Product Development Risk: A Bayesian Network Approach

The Unbabel Case Study

Student: Gil Afonso Correia Coelho
gil.coelho@tecnico.ulisboa.pt

Supervisor: António Manuel da Nave Quintino
antonio.quintino@tecnico.ulisboa.pt

***Abstract*—Highly innovative technology startups and scale-ups rely on launching disruptive products. Despite facing very uncertain scenarios these companies rarely adopt ”formal” and ”sophisticated” risk analysis tools. There are challenges related to engineering, user experience, and business models that make it so that only a small percentage succeeds in the market [12]. The high risk, high return nature of these ventures makes it so that any improvement in the low success rate can bring great benefits, namely to investors and to the economy.**

While there is a growing number of books and articles on how to launch technological products [3][5] - there are few examples in the literature aligned with these new ideas. When it comes to product development risk there is a clear gap between what is being adopted by the industry and what the academia has already studied.

Bayesian Networks are a powerful technique to create visual probabilistic models which can be used for multiple applications.

The employed methodology uses Unbabel, a Lisbon-based technology scale-up, as a case study on how modern technology companies think about product development risk. Through expert interview 4 different bayesian networks are generated to model and predict Feasibility, Usability, Value and Viability Risks. Findings suggest that the method creates models that behave consistently under different scenarios and that are suitable for many business applications such as decision making and risk analysis. Furthermore, it is concluded that bayesian networks can formalize industry practices and concepts, bringing academia and business closer together.

***Index Terms*—Product Development Risk; Bayesian Networks; Risk Management; Product Management; Innovation**

I. Introduction

While there is a wide body of knowledge around Risk Management, there is evidence [8] [7] that technology startups and scale-ups [1] fail to adopt these practices in their businesses. The key activity behind their success (or lack of) is the development and launch of innovative products, which is a highly uncertain endeavour with a low success rate. While some methodologies are solidifying in the startup ecosystem as product management ”best practices” [2] these concepts are yet to be studied by academia. This lack of research opens an opportunity to formalize modern product management practices and explore how they can be combined with current risk management methods.

This project is divided into 3 stages:

- Study of Product Management as adopted at Unbabel
- Generation of Bayesian Networks through expert opinion
- Evaluation of the Networks and discussion on the findings

II. Background and Related Work

A. Unbabel

Unbabel combines state of the art Artificial Intelligence with a global community of translators to provide a ”Translation as a service” solution to modern enterprises. It combines the advantages of Machine Translation (such as speed and cost) with the expertise of native level speakers to provide a high quality translation service up to the standards of modern brands. It is a high growth, highly technological company that is a great example of a scale-up looking to grow by launching innovative products. Its organization is composed of teams containing engineers, designers and product managers who are responsible for designing, building and launching successful products.

B. Critical Risks in Product

Similarly to other startups and scale-ups, Unbabel has a product management function whose job is to ”combine technology and design to solve real customer problems in a way that meets the needs of the business”. This function, originally found within the marketing organizational structure, has developed in the past few decades to integrate the Research and Development departments of technology companies.

While in the past companies had very long release cycles, where it could take months or even years to go from idea until product, software has enabled companies to continuously develop and ship product. These faster cycles allow for an approach that is based on smaller releases and increments to the product, as the product managers (and product team) quickly experiment with new ideas. This means that the teams are continuously having to make decisions on where to invest limited resources (usually developer hours) so that they get the highest return on investment (customer value aligned with the needs of the business).

In this context the concept of ”Product Discovery” [3] be-

[1]Millers and Friesen [6] define a scale-up as a company that is going through a growth phase. It already has some market success but is focused on quickly scaling in terms of sales and resources

[2]Inspired [3] and Continuous Discovery Habits [5] describe ”How to create tech products customers love” and how to ”Discover products that create Customer and Business value” respectively

[3]Product organizations commonly use the term Discovery (figuring out what is the ”right product” to build) to distinguish from Delivery (building the product).

comes very important: decision makers are always trying to ”ship the right product” which means minimizing 4 key product risks:

- Feasibility - the risk of not being able to build the product/feature
- Usability Risk - the risk of the user not understanding how to use it
- Value Risk - the risk of not providing enough user value so that he chooses to buy/use the solution
- Viability Risk - the risk of not generating revenue sufficiently higher than the costs

These 4 concepts were formalized by Marty Cagan in his best-seller ”Inspired” [3], which is one of the main frameworks adopted at Unbabel by the product management function.

C. Bayesian Networks

A bayesian network (BN) is a probabilistic graphical model that is composed of nodes (corresponding to random variables) and arcs (expressing conditional dependence). Bayesian networks use Bayesian inference to compute probability and are particularly effective to represent problems in domains where expert knowledge is probabilistic and easily modelled into a network structure. Because of their characteristics BNs are a viable method to perform risk analysis. They are particularly suited to deal with very innovative projects because they do not necessarily rely on a detailed data base - as they can calculate probability by taking advantage of expert opinion.

Another advantage of bayesian networks is that they are flexible in accepting inputs on the state of any node (variable) while producing a new outcome. In other words, if we correctly model the risk factors of a product and now know the state of any variable - the network will update its output. Finally, updating the conditional probabilities of a BN is a simple task (given that the correct software is available) - so they can be continuously improved as new data is available.

D. Bayesian Networks for Product Development Risk

Bayesian Networks have been successfully applied to model and analyse the risk of launching new products [10] [9]. Some research has even been done to advance the conditional probability generation [11] in this concept. While these developments advance the use of BNs for product application they were not focused on highly technological software products and use concepts that are now outdated in the industry. None of the examples use a framework that is broad enough to cover the risks across the different product related areas seen at Unbabel - engineering, design and revenue. We propose expanding on this research by combining BNs for product development risk with the modern concepts outlined in II-B.

III. Methodology

In this section it will be explained how to generate a bayesian network from expert opinion. Figure 1 contains the main steps that were followed.

Netica software, by Norsys, was used to build the bayesian networks. It contains all the capabilities that are required to build the acyclic graphs, input conditional probability tables and perform several types of tests.

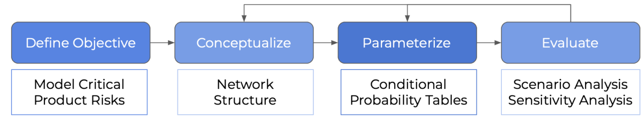

Fig. 1. Main Steps in the creation of a Bayesian Network

A. Conceptualize - Acyclic Graph

After a goal has been set the creation of the model continues by identifying the structure of the network. In this step we are building a visual representation of the concept by identifying all the risk factors (represented through nodes) and the causal relationships between them (represented by the arrows that connect the nodes). In figure 7 Risk Factor A and B affect the Risk Level (they are the ”parent nodes” of risk level.)

The first session with each expert started with an introduction to the goal of the project, bayesian networks (and its ”rules”) and by going over an example [11]. The interviewer then guided the expert in first listing all risk factors and then identifying the causal relationships between them. Finally, the resulting structure was reviewed in a ”conversational format” to find inconsistencies. This process was repeated until the expert was satisfied that the model was an accurate representation of how he thinks about risk.

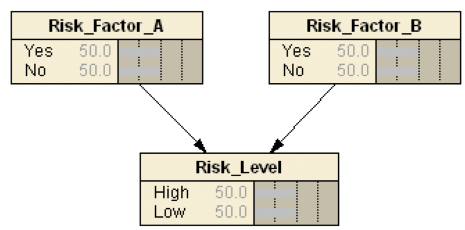

Fig. 2. Simple example of a network structure

A virtual whiteboard tool was used to conduct this exercise collaboratively with the expert.

B. Parameterize - Conditional Probability Tables

After all parent and child nodes and their relationships have been mapped then conditional probability tables model how the state of each node affects the others. This process is done with the expert by presenting scenarios for each possible parent state combination. This is, if a node C has two parent nodes (A and B), all with two possible states (A1,A2 and B1, B2 and C1, C2) - the expert will provided a judgement for the state of node C under each possible combination (A1,B1 / A2,B1 / A1,B2 / A2,B2). All these values were then input into the Netica software.

Taking into consideration the number of scenarios that must be estimated and the limited nature of the expert’s time - a direct elicitation method was used. This is, the expert was asked to directly produce the likelihood (in %) of a given state given the state of the parent nodes.

The method was very practical to quickly generate the conditional probability tables but it faced challenges when a

	B1	B2
A1	P(C1) = X P(C2) = 1-X	P(C1) = W P(C2) = 1-W
A2	P(C1) = Y P(C2) = 1-Y	P(C1) = Z P(C2) = 1-Z

Fig. 3. Simple example of a Conditional Probability Table where A and B are parent nodes for node C

node had more than 3 parent nodes. A node with 4 parent nodes already requires providing judgements on 16 different possible scenarios - which the interviewees struggled to do. Another simplification that was used was considering that all nodes had only 2 possible states (usually High or Low). A bayesian network approach would allow nodes to have many mores states which could potentially generate more sensitive and nuanced models - but it would also exponentially increase the complexity of the elicitation process to a level that would not be practical under most research conditions.

C. Evaluate - Sensitivity Analysis, Scenario Analysis, Stress Testing

Netica provides the ability to perform a sensitivity analysis on all or a few select nodes. It provides useful information by automatically calculating the Mutual Info Score for each node which is defined as the expected reduction in entropy of node Q due to findings at node F. Another way to think about it is as the reduction in uncertainty about one random variable given knowledge of another - it quantifies the "amount of information" obtained about a random variable from observing another random variable.

Defining **H** as Entropy and **I** as Mutual Information we can calculate:

$$I(Q,F) = H(Q) - H(Q|F) \tag{1}$$

$$I(Q,F) = \sum_{Q,F} P_{QF}(Q,F) \log \frac{P_{QF}(Q,F)}{P_Q(Q)P_F(F)} \tag{2}$$

Where Q is the query variable and F the varying variable with $P_{QF}(Q,F)$ joint probability distribution. $P_Q(Q)P_F(F)$ are the marginal probability distributions.

The results from 2 will come in bits so they have little intuitive meaning but they are useful for comparison purposes.We can compare and see if the order of magnitude is as expected. For example, if we expect Team to be much more important than Time Frame for the feasibility of a project, then we expect it to have a much higher score.

Scenario Analysis is a process in which the outcomes and consequences of possible future conditions are studied. Through it we can evaluate the model predictions under different probable scenarios and compare them against intuition and logic. Stress testing is a related but different technique. It focuses on assessing the consequences of extreme impact (and low likelihood) events on the model. Each network was submitted to a combination of both techniques in order to assess its robustness and the results are discussed in section V.

IV. Models

After outlining the 4 key product risk areas as defined in the framework presented by Cagan [3] and adopted by Unbabel (Feasibility, Usability, Value and Viability) 4 experts were identified to provide judgments on each network. They were chosen due to their expertise in the respective area: an Engineering Manager for Feasibility, a Product Designer for Usability, a Product Manager for Value and a Product Marketing Manager for Viability.

A series of 2 to 5 one hour long interviews was required with each expert to conclude the process outlined in III. Each network was versioned as it was continuously iterated between researcher and experts. So that Feasibility 2.0 represents a major iteration on network Feasibility 1.0 and Feasibility 2.1 represents a minor iteration on Feasibility 2.0.

A. Feasibility

"Can we build it?" - this is the question Marty Cagan asks when evaluating Feasibility risk for a product. He goes on to elaborate in his book, Inspired[3], that engineers have to consider a number of questions when trying to come up with an answer, such as evaluating the team knowledge and skills, auditing the architecture of the product, understanding the full requirements and scope, and many more.

It is not worth investing resources into starting to build a product that will not be completed. While in some scenarios it is instantly obvious that something is not achievable because the technology does not exist or the team does not have the capability in most cases expert opinion is required to predict an uncertain outcome. This prediction from the expert will be useful to negotiate with leadership and decide if a project should be pursued or not.

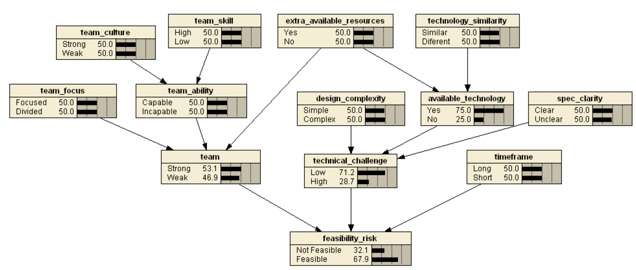

Fig. 4. Feasibility Network version 2.1

We can see from image 4 that the expert identified three main components that affect the probability of the product being built in due time - the team, the technical challenge and the time frame they have to work with. These are not surprising risk factors but it is interesting to understand which, in the opinion of the expert, has the highest impact on the feasibility risk.

We can measure this through the Mutual Info Score defined in 2 which are automatically calculated by Netica from the Conditional Probability Tables. Looking at table I we have that team is the most important factor by a large margin while time frame and technical challenge have a similar impact. It can be surprising to see that technical challenge ranks last -

TABLE I
MUTUAL INFO SCORES FOR MAIN FEASIBILITY NODES

Node	*Mutual Info Score (I)*
Team	0.08267
Time Frame	0.04786
Technical Challenge	0.03728

but this is most likely due to the past experience of the expert and the typology of the product he works with. An expert that works with complex infrastructure or artificial intelligence technology stacks might rank technical challenge higher but the interviewee works mostly with a visual interface product where there is less variability on the technologies that need to be used.

Finally it is interesting to note that under a full middle-term uncertainty [4] scenario the expert expects the product to be feasible 67.9% of the time. This reveals some inherent optimism by the expert - which, when compared to the typical delivery completion ratios for technology companies, is actually consistent with historical behaviour (industry experts [13] recommend aiming for a 70% completion ratio on goals).

B. Usability

"Can the user figure out how to use it?" - is a crucial question to ask when assessing if a product will be successful, as outlined in the book Inspired [3]. Usability is often overlooked when compared to Value or Feasibility due to its more subjective nature but the success of many digital products over their competition can be explained, not by better functionality or engineering, but by easier and more pleasant user experiences. It can be argued that Apple, one of the largest and most successful companies in the world at the time of writing, is best known for its design excellence (even over technological superiority).

A highly usable product will help the customer learn about the domain, navigate functionalities, self-serve on value and prevent him from giving up on the product. A Product Designer, whose job is to minimize the usability risk, was selected to be interviewed, results can be seen in figure 5.

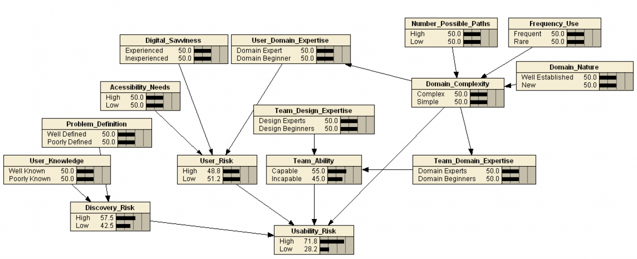

Fig. 5. Usability Network version 2.0

The expert started by giving examples of scenarios where the risk increased - for example if a product was designed with a young user in mind but was actually being picked up by an elder person. This scenario was then broken down into two separate risks: the poor identification of the user (which is included in Discovery risk) and the different usability needs that come from different personas (young vs elder user, surfaced under the User risk). Discovery is a "term" that is very commonly found in product organizations, particularly related to Product Designers and Product Managers. It represents all the work that needs to be done to "discover" what is the *right product* that should be built, in opposition with delivery which is the work to build what was decided (mostly owned by engineers).

[4]Middle-Term Uncertainty happens when all the parentless nodes of the network have 50% chance of being in either state (consider that team is either weak or strong, both with 50% probability). It is the "default state" of the network if the decision maker has absolutely no knowledge over the state of any factor.

TABLE II
MUTUAL INFO SCORES FOR MAIN USABILITY NODES

Node	*Mutual Info Score (I)*
Discovery Risk	0.119
Team Ability	0.043
Domain Complexity	0.014
User Risk	0.012

The other important factors are then team and domain complexity. The latter factor is best explained through an example: Neurosurgery is a highly complex domain while transportation from one place to another is a simple domain. The first domain has a myriad of different possible scenarios and requires a very specific knowledge that few people in the world have. Transportation - which is currently being solved by Uber - is a task that most people in the world have to accomplish on a frequent basis and doesn't involve many degrees of freedom (a user can choose a pick-up and a drop-off location, not much more).

From II we have that Discovery Risk is likely the most important factor for the expert. This finding is consistent with the intuition from the expert who said (quoting) : *"In any scenario where the Discovery Risk is high it will always be very likely (above 80%) that there is a high Usability Risk. Because if you don't have a strong understanding of what you're trying to solve - you'll most likely not arrive at a usable product"*. This is also consistent with field observations at Unbabel where discovery is mostly led by Product Designers and Product Managers - who deeply value a sound understanding of the problem and user.

Team still ranks significantly higher than Domain Complexity - while it can be explained that an expert design can quickly acquire domain knowledge or can compensate through research skills - it is also likely that there is some confirmation or motivational bias from the expert. Given that an expert designer is being interviewed - it is normal that they might value their own skills (design) over domain knowledge skills.

User Risk has the lower Mutual Info score - which can be due to historical reasons (Unbabel does not sell a consumer product with millions of users - so it is less likely to find accessibility needs amongst its users.)

C. Value

"Will the customer buy, or choose to use, this product?" - is how Cagan [3] evaluates if a product has value. It is important

to define that Value within the Product industry is centered around the user. It is not the financial value that the company is able to generate but rather the benefit that a customer gets by using a product to satisfy their needs, minus the costs (or effort). The business (financial) value that a company is able to extract from users is further explored under the definition of Viability IV-D.

Value is, at the same time, most likely the easiest risk to understand and recognize but one of the hardest to measure. There are very specific scenarios in which one could precisely quantify the value a product brings: when talking about advertising technology - we could use the revenue generated as a key indicator. But this is not the reality for most products. Many either don't directly affect revenue/costs and/or also have other unquantifiable effects.

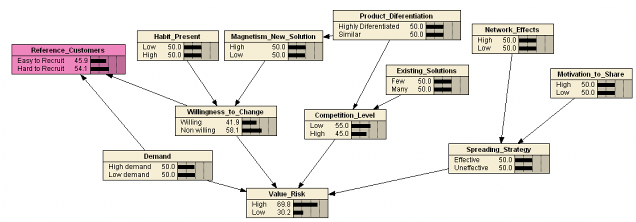

Fig. 6. Value Network version 2.1

Users cares about a product not just because of its main features but also because it helps them achieve different "jobs". These jobs can be the core functional job (Google Maps helps decide what is the best trajectory from A to B) but also related jobs (Google Maps helps find parking spots) or even social and emotional jobs. All of these factors play a role in the customer decision of moving into a new product - there is a magnetism pushing towards the new solution but a force pulling to stay in the current habit. All of these concepts were popularized by Christensen[14] and the "Jobs to be done" theory.

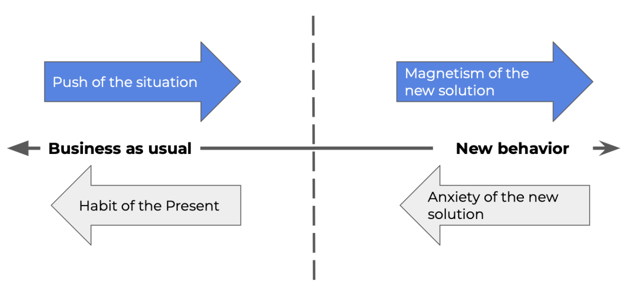

Fig. 7. Forces pushing into and pulling away the user from adopting a new product. A valuable new product will have a higher push than pull force.

The interviewee is aligned with this framework as he outlines the willingness to change from the user as a very important factor to understand the value of the product. This factor is broken down into two factors - the habit of the present (pulling) and the magnetism of the new solution (pushing). Looking at examples: people were used to opening bank accounts in-person with an specific account manager in an office, the habit of the present slowed down their adoption of bank apps (even if they are more convenient); on the other hand, Spotify's approach to music streaming was so different and innovative at instantly putting "millions of songs" available that users were eager to move away from locally stored files.

Another particularly interesting insight is the identification of the "Reference Customers". These are defined as early "real customers" (with no hidden motivations) that are willing to run a product in production and both make some sort of investment (monetary or in time) and willing to tell others how much they love the product. An analysis of the logic behind the causal relationships of the Reference Customer node reveals that it causes neither Demand nor Willingness to Change. It is actually the reverse - if there is a high demand and a lot of willingness to change then it will be easy to recruit people as reference customers. This is actually a "symptom" node, a signal, that updates our confidence on the state of Demand and Willingness to Change. This relationship is represented by the direction of the arrows (which point towards Reference Customers) and by the different coloring of the node. Note that within all the networks this is the only "symptom" node - most likely because the interview script and elicitation method were not structured to find them. It is an interesting area to further explore in future work VII-B.

Finally, it is important to note how the "optimism" of this network is so different from the feasibility - the expert actually expects the product to provide little value more than 70% of the time. Once again it is useful to revisit the nature and history of the expert - he is an experienced executive in the "innovation space" (disruptive rather than iterative improvements) - where a higher failure rate is expected. The motto *"If less than 50% of the products fail, we are not being ambitious enough"* is common amongst the team he leads. This historical factor and the different nature of feasibility and value might explain the discrepancy between the optimism of the networks.

D. Viability

"Does this solution work for our business?" Assessing viability means understanding if a business is able to sustainably make a profit out of a product. It includes considerations on the costs of producing, marketing and selling the product which also relates to how the product fits into the overall brand, strategy and positioning.

There are components to business viability that are well suited for objective, quantitative study - such as the cost of manufacture and the retail price. The profit or break-even point for a given product can be objectively calculated using data analysis tools such as Microsoft Excel when given the price and cost functions. We will not focus on these components of business viability because a bayesian network approach is not particularly well suited for factors that are not uncertain. We will instead focus on the components that affect the business viability risk but that are harder to measure, evaluate and are inherently uncertain. These also relate to how a product affects the business viability of the entire organization and not just "within the product itself". A viable product should not cannibalize other products, "dilute" the brand or confuse the customers through an inconsistent offering.

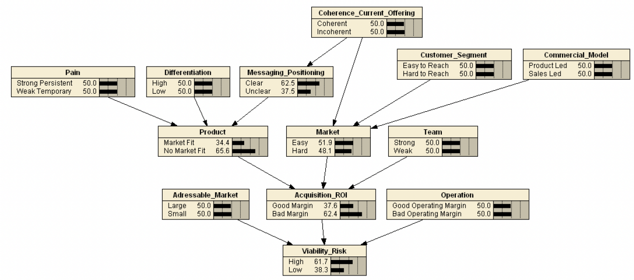

Fig. 8. Viability Network version 2.1

Looking at Unbabel - which delivers translations - the infrastructure (such as web servers and GPUs for Machine Learning applications) and translator costs are relatively easy to predict as they will be directly proportional to the number of translations delivered. Marketing and sales (customer acquisition) costs, on the other hand, are very hard to predict and will highly change depending on a number of product and market components. It is in the latter type of factors, which can account for a large percentage of the Cost of Good Sold, that we will focus our study.

As such we can see that Viability in figure 8 is directly impacted by the size of the addressable market, the operating margin and the acquisition return on investment (ROI) - but only the latter was the focus of deeper analysis. This is to avoid creating a network that is too complex and to center the model around the factors that are best suited for modeling through a bayesian network approach.

Further exploring the acquisition ROI we can see that while two other factors have a direct impact (characteristics of the market and of the go-to-market team), the product is by far the most important. This is quantified through the Mutual Info Score when calculated locally for the Acquisition ROI node and its three parent nodes (table III) - product has a score that is more than 3 times higher than market or team.

TABLE III
MUTUAL INFO SCORES LOCALLY CALCULATED FOR THE ACQUISITION ROI PARENT NODES

Node	*Mutual Info Score (I)*
Product	0.177
Market	0.053
Team	0.032

An insight from the interview process is that viability was more challenging to model than the other three critical product risks. More hours were needed to arrive at a model that was easy to understand - and even then a decision had to be made to just prioritise and model acquisition return on investment.

It is important to highlight that the overlap between the definition of Product Market Fit and Product Value IV-C is significant, particularly in the fact that both identify similar factors such as demand (addressable market), willingness to change (pain) and differentiation. It makes sense that it is much easier to drive a customer to acquire a product that clearly provides value than otherwise. From the insights coming from the experts it can be stated that a Product has achieved Market Fit if it provides Value (by solving a strong and persistent pain in a differentiated way) and has a clear messaging and positioning ("is made available in a clear and accessible way"). From this we can conclude that there is no conditional independence between the 4 critical risks in product - and exploring their relationships is an area for further research VII-B.

V. SCENARIO ANALYSIS AND STRESS TESTING

Scenario Analysis and Stress Testing are powerful techniques to understand if a model is behaving as expected or to find any inconsistencies with logical reasoning. In this section each of the 4 networks will be submitted to 5 different "scenarios" (either under probable or extreme/stressful conditions) and the results will be summarized in table IV using the following useful concepts:

- "Parentless node" - A node is parentless when there are no other nodes "pointing to it", it means that we cannot infer its state unless some knowledge is input into the network. Parentless nodes are all conditionally independent from each other. Examples from figure 4 are Team Focus, Technology Similarity or Time Frame (note how knowing if a team is focused provides no knowledge over how similar the technology will be).
- Feasibility (Usability/Value/Viability) Risk Level - It is defined as the probability of the Feasibility (Usability/Value/Viability) risk being high under the scenario. It can be thought of as a measure of how likely the product or feature is to fail under that specific area. If a scenario risk level is higher then it means there is a higher risk of it not being Feasible (Usable/Valuable/Viable).
- Risk Delta - It is the difference between the scenario risk level and middle-term uncertainty risk level. This is a measure of how much the inputs will affect the risk of the scenario (compared to the default state of middle-term uncertainty for all parentless nodes). A high risk delta means that the node(s) that were changed have a very large impact on the network.

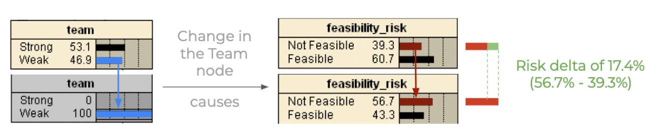

Fig. 9. Illustration of the scenario risk level (56.7%), middle-term uncertainty risk level (39.3%), and risk delta (17.4%) when Team node was changed to "Weak" state in the feasibility network

All scenarios will start from full middle-term uncertainty 4 but one (or more) node(s) will be selected to change to 100% probability of being in one state. To make sure that results are as meaningful as possible for the limited number of tests - the nodes will be selected according to their estimated variability

[5], sensitivity to findings [6] and ”uniqueness” [7].

A. Feasibility

It is interesting to note how the model behaves in a very similar way between having a Weak Team or a High Technical challenge - the risk delta is of 17% and feasibility risk approaches 50%. This ”balanced trade-off” between team and challenge is confirmed in scenario F.4 where we can see that the risk delta is less than 5% if we challenge a Strong Team with a High Technical Challenge. We can conclude that for the expert that the risk level can be maintained as long as the ability of the team increases proportionally with the difficulty of the challenge.

B. Usability

Under scenario U.1 the importance of Discovery work becomes instantly clear. If this risk is high then Usability Risk Level gets close to 90%. We conclude that it is crucial to research the user and problem before committing to delivery - or there is a huge risk that the customer will not be able to use the product effectively. Scenario U.4 has a risk delta of almost 0% which means that designing for a highly digital, capable and expert user in a very complex domain represents approximately the same level of challenge as designing under a fully uncertain scenario.

C. Value

The instant conclusion we take from the Value Scenarios is that it is not worth investing in a product when there is no demand - which is consistent with any logical reasoning. Scenario V.2 is slightly more surprising because one could expect the competition to have a huge impact on risk level (how many companies were driven to failure by competitors?) - but, similarly to what was discussed in the previous section, this might be explained by the expert’s optimism and availability bias - his experience is focused on highly innovative products which are usually highly differentiated and more resilient to competitive pressures.

One of the factors we are trying to evaluate when running discovery work is whether there is Willingness to Change from users. Being certain that they are, in fact, willing to change reduces the risk to almost 50% - doing discovery work is a great investment as was concluded in the Usability Scenarios.

Understanding if it is easy to recruit reference customers is one of the ”least expensive” data points a company can capture (a landing page and a mailing list could accomplish results) while still providing a lot of information on the value risk level. We can see that if they are easy to recruit there is a risk delta of -14% - which will be very useful to make a decision. As such, it is probably a good investment of time to try to recruit reference customers before deciding on a product.

D. Viability

From Vi.2 a product that targets a small addressable market will fail to become viable more than 4 out of 5 times. This judgment is consistent with logical reasoning because software development is typically tied to large fixed costs and small variable costs. This ”cost profile” greatly benefit from scale, which is not possible if the addressable market is small.

It is interesting to note how in scenario Vi.3 (when the Acquisition ROI node is set to 100% probability being good) the risk of the product not working for the business goes down very significantly - almost 25%. When this happens it means that the costs of acquisition are significantly smaller than the Lifetime Value of the customer. While this might be an unexpectedly high risk delta might (it leads to questioning the impact of the other type of cost - service/operation) the expert explained that acquisition is typically harder to solve (and riskier) because it depends on many uncertain external factors. The operating margin, on the other hand, will usually improve with growth because software costs do not scale proportionally with revenue.

E. Conclusions on Scenario Analysis and Stress Testing

While some specific scenarios present risk deltas that are higher or smaller than what would be intuitive - these differences are almost always explained by the nature of the expert. Through the different tests all of the networks display a consistent behaviour that is aligned with the experts beliefs that were communicated during the interviews.

It should also be acknowledged that, particularly due to the direct elicitation method that was used, this project is exposed to availability [8], representativeness[9], anchoring[10], motivational[11] and confirmation biases[12].

VI. Applications

This project approaches an area that is mostly unstudied by the academia - ”modern” technology product management practices. As such, there are still a number of possible applications particularly due to the flexible nature of bayesian networks. As visual probabilistic models they can be very easily shared, changed over time and applied in different business contexts. Some possible applications are:

[5]Variability refers to how spread out a set of data is. The higher the variability the more important it is to study extreme events as they are more likely to happen

[6]Sensitivity to Findings studies how the uncertainty in the output of a model can be attributed to different inputs of a model. In this project it is measured through the Mutual Info score

[7]”Uniqueness” will be informally used by the author to identify causal factors that are less frequently found in the literature. Discovery Risk, for example, was often mentioned in the interviews but is rarely found in publications

[8]estimation of a probability will be impacted by how easily the expert can retrieve similar event

[9]estimation of a probability is made based on similarities within a group. It can bias the expert into not taking into account sample size

[10]estimation of a probability is highly impacted (anchored) on the first assessment that is made. Providing a random number as an example during the interview can influence the interviewee to provide judgments within the same order of magnitude - even when there is no logical reason to do so

[11]happens when there is a (usually hidden) incentive for the expert to provide a given assessment.

[12]illustrates a tendency from people to favor information that already confirms their beliefs - might affect the interviewer

TABLE IV
RISK SCENARIOS FOR ALL NETWORKS

Scenario (input)			Risk Level	Risk Delta
id	*Node(s)*	*State*		
Feasibility Scenarios				
F	Feasibility Middle-Term Uncertainty		32.1%	/
F.1	*Team*	100% Weak	48.8%	+16.7%
F.2	*Technical Challenge*	100% High	49.1%	+17%
F.3	*Time Frame*	100% Long	20.2%	-11.9%
F.4	Technical Challenge and Team	100% High and 100% Strong	30%	-2.1%
F.5	Time Frame and Technology	100% Short and 100% Available	41%	+8.9%
Usability Scenarios				
U	Usability Middle-Term Uncertainty		71.8%	/
U.1	*Discovery Risk*	100% High	87.9%	+15.2%
U.2	*User Risk*	100% High	78.4%	+5.7%
U.3	*Team*	100% Incapable	84.4%	+11.7%
U.4	Domain and User Risk	100% Complex and 100% High	73.4%	+0.7%
U.5	Discovery and User Risk and Team	100% Low and 100% Low and 100% Capable	17.1%	-54.7%
Value Scenarios				
V	Value Middle-Term Uncertainty		69.8%	/
V.1	*Demand*	100% Low	91%	+21.2%
V.2	*Competition Level*	100% High	74.8%	+5%
V.3	*Willingness to Change*	100% Willing	54.8%	-15%
V.4	Reference Customers	100% Easy to Recruit	56.1%	-13.7%
V.5	Habit of the Present and Magnetism of the new solution	100% Strong and 100% High	71.8%	+2%
Viability Scenarios				
Vi	Viability Middle-Term Uncertainty		61.7%	/
Vi.1	*Product*	100% Market-Fit	68.4%	+6.7%
Vi.2	*Addressable Market*	100% Small	83.7%	+22%
Vi.3	*Acquisition ROI*	100% Good	37.5%	-24.2%
Vi.4	Addressable Market and Operating Margin	100% Small and 100% Good	78.7%	+17%
Vi.5	Addressable Market (size) and Market Characteristics	100% Large and 100% Hard	46.1%	-15.6%

- Explaining how product professionals make decisions - Some frameworks, such as the one presented by Cagan [3], have been widely adopted in the industry. These concepts are still to be picked up and adopted by the academia - which is an important step in finding more business applications for research.
- Predicting risk under specific product areas - Each of the 4 network outputs a prediction on the probability of the product succeeding or failing under each key area. A decision maker could simulate different scenarios (we can imagine different team-feature allocations) to find the combination with the minimum risk level.
- Customized decision making tools that evolve over time - Bayesian networks are represented as a group of nodes with causal relationships and conditional probability tables that dictate their behaviour. Networks from this project can be used as templates and easily adapted to different professionals, industries and frameworks. As they apply and learn users can then easily change specific sections of the models to increase accuracy and robustness.
- Assessing product related investments - The Mutual Info Score [13] can be used as an indicator of which knowledge provides the most confidence in the success of a product. For example, it can be used to decide whether or not it is worth investing in paying a given amount to run a market study to acquire knowledge over the node ”Addressable Market”.

[13] Mutual Info (I) - defined as the expected reduction in entropy of node Q due to findings at node F

VII. Conclusions and Future Work

A. Conclusions

The goal of this project was to formalize modern product industry concepts by leveraging a bayesian network approach. Through direct elicitation to experts 4 different models were successfully built. These networks withstand scenario analysis and stress testing, behaving consistently with logical reasoning the majority of the time. While there are still many improvements to be made, we propose that bayesian networks have multiple potential business applications within product development such as risk analysis, prediction, and as a decision making tool that is easily customized. This is a robust approach to modeling scenarios under uncertain and unstructured domains.

B. Future Work

There are multiple possible paths as follow-up work to this project particularly related to testing the method under different scenarios, increasing the robustness of the models and expanding into new applications:

- **”Universal” product risk network** - by combining the 4 risk areas into one model it would be possible to represent all risk factors under a ”unified” model. This network could provide a ”risk score” for a product launch.
- **Exploring other elicitation methods** - Direct elicitation was used because it is the fastest method to generate conditional probabilities. Consistency could be increased while mitigating biases by adapting methods such as the one seen in the MACBETH software [1] or AHP Method [2].
- **Expanding on ”symptom nodes”** - Medical diagnostic applications of bayesian networks commonly model symptons as nodes to help disease detection. Different business indicators (such as reference customers) can be modelled and used to increase the ”knowledge” of the network in an efficient way.
- **Conditional Probability Generation through historical data** - Netica software has the ability to generate bayesian networks from case data. An organization with enough structured data could ”allow” the software to find all the nodes, causal relationships and conditional probabilities and only fine-tune the results.

VIII. References

[1] C. A. B. e Costa, J.-M. D. Corte, and J.-C. Vansnick, “Macbeth,” 2012

[2] R. W. Saaty, “The analytic hierarchy process—what it is and how it is used,” Mathematical Modelling, vol. 9, pp. 161–176, 1987.

[3] M. Cagan, Inspired: How to Create Tech Products Customers Love, 2nd ed., S. V. P. Group, Ed. John Wiley Sons, 2008

[4] K. S. Chin, D. W. Tang, J. B. Yang, S. Y. Wong, and H. Wang, “Assessing new product development project risk by bayesian network with a systematic probability generation methodology,” Expert Systems with Applications, vol. 36, pp. 9879–9890, 8 2009

[5] T. Torres, Continuous Discovery Habits: Discover Products that Create Customer Value and Business Value, 1st ed. Product Talk LLC, 2021

[6] D. Miller and P. H. Friesen, “A LONGITUDINAL STUDY OF THE CORPORATE LIFE CYCLE.” Management Science, vol. 30, no. 10, pp. 1161–1183, 1984

[7] A. Nadali, A. Grilo, and A. Zutshi, “A conceptual framework of risk identification for scale up companies in transition period,” Proceedings of the International Conference on Industrial Engineering and Operations Management, vol. 2018-March, pp. 2346–2357, 2018

[8] R. Pukala, E. Sira, and R. Vavrek, “Risk management and financing among Start-ups,” Marketing and Management of Innovations, no. 3, pp. 153–161, 2018

[9] S. Nadkarni and P. P. Shenoy, “A bayesian network approach to making inferences in causal maps,” Eur. J. Oper. Res., vol. 128, pp. 479–498, 2001

[10] L. G. Cooper, “Strategic marketing planning for radically new products,” Journal of Marketing,vol. 64, pp. 1–16, 2000

[11] S. Chin, D. W. Tang, J. B. Yang, S. Y. Wong, and H. Wang, “Assessing new product development project risk by bayesian network with a systematic probability generation methodology,” Expert Systems with Applications, vol. 36, pp. 9879–9890, 8 2009

[12] T. Eisenmann, “Why start-ups fail,” Harvard Business Review, 5 2021

[13] S. Prince, ”How to grade OKRs”, www.whatmatters.com, 2020

[14] C. M. Christensen, K. Dillon, T. Hall, and D. S. Duncan, Competing Against Luck: The Story of Innovation and Customer Choice. Harper Business, 2016