A flexible model for scheduling building processes based on graph theory and fuzzy numbers

Martin Freundt, Karl Beucke; Bauhaus University Weimar, 99421 Weimar, Germany
(martin.freundt@bauing.uni-weimar.de; karl.beucke@bauing.uni-weimar.de)

Summary

The methods currently used for scheduling building processes have some major advantages as well as disadvantages. The main advantages are the arrangement of the tasks of a project in a clear, easily readable form and the calculation of valuable information like critical paths. The main disadvantage on the other hand is the inflexibility of the model caused by the modeling paradigms. Small changes of the modeled information strongly influence the whole model and lead to the need to change many more details in the plan. In this article an approach is introduced allowing the creation of more flexible schedules. It aims towards a more robust model that lowers the need to change more than a few information while being able to calculate the important propositions of the known models and leading to further valuable conclusions.

1 Introduction – Scheduling building projects

1.1 Initial Situation

Currently the planning of large building projects is executed in two steps. In step one a schedule is worked out while in step two the available and/or needed capacities and resources are calculated and set into relation to the planned schedule and the costs are calculated. The main objective for step one is to arrange all tasks of the project in logical and chronological order. This step can be performed by project managers, building companies, construction managers and others. Information about resources often are not precisely known at this stage, but rather it is assumed that there will be sufficient capacities. The second step, the capacity and cost management, is mainly done by construction companies where information about available labor force and other resources are known. Thus separating the scheduling from the resource management is not only common practice but often a practical need, especially with the current way contracts and sub-contracts are handled in the construction industry.

1.2 State of the art

Schedules for building projects usually are modeled in network plans using the critical path method (CPM), in Gantt-charts or in other similar plans. The advantages of these methods are the arrangement of the tasks of a project in a clear, easily readable form, the ability to calculate the possible amount of time the project will need and the calculation of the critical path within the chosen arrangement of the tasks. The main disadvantage of the methods and tools currently available is the inflexibility caused by the model and the modeling paradigms. Changes of the plan during construction often have major impact on the whole model making the originally planned schedule at least partially obsolete.

Besides the deterministic methods there exist non-deterministic approaches, mainly the PERT method (Program Evaluation and Review Technique). PERT is an approach using and aiming for probabilistic information. Three durations are modeled for each task, one pessimistic, one probable and one optimistic duration. The three information are used to define a distribution function assuming a beta-distribution. Aim of PERT is to get propositions about the probable duration of each task and the probable duration of the whole project. Practical use has shown that this aim is not reached sufficiently enough to justify the use of PERT for general building projects. The reason is that the predicted probable project durations significantly differ from the real durations. Reasons for that are problems with the theory itself, i.e. the violation of the central limit theorem of the probabilistic theory which is not applicable to all steps performed, and problems with the application of the theory in building projects, i.e. the assumption of the beta-distribution which often is not applicable. Besides PERT other methods using probabilistic information have been developed, i.e. GERT (Graphical Evaluation and Review Technique), but none of these methods have been widely adopted for practical applications. Probabilistic approaches as a whole generally are not used for scheduling building projects with the exception of the use of PERT for very special projects.

Another non-deterministic approach was suggested by Mühlögger using a fuzzy network plan. This approach seems promising, especially regarding the main disadvantages of CPM and Gantt-charts. In (Mühlögger 1994) two examples show how a fuzzy extension of a classic network plan behaves compared to the classic network plan. The network model itself was not affected, but the durations have been evaluated using fuzzy numbers. The basic idea of using fuzzy numbers will also be used in the approach suggested in this article.

2 Aims

This article is aimed at introducing a flexible process model for scheduling building processes considering the typical uncertainties of information in building projects. The model will be comparable to known deterministic models and it will be possible to calculate the important and valuable propositions these models result in. Furthermore the model will be more robust to small changes frequently occurring in practical projects. It leads to extended valuable propositions and information.

The typical uncertainties of information in building projects include insufficient information at the time of planning, different durations of similar tasks on different sites due to different capacity assumptions and availabilities, the typical risks of the construction process (i.e. weather, fluctuation of the workforce, …) and a wide range of other things. In the field of reconstruction the uncertainties are even bigger due to incomplete and possibly wrong information about the object to reconstruct.

3 Model

The essential components of building projects in the terms of scheduling are events, tasks and the dependencies between these objects. The sequence of the events determines the schedule. Thus the purpose of the models is to arrange the events and tasks in logical and chronological order and to calculate the dates of the events and additional information like the critical path.

A "*schedule*" is defined as one complete workflow containing all events and tasks being part of the project for exactly one time. The dependencies between events and with that between tasks are used to define the sequence of the project. A useful and functional approach to model such a schedule is to use a directed graph with one set of nodes and one set of edges. All calculations will be done using path algebras in this graph.

All periods of time will be modeled using fuzzy numbers in order to include the described kinds of uncertainties. Instead of defining one exact, deterministic number for a period, an approximate range will be used. Giving an approximation for the duration of tasks is common usage in building industry. Thus, using fuzzy numbers for durations corresponds to the way of thinking in practical applications. The information modeled using fuzzy numbers is a statement about possible durations of a task or a dependency. This is not to confuse with a probability. Although classic risks like weather influences are meant to be included the model does not aim towards calculating probabilities of durations or dates but rather the results are much more a combination of possibilities and probabilities which leads to the term fuzziness (Klir and Wierman 1999). Fuzziness, being the more general concept for modeling uncertainty, is exactly what is appropriate here. The model presented aims towards a more realistic design of the classic network models by adding uncertainty and thus gaining more flexibility.

4 Implementation

4.1 Directed graph as base structure

Nodes and edges need to be defined in order to use a directed graph. The nodes will be used to represent events and with that to model a point in time. Possible events are the start of the project, its end, start and end of tasks, milestones or any other important event. The edges on the other hand represent periods of time constituting durations of tasks and/or dependencies.

The model has to conform to a set of rules. The rules defined are:

- each task has its own start and end events, modeled in separate nodes
- for each task one edge is modeled to represent the period of time
- for each dependency between two events one edge is modeled
- edges for dependencies are definable between any two nodes
- but: the graph has to be free of cycles with a positive length (the sum of all durations on a cycle must be zero or negative)
- without loss of generality the project must have a single one start event and a single one end event

The following sets are defined:

- SN … set of nodes containing all start events of tasks (StartNodes)
- EN … set of nodes containing all end events of tasks (EndNodes)
- ON … set of all other nodes (OtherNodes)
- $N = SN \cup EN \cup ON$ ... set of all nodes

- $TE := \{ (a, b) \in SN \times EN \mid aTEb \}$ … set of edges for tasks
- $DE := \{ (c, d) \in N \times N \mid cDEd \}$ … set of edges for dependencies
- $E := TE \cup DE$ … set of all edges

- $S := (N; E)$ … directed graph representing the schedule

Based on the properties of these sets several tests can be defined to verify the consistency of the graph and its sets concerning the defined rules.

With the rules defined all dependencies known from classic network plans are easily creatable. At the same time the model should contain all necessary elements for any further expansion for purposes like cost or capacity management. For example it is possible to assign cost information like single values or functions to each node and to each edge. With that the model should not only be able to calculate all results known from classic networks plans, but it should also be usable as base model for further calculations usually done in step two of planning and managing building projects.

Figure 1 shows possible dependencies between tasks defined following the models rules. Due to defining separate start and end events for each task it is easily possible to create the known dependencies (start-end, end-start, start-start and end-end). A single task in the project is modeled using two nodes and one edge in the graph so that the dates of the begin and the end as well as the duration of the task are described.

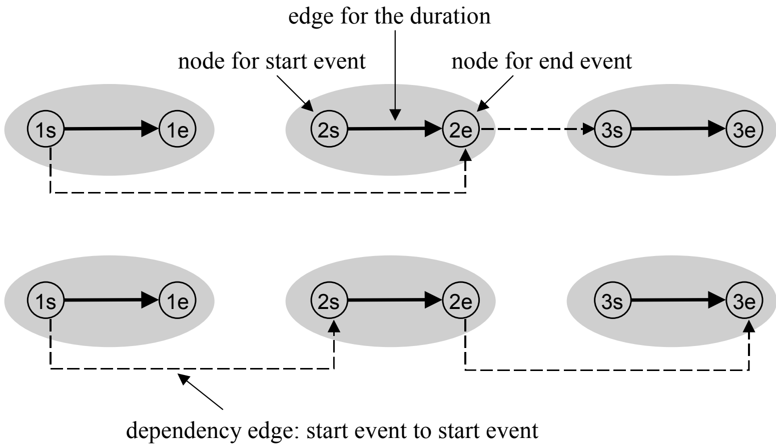


Figure 1: events, tasks and dependencies following the rules of the model

Path algebras will be used for calculating results. For this reasons specific values need to be assigned to the edges. Fuzzy numbers which are a special form of fuzzy sets are used for this purpose. Described is the degree of membership of real numbers to a fuzzy number using a membership degree function μ with $0 \leq \mu \leq 1$. One form of fuzzy numbers which is especially useful for the purpose of this model are left-right fuzzy numbers. Left-right fuzzy numbers (LR-numbers) describe the membership degree function μ using reference functions to define a left and a right span around one focal point with the membership degree 1. As such LR-numbers are a special form of fuzzy intervals.

The reference functions describing the left and the right span have to be monotonic with the left function starting at a value of 0 and increasing monotonously to the value of 1 while the right function starts at the value of 1 and decreases monotonously to 0. There must be exactly one point with a membership degree of 1. Figure 2 shows the triangular fuzzy number "approximately 3" as an example.

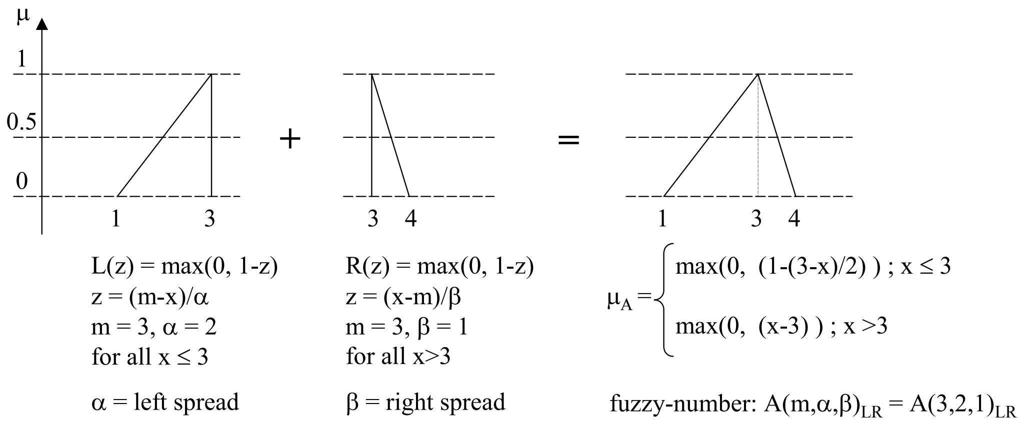


Figure 2: LR fuzzy number "approximately 3"

So called spreads are introduced in the figure 2. This figure shows the fuzzy number "approximately 3" which has a left spread of 2 and a right spread of 1, meaning that the left span has a wideness of 2 units and starts at 1 = 3-2 while the end point of the right span is at 4 = 3+1 so that the span is 1 unit wide.

No information is specified about the reasons for the spreads, namely it is no more probable to need one day or three days. Instead what the number says is that it is possible to do the task within one day and that it should not take longer than four days. In a classic, deterministic plan this task would be modeled with a duration of three days. That seems to be the best assumption possible, it is the time the company intends the task to take. Realistically it is better to assume a span for the task to take while still aiming for the intended or planned three days.

The shown number is triangular because L and R both are linear functions creating a triangle of real numbers with a membership degree other than zero. Triangular numbers are a special form of general LR-numbers. Using triangular LR-numbers is one of the easiest methods of describing uncertainty. It is easy to understand and easy to create while working with the model. That is why this form is chosen for the model presented. The duration of all tasks is modeled with such numbers so that a task will need an approximate amount of time, i.e. approximately three days. It could be less, like one day fastest, or it could be more, like four days maximal.

This also leads to a different notation for the fuzzy number A where the points a_l = begin of left span, a_m = middle point and a_r = end of right of span are given instead if m, α and β. Also L and R do not have to be specified as part of the number as long as L and R both are linear and always are the same functions for the numbers involved. It is clear that the membership function is increasing monotonously and linear from a_l to a_m and is decreasing monotonously and linear from a_m to a_r.

$$A(m,\alpha,\beta)_{LR} = A(3,2,1)_{LR} \Leftrightarrow A_\Delta(a_l, a_m, a_r) = A_\Delta(1,3,4) \qquad \text{(notations)}$$

For the use of fuzzy numbers in the field of building projects this seems much more readable and complements the way of specifying periods of time as in: approximately 3 in the span between 1 and 4. This is sufficient for many problems, for example for calculating the exact dates of the events. For this purpose the actual gradients do not matter and can be ignored – or can be simplified to be linear. Mathematically and technologically it is possible and relatively easy to use other, non-linear or polygonal functions. The actual gradient of the functions mainly matters if the calculation is aimed towards creating measures. This will not be followed in this article.

(As a remark: the need to create a measure, namely the expectation of the probability, is one of the main problems of the PERT-method. Part of this problem is that the beta-distributions used for the durations are not a good assumption in many cases but the distribution function strongly influences the actuarial expectations. Using other distributions, or other functions in the model suggested, is mathematically easy. But firstly in building projects it is very hard if not impossible to find and to justify assumptions of other distributions and secondly the necessary work to create and justify distributions for each possible task is almost impossible to do. The quantity of the parameters to include in the calculation is too big, let alone the very high complexity of the interrelations of the parameters.)

4.2 Definition of path algebras

The resulting model allows the definition of path algebras to calculate paths in the net as shown in (Pahl 2000). A generic graph solver is used to calculate specific results (Freundt 1998). Part of the generic solver is an interface for defining path algebras. Two operations need to be specified (concatenation and unification of two edges) as well as a neutral elements for each of the operations. Based on calculating sets of paths the algebras then calculate sets of values for the paths. The usage of a generic solver allows to define a set of algebras for different purposes. One specific purpose could be to calculate the critical path as done in classic network plans.

As an example the definition of an algebra for the determination of the earliest possible project end date will be shown briefly. Values are assigned to all edges in the format of triangular LR fuzzy numbers. The concatenation of two edges results in the addition of the edges' values while the unification of two edges results in the maximum of the values. Both operations have to be defined for fuzzy LR numbers. Therefore the addition and the maximum-operation for LR fuzzy numbers is shown.

The addition of two triangular LR numbers is straightforward. The rule is shown in both notations:

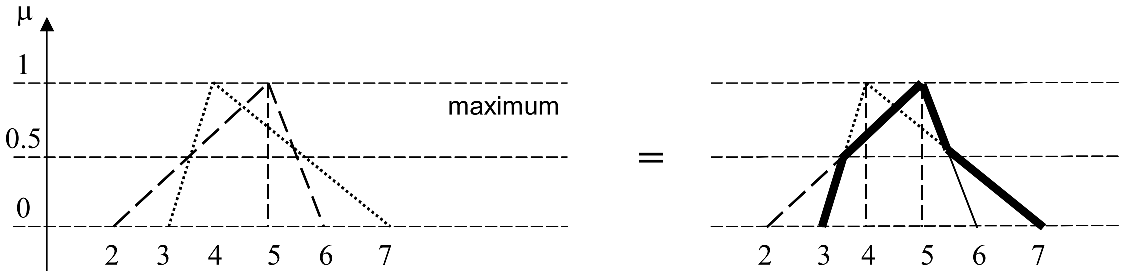
$$A + B = (m, \alpha, \beta)_{LR} + (n, \gamma, \delta)_{LR} = (m+n\ , \alpha+\beta\ , \gamma+\delta)_{LR} \qquad (1a)$$

$$A_\Delta + B_\Delta = (a_l, a_m, a_r) + (b_l, b_m, b_r) = (a_l+b_l\ , a_m+b_m, a_r+b_r\ ) \qquad (1b)$$

The result of this operation is again a triangular number. Like the middle points m and n of the numbers the left and the right spreads also add together so that the spreads of LR numbers always increase during the addition.

The result of the maximum operation for triangular LR numbers is not always a triangular number but in general is a polygonal number. The general rule (2) is gained using the general extension principle defined by Zadeh. For more information see (Böhme 1993) or other similar literature. Figure 3 shows an example.

$$Max(A,B) : \mu_{Max(A,B)}(z) = sup_{z=max(x,y)}\ (\ min(\mu_A(x), \mu_B(y)\ ) \qquad (2)$$



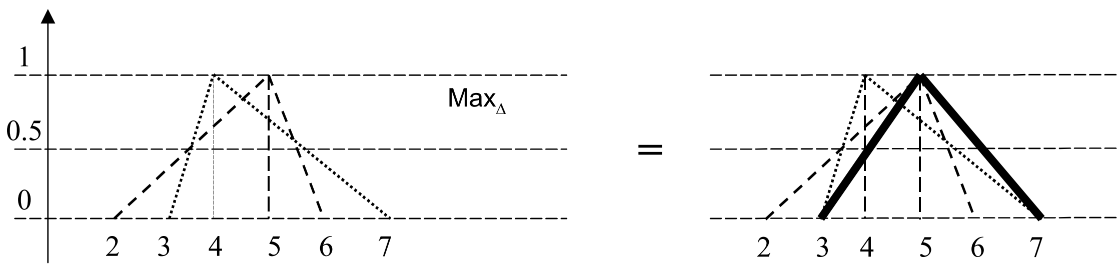
Fiugre 3 : example of the general maximum operation for fuzzy LR numbers

The new number itself is not triangular anymore. But interestingly the three prominent points c_l, c_m and c_r can be determined using the classic max operation:

$$C = Max(A, B) \text{ with } c_l = max(a_l+b_l)\ , c_m = max(a_m+b_m)\ , c_r = max(a_r+b_r)$$

Since this operation does not lead to a triangular LR number and since triangular LR numbers are sufficient for the purpose of this algebra another definition is needed. It can be found using the upper limit for the numbers A and B. The upper limit in the set of all triangular fuzzy LR numbers is the one triangular number C_Δ with $C_\Delta = Max_\Delta\ (A_\Delta, B_\Delta)$ with $A_\Delta \leq C_\Delta$ AND $B_\Delta \leq C_\Delta$. This leads to the definition of the Max_Δ operation (3), an example is shown in figure 4.

$$C_\Delta = Max_\Delta\ (A_\Delta, B_\Delta) = (\ max(a_l+b_l)\ , max(a_m+b_m)\ , max(a_r+b_r)\ ) \qquad (3)$$



Fiugre 4 : example of the Max_Δ operation for triangular fuzzy LR numbers resulting in a triangular number

Finally the neutral elements have to be defined for the operations of the algebra. The operation “concatenation of two edges“ corresponds to the addition of two triangular LR fuzzy numbers. The neutral element for this addition is the triangular number $0_\Delta = (0,0,0)$. 0_Δ can be added to any triangular number A_Δ, the result of $A_\Delta + 0_\Delta$ always will be A_Δ. The neutral element of the maximum operation (which corresponds to the unification of two edges) is the fuzzy number $-\infty_\Delta = (-\infty,-\infty, -\infty)$. For any number A_Δ the maximum $Max_\Delta(A_\Delta, -\infty_\Delta)$ is A_Δ. This is also true for the general Max-operation for LR fuzzy numbers. The neutral elements are shown in the second notation. The spread of these numbers is zero in all cases. Mathematically exact it should be a tiny positive value because a spread of exactly zero is not defined. The operations shown in (1a), (1b), (2) and (3) are not sensitive to that as the used mathematical operations “+” and “max” work well for exact zero. Thus, the notation shown is used for illustration purposes.

Now the generic solver easily can be used with the defined algebra. Calculated are all possible paths starting at the project start event (node). The algebra discussed will determine the length of the critical path of the schedule and with that the earliest possible date of the project end event. The same algebra can be used to determine the earliest possible dates for all events. The events and tasks on the critical path are not easy to specify using the shown algebra. For this purpose the edges’ values have to contain literals to identify the edges. A combined algebra for literals and fuzzy numbers is definable based on the one shown. Using it with the generic graph solver results in the calculation of the actual critical path and its length.

5 Working with the model

Due to using triangular fuzzy numbers for all calculated points and periods of time three values are given – the fastest, the intended and the slowest possible date or duration. This is independent from any buffers emerging from the network plan. The buffers also have three values. The middle value (value with membership degree 1) is exactly the same value a classic deterministic calculation would result in if the middle value of each tasks duration was used for the classic model. This way the middle version of all dates shows the “intended” version that would have been calculated classically. Additionally the fuzzy model shows two other times, the fastest possible and the slowest possible duration for each path.

Generally the three different times shown at each event (node) correspond to different paths from the start node to the actual node. This is due to the maximum operation and the unification of parallel edges. Figure 5 shows a simple example of the resulting model and of the calculation of the length of paths. It is used for determining the date for the end event E.

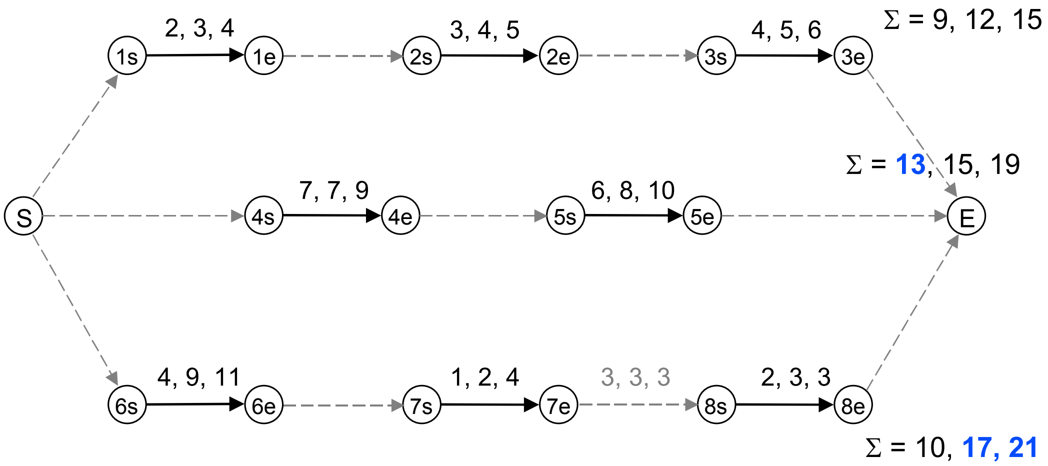


Figure 5: small example of a schedule

In this figure the single durations modeled at the edges are added for each path but the different paths from S to E are not yet united. The fuzzy number 0_Δ is applied to the edges without visual value. For the sake of ease only one kind of dependency between tasks is shown connecting the end of a task with the start of the next one.

5.1 Parallel paths

The unification of all paths leads to the triangular fuzzy number (13, 17, 21) which expresses the three different lengths of the critical paths (fast, intended, slow). To the lengths 17 and 21 units corresponds the same defining path, the path following tasks 6, 7 and 8. But in another case the critical path is the path containing the tasks 4 and 5 and its length is 13 units. This hints towards a possible change of the critical path from one path to another.

Another interesting information is how much a path can potentially be shortened (in the sense of decreasing its duration). It is assumed that the graph shown is part of a larger project with other tasks to do before S and after E. The tasks before S should have needed 100 units of time but in this assumption it took 105 units to complete this part. For the rest of the project 5 units of time are to safe so that the intended project end date can be reached. Naturally the critical path is inspected to find potential for decreasing its duration. In this case the path containing tasks 6, 7 and 8 is inspected. The resulting fuzzy number (10, 17, 21) indicates a potential to decrease the duration to 10 units while still being in the range of the possible durations of every single task and dependency. But the unification of all paths at node E shows that the fastest lowest duration is 13 units of time, caused by the path via tasks 4 and 5. So the duration of the lower path with tasks 6, 7 and 8 might be decreasable to 10 units but this doesn't solve the problem to safe 5 units as then another path becomes critical. The potential of shortening is only 4 units (intended duration of this path = 17 minus lowest possible duration of all parallel paths = 13). To solve the problem further tasks after E have to be looked at because between S and E only 4 units of time can be gained by decreasing task durations to their lowest modeled value.

5.2 Increasing spreads

The spreads of the resulting fuzzy number increases steadily during the calculation. The left spread may decrease during the unification of parallel paths due to the maximum operation (see the example in figure 5, the left spread decreases from 17-10=7 to 17-13=4 at node E). But this influence is minor compared to the many additions, so generally and as a trend the spreads will increase (see figure 6).

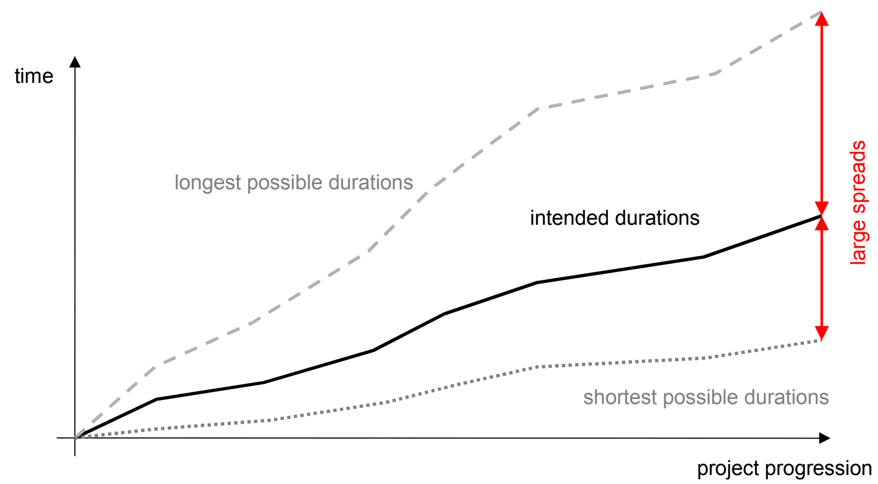


Figure 6: progression of the project over time, increasing spreads

The problem with this effect is the possible loss of the content of the conclusion. One possibility to handle this problem is to split the schedule into several parts, each of which has a much smaller spread for itself. How many parts are needed or are useful is a matter of the specific project analyzed. Some possibilities are segmentations at milestone (shown in figure 7), after a fixed amount of time like one month or at dates important for the internal management accounting of a company.

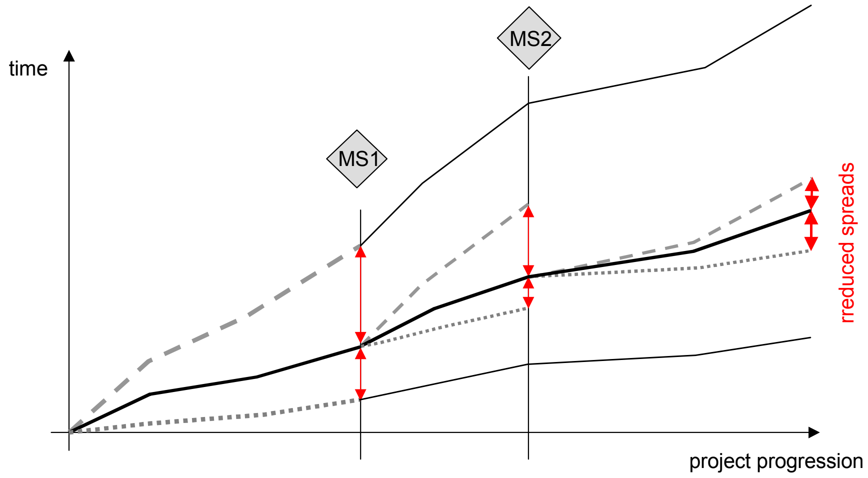


Figure 7: reduced spreads in single segments of the schedule

When splitting the schedule for each segment the included tasks have to be identified. Some tasks will be part of more than one segment. For this case methods for splitting the task itself have been specified. The single prognoses are freely relocatable in the borders of the previous segment, meaning that the prognosis for the second segment shown might as well start at the points resulting from the shortest or longest possible durations of segment one. The spreads will be different in the single segments, indicating segments with bigger and smaller uncertainty. This information can be used for a lot of interesting analyses, for example to decrease the spreads of single segments by finding alternative sequences of the tasks or to do a risk analysis to find especially sensible segments.

During the progression of the project information about the real durations of the tasks performed become available and can be included into the model. This way nominal/actual value comparisons can be done and reactions to evolving problems can be found. Since the single task durations are modeled with a span it will be not necessary to change the model in case of small deviations from the intended durations. So, in many cases the model does not have to be modified in spite of including the real information from the project progression, the need to change the model is reduced. By including the real duration of the project the spreads automatically reduce to zero at the actual point of progression and with that decrease for all future prognoses. With this reduction of the spreads the prognoses become more concrete and reliable without doing anything but registering the real project progression.

The model allows to evaluate the state of the project at every point of time. In case of arising problems like a delay in project progression the model will be able to indicate if the intended project end date can be reached by just consuming buffers (like in classic network plans) or by reducing the duration of tasks to their shortest possible value (not included in classic models) or if more drastic changes like re-arranging tasks are needed. In case of the need to accelerate tasks the model shows the potential of shortening for each task.

6 Conclusion

The model introduced is an extension of classic network models. It considers uncertainty of building projects. The modeling paradigms and rules allow to define a set of path algebras for many different purposes. It is also possible to use many different values for the edges of the schedule, for example triangular or polygonal fuzzy numbers, and to define many different rules for the unification and concatenation of the edges, shown exemplarily in form of the Max and Max_{Δ} operations.

The use of triangular fuzzy numbers to model the durations of single tasks reflects the reality better than just using single deterministic values. These numbers are easily manageable and complement the way of thinking in building industry. The resulting model includes a span for any calculated date due to the spreads of the fuzzy numbers. Due to this span the need to change the model on arising small deviations from the initial plan is reduced and partially eliminated. Of course drastic influences like major changes of the plans or long delays still lead to the need to change the model.

At the same time the increasing spreads represent a potential problem. A partial solution is shown in the form of splitting the schedule into smaller segments. The spreads of each segment are much smaller than the final spread and can be evaluated much better. They can be used for a risk analysis. For more detailed risk analyses the use of polygonal fuzzy numbers may be more appropriate than the use of the triangular numbers, the actual gradient of the membership degree functions may be of higher interest there. This can be done using the introduced model and defining a path algebra for polygonal fuzzy numbers which basically is an extension of the shown algebra for triangular numbers. Adding more complexity to the form of the membership degree functions may look promising in regards to the potential results. But there are strong limits to this approach as the experiences with the PERT method have shown.

When using triangular fuzzy numbers all dates and durations calculated exist in three forms. This also applies to the critical path. Generally different tasks belong to the three forms. The change from one critical path to another can be visualized easily. Another information gained is the potential of shortening for each path, limited by its own durations and by parallel paths.

One drawback of using fuzzy numbers is the increased effort to create such a model. For every single task and duration not only a single one duration has to be specified but now there are three values to input. That is another reason to use the relatively easy triangular fuzzy numbers compared to the more complex polygonal numbers. Furthermore triangular numbers result in a single one "intended" duration or date, this being the middle value of the number. This value is exactly the same value a classic deterministic calculation does result in. The triangular number gives two more values with the membership degree function moving linearly towards those values. The actual gradient does not matter for many applications, such as determining the earliest possible dates for events or the length of the critical path. It is sufficient to know there are two other values and that the "possibility" of all values between intended and shortest or intended and longest is decreasing in some way. The decrease expresses nothing but a deviation from the planned value.

The user of the model has to evaluate what this deviation means for him because only the user knows how to evaluate the input data. If the durations of the single tasks are modeled with large spreads then the final spreads will be very large as well. A small or medium deviation may not mean much then. If on the other hand the single durations have very small spreads then the final spread will not grow very large and small deviations are of higher interest.

Finally, the use of the model is limited to large projects or projects with large complexity. The increased effort to create the model is not justified and in very most cases not needed in smaller projects.

7 References

Böhme, G. (1993). Fuzzy-Logik, Einführung in die algebraischen und logischen Grundlagen. Springer-Verlag Berlin Heidelberg. Germany.

Freundt, M. (1998). Generische Bestimmung von Wegen in Netzen. Diplomarbeit at the TU-Berlin. Germany.

Klir, G. J. and Wierman, M. J. (1999). Uncertainty-Based Information. Physica-Verlag Heidelberg. Germany.

Mühlögger, J. (1994). Netzplantechnik mit unscharfen Methoden. Ph. D. Dissertation. University of Innsbruck. Austria.

Pahl, P. J. (2000). Mathematische Grundlagen der Ingenieurinformatik. Springer Verlag Berlin Heidelberg. Germany.