Scheduling Models and Algorithms for the Orderly Colored Longest Path

Giovanni Felici
Istituto di Analisi dei Sistemi ed Informatica
Rome, Italy

Gaurav Singh
BHP Billiton
Perth, Australia

Marta Szachniuk, **Jacek Blazewitz**
Poznań University of Technology
Poznan, Poland

Tributes

- Discrete Applied Maths, 2015
- RAIRO, 2014
- D. de Werra
- M.C. de Cola

Outline

- Introduction to OCLP
- OCLP and NMR
- OCLP in scheduling
- Complexity
- Flow based models
- Scheduling-like models
- Experimental results
- Conclusions

Definitions

Arc Colored Graph

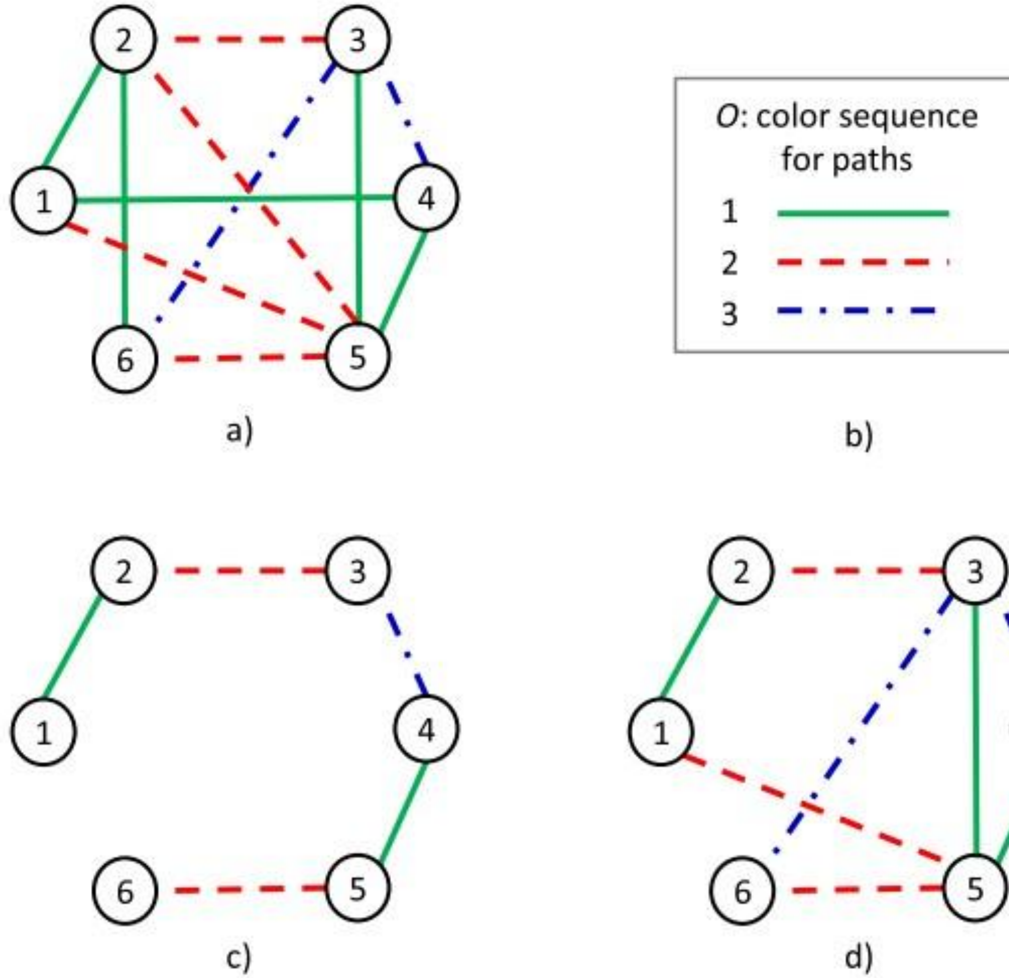
Alternating paths

Properly Colored Path

Orderly Colored Path

Shortest

Longest



Original Motivation: NMR assignment

Nuclear Magnetic Resonance (NMR) spectroscopy

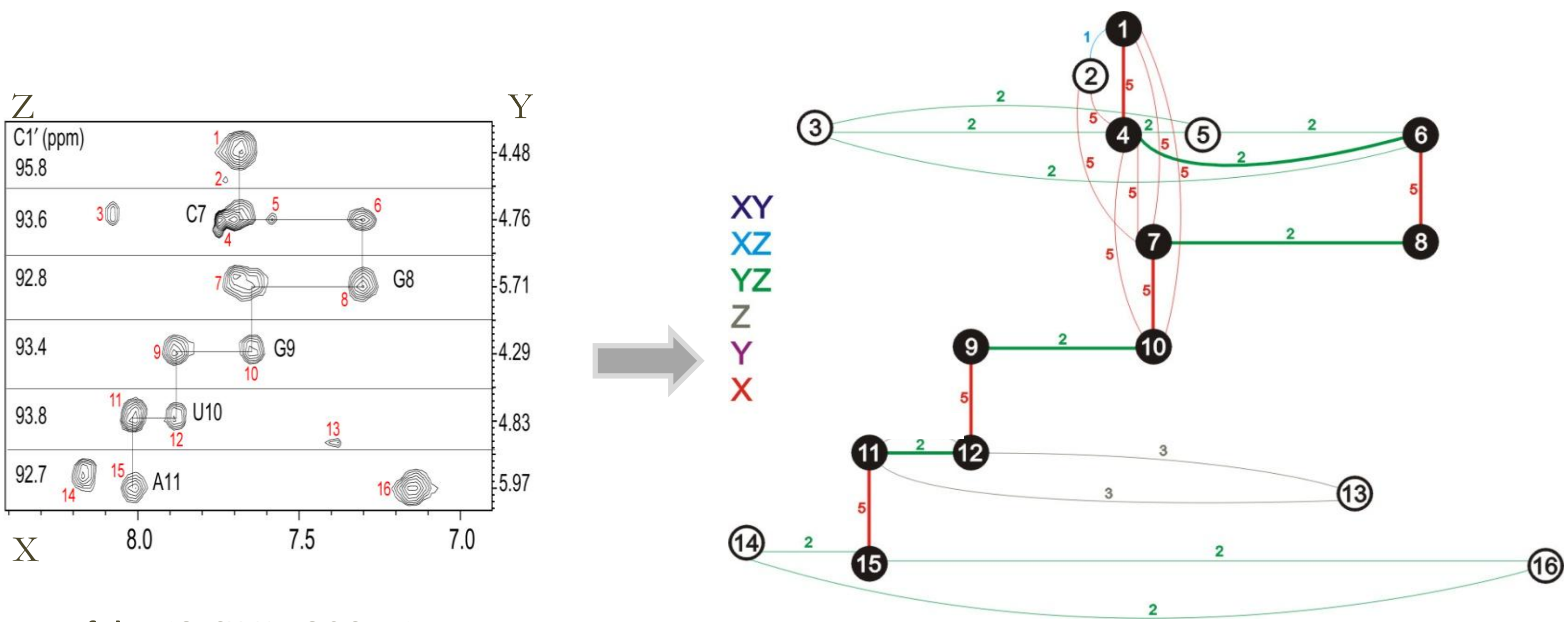
magnetic properties of certain atomic nuclei are exploited to determine physical and chemical properties of atoms or molecules in which they are contained.

- Applying a magnetic field, protons resonate
- a number of cross-peaks are generated by pairs of atoms whose protons resonate together if they are close in space.
- Starting from a trace of NMR spectrum we compute the structural parameters needed to determine the 3D structure of the molecule

M. Szachniuk, M.C. De Cola, G. Felici, J. Blazewicz, D. de Werra. **Optimal pathway reconstruction on 3D NMR maps, Discrete Applied Mathematics**, 182, (2015)

Original Motivation: NMR assignment

- Each arc is an atom involved in the crosspeaks
- The transfer of magnetization is fixed between atom types (order of colors)
- A path between the cross peaks correspond to the magnetization transfer among the atoms, important to determine the structure (folding) of the protein
- A longest orderly colored path represent the most likely magnetization transfer of the protein



Fragment of the NOESY-HMQC 3D NMR spectrum with a path (in YZ – X transition order) and the corresponding edge-colored graph

Scheduling Applications

OCLP may be suited for modeling also certain interesting scheduling problems, where:

- A number of locations (operations) must be visited (done)
- In each site, an item belonging to a class can be picked up (processed)
- An item of class i can be picked (processed) only if an item of class (i-1) has just been picked up

Waste collection

- A road network
- Colors represent time windows
- Each arc is a stretch of road where waste must be collected in a given time window
- Download arcs may be present
- Find a path that uses consequent time windows and covers largest number of arcs

Warehouse picking

- An order is composed of items of different classes (colors)
- Items are located in aisles
- Items must be picked in a given sequence one at a time
- Orders must be discharged at a sink node
- Optimise a route for a picker to assemble as many orders as possible

Complexity Results

- **Abouelaoualim et al. (2008)**: find k arc or node disjoint PEC paths is NP-complete; easy only for special classes
- **Gourves et al.(2009)**: the properly edge-colored s − t paths which visit all vertices of the graph a prescribed number of times can be found in polynomial time if the graph has no PEC cycles
- **Gourves et al.(2009)**: PEC Eulerians s−t path problem is polynomially solvable for c-edge-colored graphs, which do not contain PEC cycles
- **Bang-Jensen and Gutin (1998)**: finding a longest alternating simple path in a 2-edge-colored complete multigraph is computationally easy

- **Adamiak et al. (2004)** prove NP-hardness of the Hamiltonian path problem in 2-edge-colored simple graphs

- When n = 2, OCLP is equivalent to finding a properly colored path

- **Szachniuk at al. (2015)**: OCLP is NP-hard for n > 2

OCLP: How difficult in practice ?

Problem	Time to Solve
L0_100_3_70	3.41
L0_100_3_50	12.60
L0_100_6_70	14.88
L0_100_6_50	67.73
L0_100_10_70	3,425.09
L0_100_10_50	> 3,600.01

Previous Models based on flow formulation

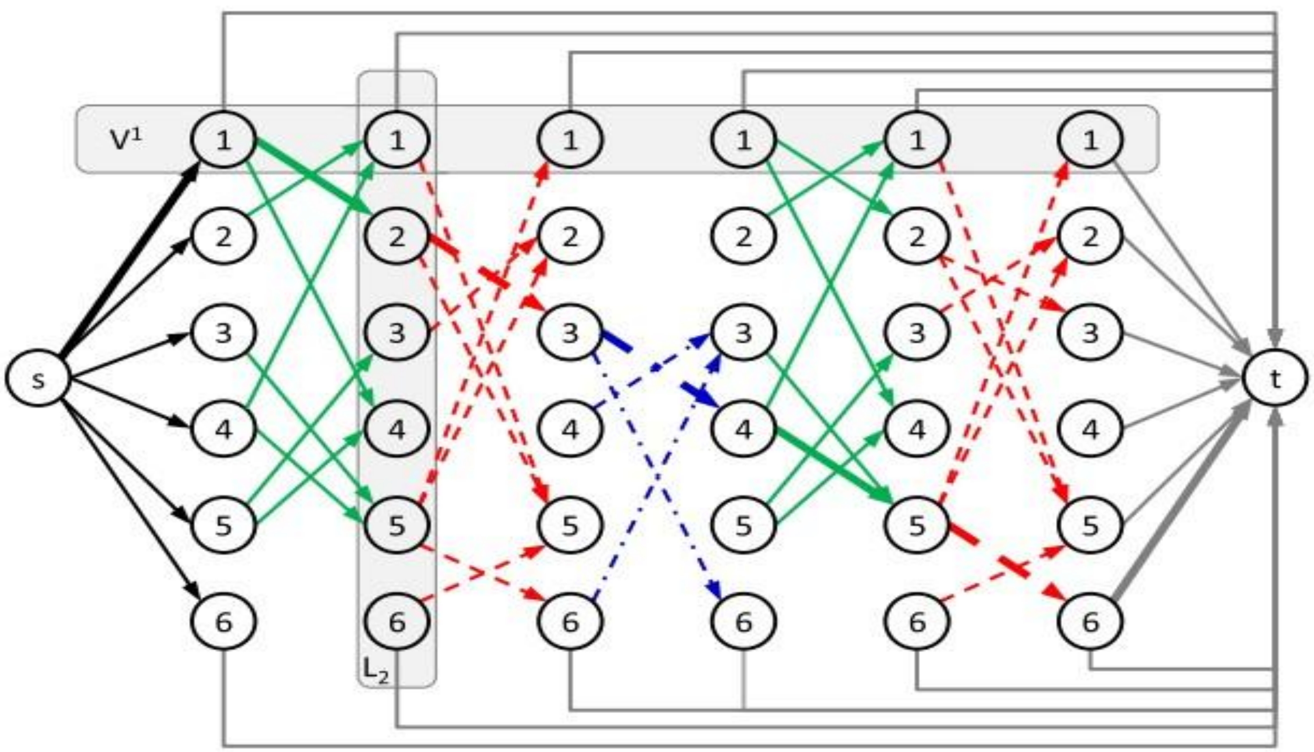
In previous work, 3 models based on the expansion of the nodes (+arcs) and on a flow-based longest path formulation are compared.

M. Szachniuk, M.C. De Cola, G.Felici, J. Blazewicz. **The Orderly Colored Longest Path Problem – a survey of applications and new algorithms**. RAIRO - Operations Research, 48-01 (2014)

M. Szachniuk, M.C. De Cola, G. Felici, J. Blazewicz, D. de Werra. **Optimal pathway reconstruction on 3D NMR maps, Discrete Applied Mathematics**, 182, 134-149, (2015)

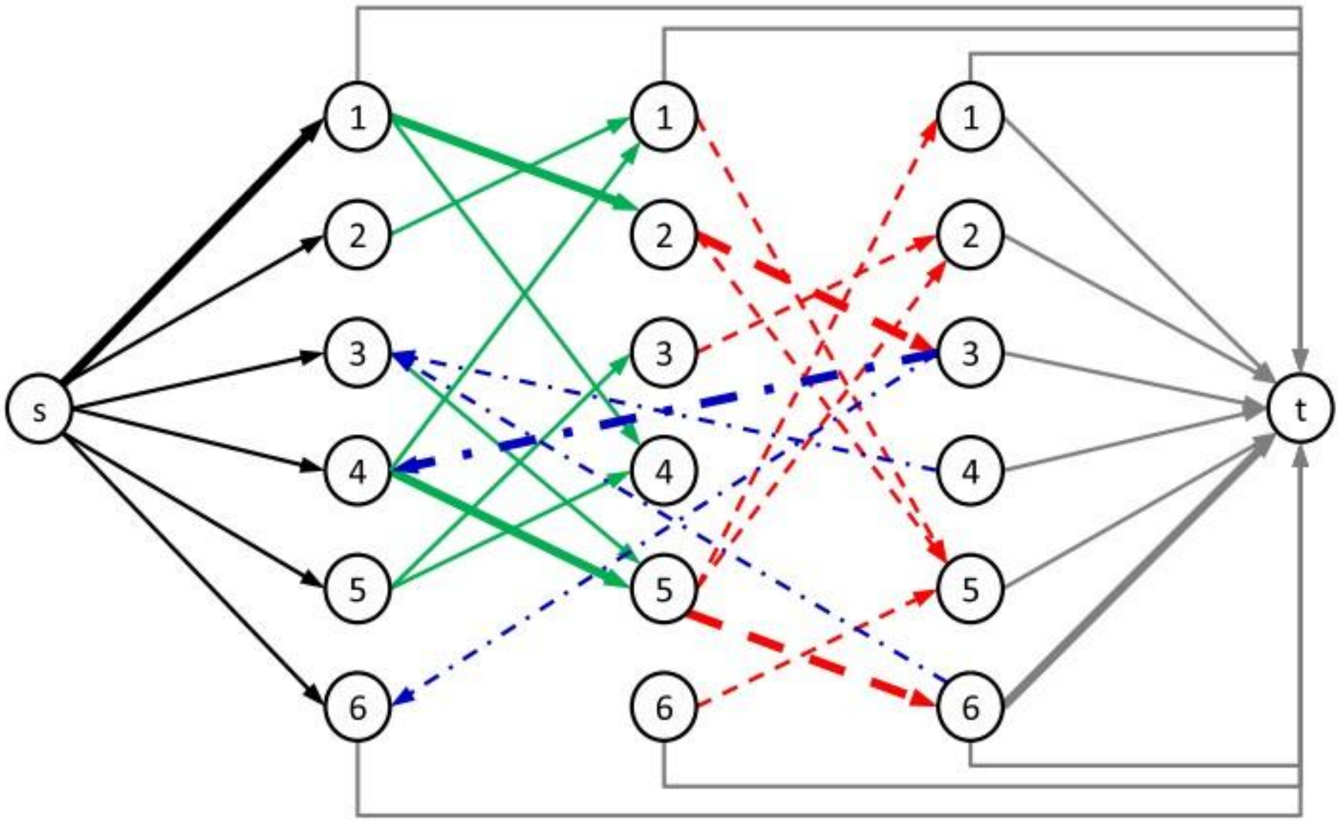
Model 1

- We use **n x n** nodes (wlog, start from a given color)
- Arcs are divided by order in the path
- Packing constraints on the n copies of the same node ensure that path does not go twice through the same node



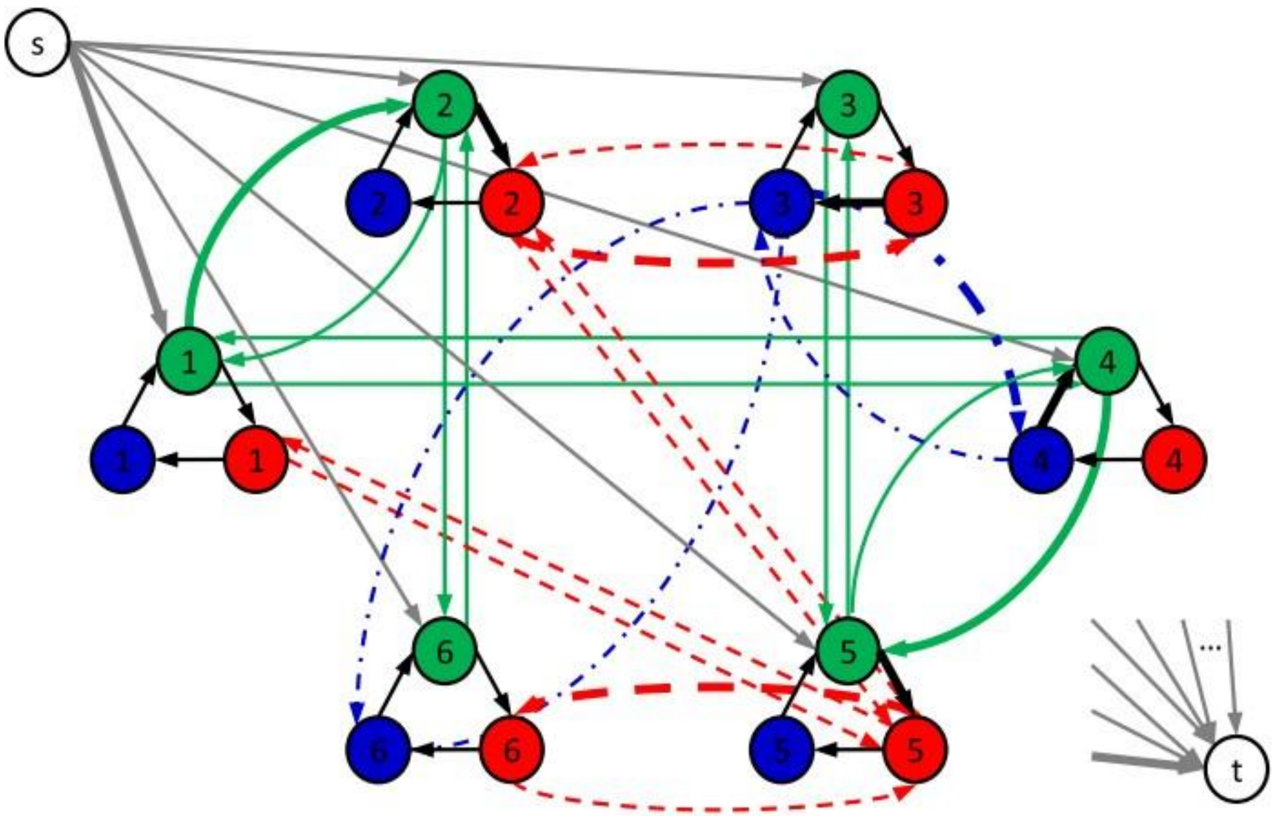
Model 2

- We use **n x c** nodes
- The graph is not acyclic, need cycle/subtour elimination constraints

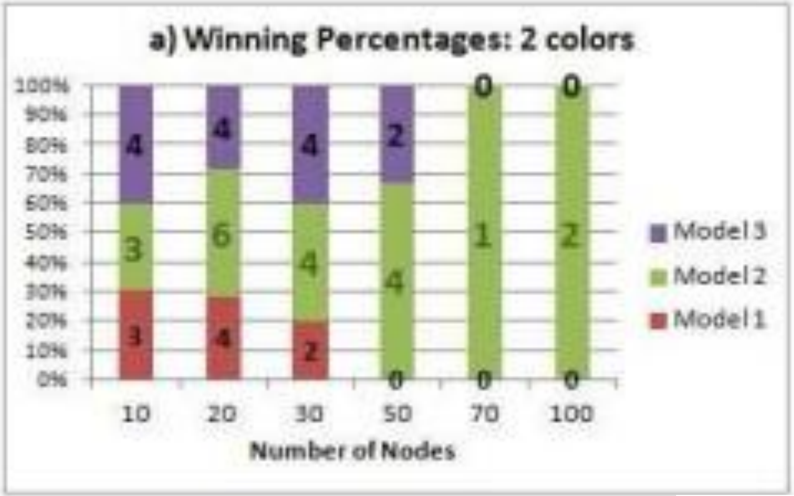


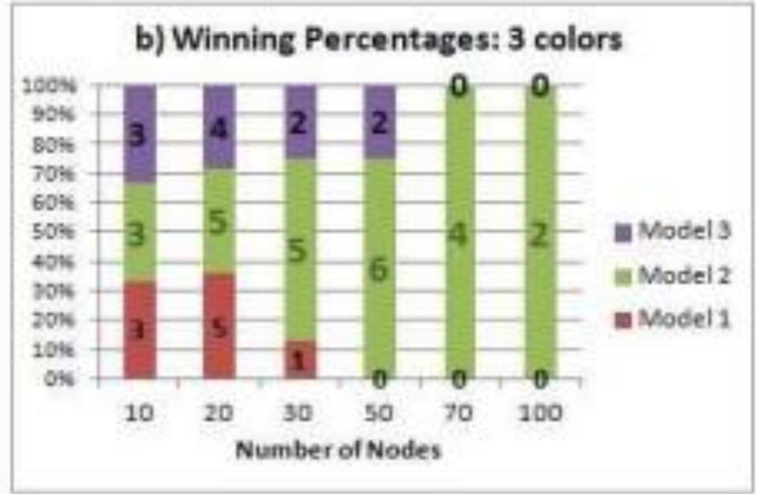
Model 3

- We use **n x c** nodes
- Each node is expanded in a subgraph
- Once entered a subgraph, can use exactly 1 arc in it
- Need cycle/subtour elimination constraints



Results on NMR-type instances





Method	Instance size 30	40	50	75	100	Total domination index
7-BC	0	1	7	4	3	15
7-IC	0	2	1	1	0	4
8-BC	0	0	1	5	7	13
8-IC	0	0	0	0	0	0
Total	0	3	9	10	10	32

Table 2: Method domination acc. instance size (instances with computing time $\geq$ 10s).

Method	Experiment type homonuclear	heteronuclear
7-BC	10	5
7-IC	1	1
8-BC	4	9
8-IC	0	0

Table 3: Method domination acc. experiment type (instances with size $\geq$ 50).

Scheduling Based Models

A different setting: assume starting color, define order of nodes according to their arcs

- **N**: set of nodes
- **i** : index of node
- **j**: index of position with color c_j associated with it
- **N(i,j)** : set of nodes k that are connected to i with an arc of color c_j
- $\mathbf{x_{ij}}$ = 1 if node i is visitied in position j, 0 otherwise
- $\mathbf{y_j}$ = 1 if the sink node is visited in position j, 0 otherwise

Scheduling Model 1

$$\min \sum_{j=1}^{n} y_j$$

Go to sink as late as possible

$s.t.$

$$\sum_{j=1}^{n} x_{ij} \le 1 \quad \forall i$$

At most a node in each position

$$y_j + \sum_{j=1}^{n} x_{ij} = 1 \quad \forall j$$

A position to each node, maybe the sink

$$y_j \le y_{j+1} \quad \forall j < n$$

Once sink, stay sink

$$x_{ij} \le y_{j+1} + \sum_{k \in N(i,j)} x_{k,j+1} \quad \forall \; 1 \le (i,j) \le n, j \ne n$$

To put node i in position j+1, I must have a node in position j that is connected with an arc of the proper color... or just go to sink ☹

Scheduling Model 2: Get rid of y's

$$\max \sum_{i=1}^{n} \sum_{j=1}^{n} x_{ij}$$

Assign as many as possible

$$s.t.$$

$$\sum_{j=1}^{n} x_{ij} \leq 1 \quad \forall i$$

$$\sum_{j=1}^{n} x_{ij} = 1 \quad \forall j$$

As before

$$x_{ij} \leq \sum_{k \in N(i,j-1)} x_{k,j-1} \quad \forall \; 1 \leq (i,j) \leq n, \; j \neq 1$$

$$x_{ij} + \sum_{k:N(k,j-1)=N(i,j-1)} x_{kj} \leq \sum_{k \in N(i,j-1)} x_{k,j-1} \quad \forall 1 \leq (i,j) \leq n, \; j \neq 1$$

The last constraints can be lifted to all other feasible candidates for position j, given (j-1)

Scheduling Model 3: new objective function

$$\max \sum_{i=1}^{n} \sum_{j=1}^{n} (j+1) x_{ij}$$

Removes simmetries by pushing solution towards positions with large index, thus mazimizing path length

Details on Experiments

1. Generate arcs according to overall graph density
2. Color arcs at random:
   a) Assign color at random to 10% of arcs
   b) Chose other colors with eq. prob. among feasible colors
   c) Inject a longest feasible path or not

- New instances with $n \geq 3$ colors
- Number of nodes (50,70, 75, 100)
- Number of colors (3,5,10)
- Density of graph (50%, 70%)
- Warrant on the presence of hamiltonian OCLP path or not

Model Sizes

- Dimension of associated MIPs: cycle based models are indeed smaller
- Size grows with n x c in Flow Models, with n^2 in scheduling like models.

	Variables			Constraints		
	mod1	mod2 mod3	Cycle	mod1	mod2 mod3	Cycle
L0_100_10_50	10,100	10,100	6,788	10,200	10,100	1,202
L0_100_10_70	""	""	8,796	""	""	""
L0_100_3_50	""	""	5,388	""	""	502
L0_100_3_70	""	""	7,396	""	""	""
L0_100_6_50	""	""	5,988	""	""	802
L0_100_6_70	""	""	7,996	""	""	""

	Solution times (secs)			
Problem	mod1	mod2	mod3	cycles
L0_100_10_50	3,600.00	3,600.00	3,600.00	3,600.00
L0_100_10_70	3,600.00	3,600.00	3,425.09	3,600.00
L0_100_3_50	12.60	14.10	18.56	7.67
L0_100_3_70	3.46	3.41	4.07	8.66
L0_100_6_50	1,014.54	1,109.70	67.73	3,600.01
L0_100_6_70	56.86	17.86	14.88	748.38
L0_50_10_50	3,600.00	3,600.00	2,047.16	3,600.00
L0_50_10_70	656.58	279.29	715.92	3,600.00
L0_50_3_50	3.83	11.51	11.05	1.33
L0_50_3_70	0.50	0.82	0.63	1.01
L0_50_6_50	79.46	163.44	138.32	3,600.06
L0_50_6_70	15.87	14.68	10.87	135.04
L0_75_10_50	3,600.00	3,600.00	3,600.00	3,600.00
L0_75_10_70	1,611.62	550.86	886.34	1,800.00
L0_75_3_50	6.81	4.40	6.89	7.51
L0_75_3_70	1.64	2.16	10.44	3.91
L0_75_6_50	79.39	1,126.76	90.56	3,600.00
L0_75_6_70	11.79	45.88	48.94	209.31
L1_100_10_50	0.43	0.21	0.18	3,600.00
L1_100_10_70	1,979.43	0.26	2.48	3,600.00
L1_100_3_50	36.93	0.38	2.85	7.22
L1_100_3_70	0.56	0.46	4.15	9.84
L1_100_6_50	948.72	0.28	0.23	1,656.57
L1_100_6_70	13.88	0.32	0.33	1,079.58
L1_50_10_50	696.84	0.06	0.32	3,600.00
L1_50_10_70	51.50	0.08	0.05	3,600.00
L1_50_3_50	1.07	0.10	0.06	0.27
L1_50_3_70	2.01	0.10	0.27	1.28
L1_50_6_50	0.14	0.06	0.07	315.76
L1_50_6_70	33.35	0.06	0.24	67.55
L1_75_10_50	3,600.00	0.11	0.11	3,600.00
L1_75_10_70	569.61	0.13	0.14	3,600.00
L1_75_3_50	4.51	0.21	0.97	4.77
L1_75_3_70	7.35	0.24	0.65	3.07
L1_75_6_50	284.47	0.15	0.12	240.82
L1_75_6_70	72.74	0.15	0.19	41.95

L0: hamiltonial path not injected

L1: hamiltonial path injected

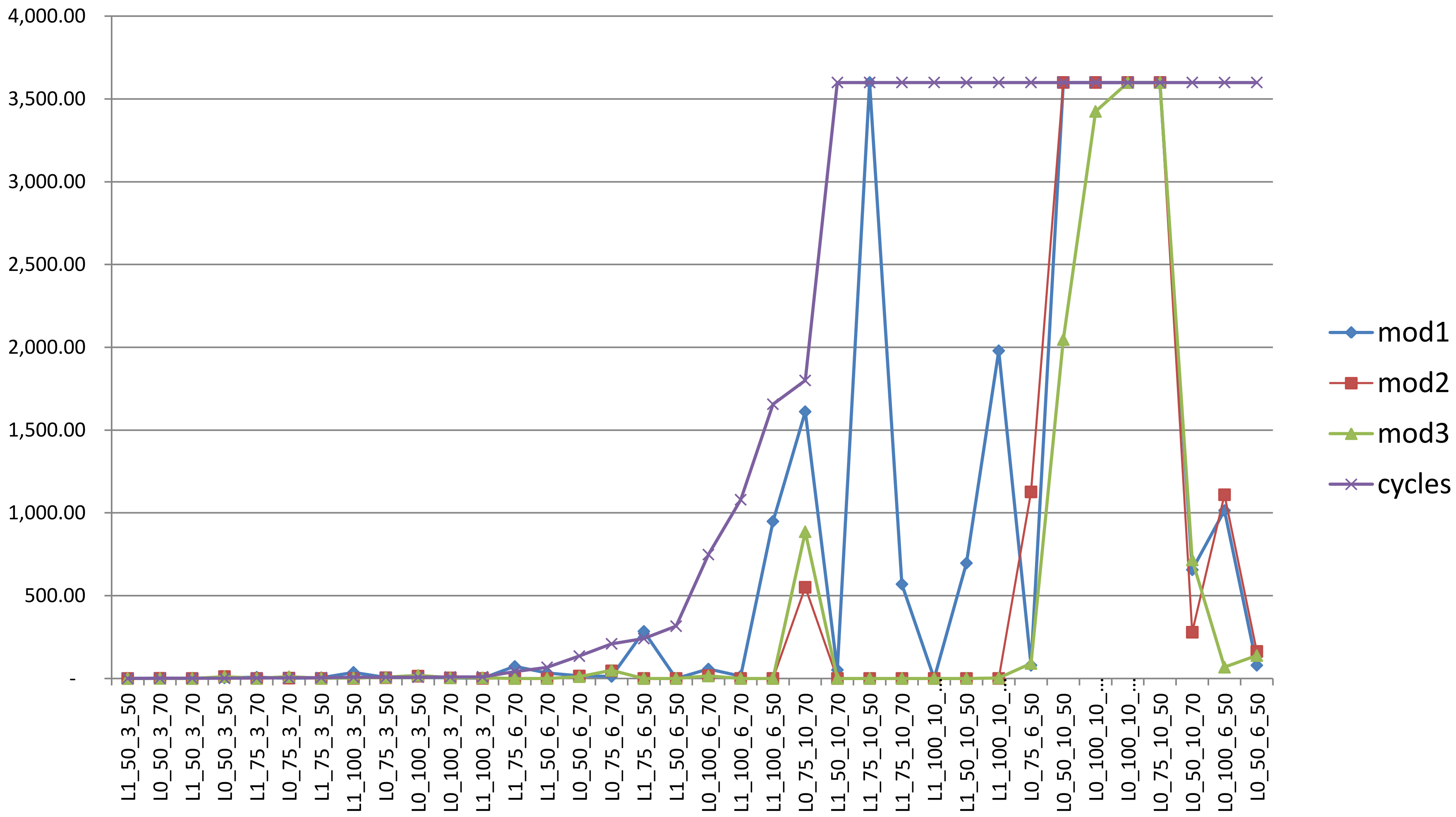
percentage not solved within 1 hour	
L0	11.11%
L1	0.00%
L0_100_10_50, L0_75_10_50	

Avg. Solution Time	
Mod1	729.4
Mod2	493.0
Mod3	408.6
Cycles	1576.4

percentage not solved	
Mod1	13.89%
Mod2	11.11%
Mod3	5.56%
Cycles	30.56%

Fractional cycle separation not convenient when we have many colors

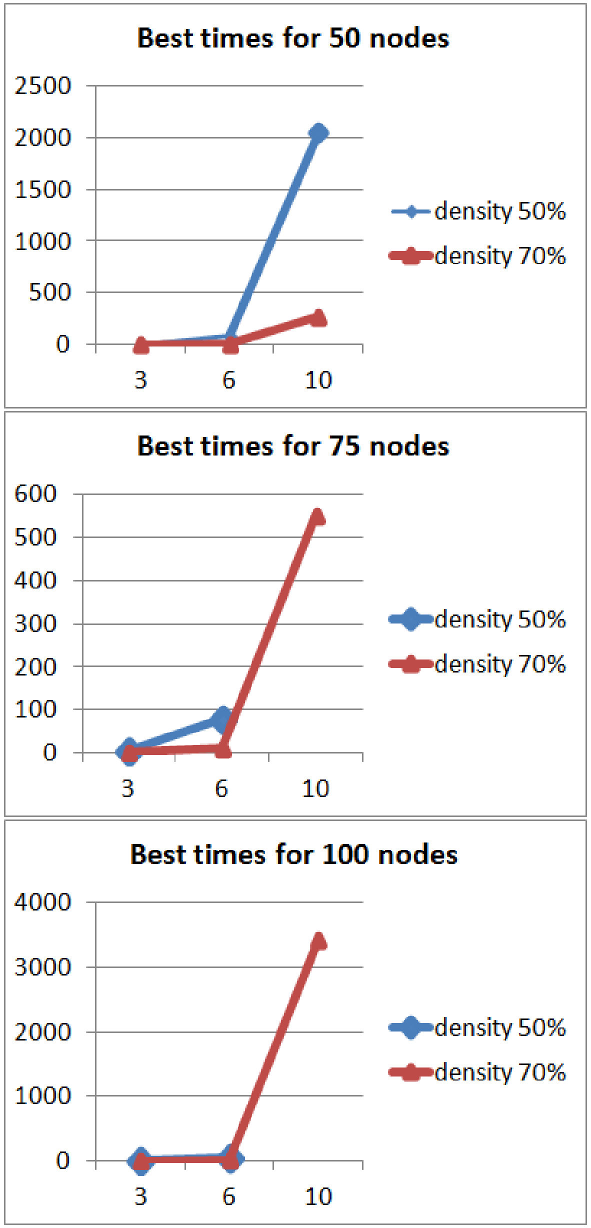
Scheduling – like models are larger but faster



Scheduling – like models are larger but faster

Problems without injected longest path are more interesting

- Denser problems are easier
- Problems with less colors are easier



Conclusions

- Orderly Colored Path Problems has been introduced
- Original motivation to be found in NMR spectra analysis
- Could be used to model complex scheduling problems?
- Problem is in NP
- Already proposed Flow based models have been described
- New Scheduling-like models are introduced and tested (successfully!)
- Model Comparisons on randomly generated instances of larger size

- New formulation performs better with many colors
- Potentially interesting for scheduling applications
- Study additional variants where path colored arcs obey general constraints

Thanks