# Today

Approximation Algorithm.

Facility Location.

# Maximum Weight Matching.

Bipartite Graph $G=(V,E)$, $w:E\rightarrow Z$.
Find maximum weight perfect matching.
Solution: $x_e$ indicates whether edge $e$ is in matching.

$$\max \sum_e w_e x_e$$
$$\forall v: \sum_{e=(u,v)} x_e = 1 \qquad p_v$$
$$x_e \geq 0$$

Dual.
Variable for each constraint. $p_v$ unrestricted.
Constraint for each variable. Edge $e$, $p_u + p_v \geq w_e$
Objective function from right hand side. $\min \sum_v p_v$

$$\min \sum_v p_v$$
$$\forall e=(u,v): \quad p_u + p_v \geq w_e$$

Weak duality? Price function upper bounds matching.
$\sum_{e\in M} w_e x_e \leq \sum_{e=(u,v)\in M} p_u + p_v \leq \sum_v p_u$.

Strong Duality? Same value solutions. Hungarian algorithm !!!

# Integer Vertex Solution.

Any "vertex" solution is integer!

Linear programming feasible region: **Polytope**.
Dimension of space: number of variables.
Vertex: intersection of $d$ linearly independent constraints.
$d$ "tight" constraints.

$$\max \sum_e w_e x_e$$
$$\forall v: \sum_{e=(u,v)} x_e = 1 \qquad p_v$$
$$x_e \geq 0$$

Dimension: $m$
Only $2n$ of the form $\sum_e x_e = 1$.
Must have $m-2n$ tight constraints of form $x_e = 0$.
Throw away variables that are 0.
Constraint matrix $C$ with $2n$ variables. $2n$ rows.

# ..and so on.

Constraint matrix $C$ with $2n$ variables. $2n$ rows.
Each variable in two constraints.
Matrix $C$ has 2 non-zeros in each row and column.
Average degree two bipartite graph.
Even cycle is linearly dependent:
Negate equations for vertices on one side and add them.
So need another constraint of form $x_e = 0$ for each cycle.
Now, matrix has degree 1 constraint:
or $\sum_e x_e = 1 \implies x_e = 1$.
This is an integer!!!
And so on.

Note:

also prove the determinant is 1 or $-1$
for the non-singular matrix.
Plus, Cramer's rule implies integrality.

That's what we did.

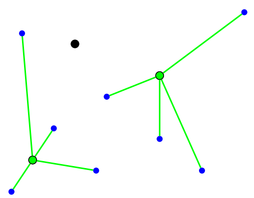
# Facility location

Set of facilities: $F$, opening cost $f_i$ for facility $i$

Set of clients: $D$.

$d_{ij}$ - distance between $i$ and $j$.
(notation abuse: clients/facility confusion.)

Triangle inequality: $d_{ij} \leq d_{ik} + d_{kj}$.



# Facility Location

Linear program relaxation:

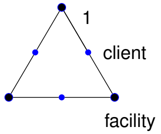
"Decision Variables".
$y_i$ - facility i open?
$x_{ij}$ - client $j$ assigned to facility $i$.

$$\min \sum_{i\in F} y_i f_i + \sum_{i\in F, j\in D} x_{ij} d_{ij}$$
$$\forall j \in D \quad \sum_{i\in F} x_{ij} \geq 1$$
$$\forall i \in F, j \in D \quad x_{ij} \leq y_i,$$
$$x_{ij}, y_i \geq 0$$

Facility opening cost.
Client Connnection cost.
Must connect each client.
Only connect to open facility.

## Integer Solution?

$$\min \sum_{i \in F} y_i f_i + \sum_{i \in F, j \in D} x_{ij} d_{ij}$$
$$\forall j \in D \quad \sum_{i \in F} x_{ij} \geq 1$$
$$\forall i \in F, j \in D \quad x_{ij} \leq y_i,$$
$$x_{ij}, y_i \geq 0$$




$x_{ij} = \frac{1}{2}$ edges.
$y_i = \frac{1}{2}$ edges.

Facility Cost: $\frac{3}{2}$ Connection Cost: 3
Any one Facility:
Facility Cost: 1 Client Cost: 3.7
Make it worse? Sure. Not as pretty!

## Round solution?

$$\min \sum_{i \in F} y_i f_i + \sum_{i \in F, j \in D} x_{ij} d_{ij}$$
$$\forall j \in D \quad \sum_{i \in F} x_{ij} \geq 1$$
$$\forall i \in F, j \in D \quad x_{ij} \leq y_i,$$
$$x_{ij}, y_i \geq 0$$

Round independently?

$y_i$ and $x_{ij}$ separately? Assign to closed facility!

Round $x_{ij}$ and open facilities?
Different clients force different facilities open.

Any ideas?

Use Dual!

## The dual.

$\min cx, Ax \geq b \leftrightarrow \max bx, y^T A \leq c.$

$$\min \sum_{i \in F} y_i f_i + \sum_{i \in F, j \in D} x_{ij} d_{ij}$$
$$\forall j \in D \quad \sum_{i \in F} x_{ij} \geq 1$$
$$\forall i \in F, j \in D \quad x_{ij} \leq y_i,$$

$$\min \sum_{i \in F} y_i f_i + \sum_{i \in F, j \in D} x_{ij} d_{ij}$$
$$\forall j \in D \quad \sum_{i \in F} x_{ij} \geq 1 \quad ; \alpha_j$$
$$\forall i \in F, j \in D \quad y_i - x_{ij} \geq 0 \quad ; \beta_{ij}$$
$$x_{ij}, y_i \geq 0$$

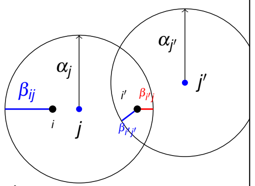
$$\max \sum_j \alpha_j$$
$$\forall i \quad \sum_{j \in D} \beta_{ij} \leq f_i \quad ; y_i$$
$$\forall i \in F, j \in D \quad \alpha_j - \beta_{ij} \leq d_{ij} \quad ; x_{ij}$$
$$\beta_{ij}, \alpha_j \geq 0$$

## Interpretation of Dual?

$$\min \sum_{i \in F} y_i f_i + \sum_{i \in F, j \in D} x_{ij} d_{ij}$$
$$\forall j \in D \quad \sum_{i \in F} x_{ij} \geq 1$$
$$\forall i \in F, j \in D \quad x_{ij} \leq y_i,$$
$$x_{ij}, y_i \geq 0$$

$$\max \sum_j \alpha_j$$
$$\forall i \in F \quad \sum_{j \in D} \beta_{ij} \leq f_i$$
$$\forall i \in F, j \in D \quad \alpha_j - \beta_{ij} \leq d_{ij} \quad x_{ij}$$
$$\alpha_j, \beta_{ij} \geq 0$$

$\alpha_j$ charge to client.
maximize price paid by client to connect!
Objective: $\sum_j \alpha_j$ total payment.
Client $j$ travels or pays to open facility $i$.
Costs client $d_{ij}$ to get to there.
Savings is $\alpha_j - d_{ij}$.
Willing to pay $\beta_{ij} = \alpha_j - d_{ij}$.
Total payment to facility $i$ at most $f_i$ before opening.
Complementary slackness: $x_{ij} \geq 0$ if and only if $\alpha_j \geq d_{ij}$.
only assign client to "paid to" facilities.



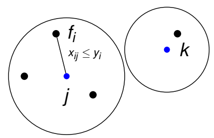

## Use Dual.

1. Find solution to primal, $(x, y)$. and dual, $(\alpha, \beta)$.
2. For smallest (remaining) $\alpha_j$,
   (a) Let $N_j = \{i : x_{ij} > 0\}$.
   (b) Open cheapest facility $i$ in $N_j$.
   Every client $j'$ with $N_{j'} \cap N_j \neq \emptyset$ assigned to $i$.
3. Removed assigned clients, goto 2.

## Integral facility cost at most LP facility cost.

**Claim:** Total facility cost is at most $\sum_i f_i y_i$.

2. For smallest (remaining) $\alpha_j$,
   (a) Let $N_j = \{i : x_{ij} > 0\}$.
   (b) Open cheapest facility $i$ in $N_j$.
   Every client $j'$ with $N_{j'} \cap N_j \neq \emptyset$ assigned to $i$.

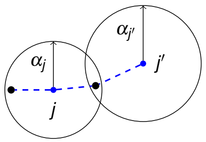



**Proof:** Step 2 picks client $j$.
$f_{\min}$ - min cost facility in $N_j$
$f_{\min} \leq f_{\min} \cdot \sum_{i \in N_j} x_{ij} \leq f_{\min} \sum_{i \in N_j} y_i \leq \sum_{i \in N_j} y_i f_i$.
For $k$ used in Step 2.
$N_j \cap N_k = \emptyset$ for $j$ and $k$ in step 2.
$\rightarrow$ Any facility in $\leq 1$ sum from step 2.
$\rightarrow$ total step 2 facility cost is $\leq \sum_i y_i f_i$.

## Connection Cost.

2. For smallest (remaining) $\alpha_j$,
   (a) Let $N_j = \{i : x_{ij} > 0\}$.
   (b) Open cheapest facility $i$ in $N_j$.
   Every client $j'$ with $N_{j'} \cap N_j \neq \emptyset$ assigned to $i$.

Client $j$ is directly connected. Clients $j'$ are indirectly connected.




Connection Cost of $j$: $\leq \alpha_j$.
Connection Cost of $j'$:
$\leq \alpha_{j'} + \alpha_j + \alpha_j \leq 3\alpha_{j'}$.
since $\alpha_j \leq \alpha_{j'}$

Total connection cost:
at most $3\sum_j \alpha_j \leq$ 3 times Dual OPT.

Previous Slide: Facility cost:
$\leq$ primal "facility" cost $\leq$ Primal OPT.

Total Cost: 4 OPT.

## Twist on randomized rounding.

Client $j$: $\sum_i x_{ij} = 1$, $x_{ij} \geq 0$.
Probability distribution! $\rightarrow$ Choose from distribution, $x_{ij}$, in step 2.

Expected opening cost:
$\sum_{i \in N_j} x_{ij} f_i \leq \sum_{i \in N_j} y_i f_i$.
and separate balls implies total $\leq \sum_i y_i f_i$.

$D_j = \sum_i x_{ij} d_{ij}$ Expected connection cost of primal for $j$.

Expected connection cost $j'$ $\alpha_j + \alpha_{j'} + D_j$.

In step 2: pick in increasing order of $\alpha_j + D_j$.

$\rightarrow$ Expected cost is $\leq (2\alpha_{j'} + D_{j'})$.
Connection cost: $2\sum_j \alpha_j + \sum_j D_j$.
$2OPT(D)$ plus connection cost of primal.

Total expected cost:
Facility cost is at most facility cost of primal.
Connection cost at most $2OPT$ + connection cost of prmal.
$\rightarrow$ at most $3OPT$.

## Primal dual algorithm.

1. Feasible integer solution.
2. Feasible dual solution.
3. Cost of integer solution $\leq \alpha$ times dual value.

Just did it. Used linear program. Faster?

Typically. (If dual is maximization.)
Begin with feasible dual.
Raise dual variables until tight constraint.
Set corresponding primal variable to an integer.

Recall Dual:

$$\max \sum_j \alpha_j$$
$$\forall i \in F \quad \sum_{j \in D} \beta_{ij} \leq f_i$$
$$\forall i \in F, j \in D \quad \alpha_j - \beta_{ij} \leq d_{ij}$$
$$\alpha_j, \beta_{ij} \leq 0$$

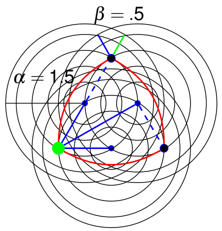
## Facility location primal dual.

**Phase 1:** 1. Initially $\alpha_j, \beta_{ij} = 0$.
2. Raise $\alpha_j$ for every (unconnected) client.
When $\alpha_j = d_{ij}$ for some $i$
raise $\beta_{ij}$ at same rate Why? Dual: $\alpha_j - \beta_{ij} \leq d_{ij}$.
Intution:Paying $\beta_{ij}$ to open $i$.
Stop when $\sum_j \beta_{ij} = f_i$.
Why? Dual: $\sum_i \beta_{ij} \leq f_i$
Intution: facility paid for.
*Temporarily open* $i$.
*Connect* all tight $ji$ clients $j$ to $i$.

3. Continue until all clients connected.

**Phase 2:**
Make "edge" between two facilities if paid by a common client.
Permanently open an independent set of facilities in common client graph.

For client $j$, connected facility $i$ is opened. Good.
Connected facility not open
$\rightarrow$ exists client $j'$ paid $i$ and connected to open facility.
Connect $j$ to $j'$'s open facility.




Constraints for dual.
$\sum_j \beta_{ij} \leq f_i$
$\alpha_i - \beta_{ij} \leq d_{ij}$.
Grow $\alpha_j$.
$\alpha_j = d_{ij}$!
Tight constraint: $\alpha_j - \beta_{ij} \leq d_{ij}$.
Grow $\beta_{ij}$ (and $\alpha_j$).
$\sum_j \beta_{ij} = f_i$ for all facilities.
Tight: $\sum_j \beta_{ij} \leq f_i$
LP Cost: $\sum_j \alpha_j = 4.5$

*Temporarily open all facilities.*
Assign Clients to "paid to" open facility.
Connect facilities with common client.
Open independent set.
Connect to "killer" client's facility.
Cost: 1 + 3.7 = 4.7.
A bit more than the LP cost.

## Analysis

**Claim:** Client only pays one facility.

Independent set of facilities.

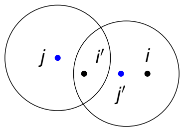
**Claim:** $S_i$ - directly connected clients to open facility $i$.
$f_i + \sum_{j \in S_i} d_{ij} \leq \sum_j \alpha_j$.

**Proof:**
$f_i = \sum_{j \in S_i} \beta_{ij} = \sum_{j \in S_i} \alpha_j - d_{ij}$.
Since directly connected: $\beta_{ij} = \alpha_j - d_{ij}$.

□

## Analysis.

**Claim:** Client $j$ is indirectly connected to $i$
$\rightarrow d_{ij} \leq 3\alpha_j$.




Directly connected to (temp open) $i'$ conflicts with $i$.
exists $j'$ with $\alpha_{j'} \geq d_{ij'}$ and $\alpha_j \geq d_{i'j'}$.
When $i'$ opens, stops both $\alpha_j$ and $\alpha_j'$.
$\alpha_{j'}$ stopped no later (..maybe earlier..)
$\alpha_{j'} \leq \alpha_j$.
Total distance from $j$ to $j'$.
$d_{ji'} + d_{i'j'} + d_{j'i} \leq 3\alpha_j$ □

## Putting it together!

**Claim:** Client only pays one facility.
**Claim:** $S_i$ - directly connected clients to open facility $i$.
$f_i + \sum_{j \in S_i} d_{ij} \leq \sum_j \alpha_j$.
**Claim:** Client $j$ is indirectly connected to $i$
$\rightarrow d_{ij} \leq 3\alpha_j$.

Total Cost:
direct clients dual ($\alpha_j$) pays for facility and own connections.
plus no more than 3 times indirect client dual.
Total Cost: 3 times dual.
feasible dual upper bounds fractional (and integer) primal.

3 OPT.

Fast! Cheap! Safe!

Check: if time.

Won't see you on Tuesday.

Guest Speaker: Tselil Schramm.

Semidefinite Programming and Approximation.