



# Robust Project Scheduling

## Motivation

Projects are often scheduled under the assumptions of complete information and a static execution environment. In reality, projects are subject to uncertainty/ disruptions which are gradually resolved during execution. Examples include: resources become unavailable, duration variability, deliveries delayed, due dates re-negotiated, and weather conditions. Such disruptions incur costs due to missed deadlines, resource idleness, higher work-in-progress inventory, and system nervousness. Thus, we seek to develop a (robust) pre-schedule that ensures stability in activity start times by incorporating uncertainty.

## Assumptions

- Activities cannot be started before their foreseen starting time in the pre-schedule.
- Proper allocation of resources has been performed, thus reflected in the precedence constraints, or that allocation will occur in a later phase based on the pre-schedule.
- The effect of one disturbance will not interact with the effect of another.

## Definitions

- A project can be represented by a directed acyclic graph  $G = (N, A)$ .  
N: Activities      A: Finish-start precedence relations
- Path  $P(i, j)$  denotes any path from  $i$  to  $j$  in  $G(N, A)$ .
- Set  $TA$  denotes the transitive closure of  $A$ , such that  $(i, j) \in TA$  if and only if  $\exists$  some  $P(i, j)$ .
- $P^*(i, j)$  is the path with the largest sum of activity durations between  $i$  and  $j$ , i.e. the longest path.

Parameters	
$d_i$	Duration of activity $i$
$w$	Final project deadline
$c_j$	Cost per unit-time overrun on start of activity $j$
$g_i(\cdot)$	Probability mass function for disturbance scenarios of activity $i$
$l_{ik}$	Disturbance duration due to scenario $k$ of activity $i$
$\lambda_{ij}$	Length of $P^*(i, j)$ not including durations of $i$ and $j$
$D_i$	Collection of disturbance scenarios for activity $i$
Decision Variables	
$s_i, s_j$	Start time of activity $i$ or $j$ , respectively
$\Delta_{ijk}$	Delay in start of activity $j$ due to disturbance scenario $k$ of activity $i$

## Formulation

$$\min \sum_{(i,j) \in TA} \sum_{k \in D_i} c_j g_i(l_{ik}) \Delta_{ijk}$$

$$\text{s.t. (1) } s_j - s_i \geq d_i \quad \forall (i, j) \in TA$$

$$(2) \, s_n - s_0 \leq w$$

$$(3) \, \Delta_{ijk} + s_j - s_i \geq l_{ik} + d_i + \lambda_{ij} \quad \forall (i, j) \in TA \quad \forall k \in D_i$$

$$(4) \, \Delta_{ijk} \geq 0 \quad \forall (i, j) \in TA$$

$$(5) \, s_0 = 0; \, s_i \text{ unrestricted} \quad \forall i \neq 0$$

*Intuition:* Objective is to minimize the expected weighted deviation of actual versus planned activity start-times.

## Current & Future Work

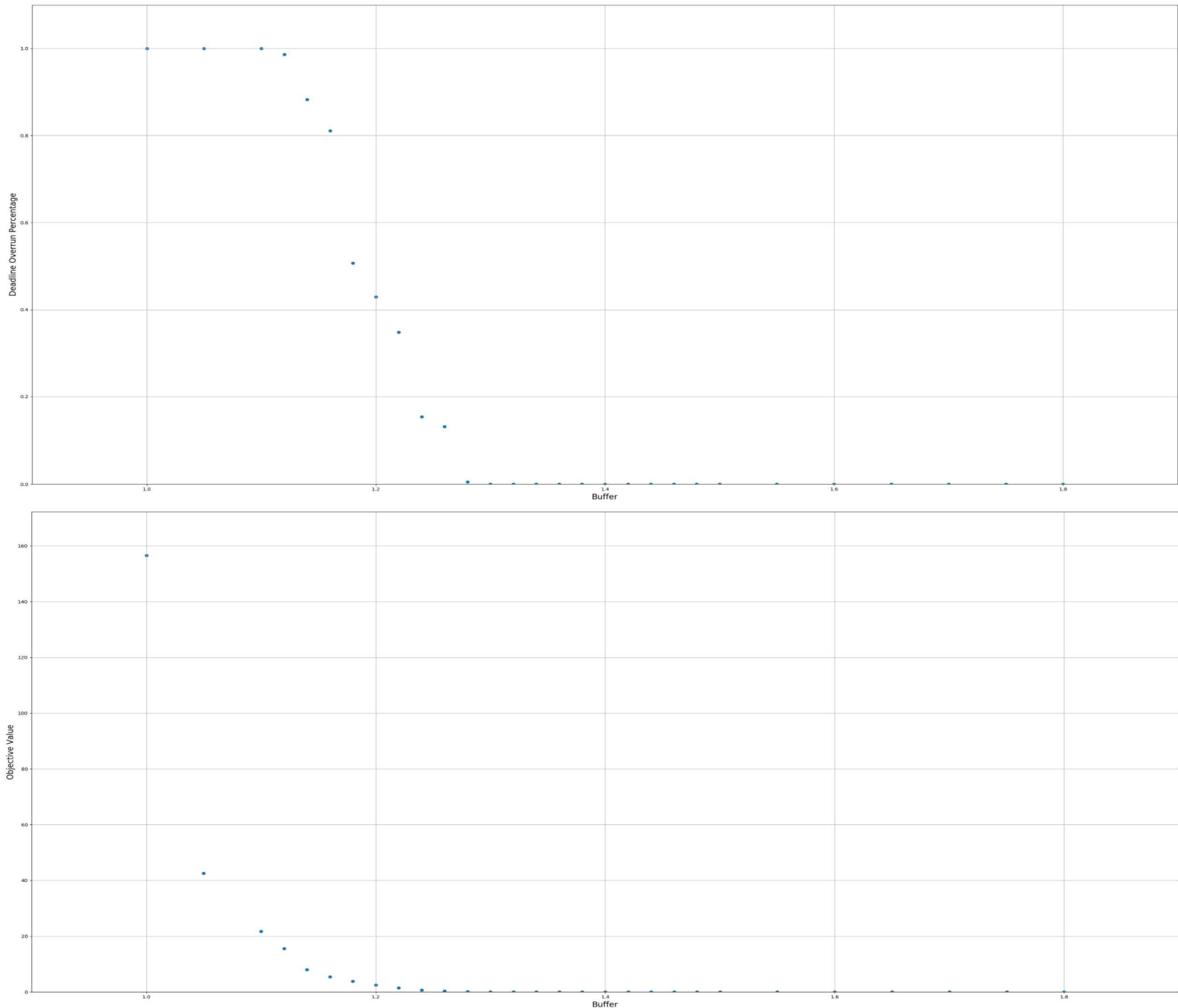
- Transforming formulation into resource-constrained integer programming format.
- Heuristic or exact procedure to restrict set  $TA$  to only include activity pairs that have close proximity within the precedence network.

## Citation

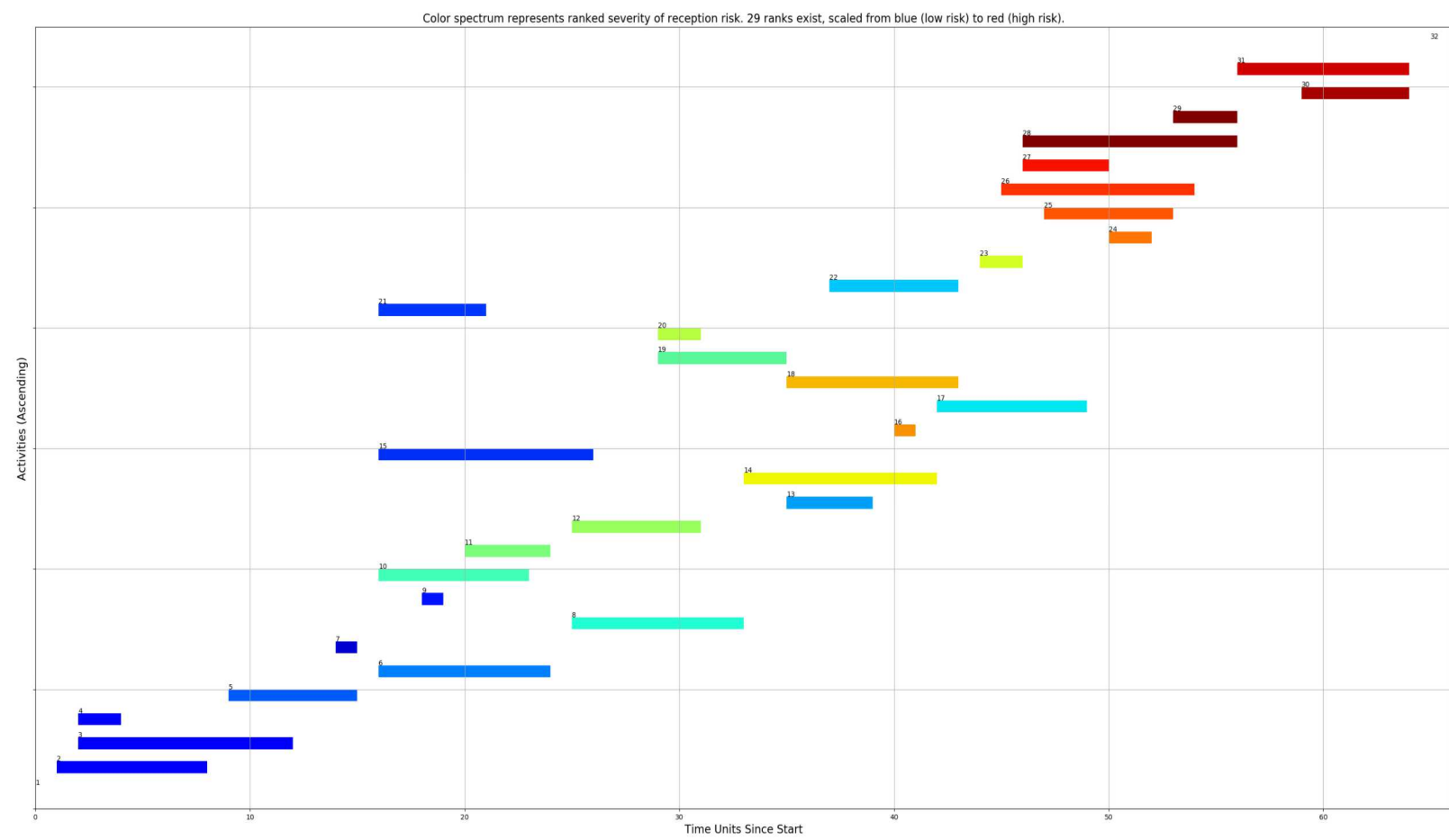
Herroelen, Willy, and Roel Leus. “The Construction of Stable Project Baseline Schedules.” *European Journal of Operational Research*, vol. 156, no. 3, 2004, pp. 550–565., doi:10.1016/s0377-2217(03)00130-9.

## Evaluation

*Illustration 1:* For a 30-node network with parameters  $c_i = U[1,2,3,4]$  and  $l_{ik} = \text{Triangular}(0, 0.15 \times d_i, 0.30 \times d_i)$  the overrun percentage of deadline  $w$  and the formulation’s objective value are observed to rapidly fall as buffer increases. Buffer is a numeric multiplier applied to network longest path  $P^*(0, n)$ . Similar patterns are evident in 60-, 90-, and 120-node networks.



*Illustration 2:* For the same 30-node network, specified parameters, and buffer set to 1.25, a project Gantt chart can be produced. Spacing is interspersed between network longest path sequence 1-2-5-6-8-18-26-31-32. Such spacing distributes risk and ensures stable start times amidst delays. This is supported by simulation – where the simulated delays exhibit a linear relationship with the  $\Delta_{ijk}$  decision variables over thousands of runs.



*Illustration 3:* For the same parameters, including buffer held constant at 1.25, fifteen distinct networks are sampled each from 30-, 60-, 90-, and 120-node network sizes to display the relationship between solution time and formulation complexity. A roughly linear relationship can be observed from lower-left to top-right in solve-time as both  $|N(G)|$  and  $|TA|$  increase in value.

