Minimum Weight Constrained Forest Problem

Slide 1

Minimum Weight Constrained Forest Problems

Xiaoyun Ji, John E. Mitchell
Department of Mathematical Sciences
Rensselaer Polytechnic Institute
Troy, NY, USA
jix@rpi.edu, mitchj@rpi.edu

2005 Optimization Days
Montreal, Canada
May 09, 2005

Slide 2

Contents

Slide 3

Problem Definition

Minimum Weight Constrained Forest Problem **MWCF**

- Given an **integer** $S > 0$ and a **complete graph** $G = (V, E)$ with weights c_e on each edge e.
- Look for a spanning forest of graph G with each tree in the forest spanning **at least S vertices**, so as to minimize the weight of the spanning forest.

NP-hard when $S \geq 4$.

[Imielińska, Kalantari and Khachiyan, 1993]

Slide 4

An Example

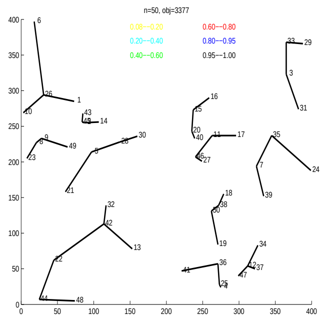

Figure 1: One Solution for a problem of size $n = 50, S = 4$

Slide 5

Applications of clustering with minimum size requirements

- Micro-aggregation for Statistical Data
  Statistical data is released to the public in small groups so that the confidentiality of respondents are protected and the informational content of the data are preserved as much as possible.
- Political Districting
  Compact, contiguous, and balanced districts
- Sports Team Alignment
- Telecommunication Network Design

Slide 6

Related Graph Partition Problems

minimum weight balanced spanning forest problem

[Ali and Huang, 1991]

the number of trees are fixed and the number of vertices in each tree cannot differ by more than 1.

- Problem is set up as an IP. Lagrangian relaxation and heuristics are used to find lower bound and upper bounds. Similar to column generation.
- $n = 100$, edge density= 20%, gap= 3%, iterations=700-1400

Slide 7

k-tree partition

[Buttmann-Beck and Hassin, 1998]

to partition V into p subsets of given size,
so as to minimize the spanning forest weight.
Heuristic Algorithm is proposed.

min-max spanning forest problem

[Yamada, Takahashi and Kataoka, 1997]

the root of each tree is given,
The objective is to minimize the maximum of the tree weights.
a branch and bound method based on a combinatorial methods to establish lower bound.

Slide 8

clique partition with minimum size requirement

[Ji and Mitchell, 2005]

Each cluster has to have at least S vertices, and the clique weight is used to measure the weight of each cluster.

- Problem is set up as an IP and solved by branch-and-cut.
- $n = 100$, gap= 3.77%, time = 176s

Slide 9

Integer Programming Formulation

Define a binary variable x_e with

$$x_e = \begin{cases} 1 & \text{if } e \text{ is in the spanning forest} \\ 0 & \text{otherwise} \end{cases}$$

Notation

- number of vertices : $n = |V|$
- number of edges: $\binom{n}{2} = \frac{n(n-1)}{2}$
- k, r : $n = kS + r$, $0 \le r \le S - 1$

Slide 10

Our IP formulation for MWCF is

$$\begin{aligned} \min \quad & \sum_{e \in E} c_e x_e & & \\ \text{s.t.} \quad & \sum_{e \in E(Q)} x_e \le |Q| - 1, & \forall Q \subseteq V \quad & (1) \\ & \sum_{e \in E(W)} x_e + \sum_{e \in \delta(W)} x_e \ge |W| - \lfloor \frac{|W|}{S} \rfloor & \forall W \subseteq V \quad & (2) \\ & x_e \in \{0, 1\} & & \end{aligned}$$

Where

$$\begin{aligned} E(W) &= \{(u, v) | u \in V, v \in W\}, \\ \delta(W) &= \{(u, v) | u \in W, v \notin W\}. \end{aligned}$$

(1) – Cycle Constraint.

(2) – Flower Constraint.

Slide 11

An Example for Flower Constraint

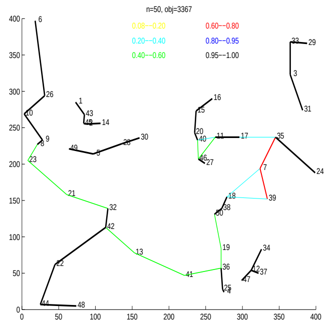

Figure 2: Flower constraint violated on vertex set {21,23}.

Slide 12

Initial Relaxation

$$\begin{aligned} \min \quad & \sum_{e \in E} c_e x_e & & \\ \text{s.t.} \quad & \sum_{e \in E(V)} x_e \le |V| - 1 & & (3) \\ & \sum_{e \in \delta(v)} x_e \ge 1, & \forall v \in V \quad & (4) \\ & \sum_{e \in E(V)} x_e \ge |V| - \lfloor \frac{|V|}{S} \rfloor & & (5) \\ & 0 \le x_e \le 1 & & (6) \end{aligned}$$

Cycle Constraint – Constraint (3)

Flower Constraint – Constraints (4) and (5)

Slide 13

The MWCF Polytope

$$F(G,S) := conv\{x \in \{0,1\}^n : x \text{ is the incidence vector of a spanning forest } F \text{ on } G, \text{ each tree in the forest has at least } S \text{ nodes}\}$$

$F(G,S)$ is full dimensional. $dim(F(G,S)) = n(n-1)/2$.

Slide 14

Facet Defining Constraints

We have shown

- $x_e \geq 0$ is a facet for $F(G,S)$.
- Cycle constraint is a facet for $F(G,S)$, where $|V| \geq 2S$, if and only if $Q = V$ or $|Q| \leq n - S$.
- Flower constraint is a facet for $F(G,S)$, where $n = |V| = kS + r$ and $|W| = pS + q$, if and only if $W = V$ or $q \geq 1$ and $p \leq k - 2$.

Minimum Weight Constrained Forest Problem

Slide 15

Branch and Cut Framework

1. **Initialize:** Set up initial relaxation.
2. **Solve the current relaxation** using CPLEX.
3. Use a **primal heuristic** to generate a feasible solution. Update the upper bound if necessary.
4. Call **separation routine** to add in violated constraints. If none is found then **go to Step 5**. Otherwise, add in constraints, and **go to Step 2**.
5. Start **Branching** using MINTO,

Slide 16

Separation Routine

1. **Cycle inequalities**:
   - can be identified in $O(n^4)$ time [Padberg and Wolsey, 1983].
   - Because of the objective function, it is not often violated.
2. **Flower inequalities**:
   - Important: 2 more constraints can change the computation time for a problem of size ($n = 50$, $S = 4$) from (1241 nodes, 40 seconds) to (3 nodes, 6 seconds).
   - Spend more effort to look for them.

Minimum Weight Constrained Forest Problem

Slide 17

Heuristic Algorithm for finding violated flower constraints

1. Consider a graph K_n with the LP solution x as edge weights.
2. Identify 2 types of vertices in this graph.
   - The vertices that are only connected with 2 fractional edges and they sum up to 1. (Eg, vertices 21, 23, etc.)
   - The vertices that are connected with one edge of value one and nothing else, at the same time, the other end point of this edge is connected to no other edges of value one. (Eg, vertices 24, 27, etc.).
3. Find the shortest path between every pair of vertices found above, and check if any path violates flower constraints.

Slide 18

Computational Results

Experiment Data

- Type I: Edge weight = Euclidean distance between data points
  - $x \sim U[1, 100]$
  - $y \sim U[1, 100]$
- Type II: Edge Weight $\sim$ uniform distribution in $[1, 100]$
- Type III: Micro-aggregation Data
  - $x \sim exp(\lambda)$
  - $y \sim exp(\lambda)$
- Type IV: Micro-aggregation Data
  - $u \sim exp(\lambda)$, $v \sim U[0, 1]$
  - $x = u$, $y = uv$

Slide 19

Experiment Setup

- Experiments on a Sun Ultra 10 workstation
- Algorithm coded in **MINTO**
- Runtime reported in seconds.
- Every set of experiments: **S=4**, r=1,2,3; **S=7**, r=1,...,6
- For problems not solved to optimality at the cutting plane stage, the branch and cut stage has a time limit of **500 seconds**.

Slide 20

Experiment Results

n	81-83	121-123	161-163	201-203	241-243	301-303	421-423
$k = \lfloor \frac{n}{S} \rfloor$	20	30	40	50	60	75	105
Heuristic Alg.							
Gap	3.84%	3.46%	4.72%	3.02%	4.21%	5.46%	5.63%
Time	0.0216	0.0429	0.0729	0.1107	0.1537	0.2376	0.4592
Cutting Plane							
Total Instances	15	15	15	15	15	15	15
Solved exactly	1	1	0	0	1	0	0
Gap	0.61%	1.15%	1.32%	1.90%	1.69%	3.64%	4.81%
Time	9.92	28.10	64.37	99.05	142.50	183.68	403.43
Cycle	35	56	84	103	125	156	187
Flower	311	515	819	1015	1183	1122	1149
LPs solved	54	60	76	81	97	64	77
B&C							
Total Instances	14	14	15	15	14	15	15
Solved exactly	14	14	15	5	7	2	0
Better Solution	9	12	11	14	13	15	6
Gap	0%	0%	0%	0.51%	0.56%	1.88%	4.06%
Total Time	17.85	53.67	171.20	401.86	403.44	492.98	513.18
Nodes	15	20	36	46	29	15	3

Table 1: Branch-and-Cut Results on MWCF Type I Problems for $S = 4$

Slide 21

n	121-123	121-123	121-123	121-123	121-123
S	7	10	20	30	40
$k = \lfloor \frac{n}{S} \rfloor$	17	12	6	4	3
Heuristic Alg.					
Gap	2.70%	2.13%	1.20%	1.10%	0.98%
Time	0.0424	0.0424	0.0428	0.0429	0.0428
Cutting Plane					
Total Instances	15	15	15	15	15
Solved exactly	1	1	1	0	0
Gap	1.56%	1.55%	1.07%	0.87%	0.77%
Time	17.46	14.96	12.62	12.93	11.44
Cycle	56	63	70	85	83
Flower	243	1837	119	105	95
LPs solved	38	30	29	32	28
B&C					
Total Instances	14	14	14	15	15
Solved exactly	14	14	13	13	15
Better Solution	14	14	10	14	13
Gap	0%	0%	0.08%	0.18%	0.0%
Total Time	121.94	222.14	177.15	201.92	61.05
nodes	94	196	250	293	85

Table 2: Branch-and-Cut Results on MWCF Type I Problems for $n = 121 - 123$ and $S = 7 - 40$

Slide 22

n	81-83	121-123	161-163	201-203	241-243	301-303
$k = \lfloor \frac{n}{S} \rfloor$	20	30	40	50	60	75
Heuristic Alg.						
Gap	2.67%	3.19%	2.67%	2.94%	2.84%	3.17%
Time	0.0214	0.0424	0.0718	0.1084	0.1551	0.2384
Cutting Plane						
Total Instances	15	15	15	15	15	15
Solved exactly	9	9	7	5	5	1
Gap	0.46%	0.25%	0.24%	0.40%	0.44%	2.08%
Time	5.61	11.52	18.41	26.95	48.28	58.60
Cycle	1	1	1	1	1	1
Flower	134	235	381	477	751	640
LPs solved	24	27	34	32	52	37
B&C						
Total Instances	6	6	8	10	10	14
Solved exactly	6	6	8	10	10	14
Better Solution	4	6	5	9	10	12
Gap	0%	0%	0%	0%	0%	0%
Total Time	5.84	12.54	24.28	34.71	60.20	104.38
Nodes	2	3	7	6	6	11

Table 3: Branch-and-Cut Results on MWCF Type II Problems for $S = 4$

Minimum Weight Constrained Forest Problem

Slide 23

n	81-83	121-123	161-163	201-203	241-243	301-303
$k = \lfloor \frac{n}{S} \rfloor$	20	30	40	50	60	75
Heuristic Alg.						
Gap	3.25%	3.57%	4.12%	3.40%	3.33%	5.05%
Time	0.0206	0.0413	0.0703	0.1067	0.1529	0.2357
Cutting Plane						
Total Instances	15	15	15	15	15	15
Solved exactly	4	4	0	0	0	0
Gap	0.98%	1.03%	1.76%	1.24%	1.25%	3.94%
Time	10.75	32.44	75.99	106.53	170.31	221.39
Cycle	36	65	100	136	170	177
Flower	284	505	952	1132	1464	1153
LPs solved	39	49	84	80	108	75
B&C						
Total Instances	11	11	15	15	15	15
Solved exactly	11	11	14	14	12	3
Better Solution	9	9	14	14	14	14
Gap	0%	0%	0.02%	0.01%	0.16%	0.68%
total time	12.84	102.90	338.29	449.86	474.85	918.58
nodes	5	38	68	44	19	20

Table 4: Branch-and-Cut Results on MWCF Type III Problems for $S = 4$

Slide 24

n	81-83	121-123	161-163	201-203	241-243	301-303
$k = \lfloor \frac{n}{S} \rfloor$	20	30	40	50	60	75
Heuristic Alg.						
Gap	3.03%	4.22%	3.47%	3.75%	3.01%	4.70%
Time	0.0210	0.0420	0.0709	0.1077	0.1540	0.2369
Cutting Plane						
Total Instances	15	15	15	15	15	15
Improved	13	15	15	15	13	12
Solved exactly	1	2	0	0	3	0
Gap	0.86%	0.79%	1.38%	1.34%	0.91%	3.60%
Time	10.75	26.69	48.56	87.48	113.29	113.01
Cycle	26	51	55	94	103	85
Flower	259	471	646	953	1014	574
LPs solved	41	40	61	76	82	34
B&C						
Total Instances	14	13	15	15	12	15
Solved exactly	14	13	13	14	5	3
Better Solution	10	7	14	14	12	15
Gap	0%	0%	0.04%	0.19%	0.07%	0.97%
total time	14.07	51.38	141.78	312.57	269.16	471.29
nodes	9	21	22	28	18	13

Table 5: Branch-and-Cut Results on MWCF Type IV Problems for $S = 4$

Slide 25

Conclusion

- Studied a new class of graph partition problem: Graph Partition Problems with minimum size requirement
- Discussed MWCF polytope structure.
- Established facets for MWCF polytope.
- Solved MWCF successfully using branch-and-cut.