Survey Design

Consider a company that sells k products and has a database with purchase history of its customers. The company wants to conduct a survey, asking a group of n customers which products they like.

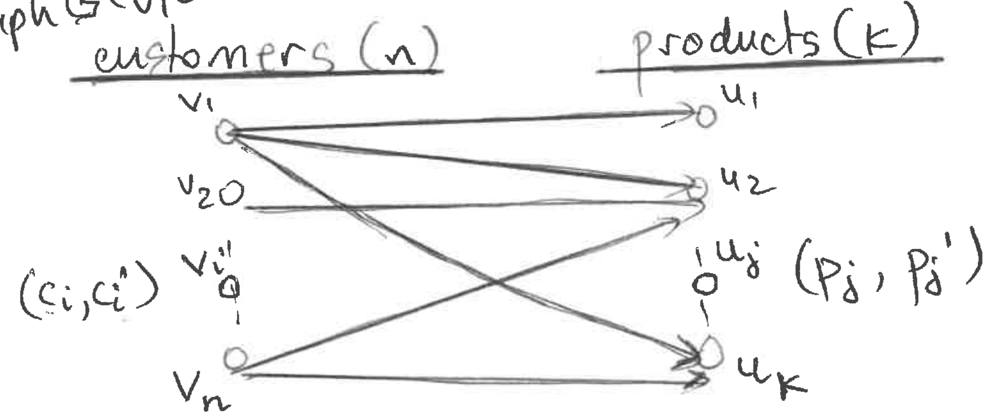
Guidelines for designing the survey:

- Each customer receives questions about a subset of the products
- A customer can only be asked about products that he has purchased
- Each customer i should be asked about a number of products between c_i and c_i'
- To collect sufficient data, between p_j and p_j' distinct customers must be asked about each product j

Decide if there is a way to design the survey to meet these conditions.

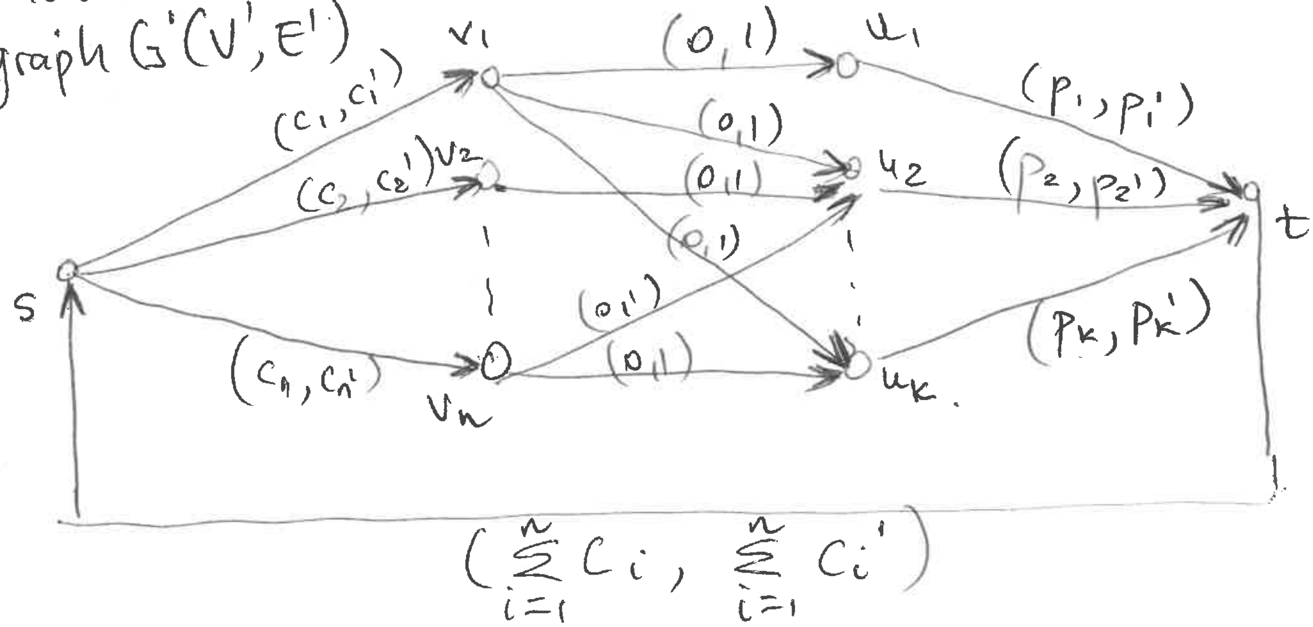
Solution

- bipartite graph $G(V,E)$



- edge (v_i, u_j) means that the customer i has purchased the product j previously
- model this problem to a circulation with the demands and lower bounds

graph $G'(V',E')$



- solve the circulation with demands and lower bounds
- all demands $= 0$
- flow on an edge (s, v_i) represents the number of questions for customer i
- flow on an edge (u_j, t) represents the number of questions about product j
- flow on an edge (v_i, u_j) is

$$\begin{cases} 0 & \text{- if customer } v_i \text{ is not asked a question about product } u_j \\ 1 & \text{- if customer } v_i \text{ is asked a question about product } u_j \end{cases}$$

- flow on the edge (t, s) represents the total number of questions in the survey.

- If G' has a feasible circulation $\Rightarrow$ we are able to design the survey meeting all the requests.
- otherwise, we cannot design the survey.

RT analysis

- use Edmonds-Karp to compute the max-flow

$$RT = O(V' \cdot E'^2)$$

$$|V'| = n + K + 2 = O(n + K)$$

$$|E'| \leq n \cdot K + n + K = O(n \cdot K)$$

$$RT = O((n+K)\, n^2 \cdot K^2) = O(n^3 \cdot K^2 + n^2 \cdot K^3)$$

- RT is polynomial!

Disaster Management

Network flow issues come up in dealing with natural disasters and other crises, since major unexpected events often require the movement and evacuation of large number of people in short amount of time.

Consider the following scenario. Due to large-scale flooding in a region, paramedics have identified a set of *n* injured people distributed across the region who need to be rushed to hospitals. There are *k* hospitals in the region, and each of the *n* people needs to be brought to a hospital that is within a half-hour's driving time of their current location. Therefore, different people will have different opinions for hospitals, depending on where they are located.

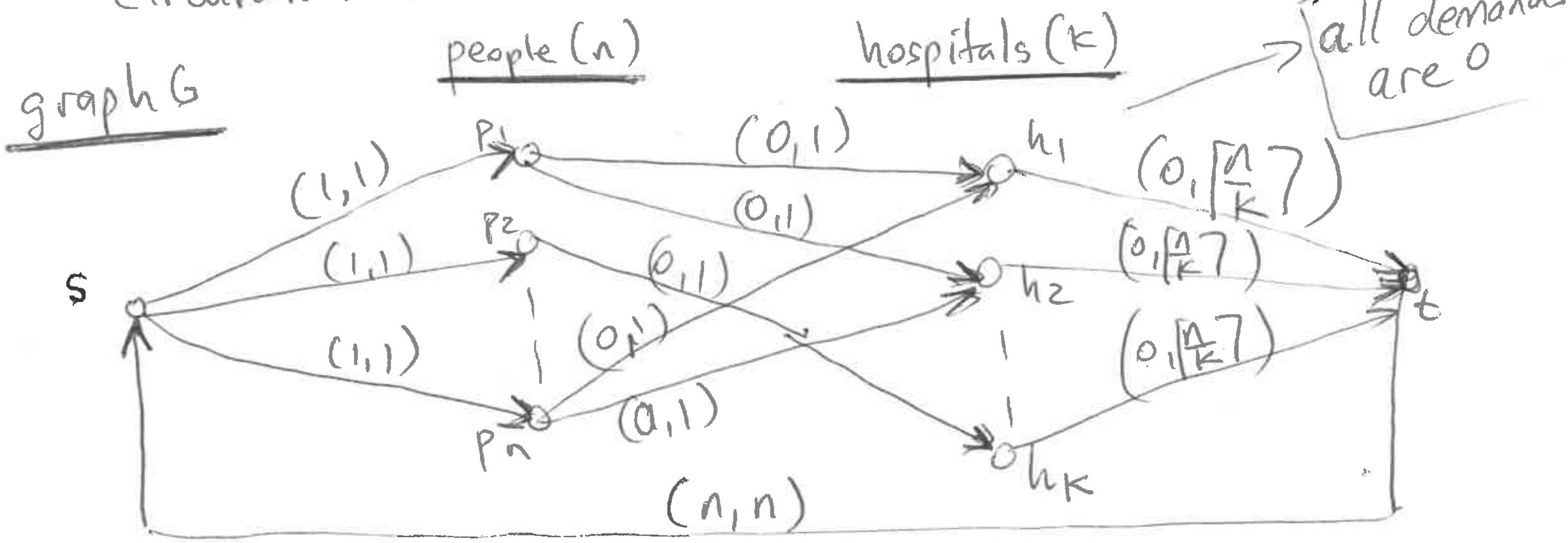
At the same time, we don't want to overload any one of the hospitals by sending too many patients its way. The paramedics are in touch by cell phone, and they want to collectively work out whether they can choose a hospital for each of the injured people in such a way that the load on the hospitals is *balanced*, that means each hospital receives at most $\lceil n/k \rceil$ people.

Give a **polynomial-time algorithm** that takes the given information about the people's location and determines whether this is possible.

Solution 1

Circulation with demands and lower bounds:

graph G



- edge (p_i, h_j) means that h_j is located within half-hour driving time from p_i's location

Semnification of the flow:

- flow on any edge (s, p_i) must be 1
- flow on the edge (h_j, t) represent the number of people transported to h_j
- flow on the edge (t, s) must be n
- flow on an edge (p_i, h_j) is:

$$\begin{cases} 0 & \text{- if } p_i \text{ is not transported to } h_j \\ 1 & \text{- if } p_i \text{ is transported to } h_j \end{cases}$$

- if G has a feasible circulation ⇒ then the people can be transported to the hospitals to meet the requirements
- otherwise, we cannot coordinate the transport of people to meet all the requirements

RT analysis

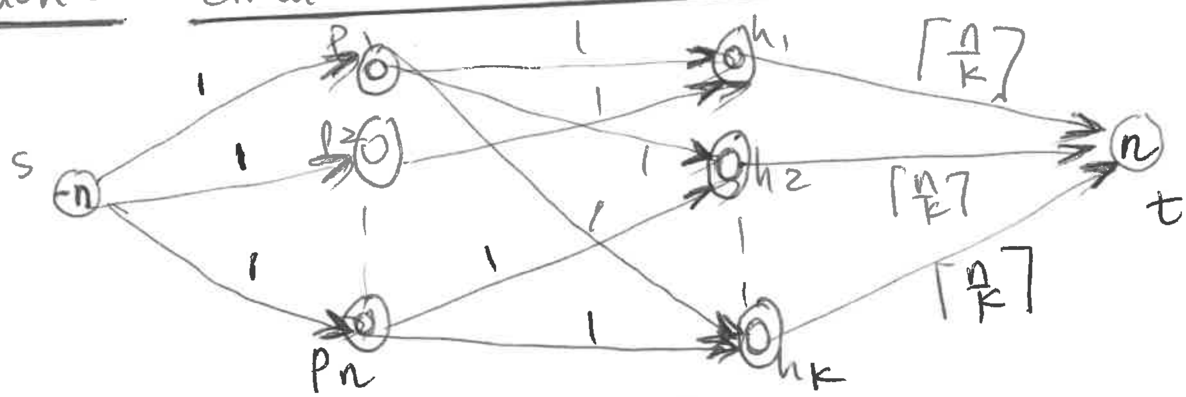
- Ford-Fulkerson ⇒ $RT = O(|f^*| \cdot E)$

  $|f^*| \leq n$

  $|E| \leq n \cdot k + n + k + 1 \Rightarrow |E| = O(nk)$

$\boxed{RT = O(n^2 k)}$ → polynomial RT!

Solution 2 — circulation with demands



Solution 3 — circulation with demand