Fine-grained dynamic graph for novel comprehension - Project Proposal -

Gaël de Léséleuc de Kérouara
gael.de-leseleuc@student-cs.fr

Alexandre Duval
alexandre.duval@student-cs.fr

Thomas Lamson
(unregistred student for NSA course)
thomas.lamson@student-cs.fr

1 PROBLEM DEFINITION

1.1 Motivation

Novels often propose a complex narrative and an elaborated plot that are sometimes difficult to grasp and follow throughout the book. Visualising different aspects of the novel and its evolution across time through a dynamic graph networks seems to be an intuitive approach to improve one's comprehension of it. Many authors using character based networks to solve higher-level problems identify the dynamics of the story as key aspect that needs to be taken into account. This especially holds in long-term narratives such as a novel saga, in which relationships and characters are likely to change over time. Yet, the vast majority of methods proposed in the literature rely on static networks which represent a novel in its entirety. As a result, an important part of the narrative story, ie the dynamic information, is lost. This motivates us to shed further light on dynamic narrative by using fine-grained graph which keep all information about temporal data.

1.2 Proposed concept

In short, we would like to create a fine-grained graph which keep the dynamic aspect of narrative and to study several key aspects to improve one's comprehension of the novel. In particular, we would like to investigate the network of characters: including the evolution of their role, their relations, their importance throughout the book. We also aim to identify structural changes in the plot, display its core component and isolate some specific segments of the story according to one's needs. We will therefore explore a wide range of graph properties and draw conclusions from them, both using supervised and unsupervised learning. An important place will be granted to the structure and visualisation of the graph, including how to represent the temporal aspect.

1.3 Related works

The majority of resources we have found closely related to our project mostly consider the static aspect. Among them, [4], which we saw in class, is interesting in the way it interprets the novel's (character) graph. Indeed, it uses centrality measures to detect main characters and apply clustering to detect the different communities. Similarly to [2], it constructs the network by applying Name Entity Recognition and drawing edges between two characters if they co-occur in window of like 15 words. In the later, they also detail a few tricks they used to solve co-reference. The aim of their paper is to build social networks from novel to quantify their plot and structure. They create static and dynamic character networks and extract features (graph density, clustering coef, central/isolate nodes proportion, weight of biggest and second biggest node, presence of character throughout novel...) to perform classification on genre and author. To do so, they collect 238 books from Gutenberg, written by a small set of authors and grouped into 11 genres. They find many interesting patterns and use purity, entropy and F1 measure to evaluate their findings. [5] study the difference of structures in novels' graphs, in rural and urban settings. It shows a different way to approach some parts described above. More precisely, it applies clustering on entities to solve the issue of co-reference. It also weights the edge not only by the number of time two entities appears together but also by the the length of dialogue between these two characters.

All these aspects, that is co-reference, interaction, classification tasks with handcrafted features, are equally detailed in the survey of methods [6]. However, the latter delves further into the visualisation techniques of dynamic networks. The one method we retain is to first plot the whole network as a shadow and then plot on this shadow network the graph obtained at each chapter. More information on nice and comprehensive dynamic visualisation is detailed in the following resource: [3]. They talk about representation of changes (marking event in the book's plot), temporality changes, structural tricks, visual emphasis, etc. Some of the challenges they faced concerned characters identification, changes depiction, order of events, spatial context and the number of elements for big novels. On a similar note, [7] use force-vector spatialization to place nodes in the space. The algorithm simulates a system of physical forces: nodes repulse each other, while edges act as springs attracting the nodes that they connect.

2 METHODOLOGY

2.1 Dataset generation

To conduct this project, we will download some targeted online books from Project Gutenberg [1] and retrieve the corresponding metadata for each book (including theme, genre, author, etc.).

2.2 Graph generation

Our idea is use to first create a graph which keep all the temporal information about the narrative dynamic. An example of such graph is represented on left-side of **figure 1**. We will now describe the plan to generate such a graph.

1. Name entity recognition for character occurrence nodes . We first plan to apply BERT NER to extract the entities present in the text, and especially the different characters'. Entities such as location or organisation do not constitute our primary focus but might still be considered later if they appear relevant.

[1] https://www.gutenberg.org/

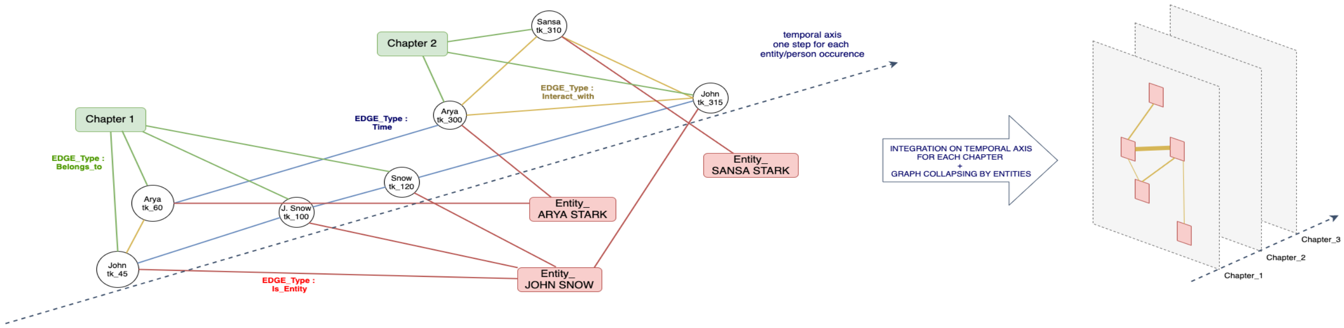

Figure 1: Left: fined-grained global narrative graph - Right: a sequence of narrative graph per chapter, obtained by integrated the global graph on the temporal dimension and reducing the occurrences nodes by using entities node

Each time a character appears, we will create a new node and keep in memory two features : the character name and the position of this occurrence in the text.

2. Co-reference for entity node.
Then, we need to tackle the issue of co-reference and link each entity to character name despite the fact each entity might be described by several names in the novel. For that, we planned to use some techniques developed in [1] to link each occurrence node to a correspond entity node. For instance link "Mme Bovary", "Emma", "Emma Bovary", "Madame Bovary" occurrence nodes to the same "Emma Bovary" entity nodes.

3. Temporal edges.
For each entity node, we will consider each related occurrence nodes and sequentially connect these by a temporal edge. For instance, if you consider the entity node ("John Snow") and the related occurrence nodes n1 : ("John", tok n°30), n2 ("John Snow", tok n°60), n3 ("Snow", tok n°80), node n1 will be link to n2, n2 to n3, etc.

4. Co-occurrence sliding window for interaction edges.
Thanks to localization features stored during first step, we will be able to link two occurrence nodes if the two character name appears in a same temporal windows of for instance 20 tokens. Of course, we will only create an edge only if the two character name correspond to distinct entities.

2.3 Fine-grained global graph to dynamic graph representation

Because our graph will be fine-grained, we can derived from it several sub-graphs of interest depending of the tasks we want to tackle. For instance, it will be straight-forward to construct from our global graph a sequence of static graphs where nodes and edges have different sizes given the importance of a node and the strength of a connection between two characters.

2.4 Graph analysis and comprehension

In addition to this, we will particularly focus on some visualisation techniques evoked in **Telling stories about Dynamic Networks with Graphs comics (2016)**, trying for instance to get nice and clear representations, preserve the order of events or show structural changes. We consider the visualisation aspect to be a key step towards understanding the novel. Finally, we will investigate how to represent this sequence of static graph into a single dynamic graph endowed with a time axis.

In this part, we shall look at different properties of the graph we have just created. To only name a few: connectedness, clustering and community detection, nodes' importance, centrality, presence and disappearance of characters throughout the novel, etc. Investigating these features can provide us information regarding the basic structure of the novel. Indeed, we can infer if there is one main character or several, what are their relation, how does it evolve across the book, how does the author construct the story and many other relevant aspects. We are planning to use this supervised approach as a baseline and complete it with some unsupervised learning through the use of graphical neural networks for some tasks.

Furthermore, we will try to create subgraphs with respect to some particular characters to visualise only its place in the story. On a similar note, we are thinking about studying a saga (e.g harry potter) to take our study to a larger scale and now compare books instead of chapters.

Finally, another idea we have concerns k-core decomposition. We could try to summarize the final dynamic graph network using its k-core decomposition and compare it with the graph of a summary of the book (that we can find on a website).

REFERENCES

[1] Mariona Coll Ardanuy and Caroline Sporleder. [n.d.]. Structure-based Clustering of Novels.
[2] Mariona Coll Ardanuy and Caroline Sporleder. 2014. Structure-based clustering of novels. In *Proceedings of the 3rd Workshop on Computational Linguistics for Literature (CLFL)*. 31–39.
[3] Benjamin Bach, Natalie Kerracher, Kyle Wm. Hall, Sheelagh Carpendale, Jessie Kennedy, and Nathalie Henry Riche. 2016. Telling Stories about Dynamic Networks with Graph Comics. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. Association for Computing Machinery, New York, NY, USA, 3670–3682. https://doi.org/10.1145/2858036.2858387
[4] Andrew Beveridge and Jie Shan. 2016. Network of Thrones. *Math Horizons* 23, 4 (2016), 18–22. https://doi.org/10.4169/mathhorizons.23.4.18
[5] David Elson, Nicholas Dames, and Kathleen McKeown. 2010. Extracting Social Networks from Literary Fiction. In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*. Association for Computational Linguistics, Uppsala, Sweden, 138–147. https://www.aclweb.org/anthology/P10-1015
[6] Vincent Labatut and Xavier Bost. 2019. Extraction and Analysis of Fictional Character Networks: A Survey. *Comput. Surveys* 52 (09 2019), 89. https://doi.org/10.1145/3344548
[7] Tommaso Venturini, Liliana Bounegru, Mathieu Jacomy, and Jonathan Gray. 2017. *How to Tell Stories with Networks: Exploring the Narrative Affordances of Graphs with the Iliad.*