a valid circulation in G if and only there is a valid survey design.

Proof: ($\Rightarrow$) Suppose that G has a valid (integer-valued) circulation. For each customer-product edge (i, j) that carries one unit of flow, customer i is surveyed about product j. By definition of the edges, customers are surveyed only about products they purchased. From our capacity constraints and the fact that demands are all zero, it follows that the total flow into each customer node is between c_i and c'_i, implying that this customer is asked about this many products. Similarly, the total flow out of each product node is between p_j and p'_j, implying that this product is involved in this many products surveys (see Fig. 6).

($\Leftarrow$) Suppose that there is a valid survey design. We construct a flow in G as follows. For each customer-product pair (i, j) involved in the survey, we create a flow of one unit on edge (i, j). We set the flow along the edge (s, i) to the number of surveys that customer i answers, we set the flow along the edge (j, t) to the number of surveys involving product j, and we set the flow on edge (t, s) to the total number of surveys. It is straightforward to see that (by the rules of a valid survey design) this is a valid flow in G, and in particular, it satisfies the lower and upper capacity constraints and the supply and demand constraints.

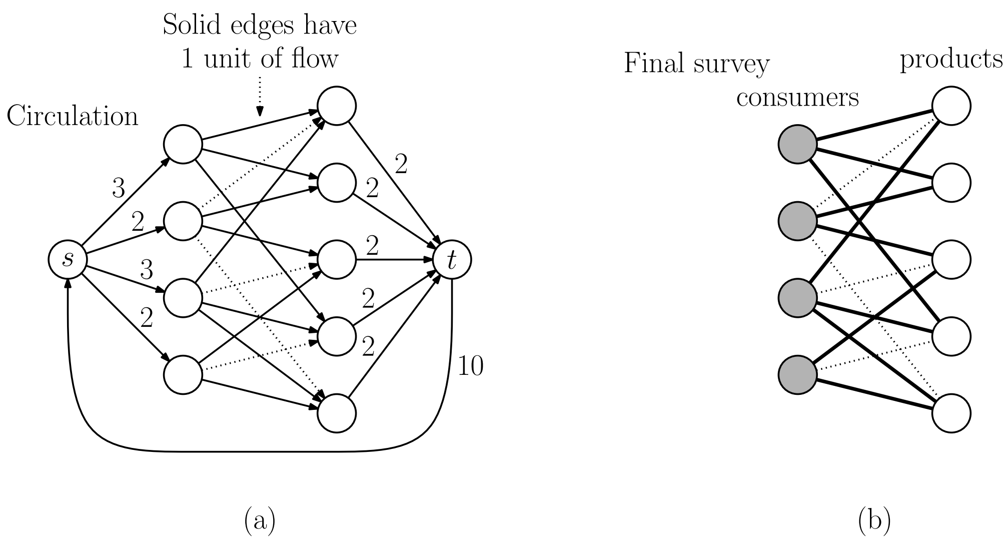

Fig. 6: Correctness of the survey-design reduction to circulation.