Ottimizzazione Combinatoria II

Il problema del Set-Covering

ANTONIO SASSANO

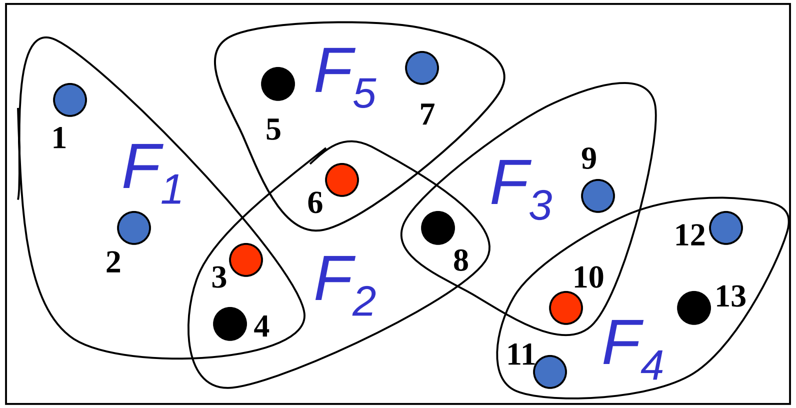
Università di Roma "La Sapienza"
Dipartimento di Informatica e Sistemistica

Roma, Aprile 2011

Set-Covering

$I = \{ F_1 , \dots , F_m \}$ Famiglia di sottoinsiemi di J

$T \subseteq J$ è un COVER di $I \Leftrightarrow T \cap F$ è non vuoto $\forall F \in I$ }



$J = \{ 1 , \dots , 13 \}$

$T_1 = \{3, 6, 10\}$; $T_2 = \{4, 5, 8, 13\}$ sono COVER di I

J è un cover di I

$X = \{1, 2, 7, 9, 11, 12\}$ non è un cover di I

Problema del *Set Covering ("Hitting Set")*

(SC) $\min c(T) = \sum_{e \in T} c_e$: $T \subseteq J$ è un COVER di I

Formulazione del Set-Covering

Problema del ***Set Covering ("Hitting Set")***

$$\text{(SC)}\ \min c(T) = \sum_{e \in T} c_e : \quad T \subseteq J, \ |T \cap F| \geq 1 \quad \forall F \in I$$

x^T **vettore di incidenza del COVER** T

$$x^T = \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \end{bmatrix} \qquad x^F = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 1 \end{bmatrix}$$

$$|T \cap F| \geq 1 \quad \forall F \in I \iff \sum_{e \in F} x_e^T \geq 1 \quad \forall F \in I$$

Ogni vettore di incidenza x^T di un COVER T soddisfa il sistema :

$$\sum_{e \in F} x_e \geq 1 \quad \forall F \in I$$

Formulazione del Set-Covering

Ogni vettore di incidenza x^T di un COVER T $(x^T \in S)$ **soddisfa il sistema :**

$$\sum_{e \in F} x_e \geq 1 \quad \forall F \in I$$

Il vettore di incidenza di un insieme $A \subseteq J$ che non è un COVER $(x^A \notin S)$ NON interseca (almeno) una $\bar{F} \in I$ e dunque **non soddisfa** la disequazione:

$$\sum_{e \in \bar{F}} x_e \geq 1$$

Il Poliedro

$$\boxed{\begin{array}{l} \sum_{e \in F} x_e \geq 1 \quad \forall F \in I \\ 0 \leq x_e \leq 1 \quad \forall e \in J \end{array}}$$

E' una FORMULAZIONE di S

Formulazione del Set-Covering

(SC) $\min \{ c(T) : T \subseteq J$ è un **COVER di** $I \}$

FORMULAZIONE NATURALE

$$P_{SC} = \begin{cases} \sum_{e \in F} x_e \geq 1 & \forall F \in I \\ 0 \leq x_e \leq 1 & \forall e \in J \end{cases}$$

$x \in P_{SC} \cap \{0,1\}^{|J|} \iff x$ vettore di incidenza di un ***COVER***

$$\text{(SC)} \begin{cases} \min c^T x \\ \sum_{e \in F} x_e \geq 1 & \forall F \in I \\ 0 \leq x_e \leq 1 & \forall e \in J \\ x_e \in \{0,1\} & \forall e \in J \end{cases}$$

Problema del Set-Covering

$$\text{(RSC)} \begin{cases} \min c^T x \\ \sum_{e \in F} x_e \geq 1 & \forall F \in I \\ 0 \leq x_e \leq 1 & \forall e \in J \end{cases}$$

Rilassamento Lineare (Naturale)

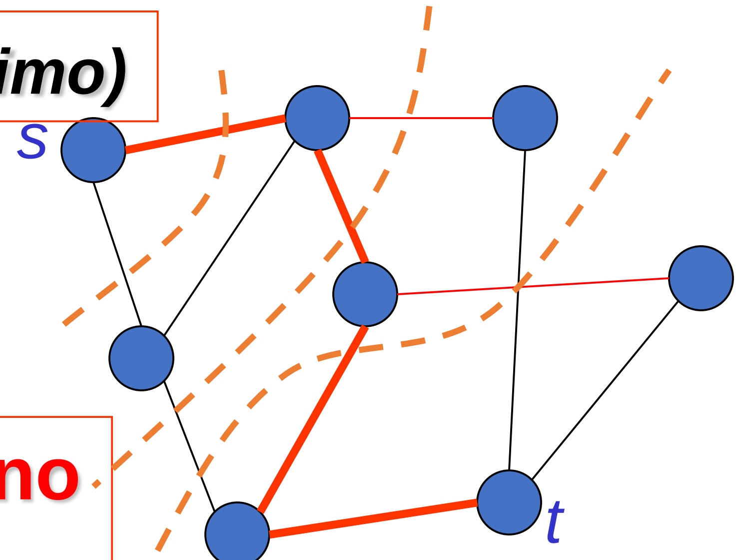
Set Covering: casi particolari

Grafo s-t connesso (di peso minimo)

J = Archi del grafo

$I = \{ F_1 , \dots , F_m \}$ Tagli s-t

"Ogni ***taglio s-t*** deve contenere **almeno un** arco di un ***cammino s-t*** "

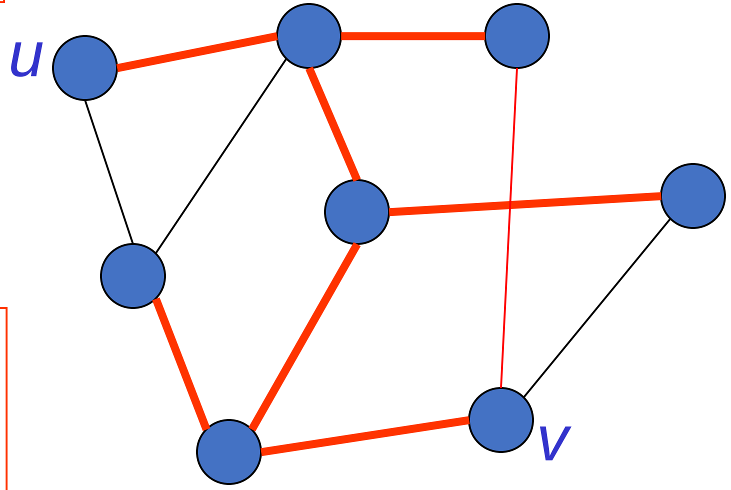


Grafo connesso (di peso minimo)

J = Archi del grafo

$I = \{ F_1 , \dots , F_m \}$ Tagli

"Ogni ***taglio*** deve contenere **almeno un** arco di un ***albero ricoprente*** "



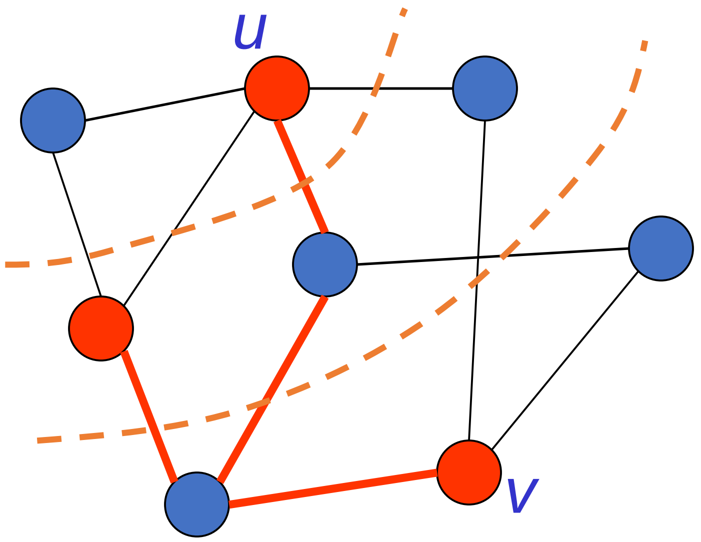
Set Covering – *Casi particolari*

Grafo di Steiner (di peso minimo)

J* = *Archi del grafo

T* $\subseteq$ *J*   *Insieme di nodi

$I = \{ F_1, \ldots, F_m \}$ ***Tagli u-v*** *per ogni coppia* ***u,v*** $\in$ ***T***



"..Un cammino tra ogni coppia di nodi di ***T*** "

Set Covering – *Crew Scheduling*

Servizi* (voli, tratte ferroviarie etc.) *da effettuare

Flight segments to be **covered** (Airline crew-scheduling):

1. **DFW** 9-12 **LGA**
2. **LGA** 13-15 **ORD**
3. **ORD** 16-18 **RDU**
4. **ORD** 17-19 **DFW**
5. **RDU** 19-21 **LGA**
6. **RDU** 19-21 **DFW**
7. **LGA** 14-16 **ORD**
8. **DFW** 16-18 **RDU**

***Sequenze di servizi* (pairings) *ammissibili per un equipaggio* (autista, macchinista etc.)**

Pairings:

1. **DFW** 9-12 **LGA** 14-16 **ORD** 17-19 **DFW** (1,7,4)
2. **LGA** 13-15 **ORD** 16-18 **RDU** 19-21 **LGA** (2,3,5)
3. **ORD** 16-18 **RDU** 19-21 **DFW** 9-12 **LGA** 13-15 **ORD** (3,6,1,2)
4. **DFW** 16-18 **RDU** 19-21 **DFW** (8,6)
5. **DFW** 16-18 **RDU** 19-21 **LGA** 14-16 **ORD** 17-19 **DFW** (8,5,7,4)

American Airlines

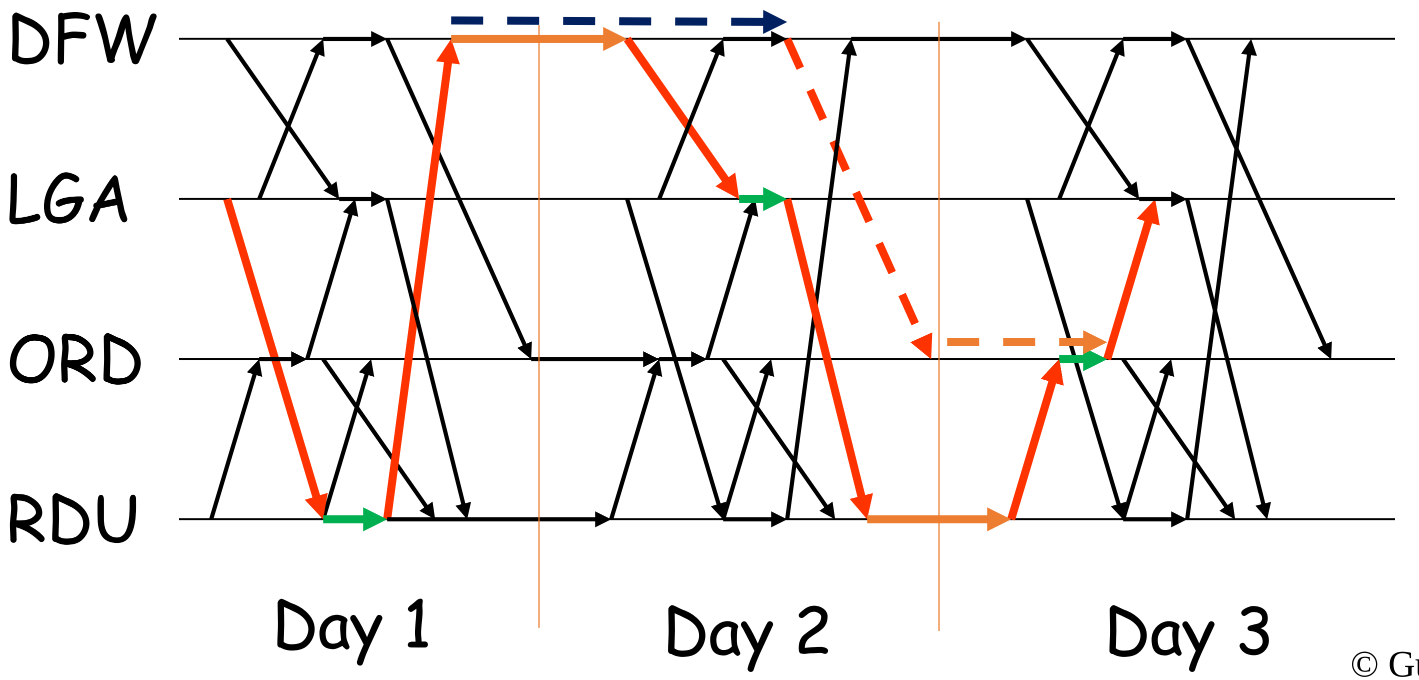
- **Problem**: Schedule crew (pilots and flight attendants) on **flight segments** to minimize cost.
  - over **25,000 pilots and flight attendants**
  - over **$1.5 Billion/year** in crew costs
- Assumes the flight segments are already fixed.
- **Methods described today**:
  - 1970-1992: **TRIP** (Trip Reevaluation and Improvement Program). **Local optimization** given an **initial guess**
  - 1992-present? **Global optimization (approximate)**

Crew Scheduling: sequenze di servizi ammissibili

Union and Federal Aviation Agency (FAA) rules

Some **example constraints**

- **8 hours** flying per duty period
- **12 hours** total duty time
- **Minimum layover time** – depends on hours of flying in previous duty period
- **Minimum time between flights** in a duty period



Crew Pairings

- **Example**: A 2 day **crew pairing** with DFW (Dallas-Fort Worth) as the base.

Duty period 1

Sign in: 8:00
DFW 9:00-10:00 AUS (segment 1)
AUS 11:00-13:00 ORD (segment 2)
ORD 14:00-15:00 SFO (segment 3)

Overnight in SFO

Duty period 2

Sign in: 7:00
SFO 8:00-9:00 LAX (segment 4)
LAX 10:00-11:45 SAN (segment 5)
SAN 13:00-19:30 DFW (segment 6)
Sign out: 19:45

Crew Scheduling: *Formulazione*

Come formulare il problema ?

Quali sono gli «eventi elementari» ?

Due *eventi elementari****:***

1. ***Il servizio i appartiene al «pairing» j*** ($x_{ij} \in \{0,1\}$)
2. ***Il pairing j è attivato oppure no*** ($y_j \in \{0,1\}$)

- *Simile al problema di* ***Localizzazione Impianti ma ..***
- ***Condizioni di appartenenza complesse*** *da esprimere ..*

.. e soprattutto

.. Il ***costo del pairing dipende dai servizi che lo compongono***

Soluzione: generare molti pairing ammissibili *con i rispettivi costi e poi scegliere* ***l'insieme ottimo di pairing che «copre» tutti i servizi*** *risolvendo un problema di* ***Set-Covering***

Set Covering – Crew Scheduling

J = Sequenze di servizi (pairings) ammissibili

Pairings:

1. DFW 9-12 LGA 14-16 ORD 17-19 DFW (1,7,4)
2. LGA 13-15 ORD 16-18 RDU 19-21 LGA (2,3,5)
3. ORD 16-18 RDU 19-21 DFW 9-12 LGA 13-15 ORD (3,6,1,2)
4. DFW 16-18 RDU 19-21 DFW (8,6)
5. DFW 16-18 RDU 19-21 LGA 14-16 ORD 17-19 DFW (8,5,7,4)

$I = \{ F_1 , \dots , F_m \}$

$F_k \subseteq J$ = ***insiemi di «pairing» che contengono lo stesso servizio** k*

$F_1 = \{ 1, 3 \}, F_2 = \{ 2, 3 \}, \dots , F_8 = \{ 4, 5 \}$

Cover = ***Insieme di «pairing»** $T \subseteq J$ **che interseca tutti gli insiemi** F_k* **<u>ovvero</u> con la proprietà che *ogni servizio è contenuto in qualche «pairing» di* T**

Crew Scheduling: cover e loro costo

Costo di un «pairing» $j \in J$

Cost can include both direct and indirect costs (e.g. employee satisfaction).

Example contributions to "cost".

- Total duty period time
- Time away from base (TAFB)
- Number and locations of layovers
- Number of time zone changes
- Cost of changing planes

Costo di un cover T

$c(T)$ *somma dei costi dei «pairing» in* T

Old System, TRIP (->1992)

1. Select an **initial solution (set of pairings)**
   Typically a modification from previous month
2. Repeat the following until no more improvements:
   - **Select small set of pairings** from current solution
     Typically 5-10 pairings from the same region
   - **Generate all valid pairings** that cover the same segments, and cost for each
     Typically a few thousand
   - **Optimize over these pairings** using the set-partitioning problem.

- **Advantage: Small subproblems**
- **Problem:** Only does **local optimization**

Generating the Pool of Pairings

- **<u>Disclaimer</u>**: *This is speculative since the authors say very little about it*.

- **Generating *6 million initial pairings* out of billions of possible pairings**
  1. **Generate graph**
     - **Vertex: time and airport**
     - **Edge: flight-segment, wait-time, or overlay**
     - **Edge weight: "excess cost" of edge**
  2. **Find <u>6 million shortest paths</u> in the graph that start and end at a crew base**

- This is a **heuristic** that prefers **short TAFB**.