Project Management Method Selection using Bayesian Networks: a Novel Approach

Carlijn Kokkeler
University of Twente
The Netherlands

ABSTRACT

Selecting an appropriate project management method is important for achieving project success. There are two main project management methods: traditional and agile approaches. Traditional approaches are characterised by their adherence to a plan, established at the start of a project. In contrast, agile approaches emphasise adaptability and flexibility. The appropriate method for a project is dependent on project characteristics. Selecting a suitable method is a difficult task, so a decision model is desired. Considering that not much quantitative data is available, such a model could be constructed using Bayesian network modelling. Bayesian networks are probabilistic graphical models that express uncertain relationships between variables. These models can often be built in the absence of data. The purpose of this research is to investigate whether such a model for project management method selection can be built. It was found to be possible to construct a Bayesian network for selecting the appropriate approach, even without the availability of quantitative data.

Keywords

Project management methods, Bayesian network, project decision-making.

1. INTRODUCTION

Project management (PM) is concerned with the planning and control of a project, with the goal of achieving project objectives [30]. The selection of an appropriate project management method (PMM) plays an important role in achieving desired project results. Many studies have confirmed that alignment between a particular project and the chosen PMM is essential for project success [3, 5, 30]. Moreover, it has been argued that an inappropriate choice of PMM has been a critical factor in project failure [34].

Regarding the choice of PMM, there are two main streams: traditional and agile approaches. Typically one of these two methodologies is chosen for managing a project [3, 33]. Traditional methodologies are plan-driven, ensure that all requirements are established at the start of the project, and typically follow a sequential design process [12]. In contrast, agile methodologies are characterised by their flexibility and continuous adaptations to clients' requirements [12, 35].

Although some argue that agile methodologies are superior to traditional methodologies [2], the most appropriate PMM for a project is dependent on project characteristics [36]. This is confirmed by project contingency theory, which states that the best PMM depends on the context [19].

Selecting the appropriate method has been identified as a major challenge [41], and thus it seems that a decision model is desired. However, such decision models are very scarce [2, 19, 41]. In [41], a simple decision model is constructed, but this model is mainly informative, without yielding an immediate suggestion regarding the preferable PMM. Moreover, this model does not show relationships between chosen variables.

Since there is little quantitative data available to construct such decision models [41], Bayesian network modelling could show to be useful. Bayesian networks are probabilistic graphical models that allow for reasoning under uncertainty [7, 20]. They consist of nodes, representing random variables, and edges between nodes, representing probabilistic dependencies. They can be built from incomplete data [16, 24, 27, 29, 39], for example by exploiting expert knowledge [24, 27, 29]. Bayesian networks are easily understandable by humans due to their semantic clarity [24]. Moreover, relationships between variables are represented explicitly.

In the domain of PM, some Bayesian networks have been constructed regarding project cost, risk, and benefit, such as in [45]. However, no Bayesian networks have been constructed to serve as a decision model for selecting the appropriate PMM. This yields the following research question:

How to build a Bayesian network for selecting the most appropriate project management approach in the absence of data?

Although it is stated that Bayesian networks can be built with missing data [16, 24, 27, 29, 39], this remains a challenge, since there is a risk of having too little information to construct a valid model. Therefore, part of the research will consist of exploring whether it is even possible to develop such a network. The PM approaches that are considered are the ones mentioned earlier; agile and traditional methodologies. To answer the research question, the following sub-questions can be defined:

1: What are the characteristics, benefits, and constraints of agile and traditional PMMs?

2: What are the qualitative relationships between the chosen variables?

28th Twente Student Conference on IT Febr. 2nd, 2018, Enschede, The Netherlands.

3: What probabilities can be assigned to the relationships between the chosen variables?

2. REQUIREMENTS

Several requirements for the Bayesian network can be formulated. First of all, the decision model should yield an immediate suggestion regarding the appropriate choice of PMM. Secondly, the model should clearly show the relationships between variables, such that it is easily understandable. Lastly, the relationships between variables presented in the model should be consistent with existing literature.

3. RELATED WORK

The importance of having a decision model for selecting the appropriate PMM is emphasised in [2, 19, 41]. In [41], it is acknowledged that no such model exists yet.

In [2], several critical success factors (CSFs) for software development projects are identified, and these are compared between agile and traditional approaches. These CSFs can be used to select variables for the Bayesian network. In [3], the same authors compared the differences in CSFs between agile and traditional approaches empirically. This will be used for deciding on the most significant variables for the construction of the network.

In contrast to the requirements as given in section 2, in the decision model constructed in [41], relationships between variables are not shown, and no immediate suggestion regarding the appropriate methodology is yielded. Nevertheless, the criteria defined in the paper can be used for the determination of variables for the network, as well as to compare these criteria to the factors found in [2].

In [46], the lack of knowledge about the difference in performance factors between agile and traditional firms is emphasised, and these factors are investigated relating to human resource management (HRM). Although HRM is not the focus field, some factors may appear to be useful.

Strategies and challenges regarding agile and traditional PMMs are defined in [12]. While it is stated in the paper that the examination of success factors regarding the introduction of agile methodologies into the traditional field would be helpful for future research, the strategies and challenges mentioned in the paper could already be useful for the construction and support of the Bayesian network.

Lastly, in [19, 22, 28, 35], several success factors regarding PM, but not specifically relating to the comparison between agile and traditional PM, are stated. This could be helpful for the formulation of the variables.

Regarding Bayesian networks, there are several papers in which these models are constructed, such as in [9, 15, 21, 45]. These papers will be used as examples for the design procedure of the network. Moreover, in [26], a design principle for the construction of Bayesian networks is detailed. This design principle will be followed, and is explained in the next section.

4. METHODOLOGY

4.1 Notation and Representation

A *Bayesian network* (BN) is a graphical model that represents both qualitative and quantitative information about probability distributions. The qualitative information is given by a graph. Following the structure of the graph, numerical probabilities can be assigned, such that the quantitative information is represented.

To represent *random variables* in the network, upper case letters, such as X, will be used. The *domain* of a random variable X, which indicates the set of values that X may have [13], is represented by $\text{dom}(X)$. Random variables may be either *discrete* or *continuous*. In this thesis, all random variables are discrete, meaning that their domain is finite or countably infinite. Moreover, the domains are ordered, meaning that, if $\text{dom}(X) = \{a, b, c, d\}$, then $a < b < c < d$.

To denote the *probability distribution* of a random variable X, the notation $P(X)$ will be used. The probability of a certain value $x \in \text{dom}(X)$ is given by $P(X = x)$, or, when it causes no confusion, $P(x)$. The notation for a set of random variables is a bold face letter, such as $\mathbf{X} = \{X_1, \ldots, X_n\}$. The probability distribution of such a set of random variables is denoted by $P(X_1, \ldots, X_n)$ or $P(\mathbf{X})$, and is called a *multivariate* or *joint* probability distribution. Given a joint probability distribution, *any* other (conditional) probability distribution can be computed by combining summation (also called *marginalisation*):

$$P(\mathbf{Y}) = \sum_{\mathbf{Z}:\mathbf{X}|\mathbf{Y}} P(\mathbf{Y},\mathbf{Z})$$

with disjoint sets $\mathbf{Y}$ and $\mathbf{Z}$, $\mathbf{X} = \mathbf{Y} \cup \mathbf{Z}$. and the definition of *conditional probability distributions*:

$$P(\mathbf{Y} \mid \mathbf{Z}) = \frac{P(\mathbf{Y},\mathbf{Z})}{P(\mathbf{Z})}$$

with $P(\mathbf{Z} = \mathbf{z}) > 0$ for any value of $\mathbf{z}$.

A Bayesian network consists of a joint probability distribution P and an associated *graph* $\mathcal{G}$, that is defined as a pair $\mathcal{G} = (\mathbf{V}, \mathbf{A})$, where $\mathbf{V}$ is a set of objects $i \in \{1, \ldots, n\}$, $n = |\mathbf{V}|$, called *nodes*, and $\mathbf{A} \subseteq \mathbf{V} \times \mathbf{V}$ is a set of node pairs called *edges*. If $\mathcal{G}$ is a *directed graph*, then each edge of A is an ordered pair (i, j), also represented by $i \to j$, such that $(j, i) \notin \mathbf{A}$. The edges are then called *directed edges* or *arcs*. A certain node i is called the *parent* of a certain node j if $i \to j \in \mathbf{A}$ is an arc, and in that case j is called the *child* of node i.

In this thesis, $\mathcal{G}$ is a directed acyclic graph (DAG). This is a directed graph that contains no directed cycles, i.e., there is no sequence of arcs of the form $i \to j \to \cdots \to i$ (first and last node in the sequence are the same). As usual, each node i in the DAG with $V = \{1, \ldots, n\}$ will be associated in a one-to-one way to a random variable X_i from the set of variables $X_1, \ldots, X_n$. In the following, nodes and variables will be referred to interchangeably, and X_i will be used to refer to both the node and the variable.

One way to define Bayesian networks is from the notion of factorising a joint probability distribution according to the structure of a graph as follows.

DEFINITION 1 (FACTORIZATION). *Let $\mathcal{G} = (\mathbf{V}, \mathbf{A})$ be a directed acyclic graph with nodes $\mathbf{V} = \{X_1, \ldots, X_n\}$. A joint probability distribution P over the same variables* factorises *according to $\mathcal{G}$ if P can be written as:*

$$P(\mathbf{X}) = P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i \mid \pi(X_i))$$

where $\pi(X_i)$ refers to the parents of the node X_i in $\mathcal{G}$, and each factor $P(X_i \mid \pi(X_i))$ is called a conditional probability table *(CPT, for short).*

DEFINITION 2 (BAYESIAN NETWORK). *A* Bayesian network *is a pair $\mathcal{B} = (\mathcal{G}, P)$, where P is a joint probability* *distribution that factorises according to a directed acyclic graph $\mathcal{G}$.*

4.2 Design Principle

The stages in the knowledge-driven development of a Bayesian network are presented in Figure 1 [26].

In the first step, a causal graph will need to be established. For this, critical variables and relationships have to be identified by using existing literature on the topics of agile PM and traditional PM. The relationships in the graph are not necessarily causal, since a causality notation is not per se embodied in the semantics of a Bayesian network. The variables and relationships will be determined by taking into account the characteristics, benefits, and constraints of both methodologies. Moreover, to determine the relations with project success, it is important that the definition of project success is discussed first.

The domains of the variables have to be indicated using qualitative order words, to show the relationships between the variables. Thereafter, the domains will have to be quantified, using literature and logic. Since not much quantitative data can be found about the relationships between variables regarding agile PM and traditional PM, it is highly likely that the exact needed information for the quantification will not be available. Therefore, it is important that substantial literature from experts regarding the relationships between the determined variables is found, and perhaps some numerical assessments for certain variables. This literature can then be compared and used to establish the quantitative relationships.

When the quantitative network is established, the network has to be evaluated. This can be achieved by comparing the results from the model to results from the literature with respect to the desired PMM.

In this research, Genie will be used for the development of the model [6], but several other tools with the same functionality are available for constructing a Bayesian network.

In this thesis, only one Bayesian network will be developed. Although it may be possible that more Bayesian networks are suitable for selecting the appropriate PMM, the model developed in this research should cover the most critical variables. Moreover, relationships between variables will be supported by literature. Therefore, it seems sufficient to construct one network, that could be developed further for future research.

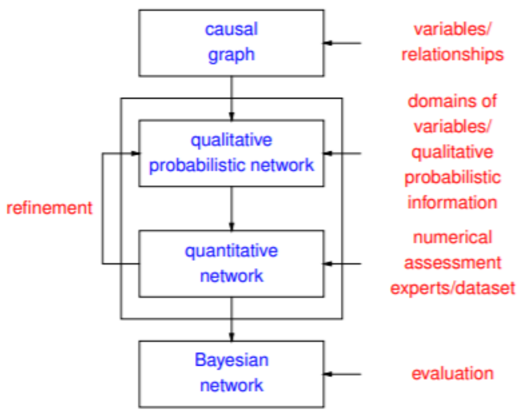

Figure 1. Design of a Bayesian network (diagram taken from [26]).

5. DEFINITIONS, VARIABLES, AND DOMAINS

5.1 Project Success

While the choice of an appropriate PMM contributes to project success [3, 5, 30], there is no direct relationship between the two [30]. First of all, there is a distinction between PM success and project success [1, 30, 32]. This distinction, as well as the definition of project success that will be used in this research, is discussed below. Secondly, although the choice of PMM plays a significant role in the overall PM performance, there are other factors influencing PM as well, such as PM staff and PM leadership [28]. However, to avoid too much ambiguity and complexity, these factors will be disregarded in the determination of the relationship between the PMM and the overall PM performance.

The main difference between PM success and project success is related to the expected total life-span of a project. Specifically, PM success generally concerns the achievement of short-term goals that can be evaluated immediately after completion of the project, while project success relates to the evaluation of the higher, long-term goals of the project overall [1, 30, 32].

Although successful PM contributes to project success, project failure is not prevented by it [30]. Moreover, a project may be successful, despite its PM being unsuccessful [32]. Thus, in this research, the concept of project success from [2] will be followed, where PM success may be part of project success, but is not equal to it. In [2], project success is composed of process success and product success. Process success in this case is measured against budget, time, and scope criteria, and product success is related to product outcomes, such as the overall quality of the produced outcome, and the user satisfaction.

5.2 Critical Success Factors

CSFs are elements that, if addressed appropriately, contribute to the likelihood of achieving project success criteria [1, 3]. For the construction of the Bayesian network, some of the CSFs defined in [2] will be used as variables in the network. To determine which variables are suitable for the network, it is appropriate to assess them based on criteria. In this paper, the following criteria will be evaluated: (C.1.1) the percentage of papers on project success in which the CSF is mentioned, as investigated in [2], should be larger than 70%, (C.1.2) if the influence of the CSF is investigated in [3], that influence is found to be significant, and (C.1.3) a domain can be indicated for the CSF at the start of the project.

Some additional CSFs, that are not identified in [2] may be formulated, if their importance is evident from other literature. In that case, the variables conform to criterion (C.2) importance evident from literature other than [2].

5.3 Variables

Factors relating to the project scope, such as technological uncertainty, are most important to consider when selecting the appropriate PMM [3]. Based on the decision variables in [41] and the CSFs in [2], the most relevant variables, relating to the project scope, can be formulated as (i) technological uncertainty (TU), (ii) technical complexity (TC), (iii) level of specification (LS), and (iv) project team's expertise with the task (PTE).

Variables (i) and (ii) meet criteria (C.1.1), (C.1.2), and (C.1.3). Variables (iii) and (iv) meet criteria (C.1.1) and (C.1.3). These variables are not specifically investigated in [3], so criterion (C.1.2) is not applicable.

In agile methodologies, the role of the customer is critical [46], because requirements set for the project are subject to change depending on feedback of the customer [35]. In contrast, with traditional methodologies, project requirements are not to be adapted during the implementation process [12]. Therefore, the role of the customer can be considered less critical for traditional methodologies [46]. The following variables related to the customer can be defined: (v) communication with the customer (CC), (vi) knowledgeability of the customer (KC), and (vii) importance of the role of the customer (IRC).

The CSF 'user participation' from [2] meets criteria (C.1.1) and (C.1.2), but fails to meet criterion (C.1.3) due to its ambiguous formulation. Thus, this CSF has been partitioned into variables (v), (vi), and (vii), such that criterion (C.1.3) is met. Moreover, the variables meet criterion (C.2).

Regarding the project team, the size of the project team plays a role in the determination of the suitability of a PMM [41]. The size of the project team is not given as a CSF in [2], but the importance of it relating to agile methodologies is confirmed in [47], and thus, (viii) project team size (PTS), is an appropriate variable for the network, meeting criterion (C.2).

Lastly, the two variables relating to project success have to be defined, as well as the variable for the preferred choice of PMM. These are the following: (ix) project success with a traditional PMM (PST), (x) project success with an agile PMM (PSA), and (xi) choice of PMM (C).

5.4 Domains

The variables and their domains can be found in Table 1, and these are explained in this section.

Table 1. Variables with corresponding domains.

Variable (X)	Domain (dom(X))
TU	$\{low, medium, high, superhigh\}$
TC	$\{assembly, system, array\}$
LS	$\{nf, nff\}$
PTE	$\{<2, 2-5, 5-10, >10\}$
CC	$\{>4, 2-4\}$
KC	$\{associate, bachelor, master, doctoral\}$
IRC	$\{low, moderate, high\}$
PTS	$\{7-9, 10-30\}$
PST	$\{failure, success\}$
PSA	$\{failure, success\}$
C	$\{agile, traditional\}$

5.4.1 Technological Uncertainty (TU)

To indicate the levels of technological uncertainty, the dimensions and definitions of technological uncertainty from [37] will be used. This results in the following categorisation: (i) low; implementing familiar technologies, (ii) medium; adaptations of familiar technologies, (iii) high; first use of new technologies, and (iv) super high; development of new technologies.

5.4.2 Technical Complexity (TC)

The dimensions of technical complexity and the corresponding definitions from [37] will be used to indicate the domains of technical complexity: (i) assembly; building a single component or a collection of components for a single unit, (ii) system; building a complex collection of interactive elements and subsystems for a single product, and (iii) array; "building a large, widely dispersed collection of different systems that function together to achieve a common purpose" [37], p.611.

5.4.3 Level of Specification (LS)

The level of specification can be measured in the level of requirements. Requirements specify the facts that must be accomplished by a system or application [18]. They can be categorised into (i) non-functional requirements, which are the basic conditions for a product or system, and (ii) functional requirements, which are higher-level requirements that may be formulated after the non-functional requirements are defined [11]. The following classification will be used to indicate the level of specification: (i) only non-functional requirements (nf), and (ii) both non-functional requirements and functional requirements (nff).

5.4.4 Project Team's Expertise with the Task (PTE)

The level of expertise with the task can be measured according to the years of experience of the team members. In [3], experience is categorised into the following: (i) <2 years, (ii) 2-5 years, (iii) 5-10 years, and (iv) >10 years. This categorisation will be followed.

5.4.5 Communication with the Customer (CC)

In agile practices, communication with the customer is an important aspect [40]. Usually, projects with an agile PMM are divided into two to three week cycles [14], after which a meeting with the customer is scheduled. In some cases, the duration of these cycles may be four weeks [10]. Thus, it can be concluded that for agile PM, the frequency of communication with the customer is every two to four weeks.

In traditional practices, communication with the customer is of less importance, because all requirements for the project are established at the start of the project [12]. Thus, although frequent communication with the customer is recommended [44], the frequency of communication with the customer in a project with traditional PMM may exceed four weeks.

Thus, the following classification will be used: (i) every 2-4 weeks, and (ii) $>$ every 4 weeks.

5.4.6 Knowledgeability of the Customer (KC)

The knowledgeability of the customer can be measured with respect to the academic level of the customer regarding the topic of the project. In general, there are four levels of degrees, from less advanced to more advanced: (i) associate degrees, (ii) bachelor's degrees, (iii) master's degrees, (iv) doctoral degrees.

5.4.7 Importance of the Role of the Customer (IRC)

The importance of the role of the customer concerns the amount of input required from the customer. In [8], the level of customer participation across different services is categorised. This categorisation will be followed, but adapted in such a way that the categories may relate to all types of project, instead of merely relating to service projects. This results in the following: (i) low; customer presence required during feedback moments, (ii) moderate; customer inputs required during creation, (iii) high; customer co-creates.

5.4.8 Project Team Size (PTS)

The exact recommended project team size for agile methodologies is 7-9 individuals [23, 31]. The reason for this is that, for agile methodologies, team efficiency is at its peak with this number of people [23]. Medium-sized teams (11-

30 individuals) are most prevalent with traditional approaches [42]. Also according to [41], a traditional approach is more appropriate when the number of individuals in a project team is larger than 10. Thus, the classification used is the following: (i) 7-9 individuals, and (ii) 10-30 individuals.

5.4.9 Project Success with a Traditional PMM (PST)

The value of project success with a traditional PMM can be either (i) success, or (ii) failure. The definition of project success as defined in section 5.1 will be regarded.

5.4.10 Project Success with an Agile PMM (PSA)

The value of project success with an agile PMM can be either (i) success, or (ii) failure. The definition of project success as defined in section 5.1 will be regarded.

5.4.11 Choice of PMM (C)

The choice of PMM can be either (i) agile, or (ii) traditional. The PMM that yields the highest percentage of project success will be the recommended choice of PMM.

6. BAYESIAN NETWORK

6.1 Qualitative Probabilistic Network

In a *qualitative probabilistic network*, signs, instead of conditional probabilities, capture the probabilistic influences and synergies between variables [43]. A *qualitative probabilistic influence* between two variables denotes how the values of one variable influence the probabilities of the values of the other variable. In this thesis, $X\uparrow \rightarrow Y\uparrow$ indicates that observing a higher value for X makes a higher value for Y more likely, regardless of any other direct influences on Y, where X and Y are variables in the network. Thus, $P(y|x, z) \geq P(y|\bar{x}, z)$, for any combination of values z, where z is a direct influence on Y, other than X, and $\bar{x}$ is the negation of x. Other influences are defined analogously, e.g. for $X\uparrow \rightarrow Y\downarrow$, the $\geq$ in the above formula can be replaced by $\leq$. The general notation for a relation between X and Y is $X \rightarrow Y$.

In Figure 2, the relations between variables in the network are depicted. These relations are explained below.

First of all, although TC $\rightarrow$ TU is not necessarily causal [37], it is likely that TC$\uparrow$ $\rightarrow$ TU$\uparrow$. The reason for this, is that when more complex systems have to be built, there is a higher probability that technologies have to be implemented that are not familiar. Some studies even consider TC and TU under the same variable [3]. Moreover, in some studies TC is considered to be part of TU [17]. Therefore, TC $\rightarrow$ TU is depicted in the Bayesian network.

Moreover, PTE $\rightarrow$ TU. The reason for this relation, is that it is probable that technologies are more familiar (TU$\downarrow$) when the person implementing those technologies is more experienced in the field (PTE $\uparrow$). This relation is also hypothesised in [3].

Regarding LS, the relation TU $\rightarrow$ LS is depicted. TU $\rightarrow$ LS in that it is likely that TU$\uparrow$ $\rightarrow$ LS$\downarrow$, since it is more difficult to establish requirements for unfamiliar technologies. In [38] it is also suggested that for high-tech projects, requirements must be specified during the project, such that advantage can be taken from the knowledge gained during the process.

With regard to CC, the relations KC $\rightarrow$ CC and IRC $\rightarrow$ CC are given. When the customer has more knowledge relating to the project (KC$\uparrow$), it is likely that more frequent communication with the customer is needed (CC$\uparrow$), so that this knowledge can be obtained when needed during the process. Furthermore, if the role of the customer is of great importance (IRC$\uparrow$), it is likely that more frequent communication with the customer (CC$\uparrow$) is necessary to obtain the required input. Also, KC $\rightarrow$ IRC, since it can be expected that the role of the customer is more important (IRC$\uparrow$) when the customer is more knowledgeable with respect to the project (KC$\uparrow$).

The variables concerning project success (PSA, PST) are directly related to the variables LS, PTS, and CC. PSA$\uparrow$ when LS$\downarrow$, such that requirements can be established and adapted during the process [3, 12, 35, 41]. The reason for this, is that agile methodologies are especially developed for projects where requirements are unknown or uncertain at the start of the project [3, 12, 35]. In contrast, projects where a traditional PMM is followed require an early specification freeze, meaning that the level of specification should be high [3, 12, 35]. Thus, LS$\uparrow$ $\rightarrow$ PST$\uparrow$.

Project success with an agile approach is higher (PSA$\uparrow$) when the project team size is small (PTS$\downarrow$), while a larger project team size (PTS$\uparrow$), up to 30 people, is preferable for traditional approaches (PST$\uparrow$) [4, 28, 41]. For agile methods, communication and coordination are of great importance, and these aspects become more difficult when the team size increases [25]. Moreover, documentation is likely to be more prevalent when the team size is larger, and documentation is an important part of traditional PM [4].

In agile practices, much emphasis is placed on communication with the customer [3, 35, 41]. In contrast, traditional approaches usually involve low customer interaction [35], and do not require full-time customer involvement [3]. Thus, in the case that communication with the customer is low (CC$\downarrow$), a traditional PMM can better be chosen (PST$\uparrow$) [3]. On the contrary, since agile methodologies allow for much communication with the customer, it can be stated that, when communication with the customer should be more frequent (CC$\uparrow$), project success with an agile PMM is likely to be higher (PSA$\uparrow$).

Lastly, C depends on both PSA and PST. The PMM that, according to the model, yields the greatest project success, will be the recommended choice.

6.2 Quantitative Probabilistic Network

Although Bayesian networks can be built from incomplete data [16, 24, 27, 29, 39], it can be difficult to establish a quantitative model without training data and test data to validate it. For the Bayesian network described in this paper, no data is available that can be used immediately. However, relations between variables in the model have been specified in the literature, as described in section 6.1. For the construction of the quantitative network, probabilities are assigned to these relations. Since no probabilities regarding these relations are known, these numbers are estimated. Because the PMMs that are investigated in this study differ quite significantly, these estimations seem to be sufficient for yielding the most appropriate approach in most cases. Nevertheless, for future research, it would be important to obtain quantitative data to make the model more precise and to validate the proposed relations further. An initial validation of the network is given in the next section.

The quantitative probabilistic network is given in Figure 2. For all nodes, evidence can be set for a certain domain, meaning that a variable is assigned a certain domain value. The bold domains indicate the evidence that is set, and

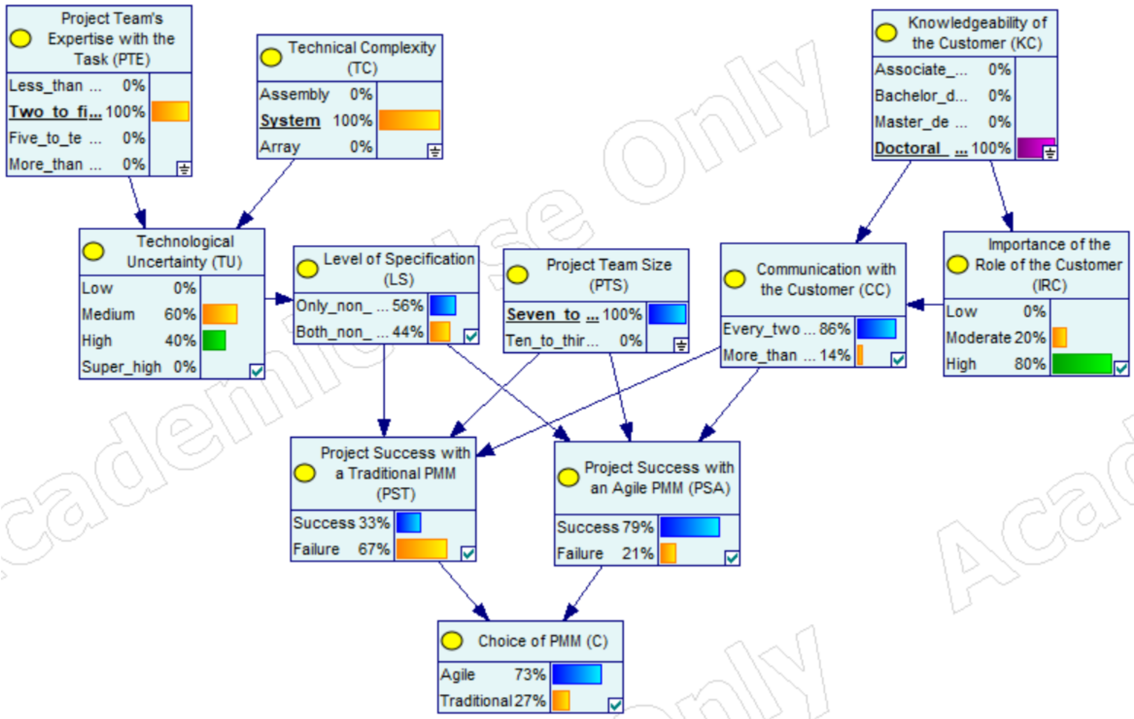

Figure 2. Bayesian network for determining the appropriate choice of PMM, given by the highest percentage for C.

the percentages indicate the probabilities for the domains, which are obtained after running the model with the given evidence.

7. VALIDATION

7.1 Results by Domain

In the network, it is possible to specify the domains of all nodes. However, when the value for a certain node X_i is set, parent nodes of X_i have no influence on child nodes of X_i. For example, when LS is set to nf, the values for TU, PTE, and TC, are not relevant for the choice of PMM. Nevertheless, to validate that the relations described in section 6.1 are satisfied in the network, it is useful to test all nodes.

For the validation, the nodes that lack a parent node have a uniform probability distribution, e.g., there are four domain values for PTE, which all get assigned a probability of 0.25. This way, it is possible to analyse the influence of all nodes and their domains.

The probability distribution when certain evidence is given is denoted by P_e, where $P_e(x) = P(x|e)$. In Table 2, all variables are shown, with all possible values of their domains. For every possible domain value, $P_e(C = agile)$ and $P_e(C = traditional)$ are given. Thus, e.g. for TU with TU $= low$, the following is given: $P_e(agile) = P(C = agile|\text{TU} = low) = 0.40$, and $P_e(traditional) = P(C = traditional|\text{TU} = low) = 0.60$.

The results from Table 2 were obtained indirectly, through estimated probabilities that were assigned to each direct relationship between variables. The results were obtained by running the model in Genie. All probabilities satisfy the relations implied in section 6.1. Moreover, from the survey results obtained from project managers, that were found in [3], it is evident that (i) PTE$\downarrow \rightarrow P_e(C = agile)\uparrow$, (ii) TC$\uparrow \rightarrow P_e(C = agile)\uparrow$, (iii) TU$\uparrow \rightarrow P_e(C = agile)\uparrow$, (iv) LS$\downarrow \rightarrow P_e(C = agile)\uparrow$, (v) PTS$\uparrow \rightarrow P_e(C = traditional)\uparrow$, (vi) CC$\downarrow \rightarrow P_e(C = traditional)\uparrow$, and (vii) KC$\downarrow \rightarrow P_e(C = traditional)\uparrow$. All these relations are satisfied. The variable IRC is not investigated in [3]. However, from [46], it is clear that IRC$\uparrow \rightarrow P_e(C = agile)\uparrow$, which is satisfied.

Table 2. Obtained probabilities for each domain value.

Variable	Value	$P_e(agile)$	$P_e(traditional)$
PTE	$<$ 2 years	0.54	0.46
PTE	2-5 years	0.53	0.47
PTE	5-10 years	0.47	0.53
PTE	$>$ 10 years	0.45	0.55
TC	Assembly	0.42	0.58
TC	System	0.48	0.52
TC	Array	0.59	0.41
TU	Low	0.39	0.61
TU	Medium	0.47	0.53
TU	High	0.60	0.40
TU	Super High	0.64	0.36
LS	nf	0.68	0.32
LS	nff	0.32	0.68
PTS	7-9	0.61	0.39
PTS	10-30	0.38	0.62
CC	$>$ Every 4	0.41	0.59
CC	Every 2-4	0.58	0.42
IRC	Low	0.43	0.57
IRC	Moderate	0.50	0.50
IRC	High	0.56	0.44
KC	Associate	0.43	0.57
KC	Bachelor	0.48	0.52
KC	Master	0.52	0.48
KC	Doctoral	0.56	0.44

As can be seen, the probabilities for either an agile approach or a traditional approach are quite close. The reason for this is that the values of the other nodes are not set. When the values of several nodes are known, the network will usually yield more distinct results. Nodes that are closer to C show a greater difference, since there is less interference with other variables.

In contrast to [41], this decision model shows the relations between variables, and yields an immediate suggestion regarding the desirable PMM. In some cases, both PMMs seem to be equally suitable. However, in most cases there is a clear preference for one of the two methodologies.

7.2 Sensitivity Analysis

For further validation of the model, a sensitivity analysis was performed. Such an analysis shows which nodes contain parameters that are important when calculating the posterior probability distributions for a certain target node. The sensitivity analysis was performed three times, for three different target nodes. The figures given in this section were obtained from the sensitivity analysis option in Genie. Lightly coloured nodes are less sensitive for the target node, while intensely coloured nodes are more sensitive for the target node. The figures serve to show the structure of the network with the sensitivity of the nodes. For a more detailed figure, where all variables and their domains are clear, Figure 2 is given.

First of all, C was set as target node, so that it was possible to analyse the sensitivity of all nodes on the final decision for the appropriate PMM. As can be seen in Figure 3, three nodes, apart from the target node (C), are coloured more intensely. These three nodes are TC, LS, and PTS, which are related to the project scope. Since variables relating to the project scope were found to be most important to consider when selecting an appropriate PMM [3], this sensitivity is as to be expected. From the factors relating to project scope, TU has the lightest colour. The reason for this, is that TU contains the largest number of parameters, meaning that a change in one of these parameters does not yield a great change in C.

Secondly, PST was set as target node, such that it could be analysed which nodes were most sensitive specifically for project success with a traditional PMM. From Figure 4 it is clear that these factors are also related to the project scope. More specifically, the most intensely coloured nodes are PTE, TC, LS, and PTS.

Lastly, to analyse which nodes are most sensitive specifically for project success with an agile PMM, PSA was set as target node. From Figure 5, it can be concluded that factors related to the customer are more important in this case as compared to the previous cases. The most intensely coloured nodes are PTS, CC, and KC. The reason for customer factors being of greater importance, is that the role of the customer is critical for agile methodologies, while for traditional methodologies this role is less critical [46].

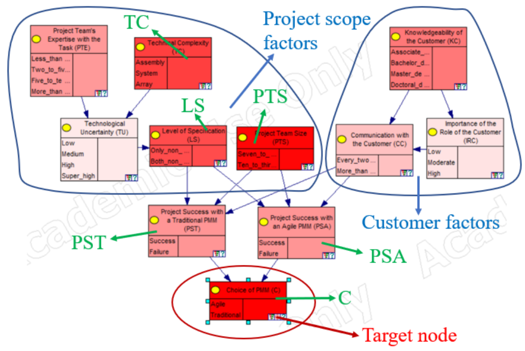

Figure 3. Sensitivity analysis with target node C.

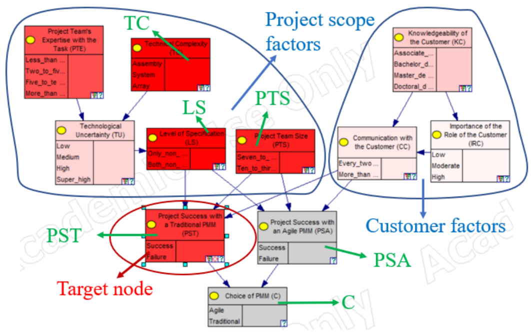

Figure 4. Sensitivity analysis with target node PST.

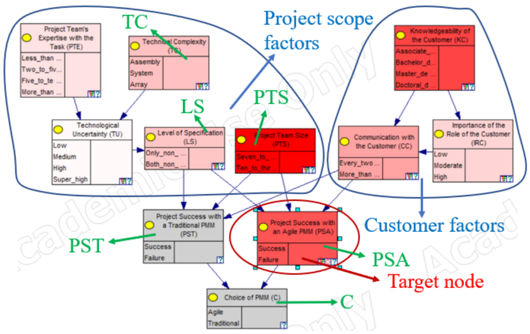

Figure 5. Sensitivity analysis with target node PSA.

7.3 Scenario Analysis

For a more intuitive interpretation of the applicability of the model, it seems useful to design some scenarios in which it is necessary to select a PMM, and to analyse the suitability of the proposed PMM.

7.3.1 Example: Assembling a Radio

The first example project concerns the assembling of a radio. In that case, TC will have the value *assembly* [37]. Since many radios have already been built, it seems likely that people working on the project have experience with the task, so the value $5-10$ will be assigned to PTE. PTS may vary depending on the type of radio to be built, but it seems likely that not many people will be needed, since the project is not very extensive. Therefore, $7-9$ seems suitable for PTS. Regarding KC, let us assume that the customer has very little knowledge about the components of a radio, and thus KC $= associate$.

After running the model, the following values are obtained: $P_e(agile) = 0.42$, and $P_e(traditional) = 0.58$. Thus, it is clear that even though PTS for this project is perfect for an agile PMM, the other variables are more appropriate for a traditional PMM, yielding a traditional PMM suggestion. This seems appropriate, since building a radio is a known task, where it is possible to establish a detailed plan at the beginning of the project. Moreover, it is expected that, unless the design of the radio is extremely important, not much communication with the customer is required.

7.3.2 Example: Designing a Combat Aircraft

The second example concerns the design of a completely new combat aircraft. Such a project can be classified as an *array* project [37], and, since the new aircraft should be very innovative, PTE $= <2\,years$. In this case, it is also possible to immediately set the value of TU to *superhigh*, since it is clear that the development of new technologies is needed. It is assumed that PTS = $7-9$, consisting of knowledgeable engineers, designing separate parts of the aircraft. Furthermore, the customer is specialised in the design of aircrafts, so KC = *doctoral*. Lastly, it is clear that IRC = *high*, since the customer would like to be very much involved in the project. The following results are obtained: $P_e(agile) = 0.84$, and $P_e(traditional) = 0.16$. Clearly, an agile approach should be taken, where the unknown requirements can be established during the project, and much communication with the customer is possible.

8. CONCLUSIONS

The purpose of this research was to construct a Bayesian network for selecting the appropriate PMM in the absence of data. Part of the research consisted of exploring whether it is even possible to develop such a model in this case. It has appeared that this is possible, since the network yields an immediate suggestion regarding the appropriate PMM, which is explainable using literature. However, it is preferable that more empirical data is available for the construction of the quantitative probabilistic model and the validation thereof.

A reason for the lack of data not being a significant problem, is that the PMMs that were analysed in this research are quite distinct. To answer the first research question, agile approaches are more applicable when there is uncertainty regarding the project scope, such as high technical complexity and little expertise concerning the task. Moreover, the team size is preferably small and communication with the customer is an important aspect. In contrast, traditional approaches are characterised by their fixed scope, where almost all requirements are established at the start of the project. This means that less communication with the customer is required. Furthermore, the preferred project team size is larger.

The second research question concerned the investigation of the relationships between chosen variables for the establishment of the qualitative probabilistic model. The variables concerning the level of specification, the project team size, and the communication with the customer are directly related to the variables concerning project success. The level of specification is influenced by the level of technological uncertainty, which itself is influenced by the project team's expertise with the task and the level of technical complexity. The level of communication with the customer is influenced by the knowledgeability of the customer and the importance of the role of the customer, where the importance of the role of the customer is also dependent on the knowledgeability of the customer.

After the establishment of the qualitative probabilistic model, probabilities were assigned to the relationships between the chosen variables, answering the third research question. Exact probabilities were not available from literature, so these were estimated. It is apparent that, for traditional approaches, the project scope should be clear, and factors relating to the customer are of less importance. Conversely, customer factors are of great importance for agile approaches, and the project scope is likely to be uncertain.

For further validation of the model it is important to obtain quantitative data. This data could be used to verify the relations between the variables presented in the model, and to refine the probabilities assigned to the relations. This could be achieved by analysing the values of the chosen variables among many projects. Moreover, variables other than those presented in this research, such as organisational culture and project criticality, could be investigated and empirically validated. This way, the model can be developed further.

9. ACKNOWLEDGEMENTS

I would like to thank my supervisor Peter Lucas for his valuable feedback and guidance.

10. REFERENCES

[1] W. Abdullah and A. Ramly. Does successful project management equates to project success. In *Proceedings of 5th IEEE International Conference of Cognitive Informatics*. Citeseer, 2006.

[2] A. Ahimbisibwe, R. Cavana, and U. Daellenbach. A contingency fit model of critical success factors for software development projects. *Journal of Enterprise Information Management*, 2015.

[3] A. Ahimbisibwe, U. Daellenbach, and R. Cavana. Empirical comparison of traditional plan-based and agile methodologies. *Journal of Enterprise Information Management*, 2017.

[4] M. Awad. A comparison between agile and traditional software development methodologies. *University of Western Australia*, 30, 2005.

[5] A. Badewi. The impact of project management (pm) and benefits management (bm) practices on project success: Towards developing a project benefits governance framework. *International Journal of Project Management*, 34(4):761–778, 2016.

[6] BayesFusion. https://www.bayesfusion.com.

[7] I. Ben-Gal. Bayesian networks. *Encyclopedia of statistics in quality and reliability*, 1, 2008.

[8] M. Bitner, W. Faranda, A. Hubbert, and V. Zeithaml. Customer contributions and roles in service delivery. *International journal of service industry management*, 1997.

[9] G. Büüyüközkan, G. Kayakutlu, and S. Karakadılar. Assessment of lean manufacturing effect on business performance using Bayesian belief networks. *Expert Systems with Applications*, 42(19):6539–6551, 2015.

[10] H. Cervone. Understanding agile project management methods using scrum. *OCLC Systems & Services: International digital library perspectives*, 2011.

[11] Z. Chen and Y. Zeng. Classification of product requirements based on product environment. *Concurrent Engineering*, 14(3):219–230, 2006.

[12] D. Ciric, B. Lalic, D. Gracanin, N. Tasic, M. Delic, and N. Medic. Agile vs. traditional approach in project management: Strategies, challenges and reasons to introduce agile. *Procedia Manufacturing*, 39:1407–1414, 2019.

[13] M. H. DeGroot, M. J. Schervish, X. Fang, L. Lu, and D. Li. *Probability and statistics*, volume 2. Addison-Wesley Reading, MA, 1986.

[14] M. Drury-Grogan and O. O'DWYER. An investigation of the decision-making process in agile teams. *International Journal of Information Technology & Decision Making*, 12(06):1097–1120, 2013.

[15] A. Ekici and S. Onsel. How ethical behavior of firms is influenced by the legal and political environments: A Bayesian causal map analysis based on stages of development. *Journal of Business Ethics*, 115(2):271–290, 2013.

[16] C. Fiot, G. Saptawati, A. Laurent, and M. Teisseire. Learning Bayesian network structure from incomplete data without any assumption. In *International Conference on Database Systems for Advanced Applications*, pages 408–423. Springer, 2008.

[17] S. Ghosh and B. Bhowmick. Technological uncertainty: Exploring factors in indian start-ups. In *IEEE Global Humanitarian Technology Conference (GHTC 2014)*, pages 425–432. IEEE, 2014.

[18] H. Halbleib. Requirements management. *IS Management*, 21(1):8–14, 2004.

[19] D. Howell, C. Windahl, and R. Seidel. A project contingency framework based on uncertainty and its consequences. *International Journal of Project Management*, 28(3):256–264, 2010.

[20] F. Jensen. Bayesian networks basics. *AISB quarterly*, pages 9–22, 1996.

[21] B. Jones, I. Jenkinson, Z. Yang, and J. Wang. The use of Bayesian network modelling for maintenance planning in a manufacturing industry. *Reliability Engineering & System Safety*, 95(3):267–277, 2010.

[22] N. Kandelousi. Key success factors for managing projects. *International Journal of Economics and Management Engineering*, 5(11):1541–1545, 2011.

[23] V. Lalsing, S. Kishnah, and S. Pudaruth. People factors in agile software development and project management. *International Journal of Software Engineering & Applications*, 3(1):117, 2012.

[24] W. Liao and Q. Ji. Learning Bayesian network parameters under incomplete data with domain knowledge. *Pattern Recognition*, 42(11):3046–3056, 2009.

[25] M. Lindvall, V. Basili, B. Boehm, P. Costa, K. Dangle, F. Shull, R. Tesoriero, L. Williams, and M. Zelkowitz. Empirical findings in agile methods. In *Conference on extreme programming and agile methods*, pages 197–207. Springer, 2002.

[26] P. Lucas. Building Bayesian networks, 03 2017.

[27] A. Masegosa, A. Feelders, and L. van der Gaag. Learning from incomplete data in Bayesian networks with qualitative influences. *International Journal of Approximate Reasoning*, 69:18–34, 2016.

[28] F. Mir and A. Pinnington. Exploring the value of project management: linking project management performance and project success. *International journal of project management*, 32(2):202–217, 2014.

[29] K. Mohan, G. Van den Broeck, A. Choi, and J. Pearl. An efficient method for Bayesian network parameter learning from incomplete data. In *Causal Modeling and Machine Learning Workshop*, volume 951, page 2014, 2014.

[30] A. Munns and B. Bjeirmi. The role of project management in achieving project success. *International journal of project management*, 14(2):81–87, 1996.

[31] A. Poth, M. Kottke, and A. Riel. Evaluation of agile team work quality. In *International Conference on Agile Software Development*, pages 101–110. Springer, 2020.

[32] M. Radujković and M. Sjekavica. Project management success factors. *Procedia engineering*, 196:607–615, 2017.

[33] B. Ramesh, K. Mohan, and L. Cao. Ambidexterity in agile distributed development: an empirical investigation. *Information systems research*, 23(2):323–339, 2012.

[34] B. Sauser, R. Reilly, and A. Shenhar. Why projects fail? how contingency theory can provide new insights–a comparative analysis of nasa's mars climate orbiter loss. *International Journal of Project Management*, 27(7):665–679, 2009.

[35] P. Serrador and J. Pinto. Does agile work?—a quantitative analysis of agile project success. *International Journal of Project Management*, 33(5):1040–1051, 2015.

[36] A. Shenhar. One size does not fit all projects: Exploring classical contingency domains. *Management science*, 47(3):394–414, 2001.

[37] A. Shenhar and D. Dvir. Toward a typological theory of project management. *Research policy*, 25(4):607–632, 1996.

[38] A. Shenhar and D. Dvir. *Reinventing project management: the diamond approach to successful growth and innovation.* Harvard Business Review Press, 2007.

[39] M. Singh. Learning Bayesian networks from incomplete data. In *AAAI/IAAI*, pages 534–539. Citeseer, 1997.

[40] M. Stoica, M. Mircea, and B. Ghilic-Micu. Software development: Agile vs. traditional. *Informatica Economica*, 17(4), 2013.

[41] T. Thesing, C. Feldmann, and M. Burchardt. Agile versus waterfall project management: Decision model for selecting the appropriate approach to a project. *Procedia Computer Science*, 181:746–756, 2021.

[42] L. Vijayasarathy and C. Butler. Choice of software development methodologies: Do organizational, project, and team characteristics matter? *IEEE software*, 33(5):86–94, 2015.

[43] M. Wellman. Fundamental concepts of qualitative probabilistic networks. *Artificial Intelligence*, 44:257–303, 1990.

[44] R. Wysocki. *Effective project management: traditional, agile, extreme.* John Wiley & Sons, 2011.

[45] B. Yet, A. Constantinou, N. Fenton, M. Neil, E. Luedeling, and K. Shepherd. A Bayesian network framework for project cost, benefit and risk analysis with an agricultural development case study. *Expert Systems with Applications*, 60:141–155, 2016.

[46] E. Zavyalova, D. Sokolov, and A. Lisovskaya. Agile vs traditional project management approaches. *International Journal of Organizational Analysis*, 2020.

[47] A. Zia, W. Arshad, and W. Mahmood. Preference in using agile development with larger team size. *International Journal of Advanced Computer Science and Applications*, 9(7):116–123, 2018.