Network Analysis in team sports

3UAU0 Information technologies of the future - Group 11

Authors:

Rutger de Koning	1339389	Department of Industrial Engineering and Innovation Sciences
Justin Kuppen	1262033	Department of Mathematics and Computer Science
Rick van Oosterhout	1450131	Department of Mathematics and Computer Science
Marijn Brozius	1510886	Department of Industrial Engineering and Innovation Sciences

Supervisors:

Georgios Papachristos	Department of Industrial Engineering and Innovation Sciences
Pim van der Hoorn	Department of Mathematics and Computer Science

Executive summary

The goal of this project is to discover and develop a sustainable business model for the application of network technology in team sports. The report includes an overview of the investigated innovation, network technology in team sports. This overview covers several metrics that are used within network technology and discusses the significance of these metrics with regards to team sports. Also in this report, the considered applications of network technology are explained, multiple competing technologies are analysed and a proposed business model is given.

Prior research has shown that some network metrics, on collected data of football matches of a team, might have a strong relation with the outcome of football matches. This report further explores this relation, also by providing different metrics and conclusions that can be drawn from them. The application that is analysed is that of software which helps football coaches in coaching his team. This software uses network technology and collected data to find strengths and weaknesses. The software is accompanied by consultancy, this way the coach has optimal assistance to improve his team.

Lastly, a business model and go-to-market strategy are proposed for the application. These focus on a subscription service, where multiple tiers of subscriptions offer different features. The cheaper tiers also aim to advertise the other subscriptions, in an attempt to convince the coach to use the highest tier subscription. The go-to-market strategy revolves around building a reputation and experience for a lower cost in a lower division. These proposals were influenced by competing technologies, since the aim is to acquire a part of the existing market.

Commented [Gp1]: This is one way of indicating paragraphs.
Another is to have increased line spacing between paragraphs.
Pick one and implement it throughout the report.

Table of Contents

Introduction

~~With n~~Network science can be used to analyse all sorts of networks ~~can be analysed~~. In some ways, networks act as a mathematical language, allowing scientists from many domains to communicate with one another. As a result, networks are instruments that enable us to work on a variety of challenges. Examples of networks are the world wide web, air traffic and the human brain. Using network science, certain metrics can be calculated and mathematical models can be applied.

Our project is based on network technology in team sports. In team sports, networks originate from the players being nodes and some sort of connection between them. The project goal is to answer the following question: How can analysis of team sports using network technology yield the basis for a sustainable business case? The analysis that is done is computing network metrics on a football team. By analysing this data, insights are gained on for example the strongest player, finding an intermediary role or how the team plays as a collective. In football analysis this is done on many aspects of the game. Although network science is already used, not a lot of business cases are developed around it.

The approach that ~~has been~~ is followed to ~~achieve the goal of~~ find~~ing~~ a sustainable business case that uses network technology ~~to analyse~~ for team sports analysis ~~is as follows~~involves X steps.

Commented [Gp2]: How many steps?

The first step that was taken was the gathering of information about network technology, network technology in team sports and network measures that can be applied on team sports. This information was gathered by finding and reading papers (which can be found in Appendix B) that provided the necessary information about network theory, network measures and their relations to team sports. The information that was gained this way is bundled in the Technology description chapter. In this chapter, information is presented about what a network is, how a network can be applied on team sports and a multitude of network measures that can also be applied on a network applied on team sports.

The second step ~~that followed from the gathering of information was~~is finding possible applications for network analysis in team sports. The Possible applications chapter 2?, in which this process is described, provides the reader with knowledge about what types of applications can be developed using network analysis in team sports. It is discovered that each of the possible applications has different benefits but also different needs of the network technology. In the end, even though all possible applications could be used to develop a sustainable business model, one possible application stands out and is therefore chosen. This application will be used in the following chapters to develop a sustainable business model with that application as the centre of the model.

Commented [Gp3]: If you just lay down the step in a linear way there is no need for this. Repeat for the remaining steps.

When the application was chosen, the next step was to find any (if any) competing technologies in the current market. Finding competing technologies is an important step, since it can provide necessary insights about the market and its customers, for example. It was discovered that ~~there are already a multitude of~~several competing technologies in the market ~~that~~ achieve the same goal as our application ~~that was~~ chosen in ~~the previous~~ chapter X. In the Competing technologies, this step is described. The competing technologies that have been found are carefully researched by discovering what information they can provide their users and by exploring their business models. It was seen that the chosen application would aim to replace a couple of competing technologies since it will provide at least the same amount of information.

Commented [Gp4]: Chapter number

On the other hand, it was also seen that some competing technologies would be better suited to be combined with the chosen application, such that the application would be able to provide even more insights to its users.
After gathering all necessary information about the technological and the economical aspects, it became possible to describe our own application. This is done in the Our application chapter 4?. In this chapter the reader can find the way we want to set up our business model, our business model canvas, a description of the technology used in our application and whether or not we find network technology to be sufficiently developed for our application. This chapter provides the reader with insights in how we want to differentiate from the competing technologies and gain customers. The filled out business model canvas shows all different aspects of our business model, which described, for example, our needs and our value to the customer. The subsection in this chapter about the technology used in our application reports the network measures that will be used in our application. While the technical details of these measures can be found in the Technology description chapter, this section aims to already give potential users of our application insights in what information they can gain by using our application. For every network measure used in our application it is described what outcomes of those measures could indicate. Each measure is carefully chosen such that they all provide a different insight in a sports team. Lastly, it was seen that network technology is sufficiently developed to start our application at this moment, however, new developments in network technology would be able to make our application better.
After our application is described, recommendations are given for the technological development, the most suitable business model and the go-to-market strategy of our application. Each of these subsections give information about what we would want to happen with the development of network technology and what the best way would be to bring and keep our application in the market.
Finally, a conclusion is provided that summarises the most important insights that were gained by this research and what we can conclude from those insights

Technology description

This section provides an overview of network technology and measures in said technology. These measures are ways to evaluate/distinguish certain elements in a network and could be used to develop a business model.
A network is a group of objects (referred to as nodes) that can be linked to each other based on some connection; these links are called edges. Nodes can also have a connection that is directed in only one way, these 'from' and 'to' directions will lead to a network that is called directed; edges have an arrow indicating from which node to which other node some connection goes.
Another concept on networks is the concept of weights; values related to an edge that could indicate either a probability of that connection happening, or the ~~amount~~ strength of that connection. A higher weight would indicate a higher probability of that connection happening or a ~~higher amount of that~~stronger connection ~~occuring~~ between ~~the~~ two nodes ~~that are connected~~.
Many different network measures have been developed and will be described in the next sections, but first, a network in team sports that these measures can be applied to will be defined; a passing network.

Passing network

A passing network is a network applied to team sports, in this case football. In this passing network players are considered to be nodes and a directed edge with weight 1 is defined between two nodes when a player passes to another player. For every extra pass between those same two nodes, the weight of that directed edge is increased by 1.

Density

From (Peña, Touchette, 2012) and (Heikinheimo, 2019), network density is a statistic that indicates how close the network is to being a complete network. It can be computed by taking the total number of edges and dividing it by the number of possible edges in the network. The following formula is the formula to compute the network density in a directed network:

$$D = \frac{E}{N(N-1)}$$

Where D is the density, E the number of edges and N the number of nodes.
The density in a passing network can indicate if each player can reach any other player. A higher density means that more players have successfully passed and received passes from distinct players. A higher density could indicate that there are more possible targets to pass the ball to for most players. However, having a perfect density of 1.0 would not necessarily mean that your team performs better, because in most games you do not want a striker to pass the ball to his goalkeeper.

Node degree of a vertex

From (Yamamoto, 2010) the node degree is the number of edges some node i has to other nodes, denoted with k_i. The node degree distinguishes two different types; the in-degree and the out-degree. The in-degree of a node is the amount of incoming edges, similarly, the out-degree of a node is the amount of outgoing edges. The node degree in a passing network can be a measure of interactiveness. So, players with a higher in-degree or out-degree signify that they receive and pass the ball more than others, meaning they are more interactive than their teammates who have a lower in-degree or out-degree, respectively.

Bridges and structural holes

Structural holes are places in a network where the lack of a fitting player (also called a "bridge") would result in a complete collapse of cohesiveness in the entire network as in (Lusher, Robins, Kremer, 2010). The absence of such holes is an indication of a cohesive team. The amount of structural holes can be measured by counting the bridges, this is a fairly simple and straightforward measuring method, which uses the effective size. The effective size can be calculated by using Borgatti's reformulation of Bert's formula. It states that the redundancy can be calculated as:

$$R = \frac{2t}{n}$$

Where R is the redundancy, t the number of ties in an egocentric network (excludes ties to the ego itself) and n is the total number of nodes in the network. This formula can then be used to get the effective size, given by the following formula:

$$ES = n - \frac{2t}{n}$$

Where ES is the effective size, n the total number of nodes in the network and t the total amount of ties in the network.

Transitivity

The transitivity of a node indicates how likely a node is to form a triadic relation with other nodes as in (Lusher, Robins, Kremer, 2010). For instance, if player A and B have a good chemistry between them and player B and C also have a good chemistry between them, then player A and C will most likely also have a good chemistry between them. Transitivity leads to clustering in a network, as described in the next part, and thus is a great indicator for the cohesion within a team. The transitivity T of a graph is calculated by:

$$T = \frac{3 * number\ of\ triangles\ in\ network}{number\ of\ connected\ triples\ of\ nodes\ in\ the\ network}$$

Clustering coefficient

From (Arriaze-Ardiles et al., 2018), (Peña, Touchette, 2012), (Yamamoto, 2010), (Ramos, Lopes, Araújo, 2018) and (Cotta, Mora, Merelo, et al., 2013) the information was gained that the clustering coefficient captures the degree to which the neighbours of a given node have edges between each other. This can be described as smaller groups (subsets of nodes that are connected) being formed in a network.

The formula is defined by two times the number of edges between some amount of k neighbours of some node i, divided by the degree of that node i times, the degree of node i minus one. This measure is then defined as:

$$C_i = \frac{2L_i}{k_i(k_i - 1)}$$

Another way of calculating the clustering coefficient is by multiplying the total number of triangles in the network with two and then dividing by the number of possible triangles including some node i. Then, finally dividing the outcome by the degree of node i, times the degree of node i minus one. The formula is then as follows:

$$C_i = \frac{2\Delta_i}{k_i(k_i - 1)}$$

The average clustering coefficient is calculated by multiplying the sum of the clustering coefficient of all nodes with one over the total number of nodes in the network:

$$C_{avg} = \frac{1}{N} \times \sum_{i=1}^{N} C_i$$

When these measures are applied on a passing network, the clustering coefficient of an individual player (node) could be compared with the average clustering coefficient of the network. This will then indicate which players are more likely to be connected to each other and thus which players are more likely to form an internal cluster (group).

Cohesive subgroups

Cohesive subgroups are smaller groups within a larger network that have high cohesiveness, a different name for cohesive subgroups is also a clique as in (Lusher, Robins, Kremer, 2010). These cliques can identify small groups of players that work well together, as well as identifying a small outlying group in a larger network. Clique's are calculated of a certain size with the following formula:

$$C_G(x) = \sum_{k=0}^{\omega(G)} c_k x^k$$

Where c_k is the number of cliques of size k and $\omega(G)$ is the clique number; the amount of vertices in the largest possible clique in a network.

Reciprocity

Reciprocity indicates the likelihood of mutuality in a relationship as in (Lusher, Robins, Kremer, 2010). An example would be player A passing the ball to player B, then reciprocity indicates how likely it is for player B to pass the ball back to player A. When two nodes show reciprocity they can also be called reciprocated dyads. Reciprocity has two definitions, the first is called the traditional definition and states reciprocity as:

$$r = \frac{L <->}{L}$$

Where r is the reciprocity, $L <->$ the amount of edges going both ways and L the total amount of edges in the network.

The second definition is the definition of Garlaschelli and Loffredo. This definition states that:

$$\rho = \frac{r - \underline{a}}{1 - \underline{a}}$$

Where ρ is the correlation coefficient (with $\rho > 0$ meaning reciprocal, and $\rho < 0$ meaning anti-reciprocal), r being the reciprocity and $\underline{a}$ being defined as:

$$\underline{a} = \frac{L}{N(N-1)}$$

Where L is the amount of edges and N is the amount of nodes in the network.

Variability of nodes and networks

Measuring the variability of nodes and networks is another method of analysing performance in team sports. In (Martins et al., 2021) it was shown that the variability of a node or a network shows the predictability of a node or a network.
In the case of passing, the rate of possible variability in passing from this node to all other nodes in a network is defined as follows:

$$R_i^{out}(n_i) = -\sum_{j=1}^{n} \left(\frac{\sum_{j=1}^{n} w_{ij}}{L_D^w} \times m_{ij} log_2 m_{ij} \right)$$

This uses a Markov chain; a model that describes a sequence of possible events where the probability of an event depends on the state obtained from the previous event. In this Markov chain, m_{ij} are the elements associated with the weighted adjacency matrix of a weighted directed graph with n nodes.
With the rate of passing of a node, the index rate of passing of a node is determined by:

$$IndR_i^{out}(n_i) = \frac{R_i^{out}(n_i)}{log_2 n}$$

To convert these formulas from passing to receiving, w_{ij} should be changed to w_{ji}, and in the same manner m_{ij} should be changed to m_{ji}. Doing this gives the rate of reception of a node and the index rate of reception of a node, respectively.
The rate of passing of a network is calculated by summing the rate of passing of every node. More formally, the rate of passing of a network is calculated as follows:

$$R_N^{out} = \sum_{i=1}^{n} R_i^{out}(n_i)$$

In the same way as for calculating the index rate of passing of a node, the index rate of passing of a network is determined by:

$$IndR_N^{out} = \frac{R_N^{out}}{log_2 n}$$

With these calculations, information about a network can be obtained. For example, one could look at the rate of passing variability and the rate of receiving variability. If the distribution of these values over all nodes is relatively uniform and the nodes show little capacity of variability, it can be concluded that the network has well established passing patterns because, with a high accuracy, it can be predicted to which player, some player is going to pass.

In general terms, what can be concluded from the rate of variability of a node, is that when the rate of variability of a node is high, the node is more unpredictable. What this means is that it is more difficult to predict the edge that will be 'chosen' by a node i.e. to predict to whom a player will pass the ball. An unpredictable node can work two ways, on the one hand since it is harder to predict a player, it is harder to play against that player. On the other hand, it might be harder to create synergies between players when at least one of the nodes has a high rate of variability, which means at least one of the players is unpredictable.
The rates of passing (and receiving) of a node are not easily comparable between nodes in a network, that is the reason the index rates of passing (and receiving) of a node were given. This scales the rates between 0 and 1, the least and most variability, respectively.

Possible applications

In the world of (team) sports, a lot of money is invested in analysing matches and teams. Journalism and gambling are part of the reason why analyses are made. However, teams themselves are also interested in the analyses. A team can gain insights into their own performance, or analyse an opposing team to prepare for a match against them. In most sports, analyses are made based on statistics that fit the sport well, i.e. possession of the ball, attempts (on goal) and fouls, in soccer. This section will provide an overview of three applications for network analysis in team sports; sports analytics, betting and journalist sports reports.

Sports analytics

The first possible application of network technology is assisting football coaches by identifying the weakest link in the opponents' team (or in their own team). By picking out this link, an advantage could be gained by the home team. This advantage results from the weakest link making the most mistakes. These mistakes can be anticipated. Methods to maximise these mistakes are for example, putting this player under more pressure or shielding his or her passing directions. Needless to say, this is the coach's task.
A passing network is a useful network to analyse, as numerous network theories can be applied to determine the weakest player. Regarding performance requirements for the application of determining the weakest link, the difference between the in-degree and out-degree is used. If the in-degree of a player is a lot higher than the out-degree, it could signify a higher percentage of loss of the ball compared to other players in the team. In simple terms, this means that the player receives the ball more than he gives the ball away.

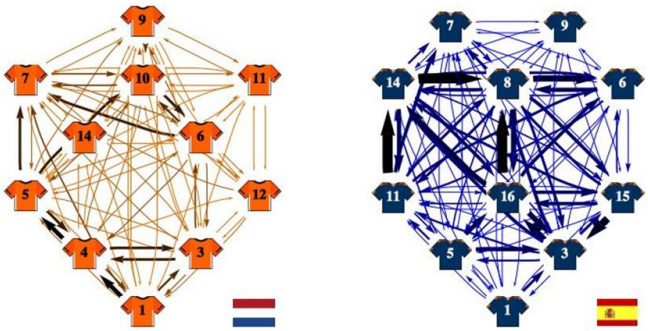

Figure 1: Passing Network Netherlands & Spain (Peña, Touchette, 2012)

Formatted: Font: 10 pt

Another explanation for a higher in-degree than out-degree is that the player is perhaps an attacker and makes a goal attempt. To illustrate this, in figure 1 the passing networks are given for the Netherlands and Spain during the semi-finals of the world cup of 2010. In these networks the edges are directed, and the thicker arrows mean that the ball is more frequently passed in that direction. For player 9 of Spain, it is remarkable that he receives the ball more than he gives the ball away. From this it could be concluded that this is a weaker player as he loses the ball more often. However, this is an attacker that scores a lot, so the method of identifying the weakest link is not applicable in this instance.
Because football is the biggest sport in the world, a lot of data is gathered on matches. Concerning network science in general, it is already used to analyse football matches. However, no definitive applications are used to implement in a football game. Measuring the difference of the in- and out-degree relates directly to the stage of development network technology is in, namely the analysis stage. (We, the research group, are going a step further and implement this network theory for football coaches to use)

Betting

An example of an application for network theory is to use it in the computations for the odds and payout of betting sites. If you have a metric which provides insights into both teams' strength, you can use it to compute the odds of a specific team winning the match. With these odds you can calculate the payout for certain bets. The problem of this application is determining one or more good metrics to base the computations on.
In (Peña, Touchette, 2012), research is conducted on passing networks within football matches. The paper suggests a connection between a high number of passes, clustering and clique size, which are described in their respective subsections in the section Technology description. If this suggestion can be proven, this application will have a lot of potential. Using multiple metrics might improve the quality of the odds and payouts, since there is not being focussed on a single aspect of the game.
Many modern betting sites have multiple betting options for a single match, these range from picking the winning team, score of the match, player to score goals and many more. All these bets have their own payout rates, which are calculated by the betting site. Some options, 'player will score' or 'player will give an assist' for instance, might be connected to passing networks in some way. If an offensive player receives a lot of passes, then he is more likely to have attempts on goal for instance and if a player passes to a player who scores a lot, he is more likely to give an assist.
The required technology for this application is developed, since a lot of theory and characteristics are well-defined. However, it is still unclear whether it is possible to compute these odds using existing network characteristics. Network theory is still a field where a lot of active research is conducted, this might lead to new calculations or simulations that can provide useful metrics for the application. Using different ways of defining the network could also provide metrics that are useful to compute the betting odds.

Journalist sports reports

Another possible application of network theory would be in journalism. Sports journalism is an important aspect of sports as it makes difficult sports theory understandable, highlights important information and points out missed information to the public.
A possible application would be to highlight the most interactive player by using node degree of a vertex as described in the section technology description. Doing this will allow the journalist to easily select an impactful player for an interview or article about the match.
Network Analysis can also be used to show journalists how a team plays by highlighting important players, showing player passing triangles and showing how the ball moves around the team, aiding them in analysing and reporting why a team did good or bad and showing their overall capabilities.
Sports journalists also need to be able to follow leads for interesting stories. Here, network analysis can aid by showing trends. This can be achieved by analysing the same team/player multiple matches in a row and can show certain players being impactful for multiple games in a row or a team's strategy being useful against different types of opponent's strategies.
Finally, network analysis can also aid in showing how a team strategy transitions over time. Comparing different network analyses of the same team over time will show the small differences made to tackle different opponents. This gives the journalists a view into how well the coach knows the opponent and how good the coach/team is at adapting to different opponents, allowing for an insight in a coach's ability.
Currently network analysis isn't sufficient enough for journalists yet, as to date only a limited amount of sports have been analysed like this. Another issue is that most research has been put into application of small network metric sets, as seen in (Korte, Lames, 2018). A final issue is that this network analysis is done as scientific research which takes a while to be published. For network analysis to be used practically by journalists it needs to be able to be used easily and results should be published quickly so that journalists will be up-to-date with their information. As such a reliable form of fast network analysis needs to be developed in order for journalists to profit fully off of its benefits.

Chosen application

While all three applications have different benefits and could all be used to develop a sustainable business model, it was decided to further investigate the sports analytics application.
In comparison to the betting and journalist sports reports applications, the sports analytics application is in a better stage concerning the current stage of development of network analysis technology as was described in their respective subsections. Considering the betting application, it is quite unclear how the multipliers and odds are calculated, while, for the sports analytics application it is quite clear what some measure indicates.
Also, even though only the node degree of a vertex was described in the sports analytics subsection, many other measures (as seen in the section Technology description) can be applied on a passing network. These measures can provide helpful information concerning player or team strengths and weaknesses.

Competing technologies

To develop a sustainable business model, it is important to analyse the current market first. By exploring current competing technologies, insights can be gained in the market and customers, from which valuable lessons can be learned for our new business model. The following sections will explore existing technologies that assist football team coaches. This market already contains multiple different solutions, for our business model it is important to differentiate from the existing solutions. Therefore, the business models that are connected to these solutions are explored.
Our application will be aimed at teams who want to improve their coaching abilities using technology. For this reason, five competing technologies were researched that are already in use in this field. Our technology will aim to replace most of the competing technologies, since our technology provides more information than the existing technologies.
Nevertheless, some of the current technologies can also be combined with our technology, this way a coach can gather even more information about his players and team performance.

STATSports tracker

The first competitive application is that of STATSports Tracker. STATSports tracker is a product that consists of a GPS tracker, to be worn during training and/or matches, and accompanying software. Using this software the data can be accessed and processed to generate heatmaps which will give an insight into how the player moves around the field. Various other metrics like Total Distance, Max Speeds, High Speed Running, Intensity and Strain Levels are also calculated by the software.
It is a competitor as it offers a companion app that allows for analysis of data the GPS tracker gathers. What sets us apart is our use of Network Analysis to go a step beyond the analysis STATSports offers, as well as our consultancy helping a coach to interpret the analysis.
The target customers for this product are sporting clubs, national teams, coaches and individual players. The demand for such a product is high since it helps develop players optimally while giving coaches insight into how players perform, allowing them to make a better training schedule.
STATSports tracker works with a bundling based business model, which means that they sell multiple products as one package, in STATSports case the software is bundled with the GPS tracker. The customer makes one purchase and owns the product forever.

Catapult One tracker

The second competitive application is that of Catapult one tracker. Like STATSports, Catapult One offers a product that consists of a GPS tracker, to be worn during training and/or matches, and accompanying software to access and process the data given by the tracker.
It is a competitor as it offers software that allows a coach to analyse team performance in multiple ways. What sets us apart is our use of Network Analysis to go a step beyond the analysis Catapult offers, as well as our consultancy helping a coach to interpret the analysis.

Catapult One's target customers are players and coaches. Like STATSports the demand is high since it helps develop players, reduce the risk of injury, provide player comparisons and improve positioning. This allows for an optimal training schedule for better results on the pitch.
Catapult One works, opposed to STATSports, with a subscription and bundling based business model which means that they sell a bundle as a subscription. The user pays a monthly fee to become a member and be able to use the product. This fee is lower than the cost of outright buying the equipment, however, the customer does not have full ownership of the bought product.

Wireless network technology

Tracking player performance is getting more and more popular in team sports. The rising use of global positioning systems (GPS) and wearable sensors can be seen as these systems have the ability to gather large amounts of data. These systems have altered how players are evaluated, and nowadays tracking systems are heavily used by a lot of professional football clubs (James et al., 2013).
Wireless sensor networks (WSN) are networks that are not connected via cables. These networks are used to send data from point A to point B. It can be used to send the data of acceleration of a player to a computer for example. WSN is a technology that is used for GPS tracking, it is also used for data analysis, with this data certain characteristics of a player can be measured directly, e.g. acceleration. So, it is not a competing technology or similar technology to network technology but more of a complementary technology for GPS tracking.
Technical specifications of this technology are:

- Throughput (Amount of data send)
- Power consumption
- Sturdy against severe weather
- Fix errors with own system
- Low cost

The target customers of this technology are for example football clubs, hospitals, vehicle fabricators or the military. WSN has such a high number of customers because almost all companies and institutions need data to be transferred from one place to the other. The crucial problem faced by these customers is that they need data to be transferred wirelessly. The unique value proposition of WSN is that the technology is able to send large amounts of data without the use of unhandy cables. This is also the solution WSN provides.

Simulation

Another competing technology is the use of simulation. It is a competitor as we do not use simulation for network technology. With simulation a situation is created that mirrors a team sports setting. In our case that would be a passing network in a football pitch with nodes. First, data must be analysed of how the team functions, for future simulations. This is a combination of physical data (how the players move) and passing data (where the ball is passed). When this data is collected it can then be used to simulate the nodes.
By doing this, simulations can be done as many times as needed and possible outcomes of playing tactics can be found.

The technical specifications of simulation are:

- A device/computer (that is able to simulate)
- Physical data of the players
- Form of results

The target customers of simulation can be very broad, from football clubs to video game producers. The crucial problem is that for some ways of gathering data it needs specific circumstances. The unique value proposition is that these specific circumstances can be simulated, so there is no need to do it live. The solution is then that all possible outcomes can be derived, which can for instance be used for coming up with new tactics in football.

uPATO

Ultimate Performance Analysis Tool (uPATO) is a tool that already allows network analysis (of soccer matches). Besides network analysis, it also has some functionality for GPS analysis. The tool is free to use and can be used both as an Android app and a website. If you sign up for an account, you can use all of the services for free. However, there are some downsides to the way the system works. Firstly, the data has to be provided either in a specific format or filled in manually using their graphical interface. In this interface you have to manually register every pass and play that is made throughout the game. If a coach has to go through this process, it is a lot of effort to get the statistics. Secondly, the metrics are computed on the servers of the service. For this reason you can not control what happens with the data you submit. If data is published or released in any way, it might result in other teams using your data. Also the processing speed cannot be controlled, which might result in rather long waiting times to compute certain statistics. Another downside to this system is that it might be difficult to draw conclusions from the analysis. The reason for this is because all the statistics are listed as options, where no further explanation is provided for any metrics. The context of the network is not taken into account when computing the metrics, meaning that the tool does not tell you what the value tells about the specific network. So although this tool is quite powerful to compute a lot of statistics on networks, it does not necessarily help the coaches of a team sport, because no conclusions are provided.
Besides, the site lists all known network metrics, which the user can select and compute on the network. So the coach, unless he is an expert on network theory himself, will not know what metrics to compute and look at.
Using uPATO is entirely free, after you register an account. So they currently do not have a commercial business model which charges customers to make profit. If the service remains entirely free, then the creators do not have a commercial goal for this service. Or they might have some form of business model that is not directly paid for by the users. An example of this might be selling data that is obtained from the users. Although, in many cases this is a sensitive topic, where privacy laws do not allow this. It could also be the case that statistics are gathered on a large scale, over many networks from users, which might give insights in their networks and might be useful for research purposes for the creators.

Our application

Our business model will be based on a subscription service business model. Our business model will have 3 different packages, with 3 different prices for each package. The most important aspect of our business model is the consultancy that is included in the most expensive package. To differentiate from the closest competitor uPATO, we aim to provide software that the user runs locally, this way they do not have to share their relatively sensitive data and do not have to wait for processing on other hardware. To use the software, the teams will need an account which will also be connected to the subscription information. This allows the software to differentiate users and limit functionality for each package separately.

Business model canvas

Key Partners

We have several Key Partners. The first one is an App Developer, for this we look for a partner that is proficient at App Development and has some experience in the area of Network Analysis. This would speed up the development process, reducing development costs. The second partner is a Web Hosting service, for this we look for a secure service with low cost. The third partner is an Advertising bureau, for this we look for a bureau that is knowledgeable in the subject of football and can offer advertisement spots in places where target customers can see them. A potential fourth partner would be a GPS tracking manufacturer, for this we look for a manufacturer that does not offer an advanced application to process data such that we can add value to the product.

Key Activities

The Key Activities exist out of Producing, as we take raw data and produce graphs and workable information from it, and Problem Solving, as we offer consultancy to help customers find solutions to their individual needs.

Key Resources

Our Key Resources are Human, in the form of consultants and customer support, and Intellectual Property, protecting our app's inner workings from being used by other competitors.

Commented [Gp5]: IP does not emerge spontaneously, you have to file for it and this takes time. It is good practice to have IP rights and then go to the market. So, this relates to the go to the market strategy.

Value Propositions

Our first value proposition is Ease of Use as the app is plug and play. Upload any data and get right to work by simply clicking a button to get the desired analysis. The app does the rest.
Our second value proposition is Convenience as we and the app do most of the work. All the consumer has to do is click a button in the app (to gain information from data) or make an appointment with one of the consultants for quick help with a problem. Meanwhile the client does not need to have any Network Analysis knowledge.

Commented [Gp6]: Then, how does he know what he is looking at? What is the output going to be? Discuss in description of application below.

Customer Relationships

Our customer relationship is based on a package. The free package is a Self-Service relationship as the customer can download the app and start right away without any communication towards the company. However, the free package is also a Personal Assistance relationship as customer support can help the user with any questions/issues they might have.
The second package is a Personal Assistance relationship as customer support can help, as well consultants being available during the first few weeks can help and give an impression as to how the full service works.
The third package is a Dedicated Personal Assistance relationship as coaches have a certain consultant dedicated to them. They will work together closely to best help the coach with his/her problems and to get the most out of our product.

Channels

Multiple channels are used to communicate with (potential) customers. The first one is the Free version which serves as an introduction to the product and can attract new customers. The second channel is the App itself through which customer support can be accessed. The third is a website on which customer support can be accessed and information on the product can be found. The fourth is Email through which the customer can communicate with us. It can also serve as a way to contact potential customers, for example by reaching out to a football club. The fifth is the telephone through which we can contact potential customers. The sixth is advertisement in a football environment which is used to attract new customers. The seventh and final channel is social media through which new customers can be reached (for example YouTube promotional videos or Twitter posts) as well as allowing customer support (for example via Twitter or Instagram's direct messaging feature).

Customer Segment

The main customer segment is international football coaches at any level of play, this is our focus demographic. However, other parties like coaches from different sports, journalists or gamblers may happen to find use in our product in which case we will assist them the best we can.

Cost Structure

The cost structure exists out of employee salaries, marketing costs for advertisement, web hosting for the website (51.87eu a year with GoDaddy Economy plan), E-mail service costs for customer contact (Included in GoDaddy's subscription) and app development costs (Cytospace is free to use, however someone would need to be trained or an App Developer has to be hired, average app developer hourly rate is about 38eu in the Netherlands).

Revenue streams

The revenue stream exists out of the Subscriptions sold to the customers. A potential revenue stream would be from bundling the product with a GPS tracker provided by a different company like Adidas Runtastic.

Description of Application

Our application can apply multiple network measures on a passing network as described in the Technology description section. In short, a passing network is a network where players are nodes and there are weighted, directed edges between players that indicate the amount of passes from a player to another player. The set of different network measures that our application uses can each provide a different, useful insight in the passing network.
The first network measure is the density of a network. As also described in the Technology description section, the density of a network indicates the proportion of possible edges that are actual edges. While just a number would not be that useful, when combining it with the theory behind it and looking at the network itself, multiple insights can be gained.
For example, a network with a high density (close to 1) indicates that more players have successfully passed and received passes from distinct players. Similarly, since a high density means there are many edges going from nodes to other nodes, it indicates that players have many possibilities in choosing who to pass the ball to. However, note that having a complete network (a density of 1) is not necessarily optimal, since it indicates that passes have been made from strikers to the goalkeeper as well, which may or may not be good depending on the situation.
Considering a network with a low density (close to 0) and possibly looking at the corresponding network, different insights can be gained. For example, not having many edges could mean the ball is lost during a pass often (which negates getting an edge between two nodes) or a player is at the wrong place on the field many times, which makes it (close to) impossible to pass to that player.
Another network measure that can be applied on a network in our application is the measure of bridges and structural holes. As also described in the Technology description section, structural holes are places in a network where the lack of a fitting player, a bridge, would result in a complete collapse of cohesiveness in the entire network. Naturally, the goal is to have as little structural holes as possible, preferably zero.
A structural hole can be imagined as a player that is the centre of a team. For example, a structural hole could be a midfielder that connects the defenders with the attackers. Logically, when that player is removed somehow (i.e. a red card), the cohesiveness of the team goes down immensely. To achieve the goal of having the least amount of structural holes it is important to apply this measure to see if there are any structural holes.

When structural holes are found they can be solved by making sure there are still edges going from one side of the structural hole to the other side.
Consider the passing network with the midfielder(s) as (a) structural hole(s). To remove this/these structural hole(s) it could be an option to have defenders and attackers also pass between them such that they are not completely reliant on the midfielder(s).
Another example is the passing network where the attackers are the only players trying to make goals, i.e. they are the structural holes in the connection of the team to the opponent's goal. To remove these structural holes there should be other players that also try to score goals when (a) attacker(s) are removed from the game somehow (i.e. an injury).
The final network measure that can be used in our application is the measure of variability of a node. The variability of a node indicates how predictable a node is, i.e. how likely it is that some player passes the ball to some other player, as also described in the Technology description section.
For example, consider a node with a high rate of variability. This can work two ways; on the one hand a player with a high rate of variability can be hard to predict, which makes it difficult for the opposing team to defend a certain side of that player (since they cannot know in what direction the unpredictable player is going to pass). On the other hand it might end up being difficult to form a synergetic group with that player since that player is so unpredictable. Therefore it is important to decide what a team needs, this measure will only indicate the predictability of a player, it is up to the coach to interpret the results since every team is different.
Considering a node with a low rate of variability the opposite happens. On the one hand the opposing team can easily predict to whom the predictable player is going to pass, which might make the opposing team focus on that connection to have a higher intercept chance. On the other hand, if that player almost always passes to one specific player, it means, due to how the variability of a node is calculated, a lot of passes do connect. This indicates that there is a very strong connection between those players. A coach could, for example, make the receiving player try to score goals and the unpredictable player will get the ball to the receiving player a lot of the time.
In conclusion, our application provides different insights in a passing network by applying the following network measures:

- The density measure, to find out whether or not many passes can be made to many different players.
- The bridges and structural holes measure, to find out whether or not the team is too reliant on one or more players.
- The variability of a node measure, to find out how predictable a player on a team (or opposing team) is.

Each of these network measures can provide a coach with a different, useful insight into a team and/or player.

Readiness of FIT

The business model will be ready for market introduction when the 3 different subscription packages can be offered. The FIT, network technology, is sufficiently developed for our application. This is because enough research is done on the subject. A lot of network science formulas have been derived and are already used to analyse team sports. The

development the FIT needs is the flow of breakthroughs that could be implemented in our application.
For the first package, the free package, a limited amount of network metrics are computed with no context given. The first subscription is meant for football clubs to get to know our product. This package is automated with coding, using Python software. A data file can be read, the metrics are computed and are given as output. If the consultants are skilled enough in programming the code can be written by the consultants otherwise a programmer has to be hired. This package is free, but the club has to follow us on all social media platforms (Instagram, Facebook & Twitter) and mention us on their own platforms. This can be on their story, or a regular post. As social media is used a lot in the football community it can draw attention to us, and possibly new customers.
The second package, which is meant as a trial package, gives a more in-depth explanation and insights into the network metrics that are computed. Extra layers of code have to be programmed to automate this process as well as possible. In this code definitions of the metrics need to be given with a list of conclusions that are drawn from the data analysis.
As a company, we try to sell the third package as this is where the most money is made. The second subscription package gives a couple of hints of how the third package works, the customer's/football teams need to be in some way dependent on the consultants so that the third package is sold.
All content from the first 2 packages is included in the third package, consultancy is added to give feedback and help with interpreting the network metrics. In this way, a close relationship could arise between the company and the football club. A plan is made on which consultants work on what kind of metrics. This is for the football clubs to determine, but the company advises a season-long membership. Some other teams must be excluded from the use of the company’s software and consultancy. The consultants have to make sure that this is complied with.

Recommendations

Technological development

Technological dimensions are the concern for methods and tools to accomplish tasks. In this case, the technological dimensions of the FIT. Numerous methods and tools need to be used to accomplish market introduction. To start, data of passing networks need to be gathered. Most of the football teams already use GPS and wearable sensors to track their players. Then computing power is needed to turn this tracking data into passing networks. When these passing networks are defined, metrics can be computed to improve the performance of the team. (data needs to be found → data is processed → processed data is analysed → advice is given // the tech. dimensions that are used to accomplish this are these methods/tools: GPS and wearable sensors using WSN, consultants, computing power (computers))

The pace of development that is needed is mainly based on the programming of processing raw data and computing metrics. Other important aspects for market introduction such as social media platforms or consultancy do not require a lot of work beforehand but need consistent input. When the programming parts are done, the social media platforms need to be maintained just as the relationships with the coaches as the consultants work directly with the clients.

Commented [Gp7]: Check : different paragraph format than preceding text.

The types of cost involved in our business are marketing, materials, transport and wages. Marketing will be done mainly via social media, only advertisements cost money as registering is free. The price of advertising is calculated per click or per impression. We would use CPM cost-per-impression which costs $6.70 which amounts to €6.10 per 1000 impressions for one advertisement. Materials consist of laptops, phones and software, all of which we already own, if additional materials are required it would cost around €2000,- per consultant for a laptop, work phone and installed software. Transport is necessary to work directly with the clients. The transportation of the consultants can be in the form of a car or aeroplane. When a partnership is established between the company and football club the transportation costs are negotiated to be on the club. The employees need to be paid, the consultants are paid per hour in addition to the subscription package. A football analyst makes around €40,- per hour. A football season is 4.5 months long, so with a 40 hour work week, one consultant earns around €29.000,- per season on salary.

Proposal for most suitable business model

Based on our business model canvas, specifically our customer segment, we deem convenience to be a promising value proposition. To elaborate, being a football coach is a busy job and eliminating some of their work such that they can dedicate more time to other important aspects of their job is an interesting value to offer. As such we decided to go with the following business model.

The business model is obtaining customers through a free package, then convincing them to switch to a package with a monthly fee. The features of the paid packages are promoted to the users to convince them to change their subscription.

Commented [Gp8]: Customer acquisition and retention.

The first package that we provide is an entirely free package. This package will allow you to compute several metrics on network data that is provided by the customer. The metrics that you can compute are those that have value for the coach, however no information about the value's context is provided. So, for most coaches and teams this value will not provide a lot of insights in the team. This package aims to be a competitor of uPATO, because this tool is also entirely free. The main difference between uPATO and this package is that we will not provide all known metrics, but solely the "useful" metrics and that we compute metrics locally. Although this package does not make money directly, it will help in drawing in new customers.

The second package will provide some more insights in the team and is aimed at users who do not want to spend too much money and time on improving coaching using our technology. It also serves as a "trial" package, where users of the free package can discover some of the possibilities of our most expensive package, before paying for that price. This middle package will provide some default explanations for the computed metrics and thus allow the users to gain more insights in the way network theory can help the coaches. It also includes limited contact with consultants, where during the first few weeks the coaches can ask for advice based on their metrics.

Lastly, the most expensive package will aim to really provide the teams an advantage over opposing coaches. Most importantly, the coaches have access to professional consultants, who will provide the coach with advice for their own team. Besides this, the consultants will also look into the network of opposing teams, this way they will be able to assist the coach for each match. Consultants are connected to a team, such that a consultant will gain insight in the team and be able to understand both the coach and the metrics better. Another useful feature of this package is the exclusiveness of the system. If a team uses this package, they are allowed to exclude two competing teams from the same league from using the consultancy package. This way if they pay for this package, they are guaranteed to have exclusive access to the system. Thus they gain an advantage over competitors who are not allowed to use this package.

A different way to generate revenue is to bundle our Network Analysis app and consultancy with another product. A suited product would be a GPS tracker as these are required to obtain the data used by the app. This way customers can generate and process data with the same bundle.

To make this happen, certain partners are needed. In particular partners in App Development, Web Hosting, Advertisement and GPS tracking manufacturing. Cytospace would be a good candidate for App Development as they have great experience in the field of Network Analysis as well as App Development. For Web Hosting GoDaddy would be a good candidate as they offer VPS hosting. This means that the website will be stored in a server with other websites, but not as much as with pure Shared Hosting, which results in the other websites not having much effect on our website thus making downtime due to other websites on the same server much more unlikely. The reason for not choosing Dedicated Hosting is that this is reserved for websites with a big need for server resources (which we don't require) and as such costs more. Another reason to go for GoDaddy Hosting is that they offer excellent site security with their advanced DDoS protection and a free SSL certificate. For Advertising, Project11 would be suited as they offer advertisement during live football games but also allow for expansion to other sports for potential new customers.

For GPS tracker manufacturing Adidas Runtastic would be a good candidate as the app Adidas offers with the product is not very advanced in Network Analysis. Adding our product could open another door for Adidas to expand within the football industry. This would allow for cooperation on a service that allows data gathering and processing+consultancy, as such broadening the possibilities.

Proposal for most suitable go-to-market strategy

The go-to-market strategy is to get experience and good results by providing more services for a lower subscription fee. The package will be advertised to companies in lower divisions in the Netherlands. This advertising also includes a lower cost for all packages. This should provide a number of customers, who will be used to test and fine-tune the services and consultancy. The experience that is obtained in the first months using this strategy is very valuable. Which makes the reduced income from fees an acceptable trade-off.
Besides the experience, this strategy also serves to prove that the product is worth it. The results of the services of the first teams can be used to advertise the services to other teams. Teams in the higher leagues will very likely be harder to convince, because the stakes are higher. The results of the services in lower leagues can be used to show that the services work, thus to convince these teams.
Once the company has enough experience and multiple customers that would recommend the services, the strategy is dropped and we adopt our business model as described in the previous section.

Conclusion

Our project started off with researching Network Analysis. Ten papers were used from which we carefully selected applicable forms of Network Analysis that were usable in our design later on. Interesting to note here were that certain types of Network theory were reused throughout a number of papers. The candidates that we eventually ended up using were Network Density, Bridges and Structural Holes and node variability. These are all types of Network theory that were not used in a lot of papers (Density in two separate papers, Bridges/Holes and Node Variability in one, separate, paper) which would help us to create something unique opposed to the theory that was already extensively covered while still being relevant and usable in the desired product.

Afterwards we looked into possible applications. Here we identified markets in which our product could be used. We found Sports Analytics, Sports Betting and Sports Journalism to be suitable markets for our product. From these markets we selected Sports Analytics, and in particular international football coaches, as the current technology is in a better stage development wise for this application.

Understanding the current market was up next. Here we looked around for technologies similar to ours. This way we could find holes in the market which would allow us to enter said market. Important here was that while apps with Network Analysis already existed (STATSports tracker+app, Catapult One tracker+app and uPATO being some examples), there still was a lack of explanation of how to interpret and use the processed data

effectively within these apps. This led to the idea of offering consultancy as well which would fill the market gap and allow us to stand out from the crowd.

Now it was time to put our business together. This started off with filling out the business model canvas, giving us a framework to build our business around. What stood out here was the lack of revenue streams, namely only the Subscription. To add an additional stream we came up with the idea to bundle the subscription with a GPS tracker. What also stood out was the amount of partnerships we would require, leading us to have to find four suitable candidates to partner with. We also further detailed the technology we would end up using, in here the basic theory behind the chosen measures is expanded upon with examples and ways to implement them in analysis of passing networks. Finally, we researched how much the FIT would have to be developed in order for it to be ready for market introduction. We found that the business model will be ready as soon as the three packages we created are ready to be offered while underlying technology has been developed enough for use.

For our future plans we first described the needed technological development, which are the technological dimensions of the FIT. We found that the pace of development is based mainly on the programming of processing raw data and computing metrics. Whereas the other aspects like Social Media platforms and Consultancy setup require a relative low amount of work. Next-up a proposal for the most suitable business model was given in which the challenge was to find suitable partnerships. Cytospace, GoDaddy, Project11 and Adidas Runtastic were found as candidates, showing us that the required partners are readily available. A proposal for the most suitable go-to-market strategy was also given. We had come to the realisation that it can be hard to gain customer trust and therefore we decided to offer our products at a discount, lowering customer risk and in turn making them more likely to try out our new product. Once this turns successful we can turn to the original business plan.

Appendix A: Project organisation

For every meeting with the supervisors a chair and secretary are appointed. The chair and secretary are rotated, such that every group member fulfils each role at least once. Besides these responsibilities during the meetings, each group member has their own responsibilities throughout the project. Each group member is responsible for completing their tasks in time and with good quality. It is expected that everyone contributes equally towards the deliverables. Also, each group member is responsible for the quality of the deliverables. Thus, all members are expected to review the work of other members, to guarantee that everyone is satisfied with the deliverables.
In the table below the planning and deliverables are listed. The meetings on Thursday occur after a meeting with the group members beforehand. The deliverables are marked bold.

	Monday	Thursday	Sunday
Week 1	Intro Meeting	USE Meeting	
Week 2	Group Meeting		**Appendix A**
Week 3	Group Meeting	USE Meeting	**Draft Report Chapter 1**
Week 4	Group Meeting	USE Meeting	**Draft Report Chapter 2**
Week 5	Group Meeting & Intermediate presentation	TECH & USE Meeting	**Draft Report Chapter 3 & Peer assessment I**
Week 6	Group Meeting	USE Meeting	
Week 7	Group Meeting	TECH & USE Meeting	**Draft Report Chapter 4**
Week 8	Group Meeting	Final Presentation	**Final Version & Peer assessment II**

The goal of this project is to answer the following question:
How can analysis of team sports using network technology yield the basis for a sustainable business case?
This question will be answered by following the proposed plan below.
The goal of the project is to develop a new business model, which is based on using network theory to analyse team sports. To obtain a concrete business case, we should focus on one specific team sport. Network theory can be applied in (almost) infinitely many ways, also in team sports. Therefore, it is important that we find a suitable way, which provides new insights, which are not already covered by other methods of analysing the team sport. The analysis using network theory could provide new insights in the team's performance or other relevant statistics.

Development of this business model requires a good analysis of the possibilities of the technology within the specified domain. In this case, the possible uses of network theory in team sports. When this analysis is concluded, it is required that we look into the business cases that can be based on the technology. The first things to come to mind are mostly focussed on improving the coach's understanding of his team's performance, or the performance of the opponents, however more business cases may come up once we have done more research on the technology.
In order to obtain an overview of the possibilities in both technology and business, we should interview multiple parties. At least one of the parties should provide us with more information and ideas of uses of the network theory. This party will most likely be represented by a researcher, who works on network theory. Whereas, another party should be interviewed to determine what the possibilities of the technology might be in the area of team sports.This party will probably be a sports coach.

Appendix B: References

1. Arriaza-Ardiles, E., Martín-González, J.M., Zuniga, M.D., Sánchez-Flores, J., de Saa, Y., García-Manso, J.M., 2018, Applying graphs and complex networks to football metric interpretation, Human Movement Science, Volume 57, ISSN 0167-9457, Pages 236-243, https://doi.org/10.1016/j.humov.2017.08.022.
2. Peña, J. L., Touchette, H., 2012, A network theory analysis of football strategies, https://doi.org/10.48550/arXiv.1206.6904
3. Yamamoto, Y., 2010, Scale-free Property of the Passing Behaviour in a Team Sport, International Journal of Sport and Health Science, https://www.jstage.jst.go.jp/article/ijshs/advpub/0/advpub_IJSHS20090014/_pdf/-char/ja
4. Ramos, J., Lopes, R.J. & Araújo, D., 2018, What's Next in Complex Networks? Capturing the Concept of Attacking Play in Invasive Team Sports, Sports Med 48, Pages 17-28, https://doi.org/10.1007/s40279-017-0786-z
5. Martins F., Gomes R., Lopes V., Silva F., Mendes R., 2021, Mathematical Models to Measure the Variability of Nodes and Networks in Team Sports, https://doi.org/10.3390/e23081072
6. Cotta, C., Mora, A.M., Merelo, J.J. et al., 2013, A network analysis of the 2010 FIFA world cup champion team play, J Syst Sci Complex 26, Pages 21–42, https://doi.org/10.1007/s11424-013-2291-2
7. Lusher, D., Robins, G., Kremer, P., 2010, The Application of Social Network Analysis to Team Sports, Measurement in Physical Education and Exercise Science, Pages 211-224, https://doi.org/10.1080/1091367x.2010.495559
8. Korte, F., Lames, M., 2018, Characterizing different team sports using network analysis, Current Issues in Sport Science, 3, https://doi.org/10.15203/CISS_2018.005
9. Heikinheimo, A,, 2019, In-Game Strategy Recommendations in Association Football : A Study Based on Network Theory, https://urn.fi/URN:NBN:fi-fe201903259818
10. James, D.A., Fanella, L., Cusani, R., 2013, Near Real Time Network Simulation for Team Sports Monitoring, Procedia Engineering, Volume 60, Pages 422-427, ISSN 1877-7058, https://doi.org/10.1016/j.proeng.2013.07.049.
11. Martins, F.M.L., Silva, F., Clemente, F.M., Gomes, A.J.P., Correia, A.F.P.P., Nguyen, Q.T., Sequeiros, J.B., Ribeiro, J.S., Lopes, V.F., 2018, Ultimate Performance Analysis Tool (uPATO). Home page: http://uPATO.it.ubi.pt

Appendix C: Business Model Canvas