Why Is This So Hard? Insights from the State Space of a Simple Board Game

Mareike Bockholt(✉) and Katharina Anna Zweig

Graph Theory and Complex Network Analysis Group,
Department of Computer Science, TU Kaiserslautern,
Gottlieb-Daimler-Straße 48, 67663 Kaiserslautern, Germany
{mareike.bockholt,zweig}@cs.uni-kl.de

Abstract. Serious Games research has become an active research topic in the recent years. In order to design Serious Games with an appropriate degree of complexity such that the games are neither boring nor frustrating, it is necessary to have a good understanding of the factors that determine the difficulty of a game. The present work is based on the idea that a game's difficulty is reflected in the structure of its underlying state space. Therefore, we propose metrics to capture the structure of a state space and examine if their values correlate with the difficulty of the game. However, we find that only one of the metrics, namely the length of the optimal solution, influences the difficulty of the game. In addition, by focusing on the part of the state space, which is actually explored by human players, we can identify properties that predict the game's difficulty perceived by the players. We thus conclude that it is not the structure of the whole state space that determines the difficulty of a game, but the rather limited part that is explored by human players.

Keywords: Serious games · Human problem solving · Complexity · Rush hour · Network analysis

1 Introduction

Serious Games development has become a growing field in research and industry in the last years. The idea of embedding a purpose, like mental or physical training, into a gaming environment has proven to be a fruitful approach. The core of many Serious Games, especially those concerned with cognitive training, contains a game logic which determines how challenging the game will be for players. In the process of developing Serious Games, it is thus of essential importance to understand what makes a game logic challenging for humans. Otherwise, it would be a matter of chance to design adequately difficult games which are neither too easy – therefore boring for players – nor too hard – and therefore frustrating.

Understanding how humans deal with tasks they are faced with and what they perceive as difficult is of great relevance in the field of *complex problem solving* [1]. Complex problem solving is a broad and active research field [2,3] which

S. Göbel et al. (Eds.): JCSG 2015, LNCS 9090, pp. 147–157, 2015.
DOI: 10.1007/978-3-319-19126-3_13

has produced a wide range of results about how humans solve problems, which heuristics they use, and which properties determine a problem's complexity. The present research focuses on the question which factors determine the difficulty of a problem. There exist several research approaches to identify these factors: Halford et al. [4] propose *relational complexity* as a measure for a problem's difficulty. They introduce the idea of a relation which contains the elements that need to be processed in parallel in order to solve the problem. The dimension of such a relation, hence the number of contained elements, is proposed as a complexity measure. According to Halford et al., humans are only able to process relations of dimension *four* or less, larger relations need to be split and processed serially. Therefore, the number of elements which need to be processed in parallel, can be used as a complexity measure. However, determining the relations and their dimension for a board game, and therefore, determining its difficulty with this method, is not obvious at all.

Kotovsky et al. [5] choose a different approach and show that the difficulty of a problem is strongly dependent on the cover story under which the problem is presented to the player: they presented the players different representations of the problem *Tower of Hanoi* and discovered that the solving performance varies considerably depending on the representation. However, these findings do not explain the *intrinsic* difficulty of a problem, since the *problem* is not varied, but the *representation*. Our focus of interest lies in the difficulty of the problem itself: which kind of structure makes a problem difficult, which one makes it easy – independent of the representation or cover story? In order to address this question, we use the concept and analysis of a *state space* of games as it is defined in the following. As a proof of concept of the proposed method, it is applied to a simple board game with defined states and defined rules of how to change these states for which it is feasible to compute the complete state spaces. Nevertheless, choosing such a simple board game might be of benefit for the development of Serious Games since gaining a general understanding of what is the dominant factor influencing the difficulty of a game, is an important prerequisite for designing appropriately difficult games.

2 Approach

The present work investigates the complexity of a simple board game. Our research is based on the assumption that the structure of the state space of a problem (i.e., of a board game configuration) should reflect the difficulty of the problem. For example, we assume that the size of the state space influences the perceived difficulty of a game, or that the average number of applicable rules per state correlates with the perceived difficulty. A motivation for this assumption can be seen in Figure 1 which shows the state spaces of games with a low and with an advanced degree of difficulty. It is intuitively clear that the structure of the state space should be related to the difficulty of the game. The present work aims to systematically investigate this relationship.

Anderson et al. define problem solving as a "goal-oriented sequence of cognitive operations" that transforms a present state into a desired state [6]. From this definition, the concept of a state space, already proposed by Newell in 1979 [7], arises almost immediately: a state space is a graph $G = (V, E)$, with V the set of all game configurations reachable from the start state by a series of allowed moves, $E \subseteq V \times V$ the set of possible moves. Hence, the state space (two examples are shown in Figure 1) for a fixed game configuration contains one node for each configuration which can reached from the start configuration by moves. The connections between the configurations represent possible moves. In this setting, problem solving consists of the task to find a path through the state space from the start to a goal state, which can be seen as a searching task. The representation as a graph allows us to use ideas from *complex network analysis* [8] to assess the complexity of the structure of the state space.

However, the present work shows that simple metrics to measure the structure of a state space are either not able to capture its structure in sufficient detail, or do not correlate with the problem's difficulty at all. For this purpose, the present paper is organized in two parts: we first introduce several metrics which are based on the structure of the state space and then show that they are – except of one – independent of the problem's perceived complexity. This can be confirmed by an experiment. As a second approach, we focus on the idea that the structure of the whole state space itself is less important than the part that is actually *explored* by a player. The first (qualitative) result is that the solving methods of different players are very similar to each other. Based on this, we focus on the part of the state space that is actually used by the human players. We therefore introduce measures that quantify aspects of the paths used by the participants in an online experiment and show that these measures do correlate with the perceived difficulty of the game. We thus provide evidence that the possible ways to solve a problem are less important than the ones that are normally used to solve it.

3 Analyzing the Structure of a State Space

The board game of interest is called *Rush Hour*.[1] It takes place on a grid of 6×6 fields, representing a parking lot, with one exit (cf. Figure 1). Cars of width 1 and length 2 or 3 are placed on the board vertically or horizontally and can be moved forwards or backwards as long as the needed fields are not occupied by any other car. Cars cannot move sideways and are not allowed to change their row or column, respectively. Given a configuration of cars placed on the grid, the goal is to find a sequence of moves that allows a particular car (the rightmost car in the third row, in Figures 1a and 1c the black one) to be moved from the board through the designated exit. *Rush Hour* is well suited for this research for several reasons: it is easy to understand how to play, yet it is still possible to design arbitrarily complex games as well as very easy ones, i. e. the range of complexity of possible games is broad and diverse. Furthermore, it is not obvious at all what determines the difficulty of a game.

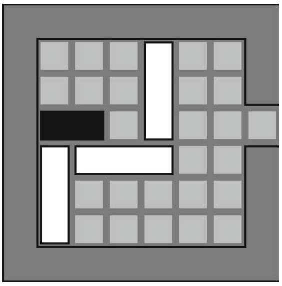
[1] The game was invented by Nob Yoshigahara and is distributed by ThinkFun Inc. and HCM Kinzel (Germany).

Figure 1 shows two examples of state spaces. The first one is that of a version of a game which is designed for children, the second one for experienced players. It is obvious that the second network is larger and more complex. This finding was the starting point for devising complexity metrics quantifying the structure of the state space.

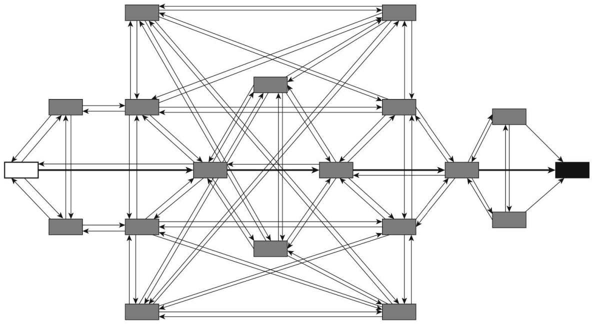
In [9], we introduce 17 metrics for Rush Hour start configurations based on the structure of the state space, following the assumption that these network analytic metrics reflect the complexity of solving the game. Out of the 17 defined metrics, only eleven are presented here, since the metrics which are left out are structurally similar to the ones presented here and do not provide any further insights. Detailed information are provided in [9]. The metrics are based on the idea that the perceived difficulty for solving a game could depend on several factors, as exemplified in (i) to (viii) where the metrics are indicated in italic. These factors are

(i) the size of the state space (number of nodes and edges), since more states may need to be explored (leads to the two metrics *nodes* and *edges*),
(ii) the number of possible moves in every state (leads to the metric *avdg* as the average node degree),
(iii) the minimal number of moves needed to reach the goal state (*lsp* as the length of the solution path),
(iv) the number of correct moves relative to the number of possible moves in every state (*br* as branching complexity), whereas correct moves means all moves which decrease the distance to a goal state,
(v) the number of possible shortest solution paths (*sp* as the number of shortest paths),
(vi) simple board game properties (*cars* as the number of cars and *fields* as the number of occupied fields on the board),
(vii) the average number of cars which can be freely moved in every state (*mc* as movable cars), since a smaller or larger number of objects which can be chosen as part of the solution way might influence the difficulty, and
(viii) the number of counterintuitive moves required in a solution path (*cm* as the weighted number of needed counterintuitive moves in a solution path and *cmpl* as this number normalized by the solution path length).

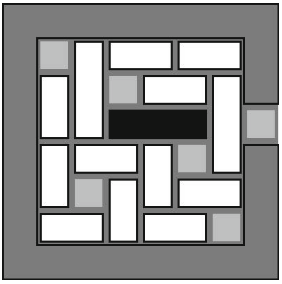
Point (viii) is based on findings from research in cognitive psychology which identify heuristics humans apply for solving a task. The most often used heuristic is called *hill-climbing* [10]: in this strategy, the current situation is compared with the desired situation, and the operator which yields a more similar situation to the solution is chosen. In our game, the goal is to unblock the black car and move it forward to the exit. A human playing the game according to the hill-climbing method will try to successively remove the blocking cars out of the way of the black car and to successively move it towards the exit. But there are starting configurations for which the solution requires moving the black car backwards or temporarily blocking the black car by another car. Because these kinds of moves contradict the hill-climbing method, we call these moves *counterintuitive moves* and suppose that a larger number of counterintuitive moves needed in a solution



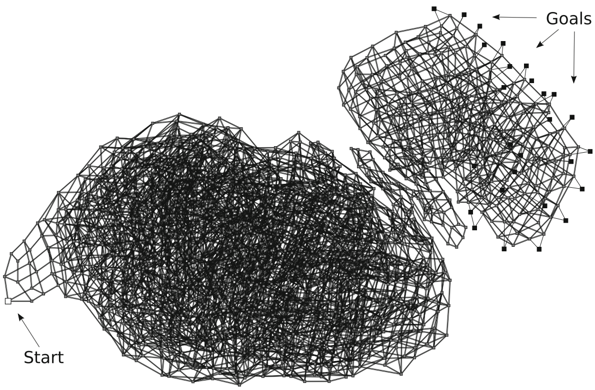
(a) A game configuration designed for children. The black block needs to be removed from the board by moving it through the exit on the right side. The white blocks are cars that can also be moved in their row respectively column. The other cells are unoccupied.



(b) The state space belonging to the board configuration in Figure 1a. Each node represents a board configuration, the edges represent changes of the cars' position. The white node on the left is the the start configuration shown in Figure 1a, the black node is the configuration in which the black car can be moved from the board. The shortest solution path of length four is indicated by bold edges.



(c) A game configuration of moderate difficulty designed for adults. As in the children's version, the black car needs to be removed from the board through the designated exit.



(d) The corresponding state space to the board configuration shown in Figure 1c. The start node is in the lower left corner, the solution states in the upper right corner.

Fig. 1. Two examples of Rush Hour game configurations and their corresponding state spaces

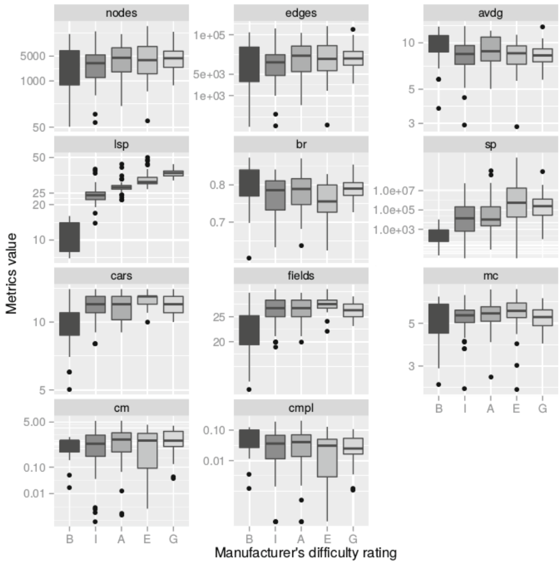
should increase the difficulty of the game (*cm*). To reduce the influence of the solution path length, the metric *cm* is normalized by the length of the shortest solution path (*cmpl*).

In order to test the introduced metrics, an objective measure for the difficulty of a board game is needed. With an objective difficulty measure at hand, one could compare the proposed metrics' values with the true difficulty of the game and check if there is a connection. Though, in general, there is no objective measure for the difficulty of a board game. Therefore, in this article, we use two different sets of difficulty measures to approximate the true difficulty. The first set contains a single measure: the categorization by the manufacturer who classified the games into five categories (beginner (B), intermediate (I), advanced (A), expert (E), and grand master (G)). The second is a set of four measures and is based on online experiments with human players: (i) the perceived difficulty as rated after solving the game, (ii) the participants' average solving time, (iii) the participants' average number of moves, and (iv) the participants' average number of moves normalized by the minimal number of necessary moves. The manufacturer of the Rush Hour game provides five game card sets with start configurations of five different levels of difficulty. Leaving identical configurations and configurations with a slightly different goal aside, 173 start configurations with the manufacturer's rating of complexity are available. The game logic and a breadth first search from every start configuration was implemented in order to create the state spaces and to compute the metrics for every configuration.

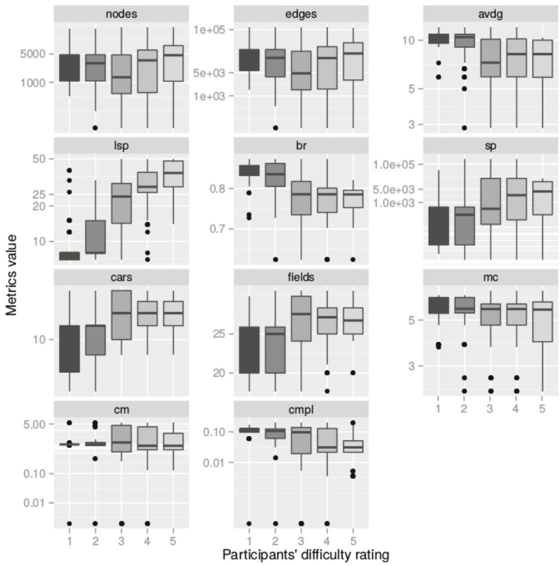
We organize the results in two different sections, depending on the set of difficulty measure the metrics' values are compared with: first, we compare the metrics' values with the manufacturer's difficulty rating. We find that there is no correlation for most of the proposed metrics between its value for the game and the game's difficulty rating. The results are visualized in a box plot diagram shown in Figure 2a in which each metric's values are plotted, ordered by manufacturer rating (beginner, intermediate, advanced, expert, and grand master). It can be seen that the only correlation is between the difficulty and the metric *lsp* which is the length of the solution path. It is a surprising finding that none of the other metrics shows any connection to the difficulty of the problem as categorized by the manufacturer.

As a second set of difficulty measure to compare the metrics with, experimental data is used: we selected 24 games of different difficulty level and conducted a study in which each of the 74 participants played at least six of the selected games. After solving a game, the participants were asked for a difficulty rating. The experiment was conducted as an online study, i. e. the game was browser-based such that the participants could participate at any time and at any place. All moves the players did were logged with timestamps. It was made sure that every participant attempted every game at most once. The majority of the participants indicated that they have not played the game *Rush Hour* before.

From this study, the second set of difficulty measures as described above can be derived. As in the previous analysis, we found that only the solution length



(a) Values of the complexity metrics categorized by the company's difficulty classification (beginner (B), intermediate (I), advanced (A), expert (E), grand master (G)). The scales are logarithmic. Games from the junior edition are excluded from analysis.



(b) Complexity metrics plotted versus the participants' difficulty rating. The scales are also logarithmic.

Fig. 2. The introduced complexity measures, compared with the manufacturer's difficulty rating and with experimental data

based metric correlated with the difficulty rating assigned by the participants; no correlation was found with any of the other metrics (cf. Figure 2b). All the other participants' based difficulty measures as defined above also do not show any correlation with the state space metrics – except for the solution length based one.

4 Quantifying Navigation Paths in the State Space

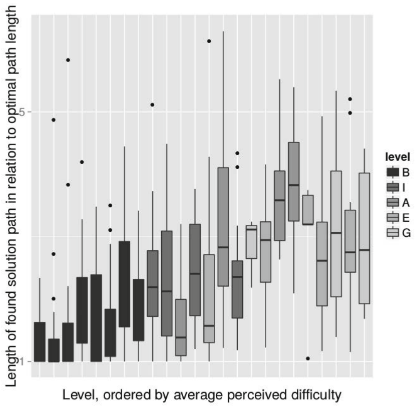
Since the proposed state space metrics, except of one, do not show a correlation with any of the difficulty measures, there are two possible explanations: either a game's underlying state space is completely independent from the difficulty of the game, or the proposed metrics are not able to capture the features of the state space which determine the game's difficulty. Thus, we decided to look more closely at how the participants navigated in the state space of the game while solving it, i.e., in the structure of the part of the space that is actually explored by the players. In Figure 3, we show the state space of a game and how the participants navigated through it. Both in this as well as in the visualizations for other games (not shown here), it is clearly recognizable that all of the participants preferred to take almost the same route through the state space although it is not necessarily the shortest one. This qualitative observation supports the assumption that the players are guided by the same heuristics in their solution strategy.

Fig. 3. Visualization of a state space and how the participants navigated through it while playing. The start node is elliptic and white (on the right side), visited nodes are black, not visited nodes are gray. Reached final nodes are black and elliptic, not reached final nodes are gray and elliptic (on the left side). The thickness of the edges shows how many from the 29 players took this transformation.

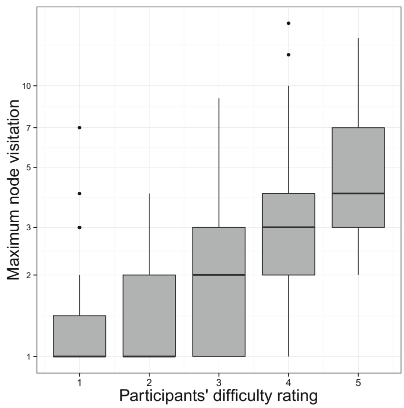
In order to approach the question of why it is more difficult to find a way through the state space to reach a solution for some versions of the games than for others, we examined the navigation of the participants through the state spaces more closely. The assumption is that participants lose their way while finding a solution, this could give them the impression that the game is harder.

Modeling that a player *gets lost* or is *losing one's way* can be done in several ways. If players struggle with finding a way to a final state, they will surely need more moves than necessary. Therefore, we consider how many moves the players needed in relation to the number of necessary moves. Indeed, a correlation between the perceived difficulty of a game and the number of used moves normalized by the number of moves in the optimal solution can be found as it can also be seen in Figure 4a.

The second approach to quantify a player's loss of orientation is to count the number of times the same state is visited within one trial of solving a game. For each state and player, we define as the *node visitation* the number of times the player visited this node while playing. As a measure for losing orientation while solving the game, the *average node visitation* and *maximum node visitation* are considered, the former being the mean value of all node visitations of all nodes visited, the latter being the maximum value for one player and one game.



(a) Relative path lengths of the selected games. The x-axis contains each of the 24 games which were played by the participants, the games' ordering on the x-axis is determined by increasing average difficulty rating by the players. The y-axis shows the length of solutions found by the players, normalized by the length of the optimal solution of the respective game.



(b) Maximum node visitation of all players and all games plotted against the participants' rating of the respective game.

Fig. 4. Two different approaches of capturing a player's loss of orientation

Figure 4b shows the *maximum node visitation* for all games and all players, ordered by difficulty rating of the players (the corresponding figure with the *average node visitation* is except of the scale very similar to this one). The figure reveals a significant relation of the node visitations to the difficulty classification of the players. Therefore, the degree of how much a player loses orientation is a good indicator for how difficult he or she will perceive the game. This leads to the conclusion that a problem's complexity does not only depend on objective properties, but there is also a high correlation with the individual performance. Note furthermore that the maximum node visitation takes on surprisingly high values, indicating that getting lost in a huge state space is a general issue in human problem solving. This could also be an important insight for the development of Serious Games since supporting the player not to visit the same state several times, can be realized in a Serious Game and avoid frustration on the players' side.

5 Conclusion and Outlook

The present work shows that although it seems obvious that the state space of a game is expected to reflect the game's difficulty for a human, the most intuitive metrics do not capture the game's difficulty. Only one not surprising metric, the length of the optimal solution path, shows a strong correlation to the game's difficulty. This finding is consistent with the results of Jarušek and Pelánek [11]. Jarušek and Pelánek also assume that the structure of the state space of a problem should determine the difficulty of the problem, and propose different metrics to measure the structure, but find poor correlations to the difficulty as well [12]. However, a qualitative observation of the participants' navigation through the state space confirmed the known fact that humans follow common methods in solving a problem and that this is constrained to a small part of the state space that is actually used by them. By focusing on this part of the state space, which is explored by the human players, we can identify properties that predict the perceived difficulty of the players. We thus conclude that it is not the structure of the whole state space that predicts the difficulty of a game, but the rather limited part that is explored by human players, which is, moreover, less individual than previously thought.

In future work, it might be interesting to consider the structure of reduced state spaces of *Rush Hour* due to the following reasons: in the course of a game, there are situations in which two or more cars need to be moved, but it does not matter in which order these moves are taken. Though, each of the possible move orders induces its own path in the state space while actually representing the same moves. In other situations, it is not of importance for the further game if a car is moved one cell more or less when it blocks the same set of other cars in both situations. Nevertheless, all future game states are multiplied by the number of equivalent move possibilities since every distinct position of a car generates a new state in the state space, even if they represent very similar situations. This is also reason for the large size of the state spaces and might explain why the size of the state spaces does not show any correlation to the difficulty of the game. Therefore, it might be worth shrinking the state spaces in the sense that states with the same meaning as described above are merged into one state. This step should decrease the size of the state spaces enormously and might change the values of our defined metrics. It is imaginable that the size of the reduced state spaces then allows a prediction of the difficulty of the game.

In addition to that, in order to generalize the results to the domain of Serious Games, other games need to be considered. The structure of state spaces might be totally different for other games which might influence the difficulty. For example, not all games generate state spaces in which all moves are reversible: there might be moves which do have a greater influence on the further game course than others since they make the player leave a part of the state space which can not be entered again. Further research needs to show if the presented results can be transferred to different games.

References

1. Frensch, P., Funke, J.: Complex problem solving: The European Perspective. Lawrence Erlbaum Associates, New York (1995)
2. Funke, J.: Problemlösendes Denken (2003)
3. Newell, A., Simon, H.A.: Human Problem Solving. Prentice Hall, Englewood Cliffs (1972)
4. Halford, G.S., Wilson, W.H., Phillips, S.: Processing capacity defined by relational complexity: implications for comparative, developmental, and cognitive psychology. Behavioral & Brain Sciences **21**, 803–865 (1998)
5. Kotovsky, K., Hayes, J.R., Simon, H.A.: Why are some problems hard? Evidence from Tower of Hanoi. Cognitive psychology **17**, 248–294 (1985)
6. Anderson, J.R.: Cognitive Psychology and its Implications. W. H. Freeman and Co., New York (1980)
7. Newell, A.: Reasoning, problem solving and decision processes: the problem space as a fundamental category (1979)
8. Newman, M.E.: Networks: An Introduction. OUP Oxford (2010)
9. Bockholt, M.: Ein netzwerkanalytischer Ansatz zur Untersuchung der Komplexität des Rush-Hour-Spiels. Bachelor thesis, Heidelberg University, Institute for Computer Science (2013)
10. Anderson, J.R.: Problem solving and learning. American Psychologist 48 (1993)
11. Jarušek, P., Pelánek, R.: What determines difficulty of transport puzzles. In: Proc. of Florida Artificial Intelligence Research Society Conference (FLAIRS 2011), pp. 428–433 (2011)
12. Jarušek, P.: Modeling problem solving times in tutoring systems. Masarykova univerzita, Fakulta informatiky, Disertacnı práce (2013)