Journal of Humanistic Mathematics

Volume 9 | Issue 2 July 2019

Choose Your Own Adventure: An Analysis of Interactive Gamebooks Using Graph Theory

D'Andre Adams
Southwestern University

Daniela Beckelhymer
Southwestern University

Alison Marr
Southwestern University

Recommended Citation
D'Andre Adams, Daniela Beckelhymer & Alison Marr, "Choose Your Own Adventure: An Analysis of Interactive Gamebooks Using Graph Theory," *Journal of Humanistic Mathematics*, Volume 9 Issue 2 (July 2019), pages 44-59. DOI: 10.5642/jhummath.201902.05. Available at: https://scholarship.claremont.edu/jhm/vol9/iss2/5

Choose Your Own Adventure: An Analysis of Interactive Gamebooks Using Graph Theory

Cover Page Footnote

We would like to thank the Southwestern University SCOPE program for the funding used to complete this research. We would also like to thank Joel Martinez and Shianne Solis who provided assistance in collecting data the first two weeks of the project. Our work was funded through NSF S-STEM grant 1458600 and funds from the H-E-B Foundation.

Choose Your Own Adventure: An Analysis of Interactive Gamebooks Using Graph Theory

D'Andre Adams

Southwestern University, Texas, USA
adamsd@southwestern.edu

Daniela Beckelhymer

Southwestern University, Texas, USA
beckelhd@southwestern.edu

Alison Marr

Department of Mathematics & Computer Science, Southwestern University, Texas, USA
marra@southwestern.edu

Abstract

"BEWARE and WARNING! This book is different from other books. You and YOU ALONE are in charge of what happens in this story." This is the captivating introduction to every book in the interactive novel series, *Choose Your Own Adventure* (CYOA). Our project uses the mathematical field of graph theory to analyze forty books from the CYOA book series for ages 9-12. We first began by drawing the digraphs of each book. Then we analyzed these digraphs by collecting structural data such as longest path length (i.e. longest story length) and number of vertices with outdegree zero (i.e. number of endings). In this paper we discuss the results of statistical analyses we used to compare books by author, year, and reader preference. We also discuss numerous errors we found in the description of certain books and the publication of others.

1. Background

From 1979 to 1998, 184 books were published in the children's interactive book series, *Choose Your Own Adventure* (CYOA) [2]. There was a large gap in publication between 1999 and 2004 while R.A. Montgomery (the series founder and main author) formed the publishing company Chooseco and in 2005 he began reprinting some of the original series and some new books [2]. CYOA books create several paths in which the reader, who is also the protagonist, gets to decide how the story will continue and eventually end.

Recognizing that these interactive books have interesting mathematical structure in the form of a directed graph (digraph), we set out to analyze various properties of a sample of this classic book series. The CYOA series we analyzed is a collection of 40 books that we assume to be a representative sample of the over 190 books that have been published as part of the original or reprinted "classic" series. We chose this set of 40 books as they are sold by the publishers of the series as a collection called the Whole Enchilada. The collection contains books originally published before and after the publishing gap and contains books written by a variety of authors. On both sides of the publication gap there are books written by R. A. Montgomery.

A digraph is a collection of nodes (dots) and arcs (lines with an arrow) that connect any two nodes in a particular order. The nodes of our digraphs represent pages of text in our books. The arcs represent possible reader moves from one page to the next based on the command(s) given on the source node page.

The first step of our research was to convert each book to a digraph that represents all possible stories. We used the app `InstaViz` [4] to create the graphs of our books. While converting the books to `InstaViz` we made several assumptions:

- Ovals represent any page of text with a single command, or an outdegree of one.
- Rectangles represent any page of text with multiple commands, or an outdegree of more than one. These are known as decision pages.
- Triangles represent any page with the command 'The End'. If the page also states a command to return to the beginning of the book, the command is ignored and the page is considered a node with an outdegree of zero. These are referred to as endings.

- All nodes (regardless of shape) are given with a number representing its page number and an optional 'p', representing whether the page is adjacent to a full picture page or not.
- Arcs (i.e arrows) represent a command from page to page.

Figure 1 shows an example digraph created by `InstaViz`. The collection of all 40 digraphs can be found at www.alisonmarr.com/cyoa [5].

While a majority of the digraphs we created are trees, there are some digraphs with cycles. Trees represent books where you never reread a page while digraphs with cycles represent books where the reader returns to a previously read page at some point in their adventure. The indegree of a node (the number of arcs going into that node) will give us information about how many pages lead to that page. The outdegree of a node (the number of arcs leaving that node) will tell us how many possible pages we can turn to from that page. For example, nodes with outdegree one are pages that have only one command at the bottom of the page; the reader has no choice on what page to turn to next. However, nodes with outdegree at least two are pages of text where the reader gets to make a choice as to what page to turn to next. We refer to these nodes as decision pages. Pages with outdegree zero are referred to as endings as the story does not continue after this page.

2. Assumptions

To convey our mathematical model of the books, we must first make several assumptions. This will help future researchers verify and expand our research. These assumptions help clarify what the properties are and how they were collected. We made many assumptions related to nodes with a p, that is, the nodes which are adjacent to a full-page picture. For all books, we assumed that page 1 appears on the right side of the book. Then, if the page number is even and a 'p' is present in the vertex, then the picture must be on the right. If the page number is odd and a 'p' is present in the vertex, then the picture must be on the left. If a book had a picture with page 1, then we add one to the total number of pages in the text as there is content on page 0. Text pages were considered to be any page with some story text on it. Some text pages include a picture, but full length picture pages with only a command were not considered text pages. In situations where the command is given on a picture page, the command is assumed to be on the page opposite the picture. Pages of text that also contained a picture were not specifically documented.

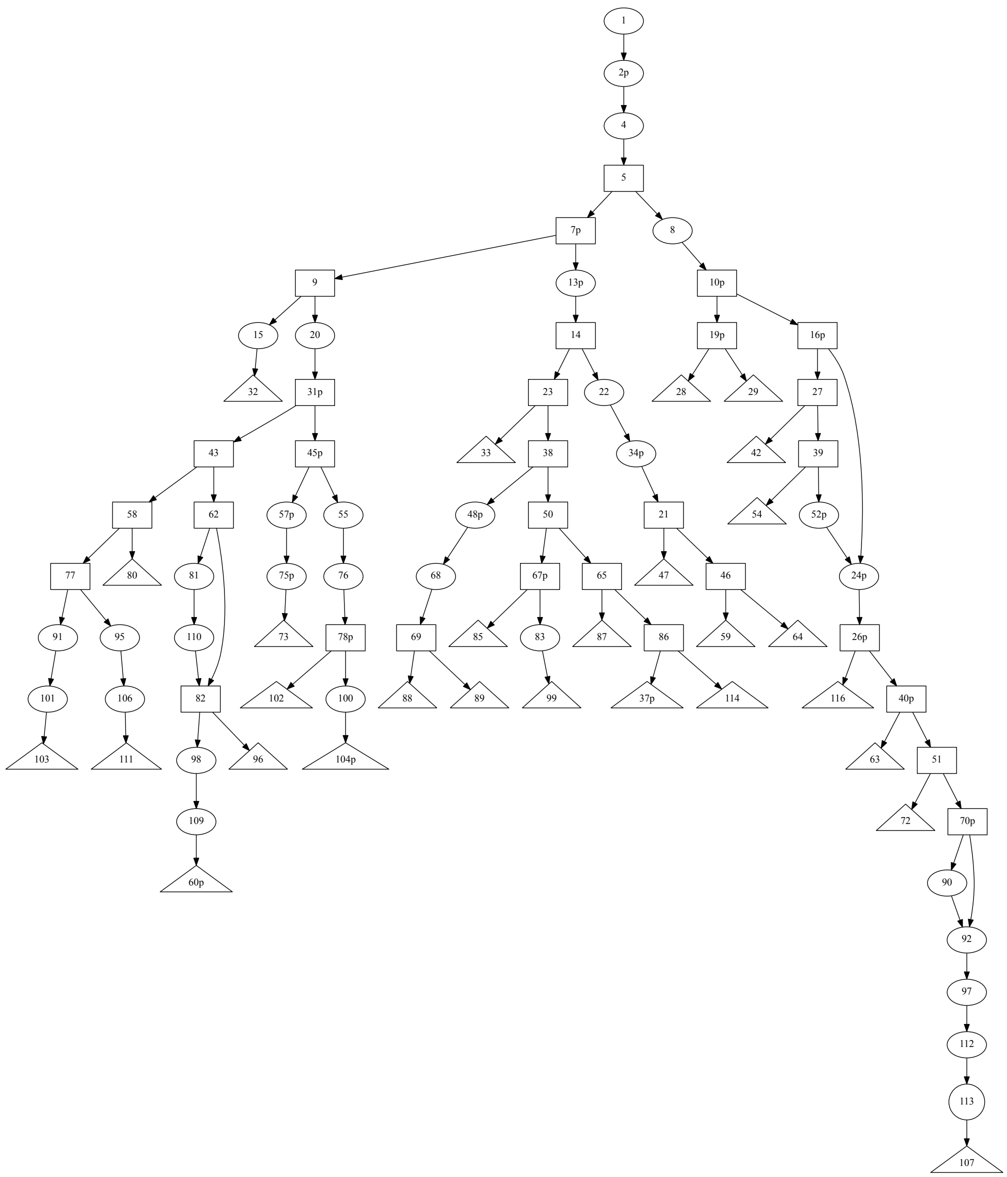

Figure 1: `InstaViz` Graph for Book 1: *Abominable Snowman*

During the graphing process we discovered a special case that had not been explained with our current assumptions. To help clarify the graphing process one more assumption was made. In Book 2, p. 75 says THE END but also has an option to go to page 106 for a different ending. You get to page 75 from page 52 (a decision page where you can go to page 74 or 75), so we drew this as 52 can go to 75, 106, or 74. This happens again on page 77 where the only way to go to page 107 is via a THE END.

3. Our Method

Our research focuses primarily on creating the digraph representation of each book, structural analyses of those digraphs, and several statistical analyses to identify differences among authors, across publication years, and in reader/buyer preference. While converting the books into digraphs using `InstaViz`, we found several errors in publication and in the advertised number of endings given on the front covers. Once we figured out what caused the errors, we were able to draw and then analyze the corrected digraph for each of these books. Then, we input those digraphs into *Mathematica* for analysis purposes. We decided to measure 24 different structural properties that we assumed were qualities that might interest a reader. For information on the errors found, see Section 5.

3.1. Data Collection

We collected data on properties we thought might be important to a reader engaged with the interactive nature of the book series. Several properties (given in Table 1) dealt with analyzing features of the nodes.

Given in Table 2 is another set of properties we measured related to path structures within our digraph.

We also measured several properties related to page turns. We were interested in not only how many pages you "skipped" but also if you were being asked to turn forwards or backwards in the book. For all of these calculations, we first labeled each of our arcs based on the difference of the labels incident with each arc. Suppose arc (n_1, n_4) goes from the node labeled 1 to the node labeled 4. Then, this arc would receive a label of $4 - 1 = 3$. These properties are listed in Table 3.

Graph theory description	**Book context**
Number of nodes	Pages of text
Number of nodes labelled with a "p"	Number of pictures
Sum of number of nodes and number of nodes labelled with a "p"	Total pages
Number of nodes of outdegree zero	Number of endings
Number of nodes with outdegree greater than one	Number of decision pages
Max indegree	Largest number of different pages that lead to same page
Max outdegree	Largest number of decision choices on any one page

Table 1: Node properties

3.2. Mathematica

After each digraph was drawn in `InstaViz`, the graphs were coded into the computer algebra system `Mathematica` [6] so that the variables described in the previous section could easily be calculated for each digraph. For many variables, there were built-in commands we were able to use to calculate their value. For example, to find the Pages of Text, we used the command "`VetexCount`". We could find the Number of Endings by using the command "`VertexOutDegree`" and counting the number of zeros present. We could also find the maximum amount of times we go to the same page by first finding "`VertexInDegree`" and then using the "`Max`" command to find the largest value in that list.

Graph theory description	**Book context**
Total number of paths	Number of stories
Length of shortest path	Shortest story length
Length of longest path	Longest story length
Length of path from node labelled 1 to first node of outdegree greater than 1	Pages to decision 1
Average length of path between any two nodes of outdegree greater than 1	Average number of pages between decisions
Average length of all paths present	Average length of all stories
For each node of outdegree zero, count the number of paths that end at that node and calculate the average	Average times you read the same ending

Table 2: Path properties

Graph theory description	Book context
Largest positive arc label	Maximum number of pages turned forward at one time
Absolute value of the smallest negative arc label or zero if no such label exists	Maximum number of pages turned backward at one time
Largest absolute valued arc label	Maximum number of pages turned (in either direction)
Average over all arc labels	Average number of pages turned at one time (where turning backwards contributes a negative number of pages)
Average over all absolute valued arc labels	Average number of pages turned at one time (regardless of direction)
Number of negative arc labels	Number of times you have a backwards page turn
Number of positive arc labels	Number of times you have a forwards page turn
(Number of negative arc labels) divided by (Number of positive arc labels)	Ratio of number of backwards page turns to forwards page turns

Table 3: Page turn properties

We wrote code to find all the possible paths in the books along with their lengths. We calculated values related to page turns by storing the differences on the labels between connected nodes and using that list (or a list of their absolute values) to calculate maximums, minimums, or averages. We calculated some properties (such as pages until the first decision) by hand for each of the 40 books. These calculations were done by two different people and double checked for accuracy.

3.3. Five Number Summary

Once all the digraphs were analyzed by *Mathematica* and data was stored, we calculated the five number summary, mean, and standard deviation for each of our variables. This information is presented in Table 4.

This analysis lead to some interesting findings. First, note that over 50% of the books have no cycles. And at least 75% have only one cycle. This means a majority of our books have a tree structure and that in a majority of books you will only read each ending once. From the data, we also know there is at least one book with no backwards page turns while at the same time there is at least one book with more backwards page turns than forwards page turns.

Property	**Min**	Q_1	**Med**	Q_3	**Max**	$\overline{x}$	s
Pages of text	88	94	98	103	115	99.6	7.44
Number of pictures	15	19.75	21.5	23.25	28	21.35	3.06
Total pages	108	115	122	125.25	137	121.05	7.57
Number of endings	11	15	20	24	42	21.18	8.49
Number of decision pages	9	12.75	17.5	26.25	47	20.43	10.03
Max indegree	1	1	1	2	3	1.5	0.64
Max outdegree	2	2	3	3	4	2.58	0.59
Number of stories	11	15.75	20	29.25	202	31.45	34.04
Shortest story length	5	9	11	14.25	21	11.8	4.16
Longest story length	14	22.5	27.5	31.25	45	27.65	7.11
Pages to decision 1	2	3.75	5.5	8.25	12	5.98	2.85
Avg. pages between decisions	1.19	2.25	3.28	4.05	6.3	3.32	1.421
Avg. length of story	11.73	15.28	19.58	24.56	31	20.26	5.46
Avg. times each ending read	1	1	1	1.36	4.93	1.33	0.77
Number of cycles	0	0	0	1	61	3.67	12.21
Max forward turns	28	56.25	67	84.75	109	68.33	20.31
Max backward turns	0	39	59	75	114	55.7	28.58
Max pages turned	32	63.25	75	88.25	114	75.05	18.92
Avg. directional page turns	0.86	3.91	5.42	8.55	14	6.34	3.34
Avg. page turns either way	2.61	6.78	9.11	16.55	30.36	11.68	6.61
Number backward turns	0	4	6.5	14	37	10.28	9.62
Number forwards turns	58	83	91.5	98	114	89.53	13.99
Ratio of backwards to forwards	0	0.04	0.07	0.18	0.64	0.14	0.16

Table 4: Descriptive statistics for all properties

After analyzing the statistics from the five number summary and recognizing that many of these properties are related or dependent upon each other, we narrowed our list of properties to analyze further to just twelve: pages of text, number of pictures, number of endings, number of decision pages, shortest story length, longest story length, pages to decision 1, average length of all stories, average times each ending is read, max pages turned, average page turns either way, and the ratio of backwards to forwards page turns. We considered these twelve properties to be good representatives of what type of structure a given book has.

4. Statistical Analyses

Once data was collected, we began analyzing the properties of the books using various statistical tests. As previously mentioned, we assumed our collection of 40 books was a representative sample of the over 184 CYOA books. All statistical analyses were done in R and statistical procedures were selected with guidance from [3].

4.1. R.A. Montgomery and Other Authors

Our first analysis was to compare the books written by R.A. Montgomery to those written by other authors. Since R. A. Montgomery was the founder of the series, he has written more books than any of the other authors combined. In our sample, he wrote 27 of the 40 books. We combined all books not written by R. A. Montgomery (13 of them) into one group so that the sample size would be large enough for statistical tests. Before running the tests, we confirmed that each group of data by author was normally distributed using a Shapiro-Wilk test at the $\alpha = 0.01$ level. All but average times we return to the same ending and the ratio of backwards to forwards page turns met this condition, so we used Welch's t-test for ten of the variables and a Wilcoxon's rank-sum test for the two non-normally distributed variables (see Table 5). We found there to be significant differences (at the $\alpha = 0.05$ level) between Montgomery's books and the others in three categories: pages of text, shortest story length, and average pages turned in any direction.

	Other Authors		R.A. Montgomery		
Variable	M	SD	M	SD	t-statistic
Pages of text	103.23	6.38	97.85	7.38	2.371*
Num. of pictures	22.61	2.69	20.89	3.11	1.805
Num. of endings	18.54	3.71	22.44	9.82	-1.815
Num. of decision pages	17.46	5.27	21.85	11.46	-1.659
Shortest story length	13.85	3.85	10.81	4.00	2.303*
Longest story length	29.30	5.38	26.85	7.78	1.162
Pages to decision 1	6.62	2.47	5.67	3.01	1.058
Avg. length of all stories	22.05	4.63	19.40	5.70	1.572
Max pages turned either direction	67.92	22.18	78.48	16.51	-1.525
Avg. pages turned any direction	8.80	5.51	13.06	6.73	-2.122*

* $p < .05$

Table 5: Results of Welch's t-test of means for author comparison

For the two non-normally distributed variables, the average time we return to the same ending did not differ significantly from R. A. Montgomery (median=1.050) to other authors (median=1.000), W=128.5, p=0.1357 and the ratio of backwards to forward page turns did not differ significantly from R. A. Montgomery (median=.088) to other authors (median=.042), W=109.5, p=0.0586.

4.2. With Respect to Time

We were also interested in what happened to our variables over time. In order to see if there were changes over time, we calculated the Pearson's correlation coefficient for each of our twelve selected variables (see Table 6). Looking at these values it seems that in general over time books have fewer endings, fewer decision pages, and fewer stories. However, the story lengths tend to generally be longer in the newer books and there are more pages between decisions. There are also fewer pages per turn and fewer backward page turns. When looking at the scatterplots of these data, the gap in publication year (from 1998-2004) is very evident as seen in Figure 2. Hence, we also calculated the correlation coefficients using just the 34 books published before 1998 to see if these books published before the gap had similar properties to the whole collection. Generally, the same relationships remained and the correlation coefficients became closer to ± 1. Only one variable (pages of text) went from a positive to negative correlation coefficient, but neither coefficient was statistically significant.

Variable	All books	Books prior to 1998
Pages of text	.188	-.251
Num. of pictures	.253	.330
Num. of endings	-.488**	-.743***
Num. of decision pages	-.524***	-.783***
Shortest story length	.461**	.758***
Longest story length	.582***	.741***
Pages to decision 1	.355*	.762***
Avg. length of all stories	.542***	.798***
Avg. time same ending read	-.297	-.448**
Max pages turned either direction	-.359*	-.399*
Avg. pages turned any direction	-.566***	-.525**
Ratio of back/for page turns	-.331*	-.216

$^{*}p < .05$ $\quad {}^{**}\, p < .01$ $\quad {}^{***}\, p < .001$

Table 6: Pearson's correlation coefficient by year (with all books and books before 1998)

4.3. GoodReads and Best Sellers on Amazon

Our final analysis involved both reader and buyer preference. We were interested in seeing if certain structural properties were related to how a book was ranked by readers on GoodReads and if these same properties were related to how well a book sold on Amazon.com. We collected the average ranking for each book as it appeared on GoodReads on May 19, 2017.

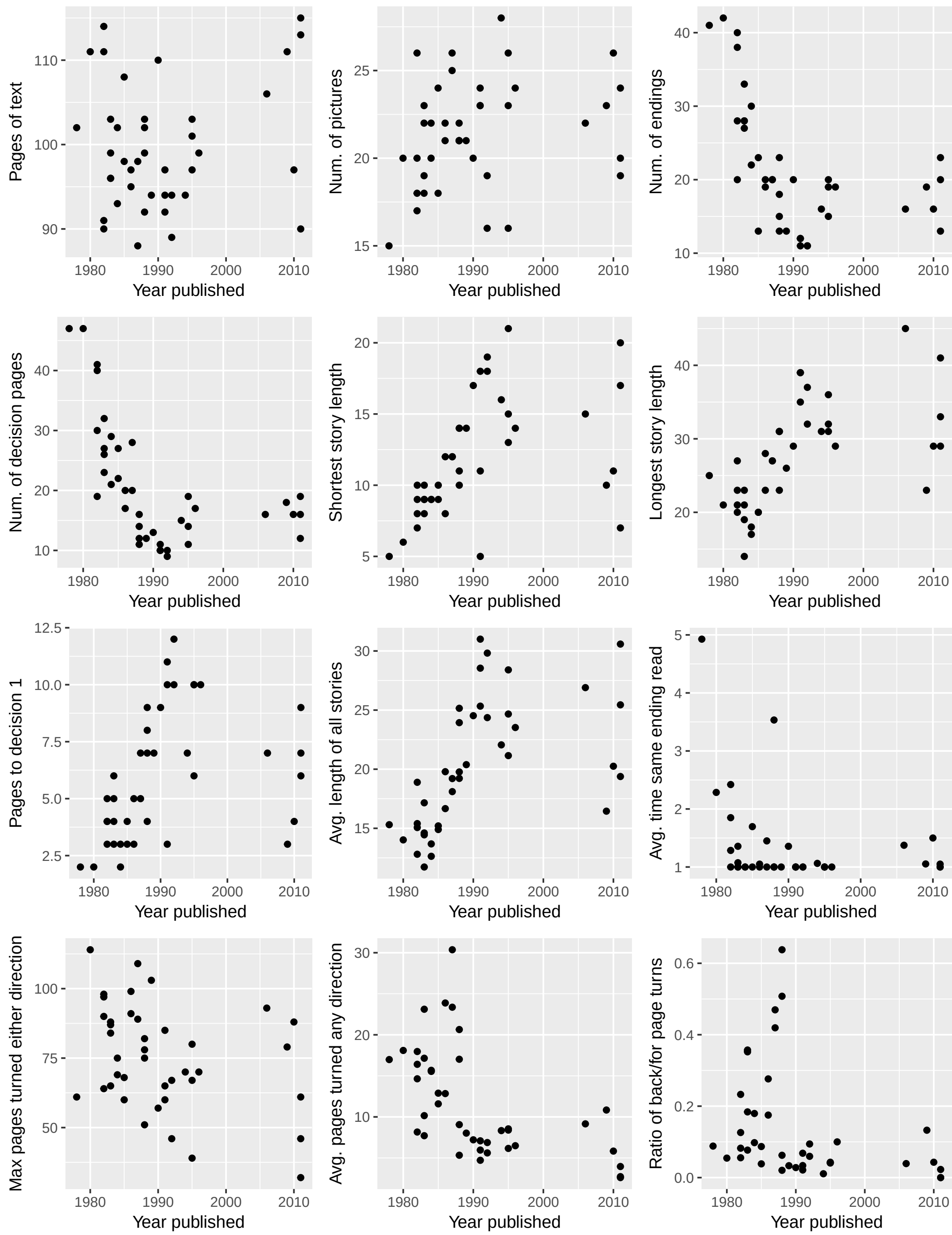

Figure 2: Scatterplots showing gap in publication years

	2	3	4	5	6	7	8	9	10	11	12	13	14
1. GoodReads	**-.35**	.07	.08	.18	**.24**	-.06	-.09	-.09	-.13	.06	**.35**	**.27**	.18
2. Amazon Ranking		-.08	.05	**-.22**	**-.28**	.13	.18	.06	.15	-.12	-.07	-.12	-.16
3. Pages of text			-.08	**.23**	.20	.03	0	-.11	-.02	.10	-.02	.04	-.19
4. Num. of pictures				-.05	-.05	.10	.06	.06	.07	-.02	.04	-.04	-.06
5. Num. of endings					**.77**	**-.45**	**-.50**	**-.43**	**-.55**	**.34**	.12	**.39**	.20
6. Num. of decisions						**-.58**	**-.59**	**-.58**	**-.67**	**.33**	**.24**	**.44**	**.23**
7. Shortest story length							**.48**	**.70**	**.61**	**-.29**	**-.23**	**-.32**	-.18
8. Longest story length								**.55**	**.82**	-.14	-.18	**-.42**	**-.34**
9. Pages to decision 1									**.66**	**-.26**	**-.23**	**-.32**	-.11
10. Avg. length stories										-.20	-.21	-.41	-.25
11. Avg. time same ending read											**.24**	**.29**	.13
12. Max pages turned												.39	**.23**
13. Avg. pages turned													**.53**
14. Ratio back/for turn													1

Bolded values indicate statistical significance at the $\alpha = 0.05$ level.

Table 7: Kendall's τ correlations amongst target variables

At this time each book in our series had at least one ranking. The current ranking on the best seller list from Amazon.com was as of June 19, 2017. We calculated Kendall's τ correlation coefficient for each of our variables as they relate to the GoodReads rating and the Amazon ranking. We used this non-parametric statistic because both the GoodReads rating and the Amazon ranking are ordinal data and many of our variables have tied ranks. Results in Table 7 show that readers generally tend to prefer more decisions and bigger page turns while buyers tend to prefer books with more endings and more decisions. For the Amazon ranking, the higher the rank, the less it sold. Thus, we conjecture that negative correlation coefficients may correspond to properties buyers prefer. Another interesting note is that in general the higher the GoodReads rating, the lower the Amazon ranking meaning books with higher ratings are generally sold more.

5. Errors Found

In the process of creating the digraphs and preparing them for analysis, we discovered several misprints in the books. The owners of the CYOA series, Chooseco, were notified of these mistakes and responded that they had also recently discovered these mistakes and that the books have been corrected in their second printing. Many of the errors seem to stem from the way that they had drawn their own graphs that they now include on the backs of some of their books. The most common misprint discovered was a publishing error with the advertised number of endings on the front of the book. Table 8 below summarizes these misprints.

Book number	**Book title**	**Advertised Endings**	**Actual Endings**
2	*Journey Under the Sea*	42	41
5	*Mystery of the Maya*	39	40
25	*Search for the Mountain Gorillas*	26	29
27	*Project UFO*	17	16
29	*Curse of the Pirate Mist*	14	13

Table 8: Books with wrong ending counts

Books 2 and 5 both had incorrectly draw graphs on their back covers. In our graphs, it is clearly shown that two different stories go to the same ending in both of these books. However, the publisher showed these two stories going to different endings (even though the last page is the same) which is how they got one more ending than we did in both of these cases.

Our graph and the publisher's graph matched for book 5, so it appears the 39 on the front cover was just a misprint. For books 27 and 29 it is unclear as to why these differences appear as the published did not have a graph drawn on the back cover for these books.

One other misprint that occurred was misprinted page numbers in the commands of two different books. These are reprints of books and hence it is likely these mistakes were made when the books were reprinted for the new series. In Book 15: *Beyond Escape* the command on page 37 should lead to page 50, not page 51. This misprint was discovered when the total number of pages in the book didn't match the count we got from looking at the graph. Book 21: *Struggle Down Under* had the most errors of any book we studied. It was clear that many errors were made in the republishing of this book. When we first drew the digraph of the book, it was a disconnected digraph, so we knew something was wrong. Hence, we ordered the original version of the book *Terror in Australia* to help us determine what went wrong. Matching pages of text from the original to the new version, we were able to correct the graph. The edits made were:

- on page 50, change decision from p. 67 to p. 51,
- on page 68, change command from p. 87 to p. 69,
- on page 69, change command from p. 33 to p. 71, and
- on page 71, change command from p. 123 to p. 87.

A figure of the original disconnected, incorrect digraph and the corrected version are given in Figure 3.

6. Conclusion

Creating the graphical representation of these CYOA books allowed us to find mistakes in the printing of the books and conduct statistical analyses to compare book structure by year, author, and reader/buyer preference. One of our initial reasons for collecting this data was to help readers choose a book that best fit their reading preferences. While the work we did in this particular paper did not provide that option, for future work we would like to upload our data set to the online data visualization tool, LineUp [1]. This tool allows readers to select preference in different properties (so maybe a reader prefers more decisions and shorter stories) and the site automatically reorders the books in order of their preferences.

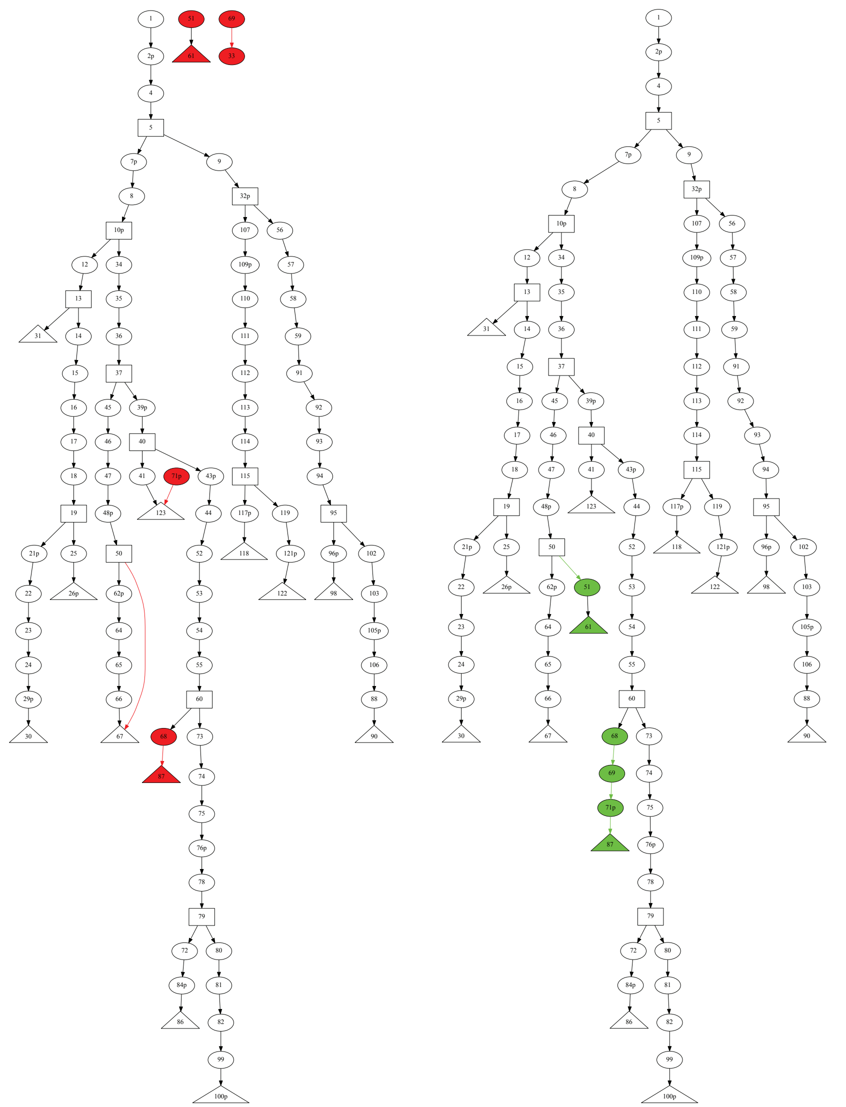

Figure 3: Digraphs for Book 21 (Original and edited)

Other options for future work include exploring the graphical structure of other interactive gamebooks such as some of the CYOA spin-off series or the Fighting Fantasy series that involves rolling dice to make decisions. Additional interesting work could be done looking at the structural differences (if any) between the original published versions of the books and the reprinted books. Netflix has also recently launched interactive video content and drawing the graphical structure of these shows would also be interesting.

Acknowledgements

We would like to thank the Southwestern University SCOPE program for the funding used to complete this research. We would also like to thank Joel Martinez and Shianne Solis who provided assistance in collecting data the first two weeks of the project. Our work was funded through NSF S-STEM grant 1458600 and funds from the H-E-B Foundation.

References

[1] The Caleydo Team, Multi-Attribute Rankings; available at http://caleydo.org/tools/lineup/, last accessed on July 11, 2019.

[2] Chooseco LLC, History of CYOA–Chooseco LLC; available at https://www.cyoa.com/pages/history-of-cyoa, last accessed on July 11, 2019.

[3] Andy Field, Jeremy Miles, and Zoe Field, *Discovering Statistics Using R*, SAGE, London, 2012.

[4] Instaviz, `Instaviz`: *Diagram Sketching, Mindmap, Flowchart, Workflow, Whiteboard App for your iPhone, iPad or iPod Touch*; available at http://instaviz.com, last accessed on July 11, 2019.

[5] Alison Marr, CYOA; available at http://www.alisonmarr.com/cyoa, last accessed on July 11, 2019.

[6] Wolfram Research, Inc., `Mathematica`, Version 11.0, Champaign, IL, 2016.