**CS256: Algorithm Design and Analysis** **Spring 2021**

# Flow Reduction Handout

*Instructor: Shikha Singh* *Source: Overleaf*

This handout steps through the survey design reduction we did in Lecture 21 in detail.

**Survey design problem.** In this problem, we want to design a survey where we ask consumers a question about a product they own. We have $n$ consumers and $m$ products. We have the following constraints on the survey design:

- We can survey a consumer $i$ about $j$ only if they own it.
- Ask consumer $i$ at least $a_i$ and at most $b_i$ questions
- Ask at least $p_j$ and at most $q_j$ customers about product $j$

Given the values $[a_i, b_i]$ and $[p_j, q_j]$ for each consumer $i$ and product $j$, determine if it is possible to design a survey that satisfies these requirements. Assume that $\sum_{i=1}^{n} a_i = \sum_{j=1}^{m} p_j$.

**Reduction to network flow.** We will solve this problem by reduction to network flow. Given an instance of the problem, we will create an instance of network flow $G$ as described below:

- Nodes in $G$: we will have $n$ nodes representing the consumers, $m$ nodes representing the products, and four additional nodes: two intermediate nodes $x$ and $y$ and a source $s$ and sink $t$
- Edges and capacities in $G$:
  - Add edge $i \to j$ with capacity 1 if consumer $i$ owns product $j$ for each consumer $i$ and product $j$
  - Add edge $s \to i$ with capacity $a_i$ for each consumer $i$
  - Add edge $j \to t$ with capacity $p_j$ for each product $j$
  - Add edge $x \to i$ with capacity $b_i - a_i$ for each consumer $i$
  - Add edge $j \to y$ with capacity $q_j - p_j$ for each product $j$
  - Add edge $y \to x$ with capacity $\sum_i (b_i - a_i)$

This flow network $G$ is described in Figure 1. (The figure is optional when describing reductions, but can be helpful in understanding them.)

Now we prove this reduction is correct by proving the following claim.

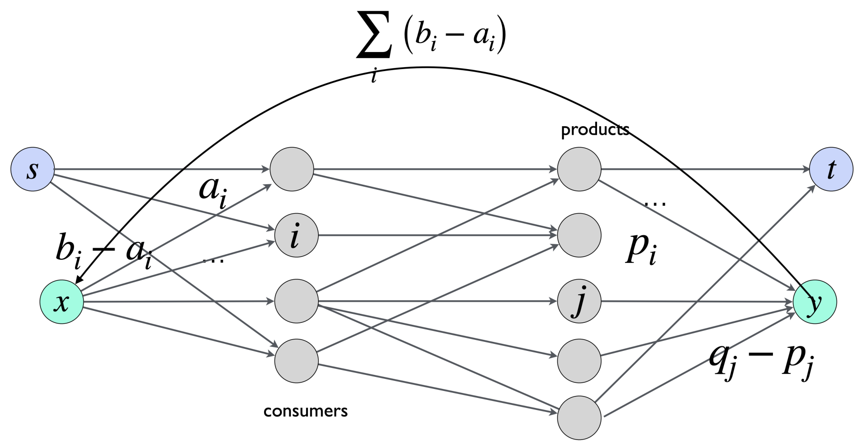


Figure 1: Figure of flow instance $G$ created in reduction

**Claim 1.** It is possible to design a survey satisfying the constraints of the problem iff the corresponding flow network $G$ has an integral max flow of value $\sum_{i=1}^{n} a_i$

*Proof.* ($\Rightarrow$) Suppose it is possible to design a survey satisfying the constraints. Let $\alpha_i$ denote the total number of questions that consumer $i$ is asked and $\beta_j$ denote the total number of consumers asked about product $j$ in this survey. We define a flow in $G$ as follows:

- Let $f(i \to j) = 1$ if consumer $i$ is asked about product $j$ and $f(i \to j) = 0$ otherwise
- Let $f(s \to i) = a_i$ for each consumer node $i$
- Let $f(j \to t) = p_i$ for each product node $j$
- Let $f(x \to i) = \alpha_i - a_i$ for each consumer node $i$
- Let $f(j \to y) = \beta_j - b_i$ for each product node $i$
- Let $f(y \to x) = \sum_{i=1}^{n}(\alpha_i - a_i)$

We verify that this flow is feasible. For any consumer node $i$, $f_{in}(i) = \alpha_i$ which is the total number of questions that consumer is asked. Conservation holds because $f_{out}(i) = \sum_j f(i \to j)$ which is exactly the total number of questions $i$ is asked. An analogous reasoning shows that conservation holds on product nodes $j$ as well. For intermediate node $x$, $f_{in}(x) = \sum_i(\alpha_i - a_i)$; $f_{out}(x) = \sum_i f(x \to i) = \sum_i(\alpha_i - a_i) = f_{in}(x)$. Since $\sum_i a_i = \sum_j p_j$, and the total number of questions asked remain the same, the conversation condition also holds for node $y$.

The flow clearly does not violate capacity constraints on edges out of the source, and consumer nodes and edges going into the sink. The flow on any edge from $x$ to a consumer node $i$ is equal to $\alpha_i - a_i \leq b_i - a_i$, because the survey satisfies the condition that consumer $i$ is asked between $a_i$ and $b_i$ questions. Similarly, we can check that the flow on any edge between product $j$ and node $y$ does not violate capacity constraints. The value of the flow is total flow coming out of $s$, which is equal to $\sum_{i=1}^{n} a_i$, which is the maximum flow as it equals the capacity of the cut $(\{s\}, V - \{s\})$.

($\Leftarrow$) Now suppose we have a feasible integral max flow $f$ of value $\sum_{i=1}^{n} a_i$ in $G$, we show that it is possible to design a survey satisfying the requirements. We ask consumer $i$ a question about product $j$ iff $f(i \to j = 1$. Since we only have edges for consumers that own products, we are only asking consumers about products that they own. We now check that we are not violating the constraints on the questions asked for each consumer $i$ and product $j$. Since the value of the flow is $\sum_{i=1}^{n} a_i$, which is equal to the capacity of the cuts $(\{s\}, V - \{s\})$ and$(V - \{t\}, \{t\})$, we can conclude that the flow on the edges leaving these cuts must be saturated, that is, $f_{in}(i) \geq a_i$ and $f_{out}(j) \geq p_i$. To check that we do not ask too many questions, we rely on the fact that the flow satisfies capacity constraints and is conserved. We observe that $f_{out}(i) =$ total number of questions consumer $i$ is asked. Since $f_{out}(i) = f_{in}(i) \leq b_i$, we know that total questions asked will not exceed $b_i$. Similarly, we can check that total questions asked about product $j$ will not exceed $q_j$. Thus, the resulting survey satisfies all the requirements. □

**Running time.** Finally, we analyze the running time of the algorithm. The flow network has $n' = n + m + 4$ nodes, and at most $m' = O(nm)$ edges. The overall running time is dominated by the running time of the Ford-Fulkerson algorithm which takes time $O(n' \cdot m' \cdot C)$ time, where $C$ is edge of maximum capacity leaving the source. Thus, the overall running time of our algorithm is $O((n + m) \cdot nm \cdot a_{\max})$, where $a_{\max} = \max_{i=1}^{n} a_i$.