H.P. WILLIAMS

LONDON SCHOOL OF ECONOMICS

APPLICATIONS AND FORMULATIONS OF THE TRAVELLING SALESMAN PROBLEM

h.p.williams@lse.ac.uk

ICORS March 2014

1. What is the Travelling Salesman Problem?

2. Applications

3. Formulations as an Integer Programme

4. Projections of different formulations into same space in order to compare strengths of different Linear Programming formulations

A Salesman wishes to travel around a given set of cities, and return to the beginning, covering the smallest total distance

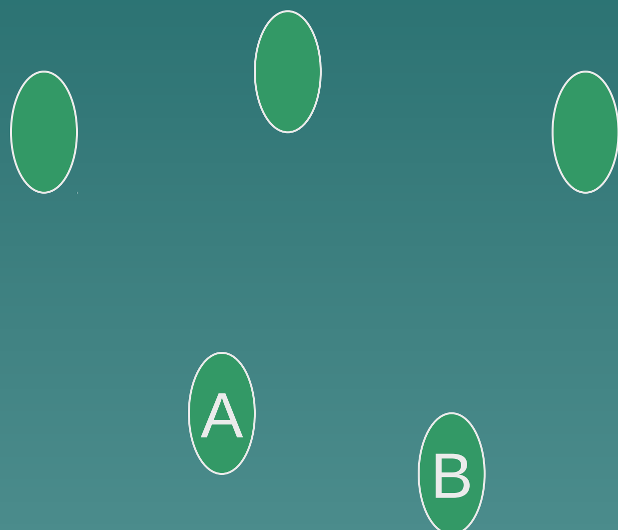
Easy to State

Difficult to Solve

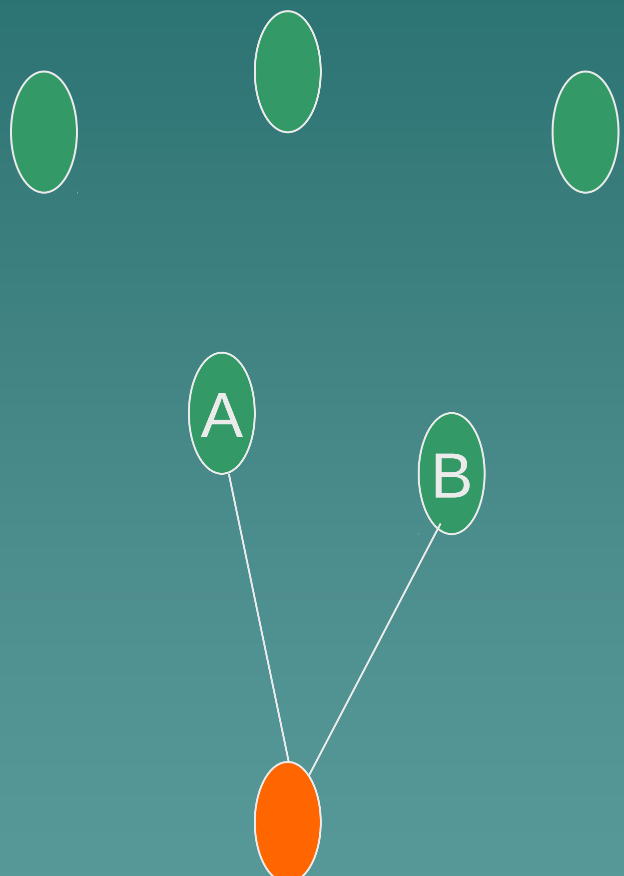
- Go around a number of cities visiting each one in such an order as to cover minimum total distance
- Distance X to Y may be same as distance Y to X (symmetric TSP) or different (assymetric TSP)

If there is no condition to return to the beginning. It can still be regarded as a TSP.

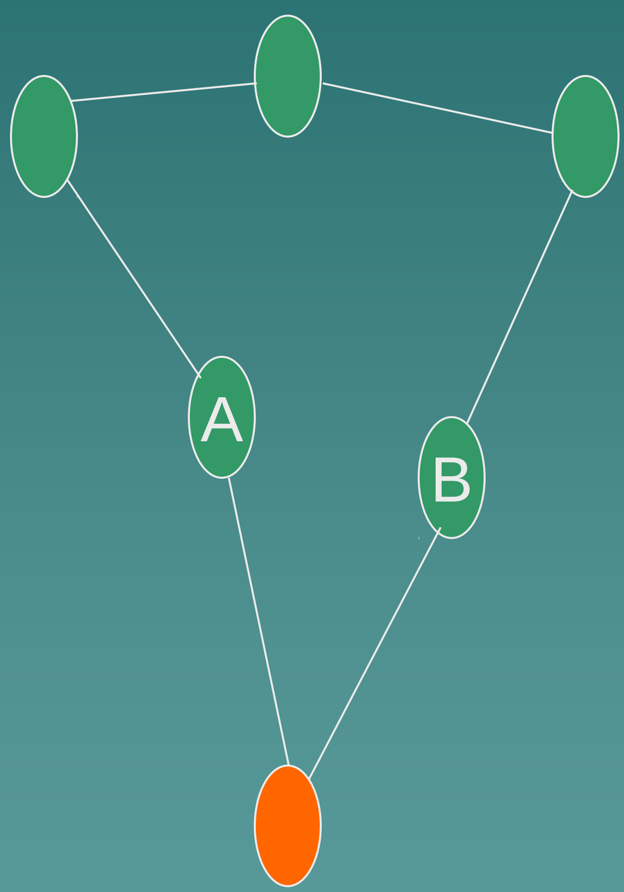
Suppose we wish to start at A and finish at B visiting all cities in between in best order.



Connect A and B to a ‘dummy’ city at zero distance
(If no stipulation of start and finish cities connect all to dummy at zero distance)



Create a TSP Tour around all cities



- If we wish to return to beginning we want
- Minimum Cost Hamiltonian Circuit

- If no requirement to return to beginning we want
- Minimum Cost Hamiltonian Path

- Essentially equivalent problems

Applications of the TSP

Routing around Cities

Computer Wiring - connecting together computer
components using minimum wire length

Archaeological Seriation - ordering sites in time

Genome Sequencing - arranging DNA fragments in sequence

Job Scheduling - sequencing jobs in order to minimise total set-up time between jobs

Wallpapering to Minimise Waste

NB: First three applications generally *symmetric*
Last three *asymmetric*

History of TSP

1800’s Irish Mathematician, Sir William Rowan Hamilton

1930’s Studied by Mathematicians Menger, Whitney, Flood etc.

1954 Dantzig, Fulkerson, Johnson, 49 cities (capitals of USA states) problem solved

1971 64 Cities

1975 100 Cities

1977 120 Cities

1980 318 Cities

1987 666 Cities

1987 2392 Cities (Electronic Wiring Example)

1994 7397 Cities

1998 13509 Cities (all towns in the USA with population > 500)

2001 15112 Cities (towns in Germany)

2004 24978 Cities (places in Sweden)

But many smaller instances not yet solved (to proven optimality)

Routing

Recent TSP Problems and Optimal Solutions

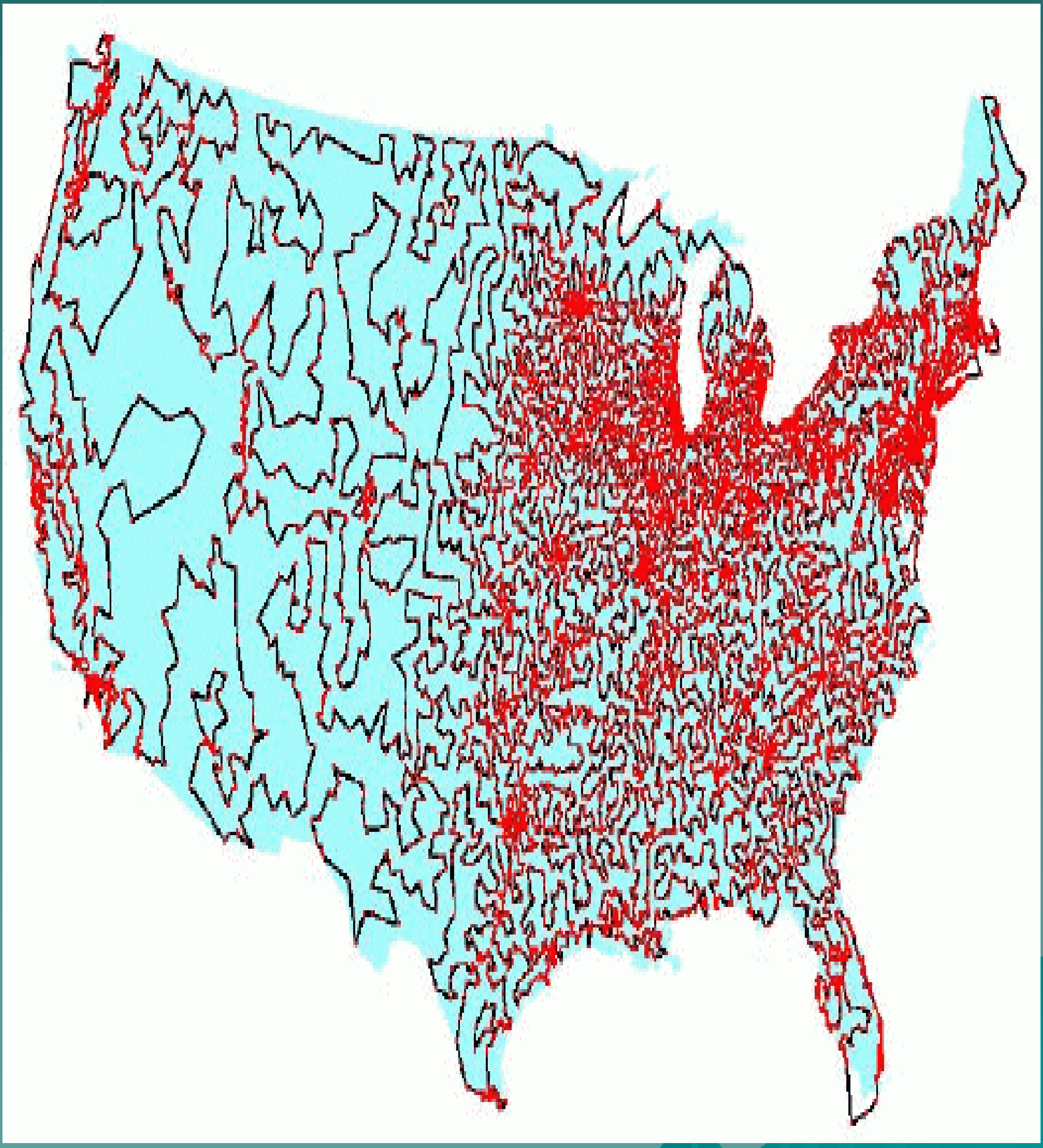
from

Web Page of William Cook, Georgia Tech, USA

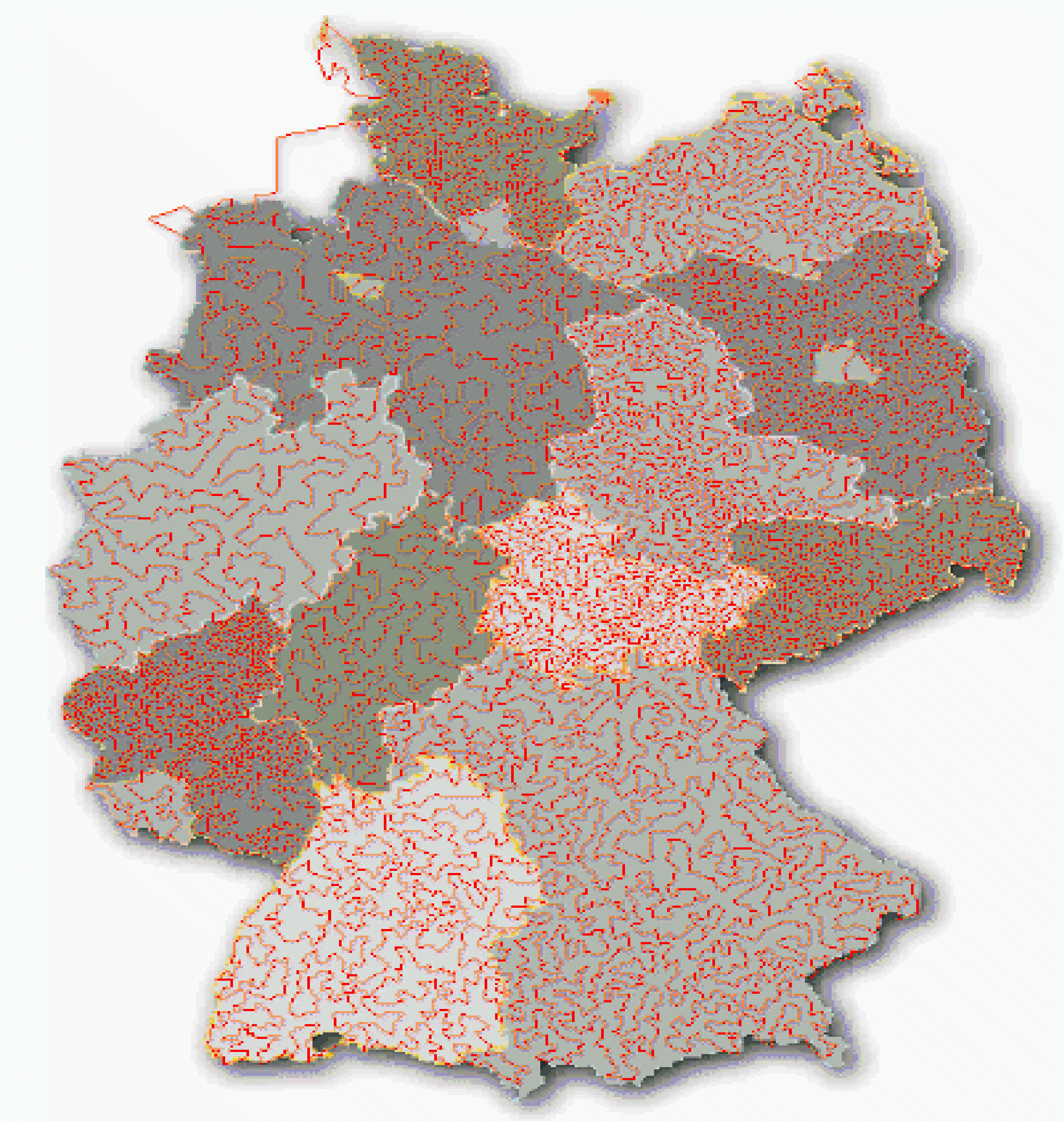
with Thanks

Routing

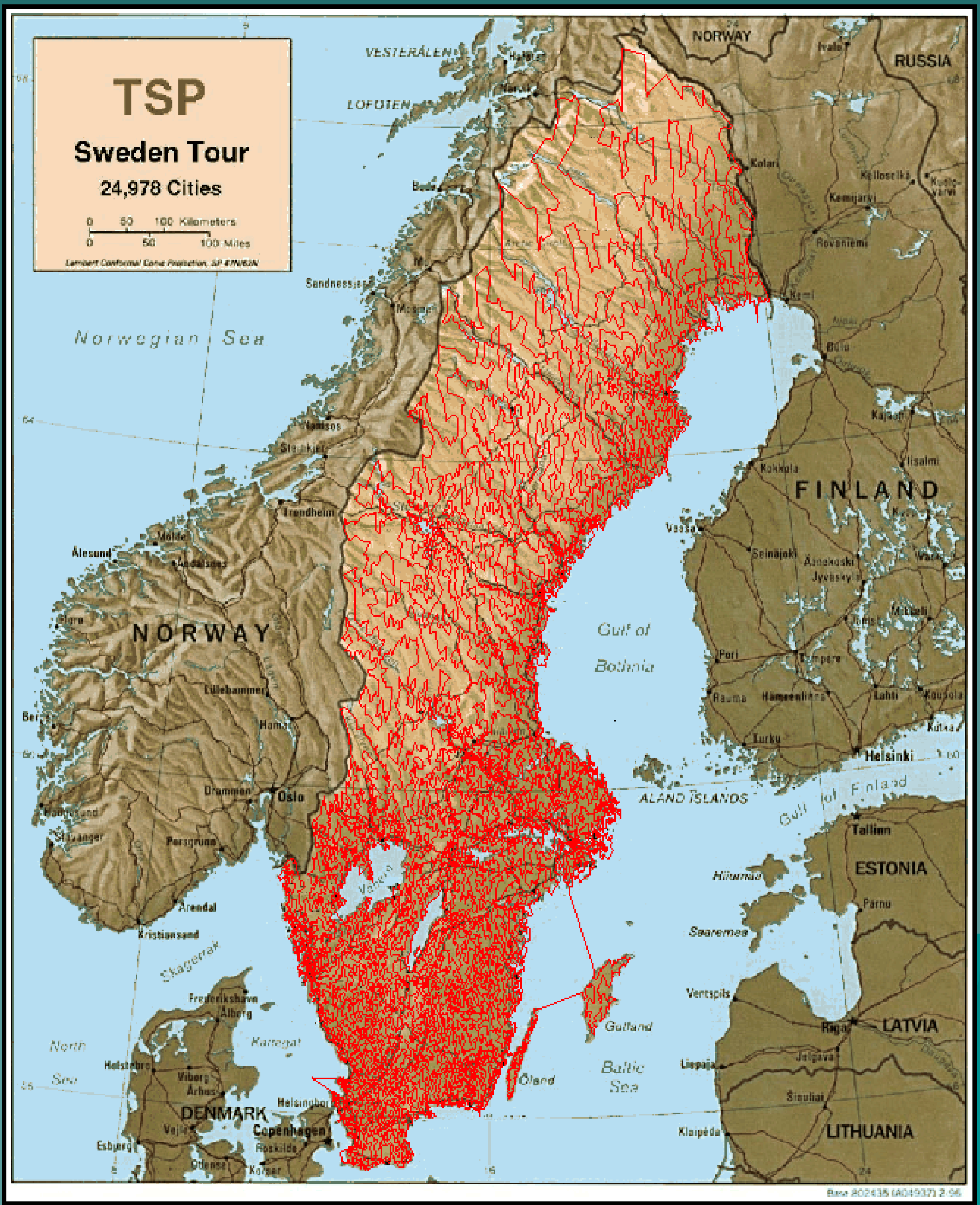
USA Towns of 500 or more population 13509 cities 1998 Applegate, Bixby, Chvátal and Cook



Towns in Germany 15112 Cities 2001 Applegate, Bixby, Chvátal and Cook

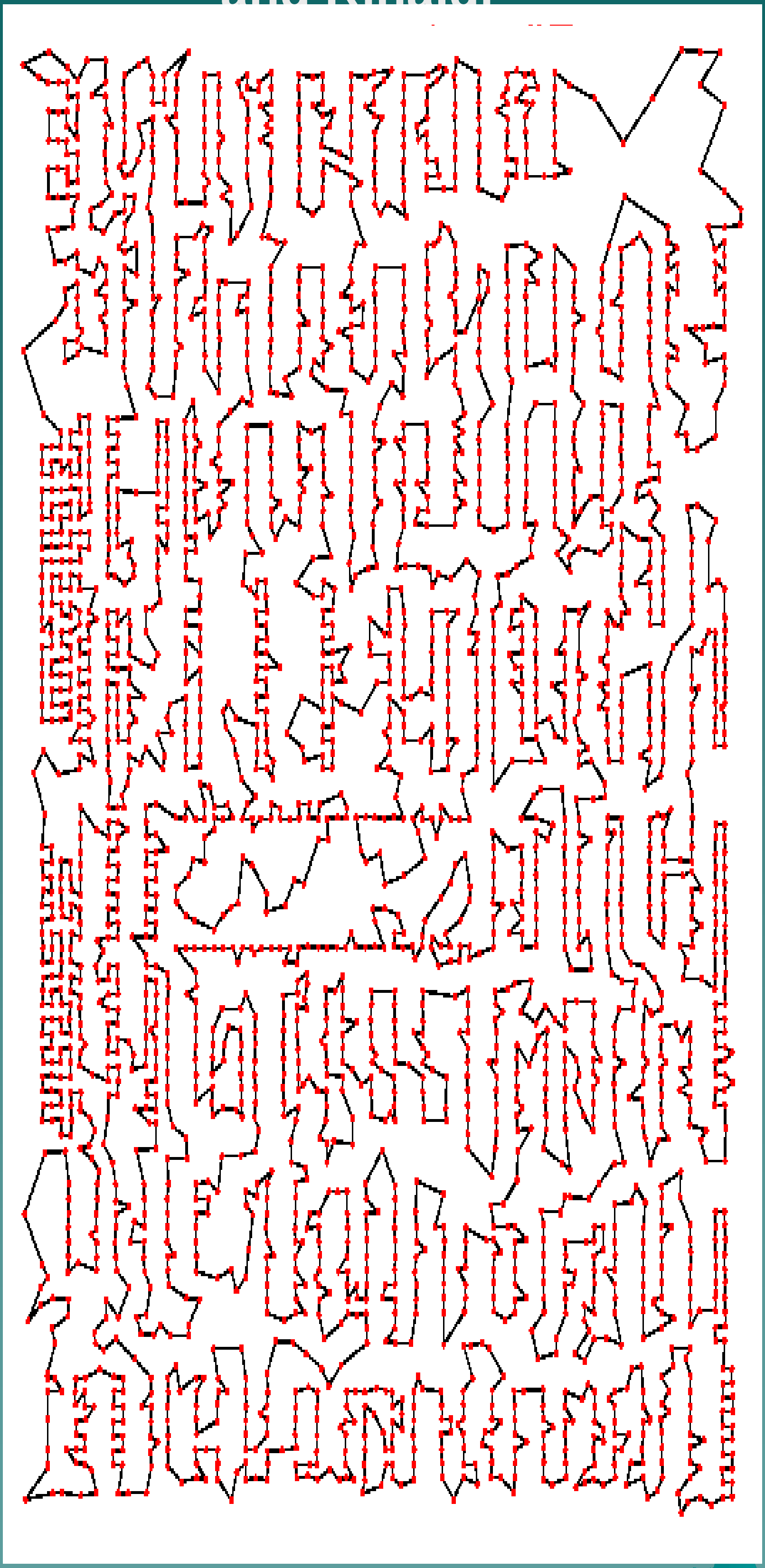


Sweden 24978 Cities 2004 Applegate, Bixby, Chvátal, Cook and Helsgaun

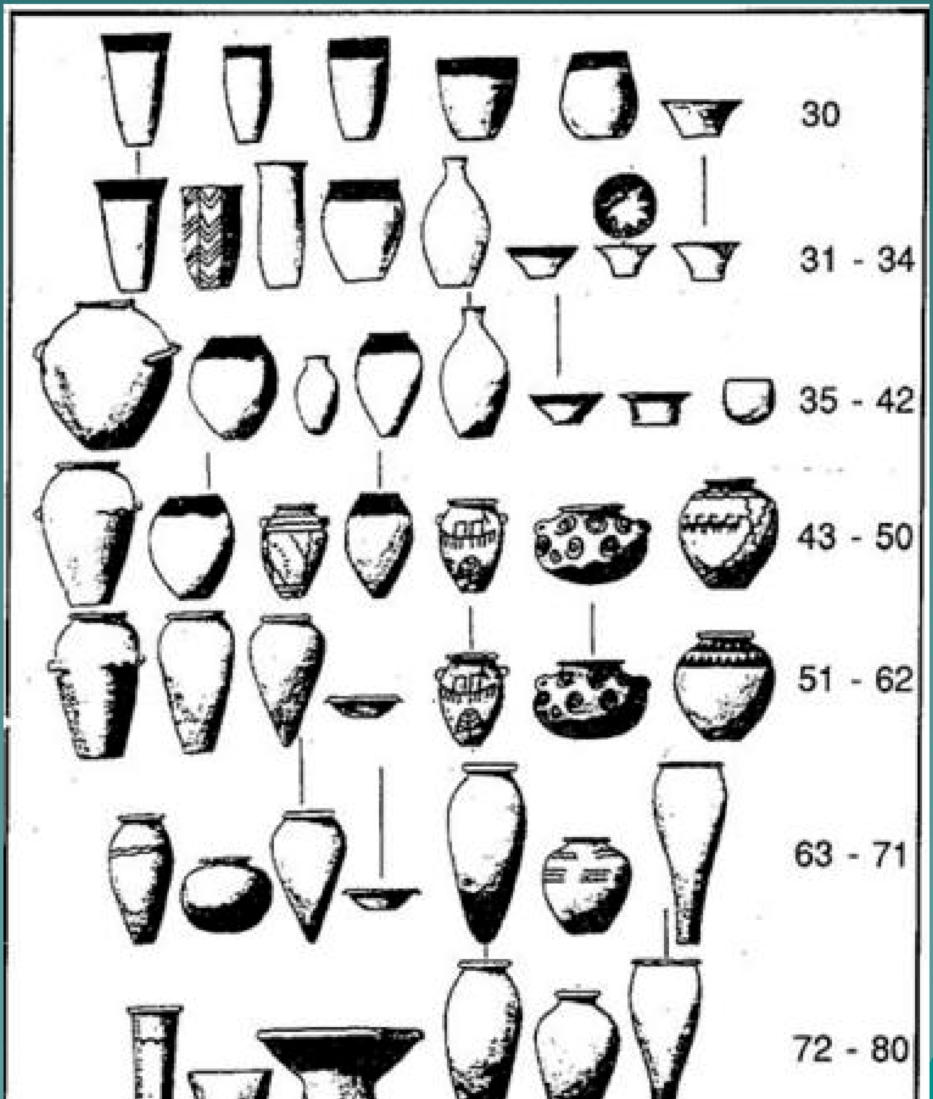


Computer Wiring

Printed Circuit Board 2392 cities 1987 Padberg and Rinaldi



Archaeological Seriation



Genome Sequences

- Made up of 4 Nucleotides
- A,C,G,T
- Sequence ‘Fragments’ of DNA

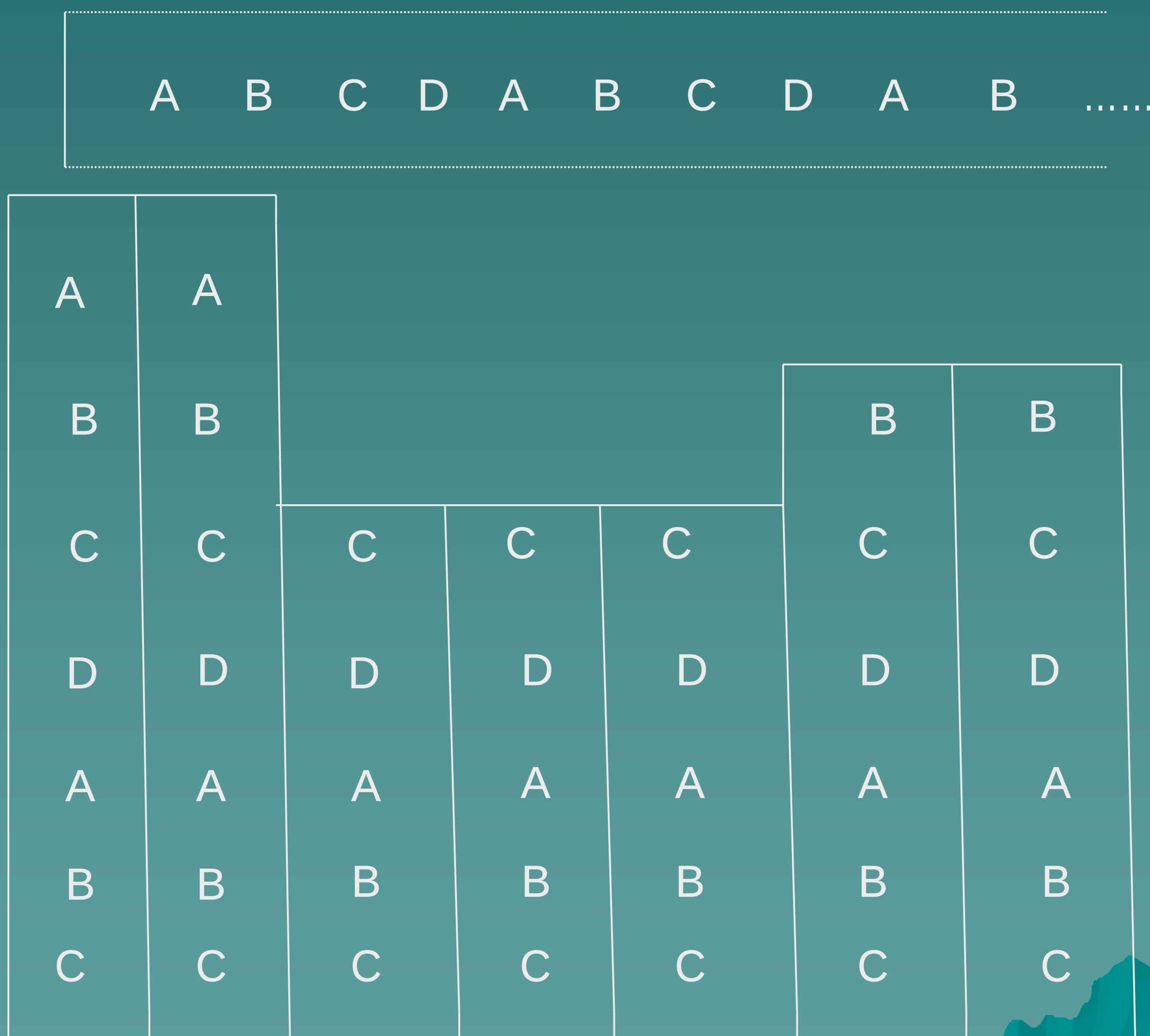
- Eg 1:CTGTAATTCGTCCAGCGA
- 2:TCCAGCGAGACGGAAGC
- 3:CGGAAGCGAGGTCCAG
- 4:CTAGCTTAGCTG

- Best Sequence
- 4---1---2---3
- Overlap 3 8 7

Job Scheduling

– ‘Distance’’ X to Y is time to reset machine after job X before performing job Y

Wallpapering to minimise waste



Solution Methods

I. **Try every possibility** (n-1)! possibilities – grows faster than exponentially

If it took 1 microsecond to calculate each possibility takes 10^{140} centuries to calculate all possibilities when n = 100

II. **Optimising Methods** obtain **guaranteed** optimal solution, but can take a very, very, long time

III. **Heuristic Methods** obtain ‘good’ solutions ‘quickly’ by intuitive methods.

No guarantee of optimality

(Place problem in newspaper with cash prize)

STANDARD FORMULATION OF THE (ASYMMETRIC) TRAVELLING SALESMAN PROBLEM

Conventional Formulation:

(cities *1,2, …, n*) (Dantzig, Fulkerson, Johnson) (1954). x_{ij} is a link in tour

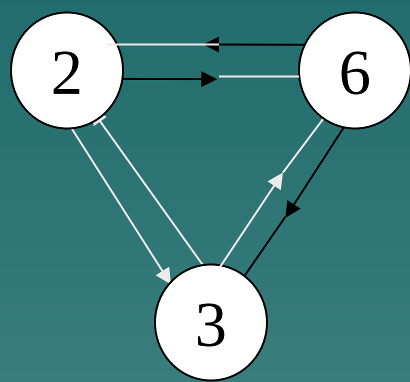
Minimise: $$\sum_{i,j} c_{ij} x_{ij}$$

subject to: $$\sum_{i} x_{ij} = 1 \quad \text{all } j$$

$$\sum_{j} x_{ij} = 1 \quad \text{all } i$$

$$\sum_{i,j \in S} x_{ij} \leq |S| - 1 \text{ all } S \subset \{2 \ldots, n\}$$

e.g.



$$x_{32} + x_{26} + x_{63}$$

$$+ x_{23} + x_{62} + x_{36} \leq 2$$

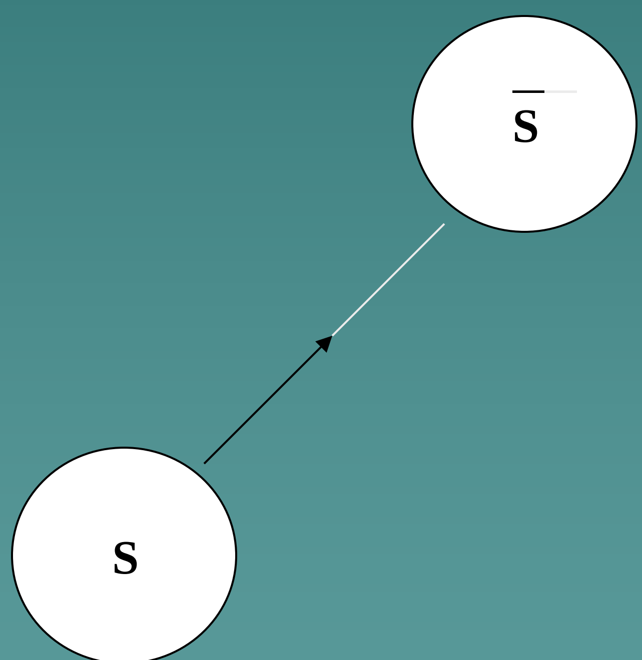
$0(2^n)$ *Constraints* $= (2^{n-1} + n - 2)$

$0(n^2)$ *Variables* $= n(n - 1)$

EQUIVALENT FORMULATION

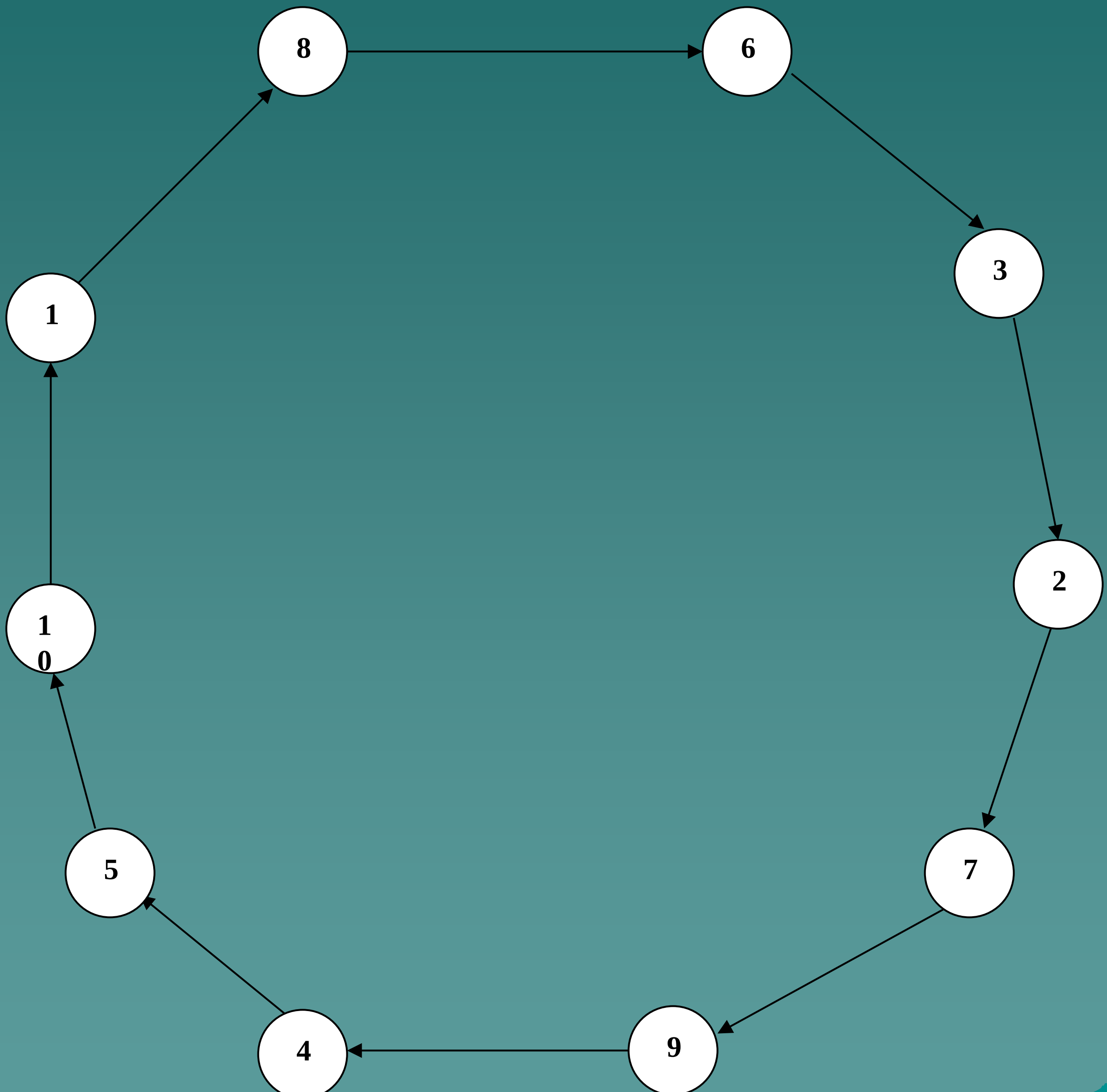
Replace subtour elimination constraints with

$$\sum_{\substack{i \in S \\ j \in \overline{S}}} x_{ij} \geq 1 \quad \textbf{all} \quad S \subset \{1, 2, \ldots, n\}$$



Add second set of constraints for all *i* in S and subtract from subtour elimination constraints for S

OPTIMAL SOLUTON TO A 10 CITY TRAVELLING SALESMAN PROBLEM

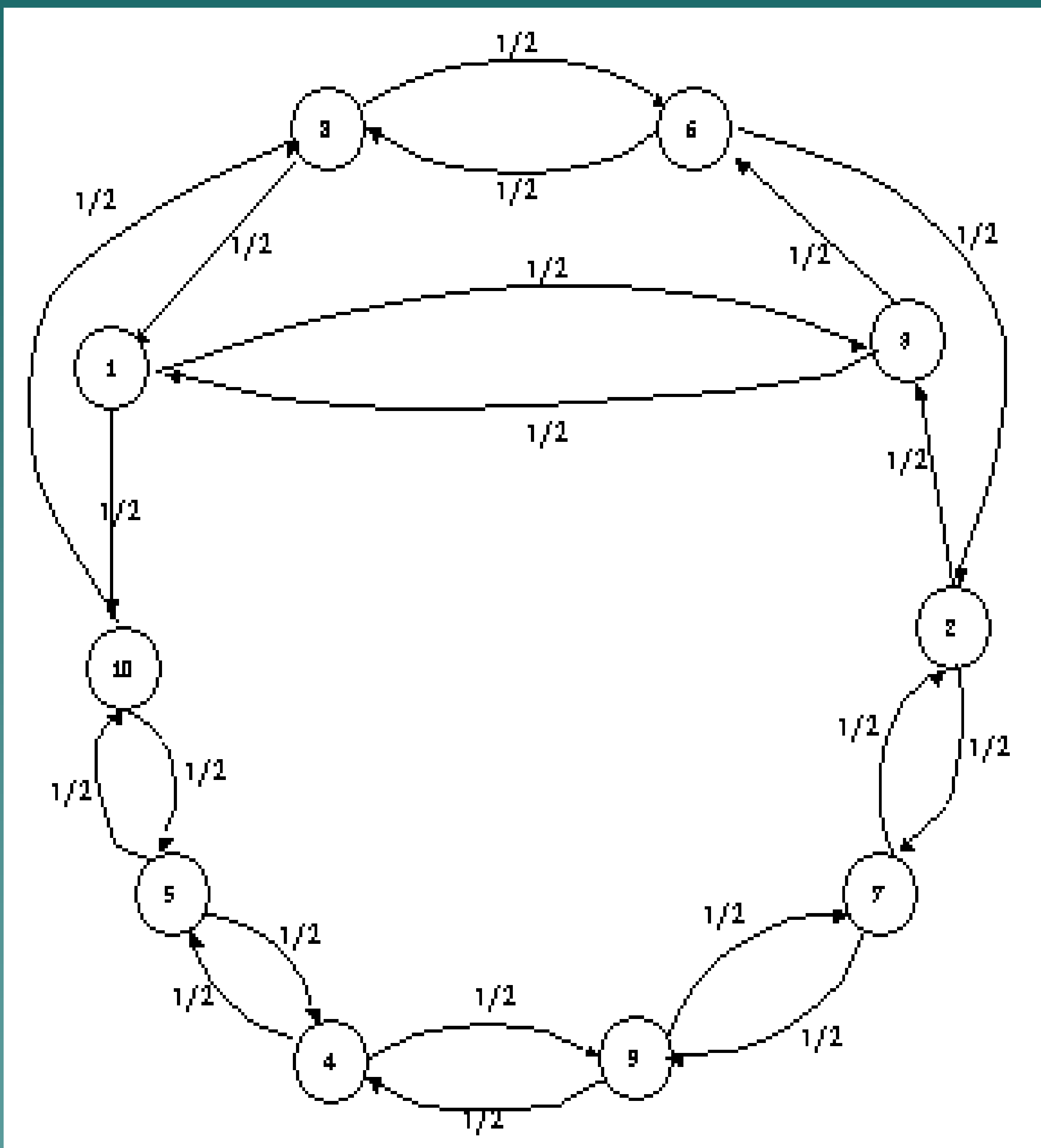


Cost = 881

Usual Method of Solution

- Solve Linear Programming Relaxation of Integer Programming model (leaving out certain constraints and integrality conditions)
- Append violated constraints (facets of convex hull)
- Repeat

FRACTIONAL SOLUTION FROM CONVENTIONAL (EXPONENTIAL) FORMULATION)



Cost = 878 (Optimal Cost = 881)

- This is only one possible Integer Programming Formulation
- For General IP models often alternative ‘better’ formulations (optimal objective value of LP relaxation closer to IP optimum)
- Is this the case for the TSP?

Sequential Formulation (Miller, Tucker, Zemlin (1960))

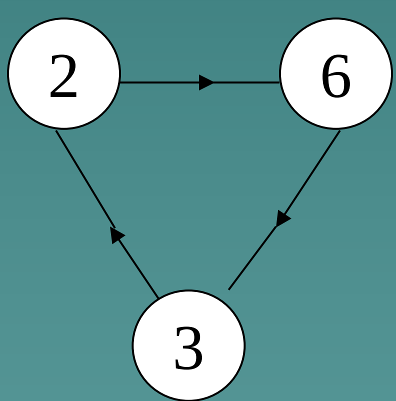
u_i = Sequence Number in which city i visited

Defined for $i = 2,3, \ldots, n$

Subtour elimination constraints replaced by

S: $u_i - u_j + n x_{ij} \leq n - 1 \quad i,j = 2,3, \ldots, n$

Avoids subtours
but allows total tours (containing city 1)



$$u_2 - u_6 + n x_{26} \leq n-1$$

$$u_6 - u_3 + n x_{63} \leq n-1$$

$$u_3 - u_2 + n x_{32} \leq n-1$$

$$\downarrow$$

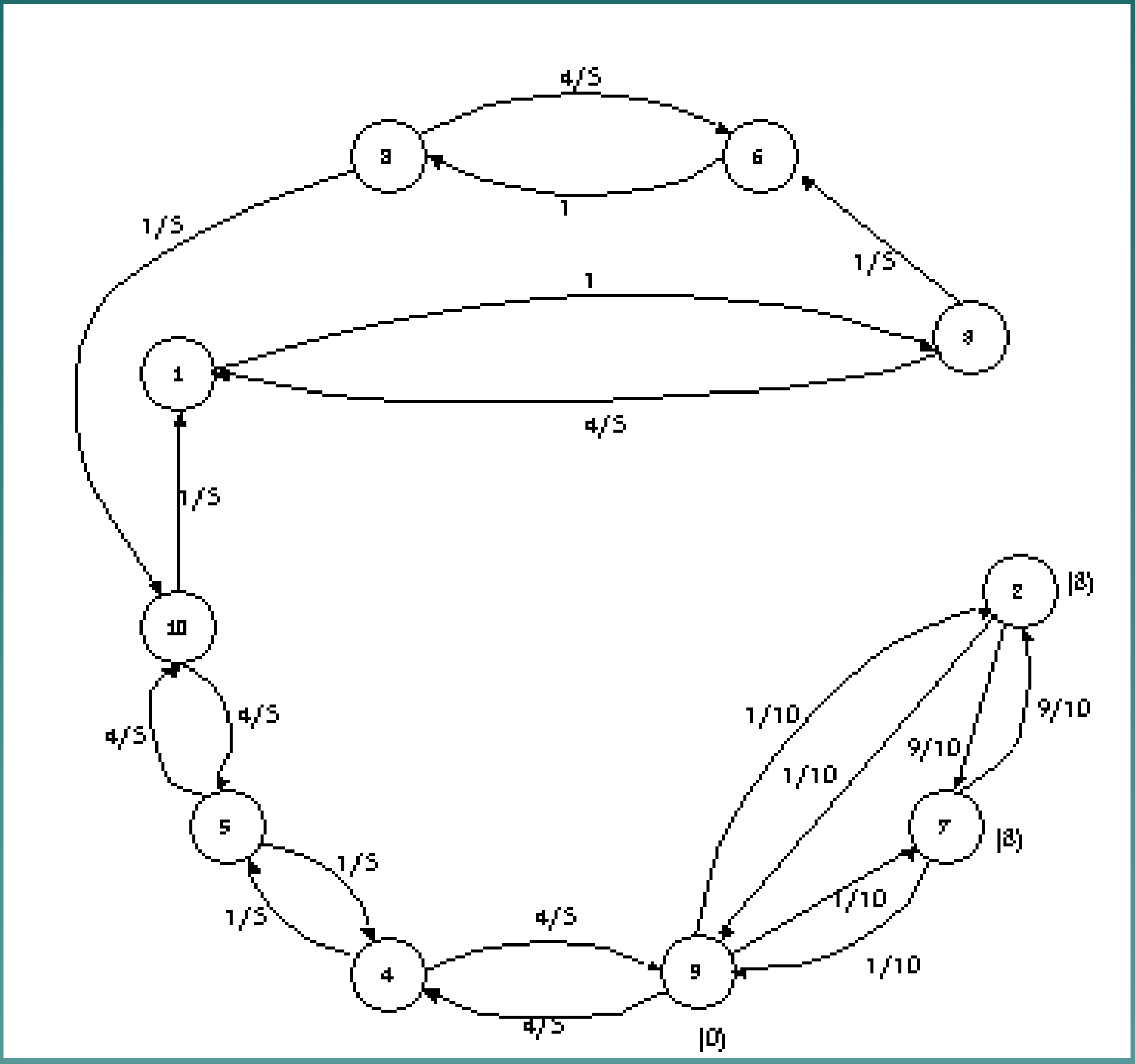
$$3n \leq 3(n-1)$$

$0(n^2)$	*Constraints*	=	$(n^2 - n + 2)$
$0(n^2)$	*Variables*	=	$(n-1)(n+1)$

Weak but can add 'Logic Cuts'

e.g. $u_k \geq 1 + x_{ij} + x_{jk} - x_{1j}$

FRACTIONAL SOLUTION FROM SEQUENTIAL FORMULATION



Subtour Constraints Violated : e.g. $\boldsymbol{x}_{27} + \boldsymbol{x}_{72} \not\leq 1$

Logic Cuts Violated: e.g. $u_9 \not\geq 1 + x_{27} + x_{79} - x_{17}$

Cost = $773\ ^{3}/_{5}$ (Optimal Cost = 881)

Flow Formulations

Single Commodity (Gavish & Graves (1978))

Introduce extra variables ('Flow' in an arc)

Replace subtour elimination constraints by

F1:

$$y_{ij} \le (n-1)x_{ij} \quad \text{all } i, j$$

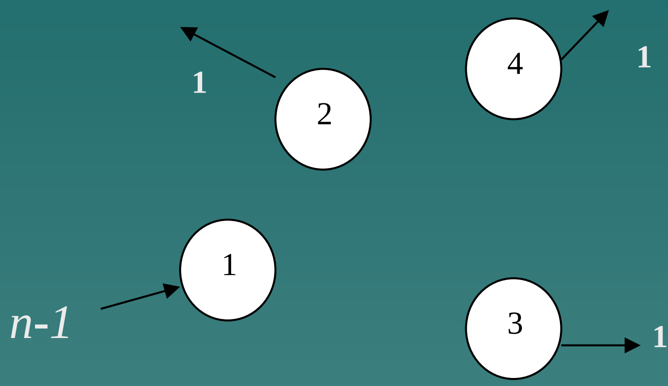
$$\sum_j y_{1j} = n-1$$

$$\sum_i y_{ij} - \sum_k y_{jk} = 1 \text{ all } j \ne 1$$

Can improve (F1') by amended constraints:

$$\boldsymbol{y}_{ij} \le (\boldsymbol{n}-2)\boldsymbol{x}_{ij} \qquad \text{all } i, j \ne 1$$

Network Flow formulation in y_{ij} variables over complete graph

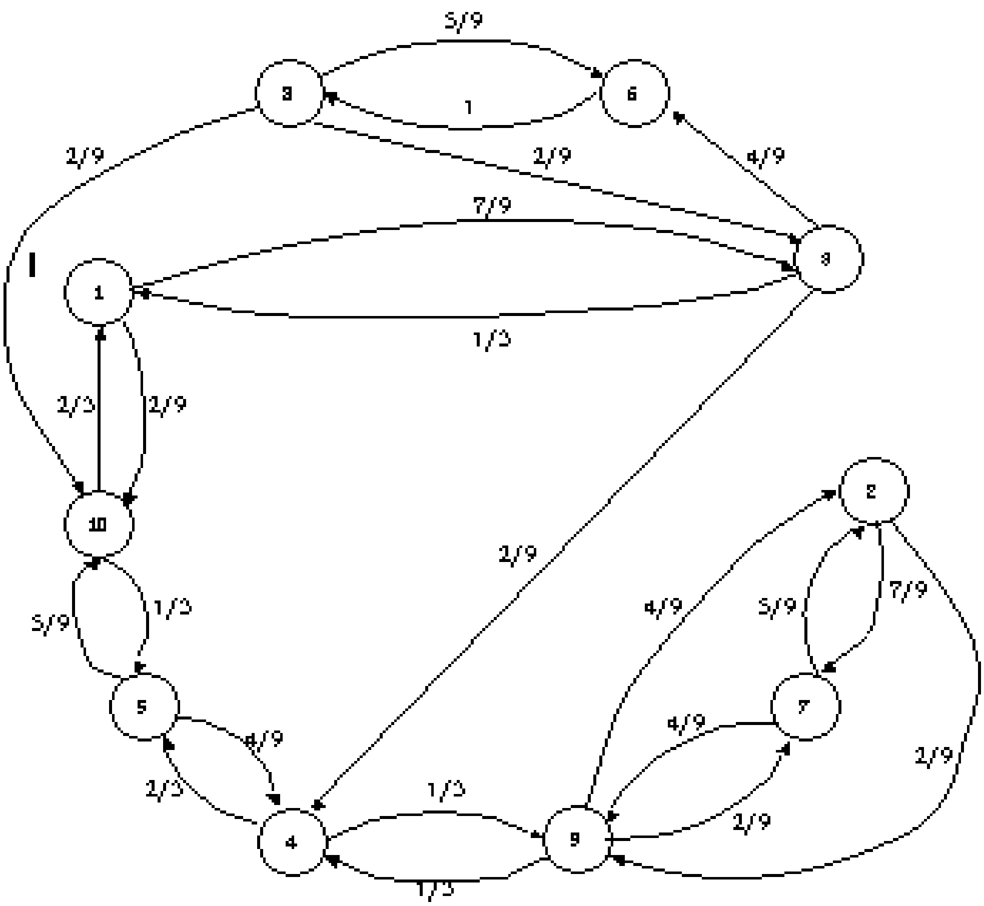


Graph must be connected. Hence no subtours possible.

$0(n^2)$ Constraints $= n(n+2)$

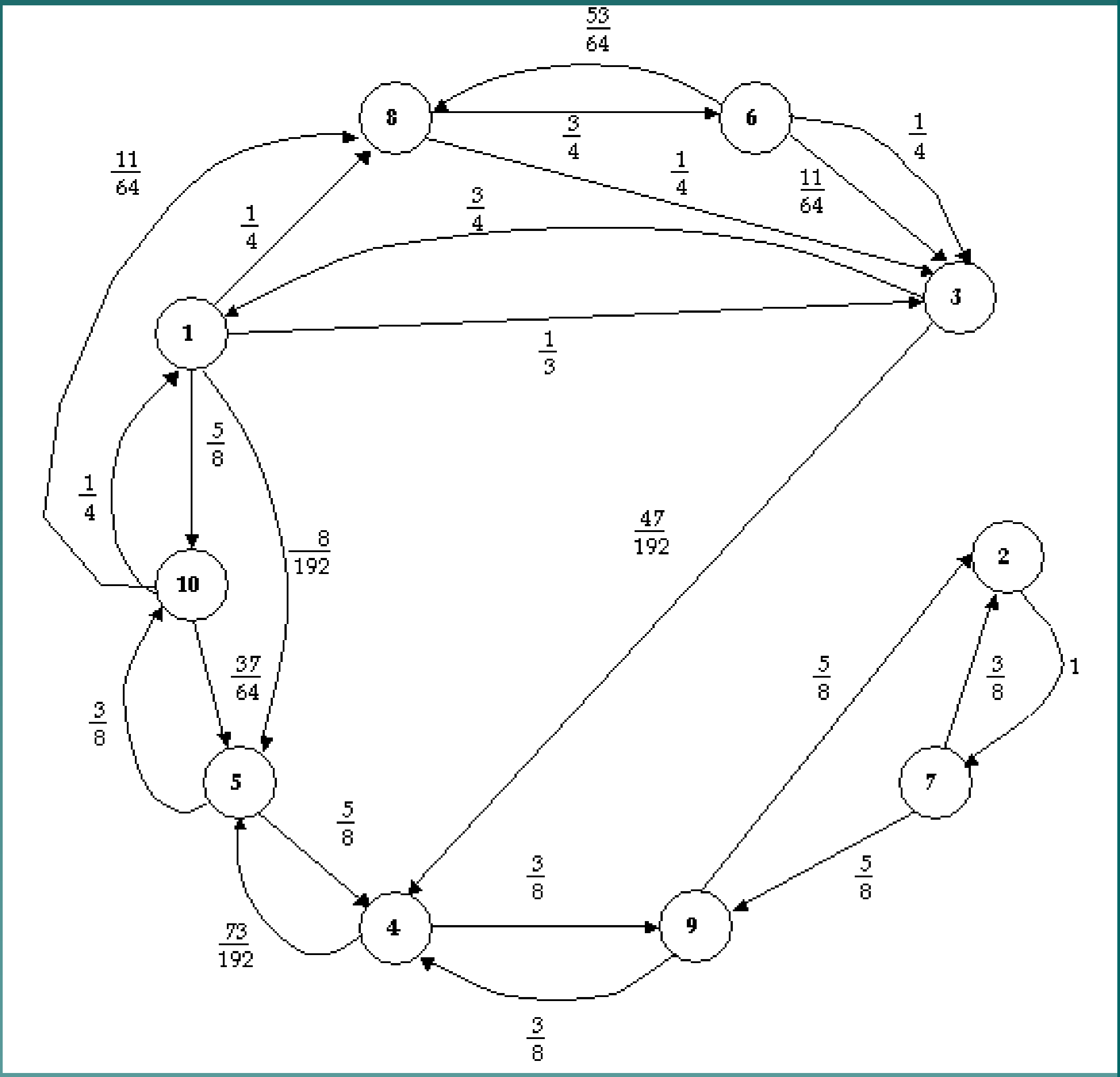
$0(n^2)$ Variables $= 2n(n-1)$

FRACTIONAL SOLUTION FROM SINGLE COMMODITY FLOW FORMULATION



Cost $= 794\frac{2}{9}$ (Optimal solution = 881)

FRACTIONAL SOLUTION FROM MODIFIED SINGLE COMMODITY FLOW FORMULATION



Cost = $794\frac{43}{48}$ (Optimal solution = 881) (192=3x64)

Two Commodity Flow (Finke, Claus Gunn (1983))

y_{ij} is flow of commodity 1 in arc $i \rightarrow j$

z_{ij} is flow of commodity 2 in arc $i \rightarrow j$

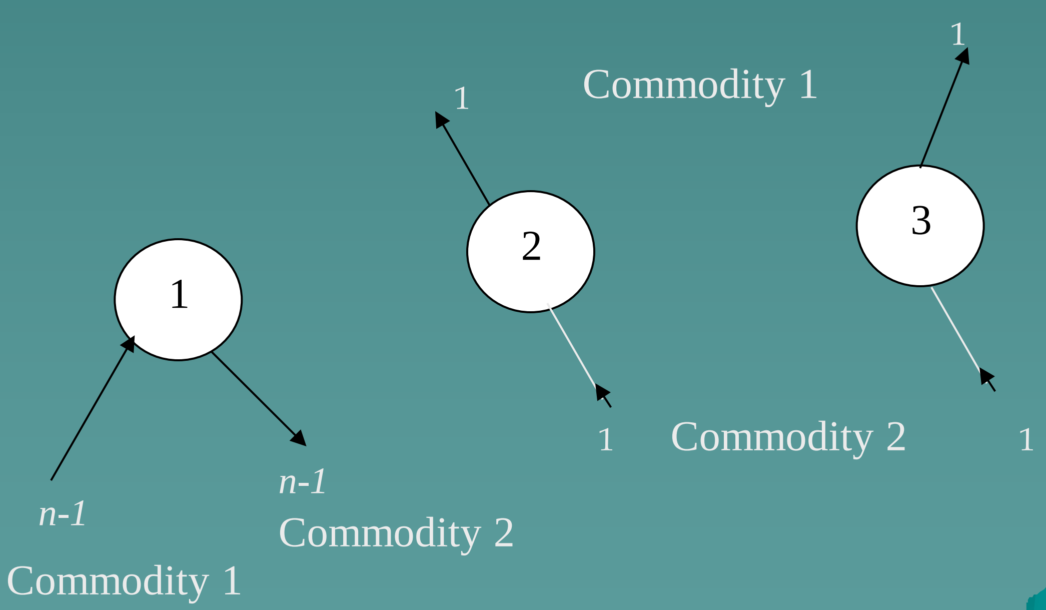
$$\sum_j y_{ij} - \sum_j y_{ji} \begin{array}{ll} = \ \ -1 & i \neq 1 \\ = n-1 & i = 1 \end{array}$$

F2:

$$\sum_j z_{ij} - \sum_j z_{ji} \begin{array}{ll} = \ \ 1 & i \neq 1 \\ = -(n-1) & i = 1 \end{array}$$

$$\sum_j z_{ij} - \sum_j z_{ji} = n-1 \quad \text{all } i$$

$$y_{ij} + z_{ij} = (n-1)x_{ij} \quad \text{all } i, j$$



$0(n^2)$ Constraints $= n(n+4)$

$0(n^2)$ Variables $= 3n(n-1)$

Multi-Commodity (Wong (1980) Claus (1984))

"Dissaggregate" variables

y_{ij}^{k} is flow in arc destined for k

$y_{ij}^{k} \le x_{ij}$ all i, j, k

F3 $\sum_i y_{ik}^{k} = 1 \quad \sum_i y_{1i}^{k} = 1 \quad \sum_i y_{i1}^{k} = 0 \quad \sum_j y_{kj}^{k} = 0$ all k

$\sum_i y_{ij}^{k} = \sum_i y_{ji}^{k}$ all $j, k, j \ne 1, j \ne k$.

$0(n^3)$ Constraints $= n^3 - 2n^2 + 6n - 3$

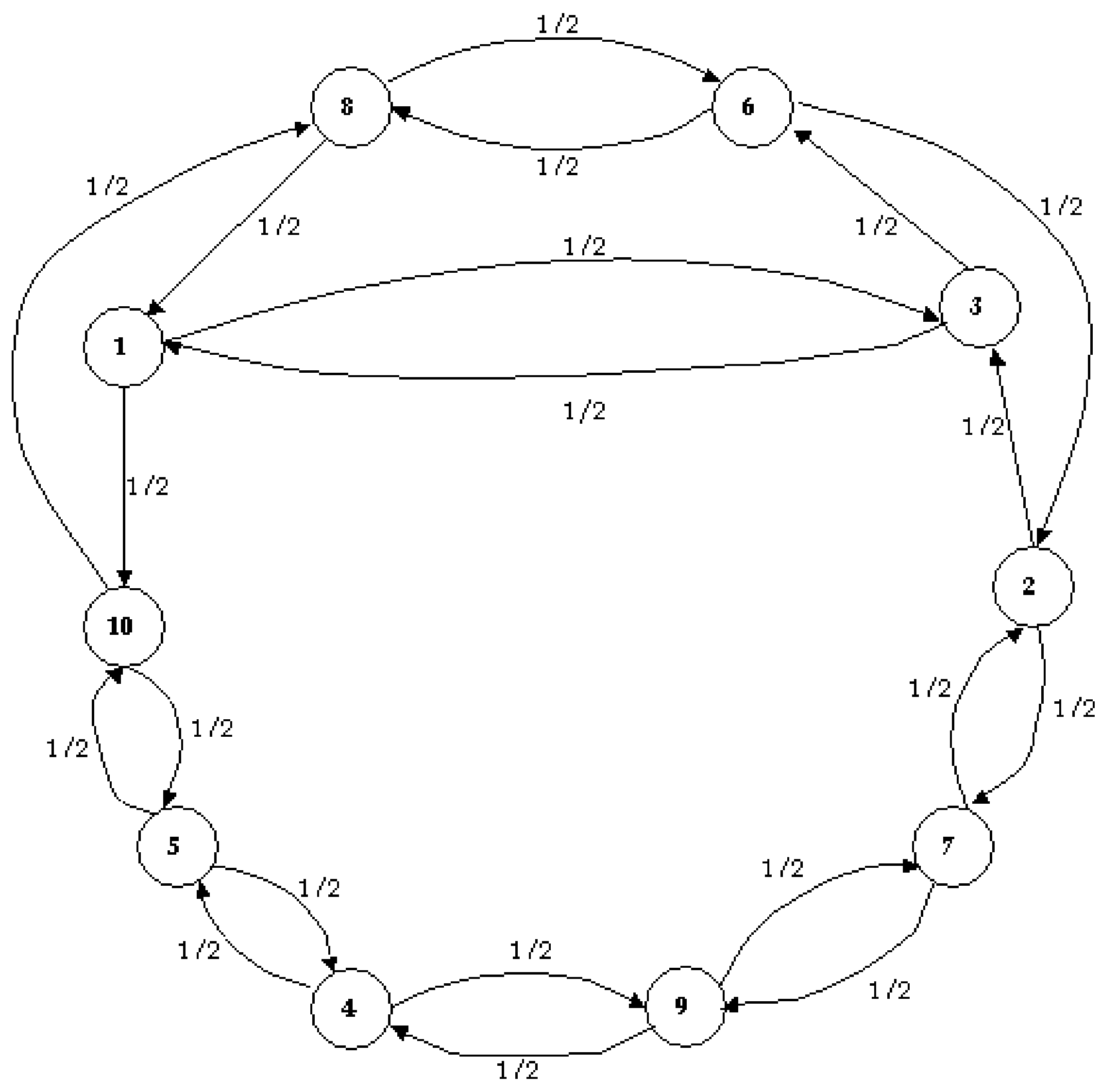
$0(n^3)$ Variables $= n^2(n-1)$

LP Relaxation of equal strength to Conventional Formulation.

But of polynomial size.

Tight Formulation of Min Cost Spanning Tree
+ (Tight) Assignment Problem

FRACTIONAL SOLUTION FROM MULTI COMMODITY FLOW FORMULATION (= FRACTIONAL SOLUTION FROM CONVENTIONAL (EXPONENTIAL) FORMULATION)



Cost = 878 (Optimal Cost = 881)

Stage Dependent Formulations

First (Fox, Gavish, Graves (1980))

= 1 if arc $\boldsymbol{i} \rightarrow \boldsymbol{j}$ traversed at stage t

= 0 otherwise

T1:

$$\sum_{i,j,t} y_{ij}^{t} = n$$

$$\sum_{j=1}^{n} \sum_{t=2}^{n} t y_{ij}^{t} - \sum_{j=1}^{n} \sum_{t=1}^{n} t y_{ji}^{t} = 1 \quad i = 2, 3 \ldots \mathrm{n}$$

(Stage at which $\boldsymbol{i}$ left 1 more than stage at which entered)

$$\boldsymbol{y}_{i1}^{t} = 0, \ \boldsymbol{t} \neq \boldsymbol{n}$$
$$\boldsymbol{y}_{1j}^{t} = 0, \ \boldsymbol{t} \neq 1$$
$$\boldsymbol{y}_{ij}^{1} = 0, \ \boldsymbol{i} \neq 1$$

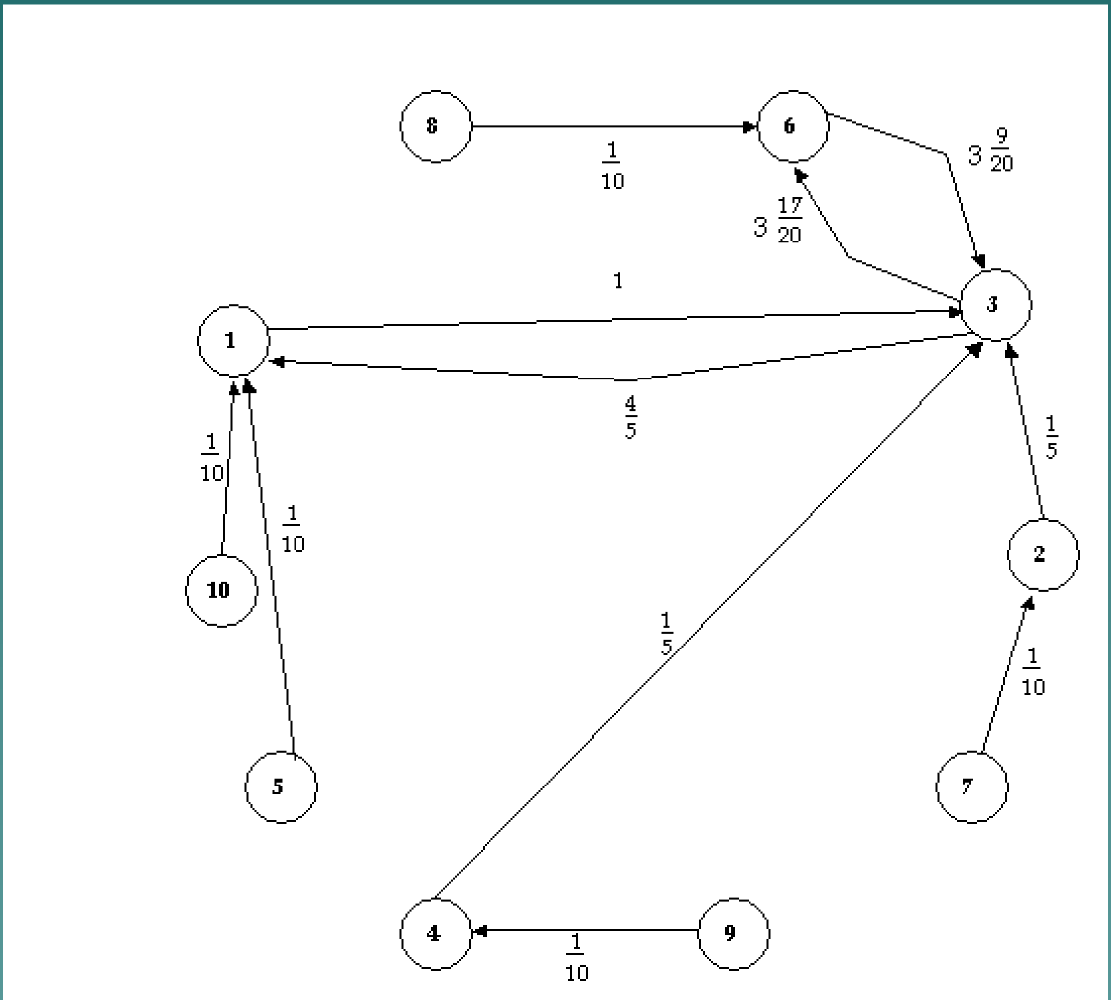
$0(\boldsymbol{n})$ Constraints $= \boldsymbol{n}$

$0(\boldsymbol{n}^{3})$ Variables $= \boldsymbol{n}^{2}(\boldsymbol{n}-1)$

Also convenient to introduce $\boldsymbol{x}_{ij}$ variables with constraints

$$\boldsymbol{x}_{ij} = \sum_{t} \boldsymbol{y}_{ij}^{t}$$

FRACTIONAL SOLUTION FROM 1ST (AGGREGATED) TIME-STAGED FORMULATION



Cost = 364.5 (Optimal solution = 881)

NB ‘Lengths’ of Arcs can be > 1

Second (Fox, Gavish, Graves (1980))

T2: Disaggregate to give

$$\sum_{i \neq j, t} y_{ij}^{t} = 1 \qquad \text{all j}$$

$$\sum_{j \neq i, t} y_{ij}^{t} = 1 \qquad \text{all i}$$

$$\sum_{i, j, i \neq j} y_{ij}^{t} = 1 \qquad \text{all t}$$

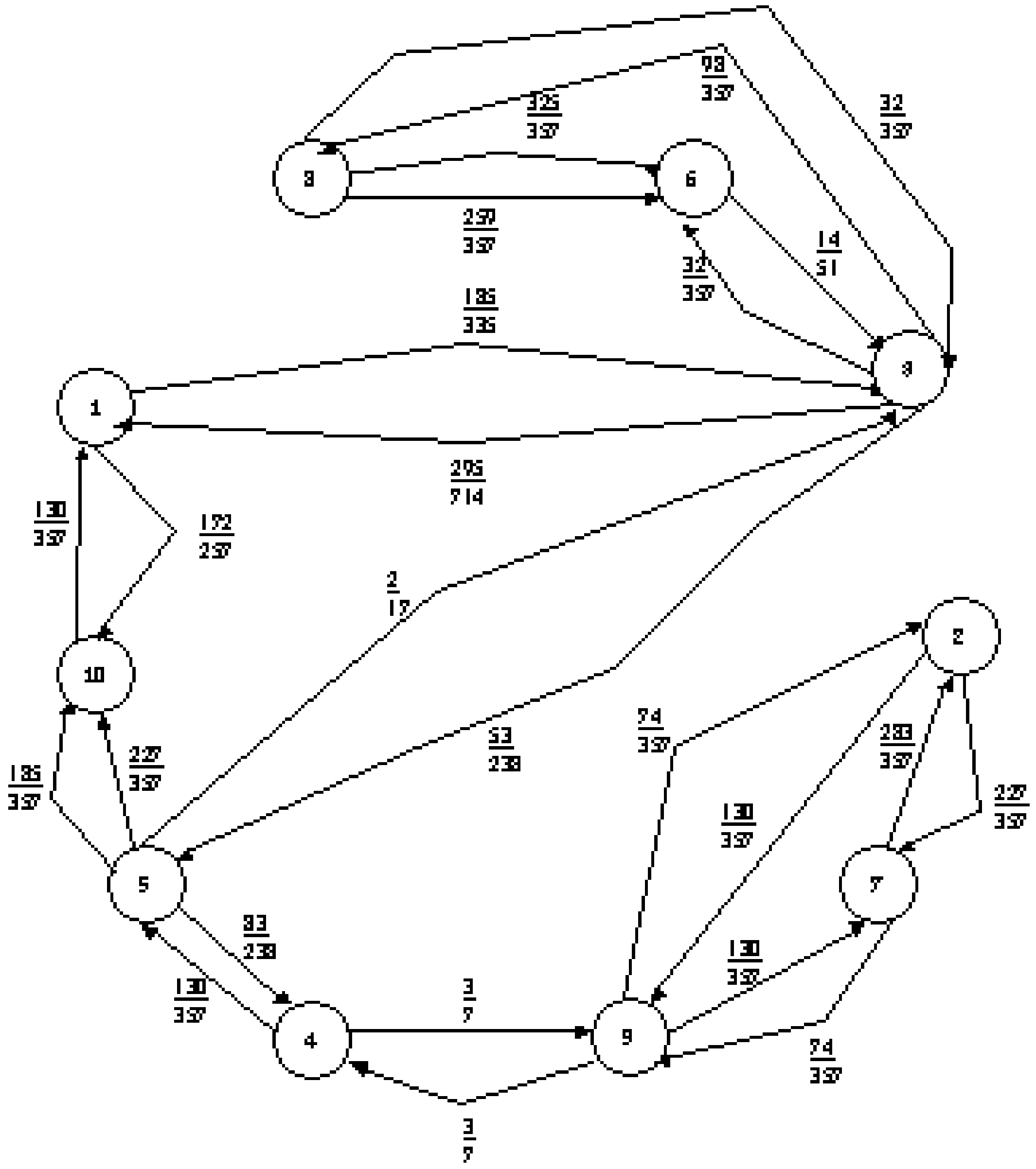
$$\sum_{j=1}^{n} \sum_{t=2}^{n} t y_{ij}^{t} - \sum_{j=1}^{n} \sum_{t=1}^{n} t y_{ji}^{t} = 1 \quad i = 2,3,\ldots n$$

Initial conditions no longer necessary

0(n) Constraints = 4 n – 1

0(n^3) Variables = n^2 (n – 1)

FRACTIONAL SOLUTION FROM 2nd TIME-STAGED FORMULATION



Cost = $799\frac{164}{357}$ (optimal solution = 881)

(714 = 2 x 3 x 7 x 17)

Third (Vajda/Hadley (1960))

T3: y_{ij}^{t} interpreted as before

$$\sum_{i \neq j} y_{ij}^{t} = 1 \quad \text{all } j$$

$$\sum_{j \neq i} y_{ij}^{t} = 1 \quad \text{all } i$$

$$\sum_{t} y_{ij}^{t} = 1 \quad \text{all } t$$

$$\sum_{i \neq j} y_{ij}^{t} - \sum_{k \neq j} y_{jk}^{t+1} = 0 \quad \text{all } j, t$$

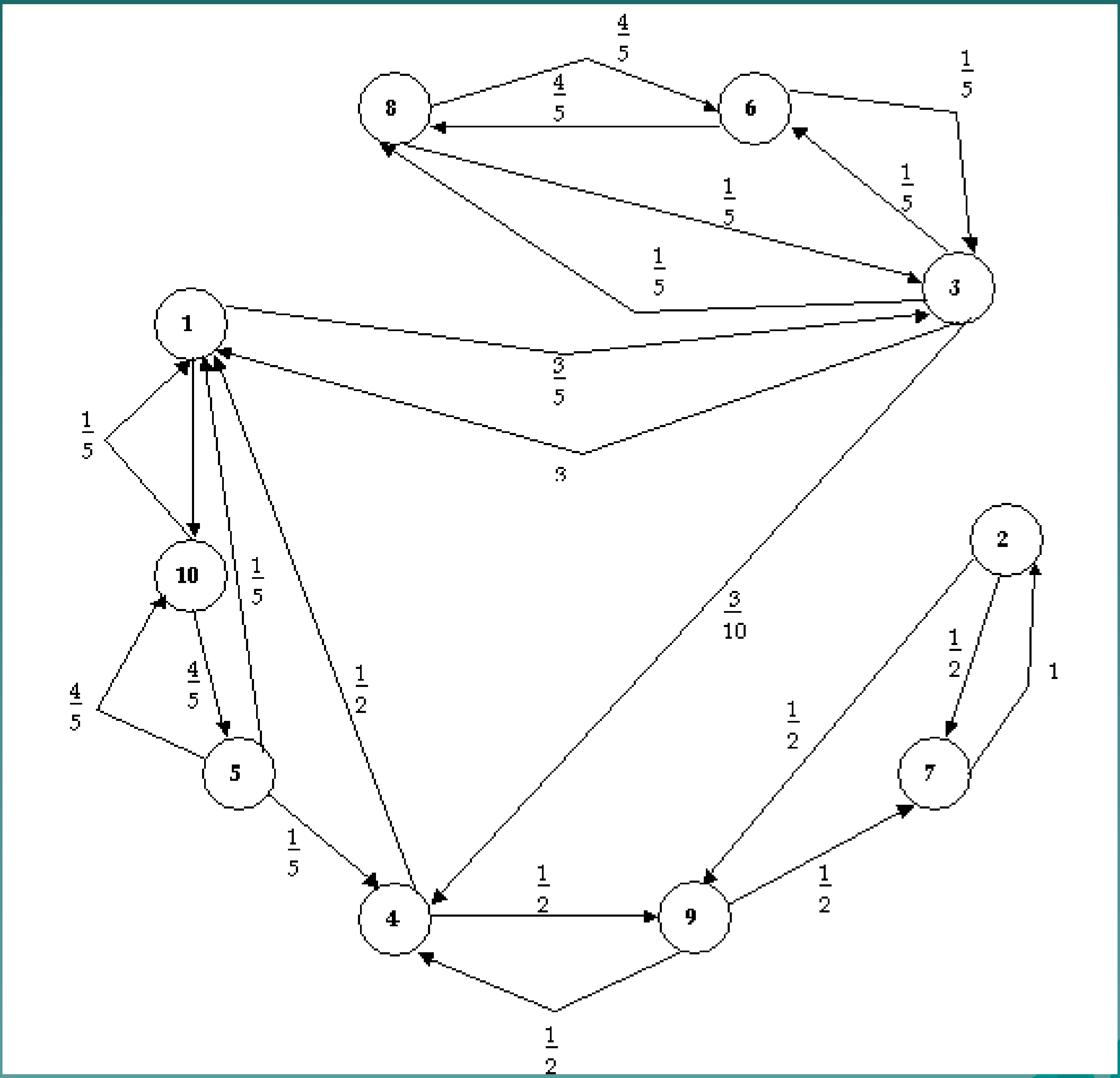
$$\sum_{j \neq 1} y_{1j}^{1} = 1$$

$$\sum_{i \neq 1} y_{i1}^{n} = 1$$

0 (n^2) Constraints = ($2n^2$+ 3)

0 (n^3) Variables = n^2(n-1)

FRACTIONAL SOLUTION FROM 2nd TIME 2nd TIME-STAGED FORMULATION



Cost = $804\frac{1}{2}$ Optimal solution = 881

OBSERVATION

Multicommodity Flow Formulation

$$\sum_i \boldsymbol{y}_{ij}^{t} - \sum_k \boldsymbol{y}_{jk}^{t} = 0$$

$\boldsymbol{y}_{ij}^{t}$ is flow $\boldsymbol{i} \to \boldsymbol{j}$ destined for node $\boldsymbol{t}$

Time Staged Formulation

$$\sum_i \boldsymbol{y}_{ij}^{t} - \sum_k \boldsymbol{y}_{jk}^{t+1} = 0$$

$\boldsymbol{y}_{ij}^{t} = 1$ ***iff*** go $\boldsymbol{i} \to \boldsymbol{j}$ at stage $\boldsymbol{t}$

Are these formulations related?

Can extra variables $(\boldsymbol{y}_{ij}^{t})$, introduced *syntactically*, be given different *semantic* interpretations?

COMPARING FORMULATIONS

Minimise: $c\,x$

Subject to: $Ax + By \leq b$

$\underline{x}, \underline{y} \geq 0$, x integer

$$W = \{w \mid wB \geq 0, \underline{w} \geq 0\}$$

W forms a cone which can be characterised by its extreme rays giving matrix Q such that

$$QB \geq 0$$

Hence $QAx \leq Qb$

This is the projection of formulation into space of original variables x_{ij}

COMPARING FORMULATIONS

Project out variables by Fourier-Motzkin elimination to reduce to space of conventional formulation.

P (r) is polytope of LP relaxation of projection of formulation r.

Formulation S (Sequential)

Project out around each *directed cycle* S by summing

$$u_i - u_j + n x_{ij} \leq n - 1$$

$$\downarrow$$

$$n \sum_{i,j \in S} x_{ij} \leq (n-1)|S|$$

ie $\sum_{i,j \in S} x_{ij} \leq |S| - \frac{|S|}{n}$ weaker than |S|-1 (for S a subset of nodes)

Hence $P(S) \supset P(C)$

Formulation F1 (1 Commodity Network Flow)

Projects to $\sum_{ij \in s} x_{ij} \leq |S| - \frac{|S|}{n-1}$ stronger than $|S| - \frac{|S|}{n}$

Hence $P(S) \supset P(F1) \supset P(C)$

Formulation F1' (Amended 1 Commodity Network Flow)

Projects to $$\frac{1}{n-1} \sum_{\substack{j \in \bar{S} - \{1\} \\ j \in S}} x_{ij} + \sum_{i,j \in S} x_{ij} \leq |S| - \frac{|S|}{n-1}$$

Hence $P(S) \supset P(F1) \supset F(F1') \supset P(C)$

Formulation F2 (2 Commodity Network Flow)

Projects to $\sum_{i,j} x_{ij} \leq |S| - \frac{|S|}{n-1}$

Hence $P(F2) = P(F1)$

Formulation F3 (Multi Commodity Network Flow)

Projects to $$\sum_{\substack{i,j \\ \in S}} x_{ij} \leq |S| - 1$$

Hence $P(F3) = P(C)$

Formulation T1 (First Stage Dependant)

Projects to

$$\sum_{\substack{i \in S \\ j \in \bar{S} - \{1\}}} x_{ij} \geq \frac{|S|}{n-1}$$

$$\sum_{i,j \in N} x_{ij} = n$$

(Cannot convert 1st constraint to ‘ $\sum_{i,j \in 5} x_{ij} \leq$ ‘form since Assignment Constraints not present)

Formulation T2 (Second Stage Dependant)

Projects to

$$\frac{1}{n-1}\sum_{\substack{i\in S\\ j\in \bar{S}-\{1\}}} x_{ij} + \frac{1}{n-1}\sum_{\substack{i\in \bar{S}-\{1\}\\ j\in S}} x_{ij} + \sum_{ijj\in s} x_{ij} \leq |S| - \frac{|S|}{n-1}$$

+ others

Hence $P(T2) \subset P(F1')$

Formulation T3 (Third Stage Dependant)

Projects to

$$\frac{1}{n-1}\sum_{\substack{i\in S\\ j\in \bar{S}-\{1\}}} x_{ij} + \frac{1}{n-1}\sum_{\substack{i\in \bar{S}-\{1\}\\ j\in S}} x_{ij} + \sum_{n,j\in S} x_{ij} \leq |S| - \frac{|S|}{n-1}$$

+ others

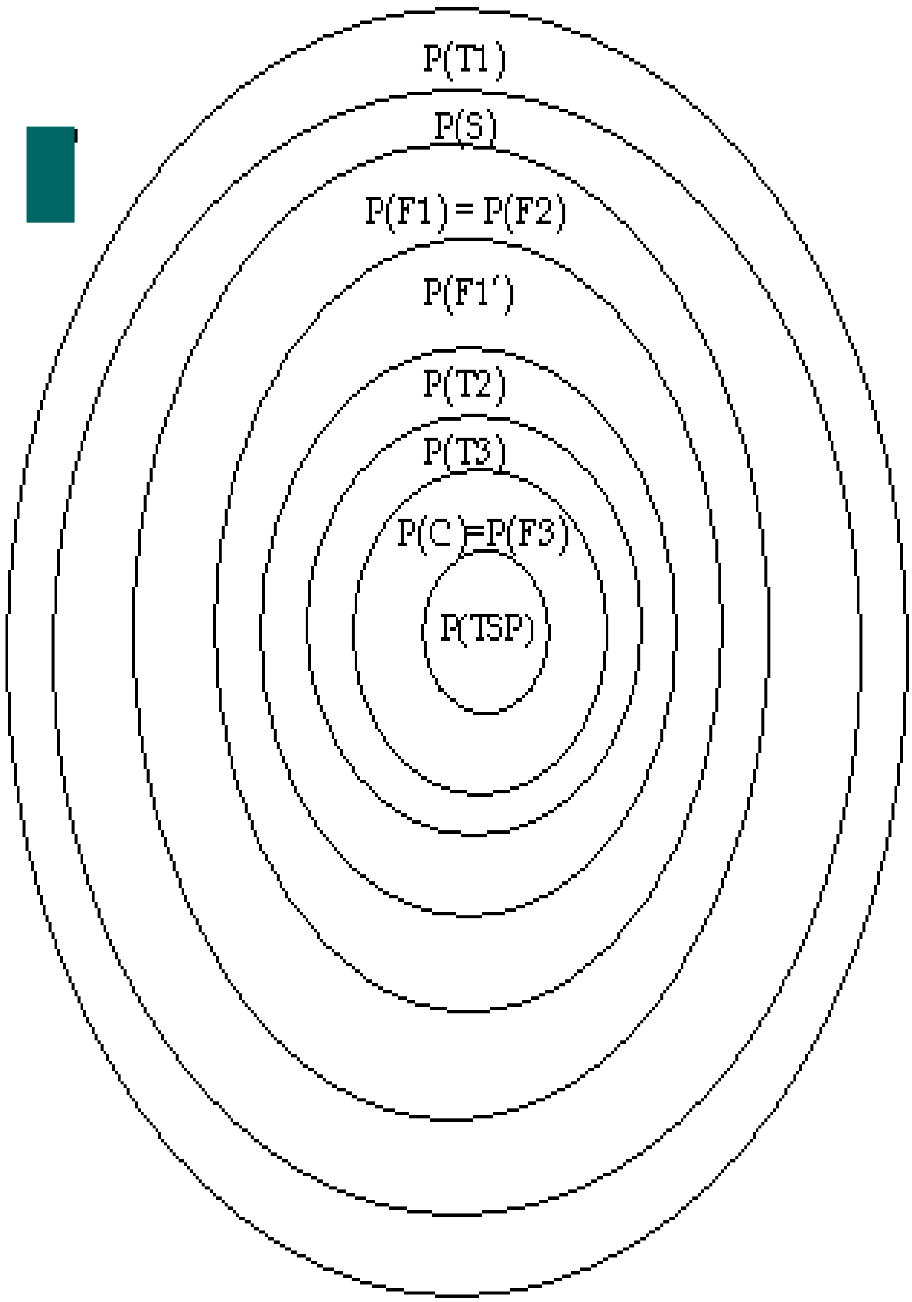
Can show stronger than T2

Hence $P(T3) \subset P(T2)$

Computational Results of a 10-City TSP in order to compare sizes and strengths of LP Relaxations

Model	Size	LP Obj	Its	Secs	IP Obj	Nodes	Secs
C (Conventional	502x90						
	(Ass. Relax	766	37	1	766	0	1
	+Subtours (5)	804	40	1	804	0	1
	+Subtours (3)	835	43	1	835	0	1
	+Subtours (2)	878	48	1	881	9	1
S (Sequential)	92x99	773.6	77	3	881	665	16
F1 (Commodity Flow	120x180	794.22	148	1	881	449	13
F' (F1 Modified)	120x180	794.89	142	1	881	369	11
F2 (2 Commodity Flow)	140x270	794.22	229	2	881	373	12
F3 (Multi Commodity Flow)	857x900	878	1024	7	881	9	13
T1 (1st Stage Dependent)	90x990 (10)x(900)	364.5	63	4	881	∞	∞
T2 (2nd Stage Dependent)	120x990 (39) x (900)	799.46	246	18	881	483	36
T3 (3rd Stage Dependent)	193x990 (102)x(900)	804.5	307	5	881	145	27

Solutions obtained using NEW MAGIC and EMSOL



P(TSP) TSP Polytope – not fully known

P(X) Polytope of Projected LP relaxations

Reference

- AJ Orman and HP Williams,

  A Survey of Different Formulations of the Travelling Salesman Problem,

  in C Gatu and E Kontoghiorghes (Eds),
  Advances in Computational Management Science 9 Optimisation, Econometric and Financial Analysis (2006) Springer