

Narrative Flow: A Formal Distant Reading Approach for Interactive Narratives

Coline Métrailler

Faculty of Arts, Department of Language and Information Sciences, University of Lausanne, bâtiment Anthropole, 1015 Lausanne, Switzerland

Abstract

Most distant reading methods are not designed to work on interactive narratives. This paper presents a model called *narrative flow*, a directed graph in which the nodes are narrative units and the edges are transitions, which can serve as an intermediate step to make those methods applicable to interactive narratives such as gamebooks or video games.

Keywords

interactive narrative, character network, video game, graph, distant reading

1. Introduction

Interactive narratives represent a challenge when it comes to analysing their narrative content: how can we compute the vocabulary size, the most frequent words, or some more complex mathematical objects such as character networks based on stories that can evolve differently every time we experience them? How do we measure the degree of variation (of vocabulary, style, topics,...) over all possible paths of an interactive narrative when the methods at disposal today are designed for linear narratives?

Video games or gamebooks are good examples of narratives which leave some degree of freedom to the player [1]. To study their scenario, choices need to be taken into account.

This paper introduces a model called *narrative flow*, which can serve as a bridge between interactive narratives and traditional distant reading methods. After a brief state of the art, we will present each element of a narrative flow, as well as a generic method to browse through it and gather all relevant information along the way. Then, we will present a concrete example and use a narrative flow on a video game called *Life is Strange*[2] to compute a character network of the first episode and make some basic analyses of the results.

2. State of the art and motivation

Distant reading, as famously called by Moretti [3] in 2000, is an approach often used to analyse books and stories from a computational angle. Observing the size of the vocabulary, the character

COMHUM 2022: Workshop on Computational Methods in the Humanities, June 09–10, 2022, Lausanne, Switzerland

✉ coline.metrailler@unil.ch (C. Métrailler)

ORCID 0000-0002-3196-481X (C. Métrailler)



© 2023 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

network (many examples are reviewed in the survey by Labatut and Bost [4]), the number of occurrences of specific words in each chapter or the embeddings of people and places (as in Grayson et al. [5]) are examples of distant reading methods that can bring valuable information when studying a work. But those methods are usually designed for linear narratives (as opposed to interactive narratives) and cannot be easily transposed to stories with bifurcations or choices, for which there is no single version of the plot to base the analyzes on.

There is a certain amount of research on interactive narratives. Among the notable results on this topic, Riedl and Bulitko have studied the way interactive narratives can be generated [6], Adams et al. have analyzed the shape of the different paths of the story [7], Hand and Varan have studied the impact on immersion and engagement as opposed to linear stories [8] and Ryan has highlighted the specific difficulties of balancing the freedom given to the player and the control on the story structure ([9]). But it seems that works like gamebooks are not yet studied from a narrative perspective, and adapting traditional distant reading methods to choice-driven stories raises some challenges.

In the specific case of video games, the question goes deeper: for years, game studies have opened a large debate on narratology and ludology, which has certainly contributed to marginalize the pure narrative content from the focus of the field, giving greater importance to game-specific elements and approaches. While this has led to fundamental results in game studies, narrative content in itself is still too rarely explored, and the lack of appropriate technical tools probably does not help to fill the gap.

The goal of this paper is to present a way to apply distant reading techniques to interactive narratives, with the help of an intermediate step called a narrative flow. This model synthesizes the shape and the relevant information of the story, and can then be used to generate one, several or all the linear narratives as experienced by different users.

3. Narrative flow

This section presents all the elements of a narrative flow: first the units and features, then the flow in itself and the traversals, and finally the aggregation function that turns traversals into exploitable data for further analysis.

3.1. Narrative units

The first building block of narrative flows is the *narrative unit*. It is a portion of a narrative, whose exact definition depends on the medium and the desired granularity of the flow. Narrative units can also be of different sizes. The union of all narrative units must represent a partition of the narrative, and one narrative can be segmented in many different partitions of narrative units. For example, a typical choice of narrative unit in a novel would be a sentence or a paragraph, but it could also be segmented in pages or chapters depending on the scale and type of analyses (paragraphs for co-occurrences of characters, whole chapters for vocabulary evolution or topic modeling, ...).

As these narrative units will then represent the smallest granularity of the narrative for further analysis, it is important to keep track of the useful information attributed to them. A narrative unit can thus be enriched with a set of *features*, i.e. pieces of information regarding

the unit that are considered relevant. Features can be explicitly mentioned in the unit, like locations or names of characters appearing in that portion of the narrative, or they can come from metadata, like the page number, the chapter or other medium-related properties that may serve as a mapping to the original content (like timestamps for movies, levels for video games or scenes for theater plays).

3.2. Narrative flow

Once the narrative has been segmented in a collection of units, along with their respective features, it can be modelled in a graph.

A *transition* is a directed link from one narrative unit to the next, representing a narrative sequence. There may be more than one outgoing or incoming transition towards or from a single unit, for example in gamebooks in which the reader has to choose between several options, or when different paths merge into a common conclusion. Each transition has a probability of being chosen, and the sum of all outgoing probabilities of a narrative unit must be 1. By default, and if there is no way to know if some transitions are more or less probable, they will be considered equiprobable.

A *narrative flow* is a directed graph in which the nodes are narrative units and the directed edges are transitions. It describes the sequencing of the narrative units. There is a large amount of possible narrative flows relative to a single narrative: it depends mainly on the chosen granularity of units, and on the desired features associated to those units. Without loss of generality, we assume that a narrative flow has only one unit with indegree 0, called *source*, and only one unit with outdegree 0, called *sink*. They will serve as starting and ending points for the whole narrative. Note that if the narrative has several endings, an extra unit (empty in terms of features) can be added with transitions from every possible ending to this new sink.

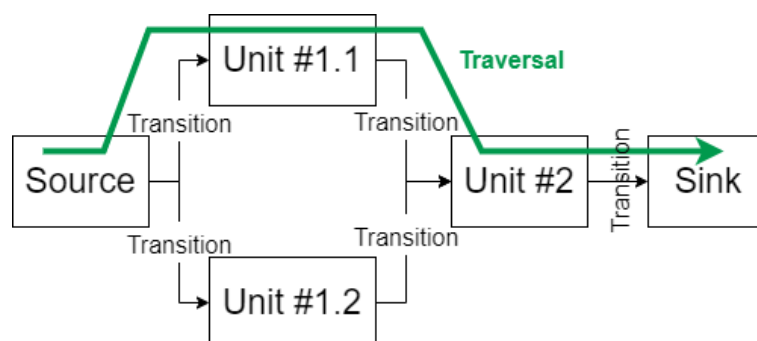


Figure 1: Structure of a narrative flow and a traversal.

3.3. Traversals

Now that the whole structure of the narrative has been considered, we can walk along this directed graph to follow the plot. This is the role of traversals: formally, a *narrative traversal* is a path from source to sink. More concretely, it represents one way the story is experienced from beginning to end (see figure 1).

In the case of a classical, linear¹ story such as a book or a movie, narrative flows will be fairly simple: every unit will have indegree and outdegree of at most 1, shaping the flow as a single line. In the most extreme case, all narrative units could be merged together, and the narrative flow would then be composed only of one unit and zero transitions, as seen in figure 2. In terms of traversals, once the flow is defined, there will be only one possible traversal from source to sink, representing one single way to experience the story from beginning to end.

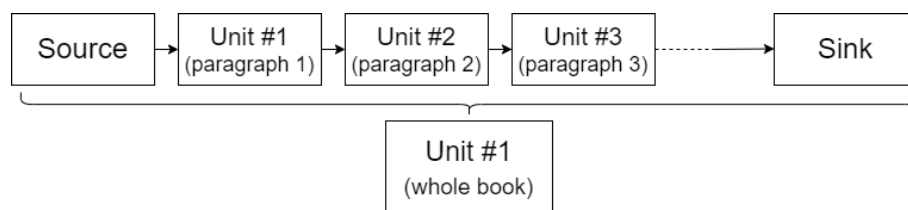


Figure 2: Narrative flow of a linear book can be collapsed in a single unit.

In the case of an interactive narrative, however, narrative flows cannot be collapsed into a single unit: each possible narrative flow must preserve the bifurcations of the story. Units linked by a single transition can still be merged, but even the most condensed version of the narrative flows will display all the transitions that show bifurcations in the narrative. This means that if at least one unit has an outdegree of more than 1, there will be more than one possible traversals for the related narrative flow. Actually, if the graph contains at least one cycle², there is an infinite number of possible traversals, as they can loop infinitely in those cycles.

3.4. Aggregation function

To gather all features on a specific traversal, we need an *aggregation function*. The definition of this function depends on the nature of the story, the format of features, and various other factors, but the general idea is to collect all relevant features on units visited by a specific traversal, in order to be able to use them in further analyses with methods initially developed for linear narratives. For example, to study the vocabulary size of one specific path in a gamebook, unique tokens present in each narrative unit can be listed as a feature, and the aggregation function will collect and merge all those lists along the desired traversal, thus forming a complete list of all unique tokens in this path of the gamebook.

4. Concrete example: character networks on a video game

The following sections are aimed at clarifying those definitions with a concrete example. We will use a narrative flow to build character networks on the game *Life is Strange*, a narrative game with a strong focus on choices made by the players.

¹“linear” here means that there are no bifurcations in the narrative, not necessarily that the story is told in chronological order.

²This can happen in narratives: one example of cycle would be dialogues with non-player characters in a video game, when we are free to interact with them as many times as we wish and they produce the same answer endlessly.

4.1. Character network generation

A character network is a graph in which the nodes represent characters in a narrative, and the edges represent interactions between two characters. It is used on various types of media (books, movies, comics, etc.) to synthesize narratives, aggregate information and highlight specific aspects in the structure of a story (e.g. Moretti [3], Amancio [10], Gil et al. [11] or the many references cited in the survey by Labatut and Bost [4]). It is easily built from a linear narrative, once the list of relevant characters is established and the type of interactions selected and extracted (co-occurrences, dialogues, mentions, etc.). But in an interactive narrative, both nodes and edges are subject to change: some interactions may or may not happen based on the choices, character occurrences may vary with each story experience and the whole character network generation therefore presents specific challenges (as stated by Rochat and Triclot [12] for video games).

In this example, we will focus on character networks based on co-occurrences³. We thus need to aggregate the narrative in blocks in which the characters are considered co-occurring, and in each block we need the list of co-occurring characters. This can be achieved with the units and features in the narrative flow: the units can be defined to divide the narrative in the desired size of blocks, and each unit will be assigned, as a feature, a list of co-occurrences between all characters present in this unit. Then, for a chosen traversal of the narrative flow, the aggregation function can gather all the co-occurrences of visited units in a single list, thus allowing the generation of a character network of this traversal.

Note that for interactive narratives, each traversal will result in a possibly different character network. Depending on the type of information we want to highlight, diverse strategies can then be chosen to study the variation between those networks. One option is to run a large amount of random traversals, generate the related networks and gather them in a new single network where the weight of an edge is the average of the weights of that same edge across all generated networks, and its standard deviation indicates how much it varies on those traversals. It can also be interesting to do some network clustering on the generated networks and to gather them into groups of networks, in order to find typical patterns and to identify how many significantly different scenarios can be experienced in the same interactive narrative.

4.2. *Life is Strange*

To illustrate the use of narrative flows and the generation of character networks as described in section 4.1, we consider the video game *Life is Strange*, developed by Dontnod Entertainment and released in 2015. This game is particularly appropriate here for several reasons. A game script has been provided by the community and can be found online [13], which is rare enough to be mentioned. *Life is Strange* is a story-driven game in which the player is often asked to choose between dialogues or between actions. Some of those decisions are considered minor and have an effect only on the next answer, but a small portion of the choices is considered crucial for the development of the story: for those choices, the percentage of players who took each option is available (the statistics are shared during the game, and can also be found on the

³Two characters are in co-occurrence if they appear in the same “window” (e.g. paragraph, group of paragraphs, chapter, etc.).

same website), which is another reason that motivated the choice of this game for our example.

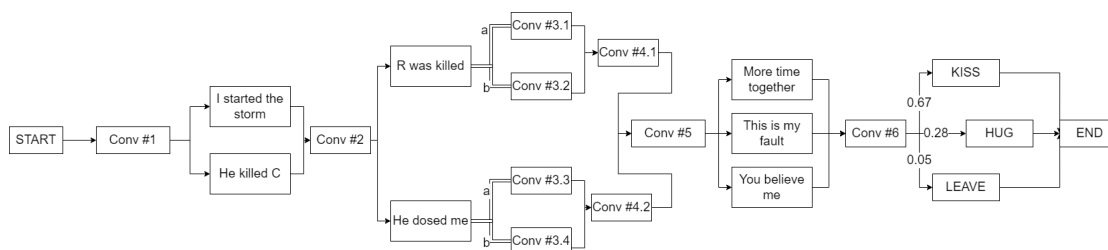


Figure 3: Narrative flow of the “Conversation with Warren” in Episode 5 of *Life is Strange*.

Figure 3 shows a narrative flow of a small part of the game, the “Conversation with Warren” in Episode 5. In that example, to simplify the graph and make it readable, each narrative unit has been chosen with the biggest possible size while preserving all the choices (meaning that no narrative unit has both indegree and outdegree 1). Double lines (e.g. leading to Conv #3.1) represent choices which follow from a previous decision (*a* or *b*, in that case, both depending on some actions made (or not) in earlier chapters of the game). By default, the outgoing transitions of a narrative unit are equiprobable and not specified in the graph. But the last three choices (written in capital letters) belong to the list of crucial choices for which probabilities are given, which is why we were able to indicate 0.67, 0.28 and 0.05 on the respective transitions.

A narrative flow following the same logic has been extracted automatically⁴ from the first episode of the game, using specific flags present in the structured text to detect bifurcations and correctly create the units and the transitions (see a sample of text and examples of flags in figure 4).

The result is a fairly big narrative flow of 855 narrative units and 1014 transitions, which we decided to simplify to accelerate the simulation of traversals (considering the computation of all traversals of a graph is factorial). To do so, we identified all choices and collapsed them into a single unit whenever the resulting lists of interactions were identical on every option. This allowed to divide by two the number of units and transitions.

Then, to keep the example as simple and straight-forward as possible, 500 random traversals of this simplified narrative flow have been generated⁵ based on the probability of each transition. Depending on the desired results, other ways to generate those traversals could prove more relevant, like looking for the shortest path or sticking to the first option at each bifurcation, or considering the flow as a Markov chain and computing global results directly on the transition matrix.

Each traversal results in a character network (computed with *charnetto*⁶) and all the networks are then gathered into a “mean” network (see figure 5) in which

- the width of edges represents the weight, computed as the mean weight of that edge over

⁴Note that all the other games in the *Life is Strange* franchise are also available in the same website with similar script structure, which motivated the automation process to extract the narrative flow, as it will then be reusable on a wider dataset.

⁵With the package *networkx* [14] in Python.

⁶<https://pypi.org/project/charnetto/>

```

====Optional Conversation with Dana====
'''Max:''' Hey, Dana.
'''Dana:''' Seriously, thanks again. I can't believe Juliet locked me in my own room. Real mature.

[Victoria is not nice]
'''Max:''' Victoria is not nice. I don't get it. She has everything. And to pull that prank on a friend.
'''Dana:''' Just because they're in the Vortex Club doesn't mean they're BFF's. I'm in it and Victoria creeps me out. Max, you're smart to be a loner here. Though Warren obviously likes hanging with you...
</blockquote>

[He's a good guy.]
'''Max:''' He's a good guy. A geek like me.
'''Dana:''' You saw the files in his flash drive? Tons of crazy shit. Films I've never even heard of. And in a special folder called "Max"?
</blockquote>

[What do you mean?]
'''Max:''' What do you mean?
'''Dana:''' Oh, nothing. When you opened the flash drive, didn't you see the special folder called "Max"?
</blockquote>
]]

</blockquote>
'''Max:''' Yeah, okay. I gotcha.
'''Dana:''' Are you blushing? Ha! Go ahead and grab the flash drive whenever...
</blockquote>

[Juliet gets her rage on.]
'''Max:''' Poor Juliet gets her rage on. I hope she doesn't kill Victoria...
'''Dana:''' Zachary should worry too. As if I'd let that ego case jockstrap touch me. But I bet Victoria would be all over that action. Or has been... Max, you're smart to be a loner here. Though Warren obviously likes hanging with you...
</blockquote>

```

Figure 4: Sample of the script of Episode 1 of *Life is Strange*, as found on the website. Characters are highlighted in yellow, flags in green represent choices, and flags in blue represent “hidden” choices: in that case, options “He’s a good guy.” and “What do you mean?” appear only if the player first selected “Victoria is not nice.”

every network

- the color represents the standard deviation of that weight (highlighting in red the edges which vary the most over all traversals).

At this point in the game, each of the 500 generated character networks has the same set of nodes and edges⁷, and the small weights of the edges shows that there are only few co-occurrences of each possible pair of characters. As this network represents only the first episode of the game, that kind of behaviour can be expected: the player controls Max, who gets introduced to everyone superficially and mainly interacts with the other main character, her best friend Chloe.

Following the same logic, almost every important edge (in terms of weight) involves Max, except the interaction between Chloe and David, which happens at a moment in the game where Max can choose to stay hidden or join the conversation. As we see in table 1, almost two thirds of the edges have no variation on the 500 traversals, which means that no matter the choices of the player, the characters involved interact the exact same amount of times. Again, as it is the first episode, the main consequences of the player’s choices may be more noticeable on the next parts of the game, causing more variation in the character networks. Here, the combination of the narrative flow and the character networks highlights the fact that despite an already fair amount of choices proposed to the player in this first episode of the game (as visible in the size of the narrative flow), those choices have few impact on the overall network of interactions, suggesting this episode to be relatively linear in terms of narrative complexity.

⁷This could be expected to vary later in the game, as the player’s choices can lead to some characters disappearing or not, for example.

Table 1

All edges of the aggregated character network, sorted by standard deviation.

edge	weight	std	edge	weight	std
Chloe – David	6.356	1.737027346	Max – Principal Wells	3	0
Max – Ms. Grant	3.524	1.499807988	Brooke – Max	2	0
Chloe – Max	48.142	1.18736515	Dana – Juliet	2	0
Luke – Max	1.874	1.114506169	David – Kate	2	0
Max – Victoria	7.42	0.873842091	Juliet – Zachary	2	0
Max – Mr. Jefferson	7.42	0.873842091	Nathan – Warren	2	0
Mr. Jefferson – Victoria	7.42	0.873842091	Alyssa – Zachary	1	0
Hayden – Max	2.948	0.72200831	Chloe – Warren	1	0
Dana – Max	5.99	0.702780193	Courtney – Max	1	0
David – Max	4.074	0.514318967	Courtney – Samuel	1	0
Kate – Max	8.494	0.499963999	Courtney – Victoria	1	0
Daniel – Max	2.494	0.499963999	Dana – Justin	1	0
Max – Warren	9.478	0.499515766	Daniel – Friend	1	0
Max – Stella	3.666	0.471639693	Daniel – Logan	1	0
Justin – Max	4.31	0.462493243	David – Principal Wells	1	0
Juliet – Max	7.008	0.379388982	Friend – Logan	1	0
Max – Nathan	6	0	Max – Taylor	1	0
Evan – Max	5	0	Max – Zachary	1	0
Courtney – Taylor	4	0	Samuel – Taylor	1	0
Max – Samuel	4	0	Samuel – Victoria	1	0
Alyssa – Max	3	0	Taylor – Victoria	1	0
Chloe – Nathan	3	0			

5. Conclusion

This example serves to show a way narrative flows can be constructed and used to run computational analyses on interactive narratives. In the case of *Life is Strange*, we only scratched the surface of what can be observed and interpreted. We plan to do a full analysis of the five episodes, in order to

- see how the network varies across all episodes
- apply clustering methods on the networks
- highlight significantly different ones
- reflect on it being a choice-driven game in terms of character interactions

among other approaches.

But this model is designed to work on other types of narratives (such as gamebooks or interactive fictions) and with other approaches than character networks, and there are many tracks to follow from that starting point. For example, many text analytic methods rely on the document-term matrix, and this object could be modeled in different ways using narrative flows. If the goal is to study the use of certain terms throughout one traversal, rows in the matrix could be the narrative units visited in that traversal. For a comparative study of the whole narrative over several traversals, to see how many times specific terms appear in different paths of the narrative, rows could be the traversals themselves, making it a traversal-term matrix.

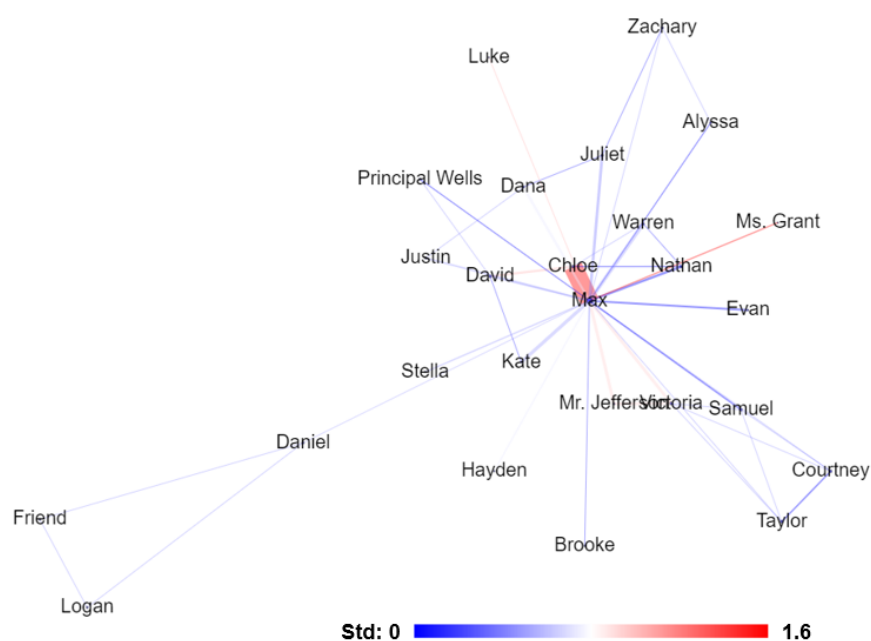


Figure 5: Aggregated character network of Episode 1 of *Life is Strange*, over 500 traversals of the narrative flow.

There are of course many more ways narrative flows could prove useful, and we hope this technique will make it easier to study interactive fictions, gamebooks or video games in terms of their narrative content, as this seems to be a largely unexplored path for now.

References

- [1] S. Domsch, *Storyplaying: Agency and Narrative in Video Games*, De Gruyter, 2013. URL: <https://library.oapen.org/handle/20.500.12657/24643>. doi:10.1515/9783110272451.
- [2] Dontnod Entertainment, *Life is Strange*, Square Enix, 2015. PC version.
- [3] F. Moretti, Network theory, plot analysis, *New Left Review* (2011) 80–102.
- [4] V. Labatut, X. Bost, Extraction and Analysis of Fictional Character Networks: A Survey, *ACM Computing Surveys* 52 (2019) 1–40. URL: <http://arxiv.org/abs/1907.02704>. doi:10.1145/3344548, arXiv: 1907.02704.
- [5] S. Grayson, M. Mulvany, K. Wade, G. Meaney, D. Greene, Novel2vec: Characterising 19th century fiction via word embeddings, in: *Irish Conference on Artificial Intelligence and Cognitive Science*, 2016.
- [6] M. O. Riedl, V. Bulitko, Interactive Narrative: An Intelligent Systems Approach, *AI Magazine* 34 (2013) 67–67. URL: <https://ojs.aaai.org/index.php/aimagazine/article/view/2449>. doi:10.1609/aimag.v34i1.2449, number: 1.
- [7] D. Adams, D. Beckelhymer, A. Marr, Choose your own adventure: An analysis of interactive

- gamebooks using graph theory, *Journal of Humanistic Mathematics* 9 (2019) 44–59. doi:10.5642/jhummath.201902.05.
- [8] S. Hand, D. Varan, Interactive narratives: Exploring the links between empathy, interactivity and structure, in: *EuroITV*, 2008.
- [9] M.-L. Ryan, From Narrative Games to Playable Stories: Toward a Poetics of Interactive Narrative, *StoryWorlds: A Journal of Narrative Studies* 1 (2009) 43–59. doi:10.1353/stw.0.0003.
- [10] D. R. Amancio, Network analysis of named entity co-occurrences in written texts, *EPL (Europhysics Letters)* 114 (2016). URL: <http://arxiv.org/abs/1509.05281>. doi:10.1209/0295-5075/114/58005, arXiv: 1509.05281.
- [11] S. Gil, L. Kuenzel, C. Suen, Extraction and Analysis of Character Interaction Networks From Plays and Movies, in: *International Conference on Digital Health*, 2011.
- [12] Y. Rochat, M. Triclot, Les réseaux de personnages de science-fiction : échantillons de lectures intermédiaires, *ReS Futurae. Revue d'études sur la science-fiction* (2017). URL: <http://journals.openedition.org/resf/1183>.
- [13] *Life is Strange* script, 2016. URL: https://life-is-strange.fandom.com/wiki/Episode_1:_Chrysalis_-_Script, last accessed: July 13, 2023.
- [14] A. A. Hagberg, D. A. Schult, P. J. Swart, Exploring network structure, dynamics, and function using networkx, in: G. Varoquaux, T. Vaught, J. Millman (Eds.), *Proceedings of the 7th Python in Science Conference*, Pasadena, CA USA, 2008, pp. 11 – 15.