Proceedings of the Fourteenth Artificial Intelligence and
Interactive Digital Entertainment Conference (AIIDE 2018)

Exploratory Automated Analysis of Structural Features of Interactive Narrative

Nathan Partlan, Elin Carstensdottir, Sam Snodgrass, Erica Kleinman
Northeastern University
360 Huntington Ave., Boston, MA 02115
{nkp, elin, ekleinman}@ccs.neu.edu, s.snodgrass@northeastern.edu

Gillian Smith
Worcester Polytechnic Institute
100 Institute Road, Worcester, MA 01609
gmsmith@wpi.edu

Casper Harteveld, Magy Seif El-Nasr
Northeastern University
360 Huntington Ave., Boston, MA 02115
{c.harteveld, m.seifel-nasr}@northeastern.edu

Abstract

Analysis of interactive narrative is a complex undertaking, requiring understanding of the narrative's design, its affordances, and its impact on players. Analysis is often performed by an expert, but this is expensive and difficult for complex interactive narratives. Automated analysis of structure, the organization of interaction elements, could help augment an expert's analysis. For this purpose we developed a model consisting of a set of metrics to analyze interactive narrative structure, enabled by a novel multi-graph representation. We implemented this model for an interactive scenario authoring tool called *StudyCrafter* and analyzed 20 student-designed scenarios. We show that the model illuminates the structures and groupings of the scenarios. This work provides insight for manual analysis of attributes of interactive narratives and a starting point for automated design assistance.

1 Introduction

The study of interactive narrative, in which players choose actions in the context of a story, combines insights from narrative theory, theater, psychology, and other fields (Murray 1997; Mateas 2000; Seif El-Nasr 2007, e.g.). Understanding how best to design an interactive narrative, balancing authorial burden, authorial intent, and player experiences, is still an open research problem (Roth and Koenitz 2017). Narrative analysis, extracting detailed meaning from and describing the possibility space of the narrative, can aid in understanding potential player experiences and design considerations. However, this analysis is most often done by hand, by a knowledgeable expert. While manual analysis is valuable, interactive or generative narratives often include many branching possibilities that render it prohibitively complicated, especially for a large corpus.

Deep analysis of the content and meaning of interactive narrative remains exceedingly difficult for a computer, requiring knowledge of context, language, culture, authorial style, and more. Though a truly comprehensive analysis requires studying content, it can begin with a simpler approach targeting one aspect of narrative: structure. In the context of this work, we conceptualize structure as the organization, flow, choices, interactions, and other elements that do not rely on a deep understanding of the culturally-informed meaning of the text. By focusing on attributes specific to the structure of interactive narrative, automated systems can augment and complement an expert's understanding of the plot, characters, dramatic arc, and other content.

A step towards building such computational techniques is to define and operationalize metrics for structural analysis. Metrics provide a concrete set of measurements to quantify and simplify interactive narrative structure. However, this subject has not been widely researched. In fact, the authors know of very few examples of clearly defined structural metrics or automated analyses for interactive narratives. One such example is the work of Szilas and Ilea (2014), who defined a small set of metrics focused on player interaction and performed a preliminary user study, finding correlations with player experiences of flow states and positive emotions. Our work incorporates and extends these metrics.

The primary contribution of this paper is to propose a set of metrics designed to quantify the following factors: *narrative structure complexity*, measuring graph structure and size (Bondy and Murty 1976); *interactive affordances*, measuring the scenario's structures of interaction and feedback as explained by Carstensdottir et. al. (2017); and *action space*, measuring the potential space of play and choice available in the scenario and related to "theoretical agency" as defined by Thue et. al. (2010) to be the real potential for the player's actions to change a story. These factors are complex, just as narrative itself is complex. We do not claim that our metrics are perfectly comprehensive in illuminating these factors, nor that they have a particular correlation with subjective player experience. However, we show that they successfully capture aspects of scenario design.

In order to derive these metrics, we developed a graph-based representation of an interactive narrative scenario, comprising multiple graphs designed to describe various aspects of its structure. Our model consists of four types of graphs: the scene flow map, layout graph, script graph, and interaction map. The metrics are computed by two methods: a static graph analysis and a randomized playthrough exploration of the scenarios' interactive possibilities, as abstracted by the interaction map.

We implemented this model for *StudyCrafter*,[1] a tool for building 2D scenarios that often employ interactive narrative (Harteveld et al. 2016). To evaluate the model, we analyzed student-created scenarios from a university research methods class.[2] We performed three analyses: a manual individual scenario analysis, an expressive range analysis, and K-Medoids clusterings of the scenarios. For the latter, the clusters are based on subsets of the metrics related to the three aforementioned factors. We show that the metrics successfully illuminate the structure and groupings of the scenarios, providing insight for manual analysis and a starting point for automated design assistance. This work has future applications in areas of mixed-initiative content creation, automated playtesting, and player experience modeling.

2 Related Work

Early work by Propp (1968) defined a framework for analyzing folktales. However, the functions he proposed are primarily content-oriented and are not directly related to interactivity. Valls-Vargas et al. (2017) developed a computational system, building on Propp's work, to extract character roles, domain knowledge, and other details from folktales. Whereas these works study and systematize the content of traditional narrative, we focus on interaction and structure.

Within the field of interactive narrative, much research focuses on the unique qualities and challenges of their content and effects on player experience, with few studies on computational analysis, and particularly structural analysis. Szilas and Ilea (2014) developed metrics for player interaction and choice and studied their correlations with player experience. They developed survey instruments to measure emotional responses, such as experiences of enjoyment, curiosity, and suspense (Vermeulen et al. 2010), which are related to narrative theory (Roth and Koenitz 2016). Using these metrics, Szilas and Ilea (2014) found a few significant correlations to player experience. Our work focuses on narrative structure, exploring factors of action space, interactive affordances, and narrative structure complexity. While the action space factor overlaps with the metrics defined by Szilas and Ilea, they did not explore the other two factors.

In terms of work on structure, Bernstein (1998) defined and described several patterns exhibited by interactive hypertext stories, a process since continued by others (Millard et al. 2013; Short 2016). Lindley (2005) suggested using graph-based models to more clearly define and detect such structures. Our work employs such an approach, and thus its graph-based representation can be thought of as an implementation of this basic idea. To our knowledge, these structural theories have not previously been operationalized.

As part of the evaluation of our model, we employ expressive range, which was designed to visualize the breadth of possibilities for a generative system's output (Smith and Whitehead 2010). It has previously been used to describe procedurally generated content (Horn et al. 2014). We perform similar analysis in the work presented here, but to visualize and evaluate the diversity of hand-designed scenarios.

3 Methods

To approach automated interactive narrative analysis, we developed a model composed of four related graph representations, each designed to enable measurement of particular metrics. These metrics are grouped into the aforementioned factors: narrative structure complexity, interactive affordances, and action space, defined and described below.

We chose a graph-based representation because it enables searching, enumeration, and relationship analysis between the elements of the scenario. The process begins with loading the scenario data, structured as files, one per scene, containing unordered dictionaries of metadata, script information, and visual layout data. We build a *scene flow map* by examining how each scene transitions to others, then transform each scene's script into a *script graph* and its layout into a *layout graph*. Finally, we build an *interaction map* from each script graph.[3]

3.1 Representation

In *StudyCrafter*, a "scene" is a single visual setting. Designers are free to break up their scenario into many scenes or to work within a single scene, modifying it with animation and scripting. The *scene flow map* contains the metadata about each scene in its vertices, and its edges represent the possible transitions between scenes. In addition to the scene metadata, each vertex holds identifiers for the associated script and layout graphs, which store most of the scenario data.

The *layout graph* represents the spatial arrangement of objects in the scene. Each visual object is a vertex, connected by edges annotated with the offset vector between them when the scene begins, similar to D-nodes by Guzdial and Reidl (2016). However, the layout graph alone cannot explain how these objects are used or changed during the scenario's operation. For that, we need the script graph.

StudyCrafter provides a visual, node-based scripting language to control its scenarios, visualized in a graph-like structure for editing. We create our *script graph* by reconstructing this structure in a more explicit, unified graph form. We represent each script node, such as a "dialogue" or a "branch," as a bag of typed properties. For instance, a dialogue node contains a "StringExpression" property for the text to be spoken and a "LayoutObject" property for the speaking character. The definition for each script node is data-driven, enabling easy expansion or modification. Edges

[1] https://studycrafter.com

[2] The projects are available in the Northeastern University Digital Repository at: http://hdl.handle.net/2047/D20291320

[3] Example visualizations of selected script graphs and interaction maps are stored with the projects in the Northeastern University Digital Repository.

Metric	Details
Narrative Structure Complexity	
Number of Scenes	Number of vertices in the scene graph
Total Layout Nodes	Number of vertices in each layout graph
Average Layout Nodes	Total layout nodes divided by number of scenes
Total Script Nodes	Number of vertices in each script graph
Average Script Nodes	Total script nodes divided by number of scenes
Average Edges Traversed	Average number of steps taken across all interaction maps, over all playthroughs
Edges Traversed Variance	Variance (average squared deviation from the mean) of Average Edges Traversed
Connected Components	In script, number of separate cliques of nodes all reachable from each other (Tarjan 1972)
Text Per Dialogue	Number of characters per dialogue node in the script
Average Outdegree	Average number of edges leaving the nodes in the script graphs (Bondy and Murty 1976)
Action Space	
Average Number of Choices	Number of choice edges in the interaction maps, averaged over all maps
Intra-Playthrough Diversity	Number of unique player actions, averaged over all playthroughs (IDIV)
Global Diversity	Total number of unique player actions found across all playthroughs (GDIV)
Renewal Rate	Inverse of intra-playthrough diversity divided by global diversity (RENEW)
Choice Range	Average number of choices presented to the player at each interaction point (CR)
Choice Frequency	Percentage of actions taken by the player, versus by the system, on average (DCFREQ)
Choice Variability	Percentage of unique versus repeated choices during the playthrough, on average (CVAR)
Interactive Affordances	
Average Interaction Points	Average number of interaction point nodes in the interaction map per scene
Choices Per Interaction Point	Average number of choice edges out of each interaction point
Feedback Per Interaction Unit	Number of feedback nodes divided by the number of interaction nodes
Feedback Per Choice	Number of feedback nodes divided by the number of choice edges
Feedback Per Event	Number of feedback nodes divided by the number of event nodes
Average Actions Taken	Average number of player actions taken during a playthrough, over all playthroughs
Variance of Actions Taken	Variance (average squared deviation from the mean) of Average Actions Taken

Table 1: The metrics for each of the three factors: narrative structure complexity, action space, and interactive affordances. The metrics based on Szilas and Ilea (2014) note their corresponding label in parentheses.

connect the script nodes to their possible run-time successors. Note that, if we were not already working with a visual, node-based scripting language, we could still build a similar graph from the "parse tree" of a program, or by other static analysis techniques. Using the script graph, we can develop a simplified graph, which we call an interaction map.

The *interaction map* is based on work by Carstensdottir and Seif El-Nasr (2018), which enables us to abstract the interaction from the scenario. It consists of three types of vertices: events, which happen without any player interaction; interaction points, which are opportunities for player choice or action; and feedback, which is any perceptible response to player actions. There are also special start and end nodes. Edges represent potential transitions and influences between the vertices, with three types: direct edges, for explicit movement from one event/feedback node to another, or to an interaction point; indirect edges, for implicit causes or influences; and options, for the possible transitions a player may choose from an interaction point. The feedback nodes influenced by a particular interaction point are part of its "interaction unit," and they are tagged as such. Thus, an interaction unit is defined as a single interaction point, the choices available there, and the feedback from that interaction.

Now that we have a clear picture of the representation, we can complete the description of the model by enumerating and explaining the metrics.

3.2 Metrics

The metrics, by factor, are summarized in Table 1. Each factor has a separate focus. Narrative structure complexity measures the size and scope of the scenario's narrative elements. Not limited to counting nodes, it measures the degree to which the script branches and loops and the average length of traversals through it. These measures are based on standard graph theory algorithms and concepts (Bondy and Murty 1976; Tarjan 1972). They are intended to represent design complexity, not necessarily "difficulty" of the experience for a player, nor sophistication of the narrative content.

Understanding the overall complexity of a scenario is different from understanding how it enables interaction and responds with feedback. These are the purview of the interactive affordances factor, based on the theory of interaction structure and feedback by Carstensdottir et. al. (2017). This factor deals primarily with the interaction map.

Separately from interaction and feedback, we can measure the possibility space of player choices and their effect on narrative progression. This is the action space factor, based on previous work on theoretical agency (Wardrip-Fruin et al. 2009; Thue et al. 2010). Theoretical agency is related to the actual effects of the player's actions, but it does not measure their subjective experience of agency. The latter is additionally affected by perceived affordances and by individual differences between players, neither of which we can fully measure with the current metrics. Most metrics in this factor are based on Szilas and Ilea's playthrough-

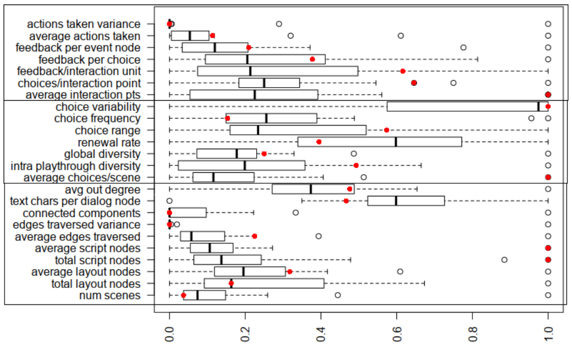

Figure 1: Box-and-whisker plots of each metric (normalized between 0 and 1). Metrics associated with a single factor are bordered. From the top, the factors are: interactive affordances, action space, and narrative structure complexity. Additionally, the solid (red) points represent the individual metrics from "Deserted Island: Cabinet Mystery."

metrics (2014). However, instead of using playtraces, we use simulated playthroughs of the scenario, which adds scalability to the approach. These are randomized walks through the interaction map, simulating a randomly-acting player in a simplified approximation of possible interactions. We average the results of many such random traversals.

4 Results

To explore the capabilities of our model, we investigated twenty interactive narrative scenarios from a university course. In the course, students used *StudyCrafter* to construct scenarios of their own design, following a prompt to create short social science experiments. *StudyCrafter* supports scenarios with 2D visuals, limited animation, and dialogue choices. It also provides a few extra game features, such as key-press and time-based events and visual object selection. In this way, it is approximately comparable to other tools, such as Twine, for visual scripting of choice-based interactive narrative.

The prompt imposed some constraints, such as randomized experimental conditions, but otherwise left significant freedom for design. Most scenarios used simple choice-based interaction, while a few used the event-based gameplay capabilities of *StudyCrafter*. In the below sections, we present several forms of analysis to illustrate potential uses of the metrics by different audiences. The analyses increase in scope from individual to cluster-level scenario analysis.[4]

4.1 Descriptive Statistics

Descriptive statistics reveal that the metrics are mostly not normally distributed in these scenarios, showing mostly positive skew. The only normally-distributed metrics are the average outdegree, text characters per dialogue, and renewal rate. We visualize the normalized box-and-whisker plots in Figure 1. A few show significant outliers, especially in variance of actions and edges traversed. The outlier scenarios in these cases had infinite potential for looping back to the content, enabling randomized playthroughs of highly variable length. Note, however, that these metrics are not actually zero for most of the other scenarios; they are simply crushed to near-zero in the box plots by the outliers.

The visualization in Figure 1 is also useful for comparing an individual scenario's metrics to a larger corpus. For example, we overlay the points for the scenario "Deserted Island: Cabinet Mystery" (described and analyzed below). Because the plots are grouped by factor, we can see how the scenario compares for each factor and for individual metrics.

Figure 2: Screenshots from the scenario "Deserted Island: Cabinet Mystery." From top left to bottom right: introducing the player to the situation, selecting an item, the response to selecting the item, and asking a demographics question.

4.2 Individual Scenario Analysis

To be useful for a designer, it is important for our metrics to illuminate the properties of a single scenario. Therefore, we present an example manual analysis, of the sort that might be automated in the future. Here, we compare with the other student scenarios; in the future and with more data, we could determine more robust ranges for the metrics. We analyze "Deserted Island: Cabinet Mystery," a scenario in which the player selects between pairs of items that magically appear from a cabinet on a deserted island, as shown in Figure 2. The scenario contains a short briefing about the experiment and a brief narrative to set the scene. Each choice between two items is accompanied by a particular drawer opening, and the drawer locks after the player makes a selection. The scenario ends with a short debriefing, as well as a final question asking whether the player is color blind.

This scenario has no connected components and a choice variability of 1.0, meaning that it has no looping or repetition of choices. It also has no variance in the number of actions taken per playthrough nor the edges traversed; all playthroughs are exactly the same length. The action space metrics are middling: choice range is fairly high (3.5), meaning that it offers multiple options at each choice point, but renewal rate is 0.46, which means that the majority of actions are the same in all playthroughs. This is likely a scenario that will feel somewhat similar every time.

In terms of interactive affordances, the scenario is high-

[4]The full metrics and clustering results are also stored in the DRS collection at: http://hdl.handle.net/2047/D20291320

est among this data set in its average number of interaction points per scene (8), and its other metrics are middling to high. Several of the metrics for complexity are very high: the scenario has the most script nodes (633), and script nodes per scene (211). Since this scenario already has high complexity, a creativity support tool might note the lack of playthrough variety and suggest reducing the overall length and number of options, but adding clearly differentiated branches of feedback for each of the player's choices.

4.3 Expressive Range

We performed an expressive range analysis on the scenarios by selected metrics, normalized between 0 and 1 in cases where the true range is unknown. An example of this is shown in Figure 3, which depicts the expressive range in terms of two metrics: choice range (factor: action space) and feedback per choice (interactive affordances). We selected this particular pair of metrics because they depict design trade-offs between two factors. This sort of analysis could, for instance, enable instructors to visualize the differing design choices between projects and determine how best to guide students. Designers could use expressive range analysis to compare their scenario to others on particular metrics.

In the figure, the "Christmas Shopping" scenario is shown to have the highest choice range, but very limited feedback. It presents many informational options to the player simultaneously, so that the player may choose an item to purchase. However, the actual feedback from selecting an item, or from making the final choice to buy it, is very limited; the scenario simply moves on without acknowledging the player's decision.

Conversely, "An Unusual Situation" is shown to have low choice range and high feedback. In it, the player makes only a single meaningful choice with two options: to punch or not punch a rude former "friend." However, the decision causes the narrative to diverge for some time (though it eventually returns to a linear plot). This scenario provides meaningful feedback for the player's choices.

As shown in the figure, none of the scenarios had both high feedback and high choice range. A manual analysis reveals that very few provide significant branching or feedback based on player choices, and that those with more choices are less likely to meaningfully differentiate them.

4.4 K-Medoids Clustering

In addition to the expressive range analysis, we performed a K-Medoids clustering on the scenarios, based on each of the three factors. This sort of clustering could aid a manual or automated analysis in categorizing a scenario, or in finding similar scenarios. We chose the number of clusters, K, based on highest silhouette width, but we required at least three clusters; two clusters do not sufficiently separate the scenarios to enable analysis.

For narrative structure complexity we chose four clusters, with a silhouette width of 0.42. The scenarios in Cluster A have a moderate amount of narration, but mostly linear progression, with few loops. For instance, "An Unusual Situation" is in this cluster. The scenarios in Cluster B are even more linear, with little to no variance in traversal length and

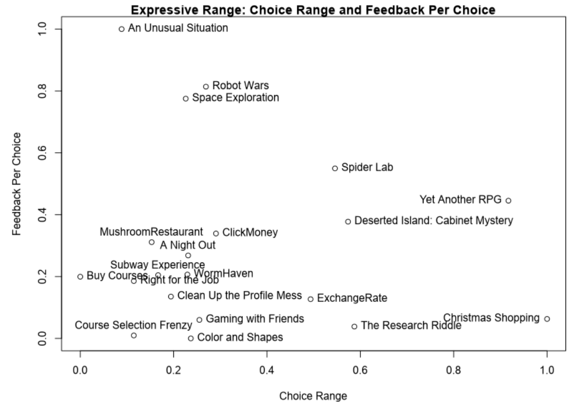

Figure 3: An expressive range analysis of the scenarios in terms of the feedback per choice and choice range metrics (normalized).

less text. Those in Cluster C are distinguished by frequent use of loops and repetition. One scenario that usually does not loop, "Clean Up the Profile Mess," is included. However, it prompts the player for input, and invalid responses cause it to repeat a question. Cluster D is less cohesive than the others, but its scenarios generally use large numbers of layout or script nodes. One fairly simple scenario, "WormHaven," is a bit of a surprise inclusion, but it is complicated by several possible paths due to initial randomization.

The interactive affordances factor's silhouette width is 0.40, for five clusters. Cluster A is the largest, and it contains a variety of moderately-interactive scenarios. It is perhaps best defined by contrasts to the extremes in other clusters. It is surprising that "Spider Lab," a scenario with rudimentary movement and combat mechanics, is included here, given that it requires by far the most actions to complete. Cluster B contains scenarios such as "Deserted Island: Cabinet Mystery" that contain the most interaction points. Cluster C's scenarios have very few interactions and little feedback. It includes two very similar scenarios where the player simply selects from a list of courses. Cluster D contains scenarios with a high number of interactions per playthrough, as in the informational popup selection in "Christmas Shopping" or the limited puzzle-like looping of "The Research Riddle," but that give little feedback. Finally, Cluster E contains scenarios with the most feedback for the player's choices.

Action space has silhouette width of 0.42, for three clusters. The clusters are somewhat unbalanced, with most scenarios in Cluster A. These have the least branching or looping. They offer only superficial choice, and their story is mostly linear. The scenarios in Cluster B contain more options for each choice, though their overall narrative is almost as linear as those in A. These include "Subway Experience" and "Deserted Island Cabinet Mystery," in which the player must repeatedly choose between various presented options. Cluster C's scenarios have the most complex gameplay and include loops, reducing their choice variability. These include, for instance, "Spider Lab" and "The Research Riddle," the latter requiring a puzzle-like search through several options, repeatedly returning to a hub for the next choice.

5 Discussion and Conclusion

The above analyses show how this work's model can be used for several purposes and by several audiences. Designers of individual scenarios, instructors assisting students, or mixed-initiative design tools could compare and examine individual metrics as in the example analysis. When analyzing a collection of scenarios, as in a review or class, a reviewer might cluster those scenarios based on the metrics, or visualize their relative positions by expressive range analyses. Such analyses might also be employed by automated tools to contextualize a scenario, find where it fits in an existing collection of artifacts of various styles, and provide suggestions based on that context.

Though this model has only been tested in *StudyCrafter*, we theorize its applicability to other similar interactive narrative creation tools, such as Twine and Ren'Py. The script graph can be adapted by defining property and script node types to fit the available operations. The interaction map is generically defined, though its implementation might need adjustment to match each platform's affordances. For text-only interactive narratives, the scene and layout graphs might still represent the designer's imagined environment. *StudyCrafter* is not designed for building parser-based interactive fiction, which may require further adaptation, especially in terms of mapping all possible interactions.

Some limitations remain, however, especially due to the small number of scenarios in this data set. For instance, the action space factor was not as clearly capable of separating the scenarios. This may be because none of the scenarios contained significant branching based on player choices. Additional refinement of the factors and metrics may better separate scenarios with similar attributes.

Moreover, this model may not always capture the nuances of scenarios that employ complex programming techniques. For instance, the scenarios with significant looping and branching, such as "Spider Lab," showed up as outliers in playthrough-length metrics. This is a limitation of our randomized automated playthroughs. Currently, the interaction map does not fully capture variable changes and conditional branches, leading to cases where random playthroughs do not perfectly simulate the scenario's real operation. Other unusual uses of the tool may also cause imperfect measurements. This can lead to, for instance, low variability for scenarios that modify and loop back to some content, e.g. in "Colors and Shapes," where the metrics detect repetition even though it varies its color and shape combinations.

There are other opportunities for short-term improvement. The factors and metrics can be expanded to explore additional attributes of interactive narrative. We could also move beyond numeric metrics calculated on the entire scenario to detect particular areas with certain structures or attributes.

Additionally, this model unlocks many opportunities for future work. A good first step would be a user study to validate the metrics and expand them to incorporate player experience. Using our playthrough approach, we can develop automated playtesting tools for interactive narrative, perhaps by adding procedural personas (Holmgård et al. 2018). Moreover, this work leads towards mixed-initiative creativity support, as has been explored for visual level design in games (Liapis, Smith, and Shaker 2016; Baldwin et al. 2017). Expanding co-creativity to the domain of interactive narrative requires metrics and programmatic analyses.

Mixed-initiative content creation is particularly exciting because it can combine the strengths of human designers with those of computational systems. People are good at noticing patterns, situating artifacts within a cultural context, and creating cohesive, meaningful work with a vision. These are all extremely difficult problems for computers. People are not, however, always good at managing complexity, visualizing structures, or envisioning how other people might experience and react to their work. Metrics, such as those proposed here, can support human designers by presenting several simplified impressions of an interactive narrative focused on specific criteria, showing the scenario to the designer in a new light. By abstracting the structure, using a representation such as ours, they can give designers opportunities to discover new insights, especially when correlated with player experience. Their value lies precisely in the contrast between the ways in which a machine can analyze a scenario and the ways in which a human designer would customarily evaluate it.

The danger of metrics, however, is also in their simplicity. By hiding complexity, and by quantifying abstract concepts, metrics can inadvertently lead to a myopic view of the artifact. Metrics can be seductive, giving a false impression of objectivity. This is exacerbated if they are unreliable, lack context, or fail to provide a sufficient variety of viewpoints. When developing and promoting metrics for interactive narrative, we should avoid encouraging over-reliance on them. In this work, for instance, our analyses included qualitative discussion of the scenarios, rather than pure reliance on metrics. In future work, we should ensure that the metrics are accompanied by other visualizations that encourage qualitative analysis, reflection, and pattern-spotting.

In conclusion, we have described how the model's graph-based representation enables metric calculation along three factors. These factors, narrative structure complexity, interactive affordances, and action space, separate and categorize scenarios by relevant and varied traits. By operationalizing these metrics and exploring their capabilities, through a single scenario analysis as well as broader analysis of several interactive narratives, we have formed a foundation to support computational interactive narrative analysis. This work represents a first step towards many applications, such as automated playtesting and creativity support, which were not possible with prior analysis tools for interactive narrative.

6 Acknowledgments

We would like to thank past and current members of the *StudyCrafter* team, and especially Dr. Camillia Matuk and Dr. Steven Sutherland. We further acknowledge the grant support from Northeastern University, NSF (IIS-1736185), and DARPA (D16AP0011).

References

Baldwin, A.; Dahlskog, S.; Font, J. M.; and Holmberg, J. 2017. Mixed-initiative procedural generation of dungeons using game design patterns. In *2017 IEEE Conference on Computational Intelligence and Games (CIG)*, 25–32.

Bernstein, M. 1998. Patterns of Hypertext. In *Proceedings of the Ninth ACM Conference on Hypertext and Hypermedia: Links, Objects, Time and Space*, HYPERTEXT '98, 21–29. New York, NY, USA: ACM.

Bondy, J. A., and Murty, U. S. R. 1976. *Graph theory with applications*, volume 290.

Carstensdottir, E., and Seif El-Nasr, M. 2018. Interaction Maps for Interactive Narratives. Technical Report NU-CCIS-TR-2018-001, College of Computer and Information Science, Northeastern University, Boston, MA.

Carstensdottir, E.; Kleinman, E.; and Seif El-Nasr, M. 2017. Towards an Interaction Model for Interactive Narratives. *ICIDS 2017*.

Guzdial, M., and Riedl, M. 2016. Toward Game Level Generation from Gameplay Videos. *arXiv:1602.07721 [cs]*. arXiv: 1602.07721.

Harteveld, C.; Stahl, A.; Smith, G.; Talgar, C.; and C. Sutherland, S. 2016. Standing on the Shoulders of Citizens: Exploring Gameful Collaboration for Creating Social Experiments. In *49th Hawaii International Conference on System Sciences*, 74–83.

Holmgård, C.; Green, M. C.; Liapis, A.; and Togelius, J. 2018. Automated Playtesting with Procedural Personas through MCTS with Evolved Heuristics. *arXiv:1802.06881 [cs]*. arXiv: 1802.06881.

Horn, B.; Dahlskog, S.; Shaker, N.; Smith, G.; and Togelius, J. 2014. A comparative evaluation of procedural level generators in the mario ai framework. In *Foundations of Digital Games*.

Liapis, A.; Smith, G.; and Shaker, N. 2016. Mixed-initiative content creation. In *Procedural Content Generation in Games*, Computational Synthesis and Creative Systems. Springer, Cham. 195–214.

Lindley, C. A. 2005. Story and narrative structures in computer games. *Bushoff, Brunhild. ed*.

Mateas, M. 2000. A Neo-Aristotelian Theory of Interactive Drama. In *Proceedings of the AAAI Spring Symposium on AI and Interactive Entertainment*.

Millard, D. E.; Hargood, C.; Jewell, M. O.; and Weal, M. J. 2013. Canyons, deltas and plains: towards a unified sculptural model of location-based hypertext. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, 109–118. ACM.

Murray, J. H. 1997. *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. New York, NY, USA: The Free Press.

Propp, V. 1968. *Morphology of the Folk Tale*.

Roth, C., and Koenitz, H. 2016. Evaluating the User Experience of Interactive Digital Narrative. In *Proceedings of the 1st International Workshop on Multimedia Alternate Realities*, AltMM '16, 31–36. New York, NY, USA: ACM.

Roth, C., and Koenitz, H. 2017. Towards Creating a Body of Evidence-based Interactive Digital Narrative Design Knowledge: Approaches and Challenges. In *Proceedings of the 2Nd International Workshop on Multimedia Alternate Realities*, AltMM '17, 19–24. New York, NY, USA: ACM.

Seif El-Nasr, M. 2007. Interaction, narrative, and drama: Creating an adaptive interactive narrative using performance arts theories. *Interaction Studies* 8:209–240.

Short, E. 2016. Small-Scale Structures in CYOA.

Smith, G., and Whitehead, J. 2010. Analyzing the Expressive Range of a Level Generator. In *Proceedings of the 2010 Workshop on Procedural Content Generation in Games*, PCGames '10, 4:1–4:7. New York, NY, USA: ACM.

Szilas, N., and Ilea, I. 2014. Objective Metrics for Interactive Narrative. In *Interactive Storytelling*, Lecture Notes in Computer Science, 91–102. Springer, Cham.

Tarjan, R. 1972. Depth-First Search and Linear Graph Algorithms. *SIAM Journal on Computing* 1(2):146–160.

Thue, D.; Bulitko, V.; Spetch, M.; and Romanuik, T. 2010. Player Agency and the Relevance of Decisions. In *Interactive Storytelling*, Lecture Notes in Computer Science, 210–215. Springer, Berlin, Heidelberg.

Valls-Vargas, J.; Zhu, J.; and Ontañón, S. 2017. *Narrative Information Extraction with Non-Linear Natural Language Processing Pipelines*. Ph.D. Dissertation, Drexel University.

Vermeulen, I. E.; Roth, C.; Vorderer, P.; and Klimmt, C. 2010. Measuring User Responses to Interactive Stories: Towards a Standardized Assessment Tool. In *Interactive Storytelling*, Lecture Notes in Computer Science, 38–43. Springer, Berlin, Heidelberg.

Wardrip-Fruin, N.; Mateas, M.; Dow, S.; and Sali, S. 2009. Agency Reconsidered. In *Proceedings of DiGRA 2009*.