



**HAL**  
open science

# Complex Network Analysis of a Graphic Novel: The Case of the Bande Dessinée Thorgal

Vincent Labatut

► **To cite this version:**

Vincent Labatut. Complex Network Analysis of a Graphic Novel: The Case of the Bande Dessinée Thorgal. *Advances in Complex Systems (ACS)*, 2022, 25 (5&6), pp.2240003. 10.1142/S0219525922400033 . hal-03694768

**HAL Id: hal-03694768**

**<https://hal.science/hal-03694768>**

Submitted on 15 Jun 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

## Complex Network Analysis of a Graphic Novel: The Case of the Bande Dessinée *Thorgal*

VINCENT LABATUT

*Laboratoire Informatique d'Avignon (LIA)*  
*Avignon Université*  
339 chemin des Meinajariès – BP 1228  
F-84911 Avignon Cedex 9 – France  
*vincent.labatut@univ-avignon.fr*

The task of extracting and analyzing character networks from works of fiction, such as novels and movies, has been the object of a number of recent publications. However, only a very few of them focus on graphic novels, and even fewer on European graphic novels. In this article, we focus on *Thorgal*, a bande dessinée, i.e. a comic of the French-Belgian tradition. We manually annotate all the volumes of this series, in order to constitute a corpus allowing us to extract its character network. We perform a descriptive analysis of the network structure and compare it to real-world and fictional social networks. We also study the effect of character filtering over the network structure. Finally, we leverage complex network analysis tools to answer two research questions from the literature, related to the similarity between *Thorgal* and the *Saga of Icelanders*; and to the position of women in the series. Our data and source code are both publicly available online.

*Keywords:* Character Networks; Graphic Novels; Complex Network Analysis.

### 1. Introduction

Character networks are a specific type of complex network built by identifying the characters involved in a narrative as well as their interactions, and modeling them through vertices and edges, respectively. As such, they constitute a model of the narrative, that can be leveraged to perform a number of tasks, such as summarizing a narrative [7], identifying its main characters [48], and separating its storylines [57]. They can also be used to answer various questions related to corpora of narratives, e.g. automatically classifying narratives by genre [22], or to specific narratives, e.g. assessing literary theories [16] and estimating the level of historicity of a narrative [33].

This modeling tool has been applied to a number of works of fiction in recent years, spanning all the main forms of these cultural objects: novels [16], tales [50], myths [33], fables [36], plays [51], movies [28], TV series [9], video games [49], and graphic novels [1] (see [27] for a complete review). However, there are only very few publications regarding the later form, and even fewer deal with *European* graphic novels. This may be because extracting character networks from comics is a difficult task to automate [27]. Yet, considering that the comic form has its own specificities but also mixes different elements from other forms [10], one may wonder how it compares to them, narratively speaking. From the modeling perspective, this amounts to finding which features are common to comic-based character networks, and what makes them different from those extracted from other forms. Of course,

## 2 Vincent Labatut

answering this question would require a larger corpus of comic-based character networks.

In this article, we add our contribution to this corpus by focusing on *Thorgal*, a bande dessinée, i.e. a comic of French-Belgian tradition. We manually annotate all the volumes of this series, in order to constitute a dataset allowing us to extract its character network. We perform a descriptive analysis of the network structure and compare it to real-world and fictional social networks. We also study how the process of filtering out minor characters, which is largely spread in the literature, affects the network structure. After this, we leverage complex network analysis tools to answer two research questions identified in the literature, and specifically related to *Thorgal*. The first concerns its hypothesized similarity with the *Íslendingasögur*, a Scandinavian corpus of medieval literature [54]. The second is to know whether *Thorgal* actually implements a form of soft masculinity [15], by studying the position of women in the series.

The rest of the article is organized as follows. In Section 2, we review the literature dealing with the character networks of graphic novels. In Section 3, we describe the methods that we use to annotate our dataset and to extract our character networks. In Section 4, we introduce the *Thorgal* series and our related research questions. We also describe the raw dataset obtained after our annotation process. Section 5 is dedicated to the descriptive analysis of the character networks, whereas Section 6 focuses on the remaining research questions. We conclude in Section 7 by summarizing our findings and contributions, and discussing the limitations and possible extensions of our work.

## 2. Related Work

There are only a very few published articles dealing with the extraction and analysis of character networks from graphic novels. As far as we know, the articles discussed in this section and listed in Table 1 constitute a comprehensive survey. Some of them deal with American super-hero comic books, Japanese mangas, Korean webtoons, or European bande dessinée.

Alberich *et al.* [1] study the *Marvel Comics* social network. They first extract a bipartite graph using the *Marvel Chronology Project*<sup>a</sup>, a fan-made database covering the 1961–2002 period, by linking each significant character to the comic books in which they appear. Then, they perform a one-mode projection to get a character-to-character network modeling same-book co-appearances. They compare the resulting unweighted co-occurrence network to a null random model and to real-world collaboration networks, such as the ones resulting from scientific co-authorship or from movie castings. Some of the studied topological properties are realistic, in the sense that they are similar to those of real-world networks (existence of a giant component, small average distance). Others are not, though: compared to the null model, the average degree is even smaller than in realistic collaboration networks, underlining the intensive re-use of the same characters in the books. Similarly, the exponent of the power law-distributed node degree is smaller than in other collaboration networks, confirming the artificiality of the network: unlike in real life networks,

<sup>a</sup><http://www.chronologyproject.com/>

Table 1: List of articles dealing with the extraction of character networks based on graphic novels.

Bibliographic References	Graphic Novel		Application
	Type	Title	
Alberich <i>et al.</i> [1] Gleiser [20]	Comics	<i>Marvel Comics</i>	Study the level of realism of the social network
Murakami <i>et al.</i> [39]	Mangas	<i>Dragon Ball</i> vol.32	Detect the most important characters
Murakami <i>et al.</i> [40, 41]	Mangas	<i>Naruto</i> vol.46–50 <i>Dragon Ball</i> vol.29–34 <i>Boys Over Flowers</i> vol.16–20 <i>Space Brothers</i> vol.1–5	Build character family trees / kinship graphs
Rochat & Triclot [49]	Bandes Dessinées	<i>Saga of the Metabarons</i> vol.1–8 <i>Worlds of Aldebaran</i> vol.6–10	Narrative classification / Study of character roles
	Mangas	<i>Akira</i> vol.1–3 (French Edition) <i>Gunnm</i> vol.1–7	
Lee & Kim [30]	Webtoons	Collection of 35 webtoons	Predict user preference

where the number of collaborations remains limited because of time constraints, the *Marvel* network is dominated by a few widely connected characters.

Gleiser [20] uses the same data and extraction method, and extends the work of Alberich *et al.* [1] by considering other network properties. First, the *Marvel* network turns out to be disassortatively mixed by degree, suggesting, like the previously discussed small power law exponent, the presence of large hubs within the network as an evidence of its artificial nature. Second, the author introduces edge weights corresponding to the number of books in which the characters co-appear. The resulting weight distribution follows a power law with a relatively small exponent, suggesting that only a few characters interact frequently. Filtering weak edges reveals a few communities centered around popular characters. Third, the clustering coefficient turns out to be much lower for widely connected characters than for more isolated ones. Besides, the central characters, mostly super-*heroes*, as opposed to super-*villains*, tend to be occasionally connected to each other, forming a *rich-club* of heroes able to triumph over evil, in accordance with the *Comics Authority Code*.

Murakami *et al.* [39] extract character networks from mangas. First, they manually annotate character occurrences in each panel. Second, they use character co-occurrence as a proxy for interaction. The intensity of this interaction is measured by a score that depends on the number of involved characters (the fewer, the better). They also assume an interaction, albeit weaker, between characters appearing in two consecutive panels. Third, they built a static network by summing these scores over the narrative to get edge weights. They apply their extraction method to Toriyama's *Dragon Ball* vol.32, and leverage vertex strength and edge weight to identify automatically the most important characters. A survey conducted over readers confirms the relevance of this approach. In [40, 41], the same team takes advantage of the text present in the speech balloons, in order to identify the nature of the character interactions. Both characters and speech text are detected manually. They assume speech is addressed to characters in the same frame, or the next one. They look for predefined kinship words in the text, such as *son* or *father*,

## 4 Vincent Labatut

and infer the relation between the speaker and the addressee. They mainly look for family relationships (father/son, spouse, etc.) over the whole narrative. They filter out infrequent relationships, considered as noise. They eventually produce a sort of family tree. They apply their method to 4 different mangas this time: *Naruto* (vol.46–50), *Dragon Ball* (vol.29–34), *Boys Over Flowers* (vol.16–20), and *Space Brothers* (vol.1–5).

Rochat & Triclot [49] do not focus specially on comics, but more generally on a *genre* of narrative: they study a corpus of science-fiction novels, video games, movies, TV series and graphic novels. The latter include mangas *Akira* (vol.1–3) and *Gunnm* (vol.1–7), and bandes dessinées *The Saga of The Meta-Barons* (vol.1–8) and *The worlds of Aldebaran* (vol.6–10). They identify character occurrences manually, and insert an edge between two characters when they appear within two consecutive pages. Its weight is 2 if both characters occur within the same page, and 1 for two distinct pages. They sum over the whole book to get a static network, and filter out edges whose weight is smaller than 3, as these are not considered as significant. Although not explicitly stated in the paper, their data reveal that the authors focus only on the main characters. The authors identify 4 types of character networks, based on their general structure: unicentric, acentric, polycentric, and kernel networks (we detail these in the Supplementary Material, cf. Section SM2.4). They additionally categorize the characters using 3 classes (scientist, politician, technician) and study how these are distributed over the networks.

Lee & Kim [30] use an embedding-based method to predict user preference regarding webtoons. Webtoons are a narrative medium popular in Korea, corresponding to online comics with music and possibly animation. A webtoon is not divided into distinct pages, but is rather a single long strip that can be conveniently read on a smartphone using infinite scrolling. It is generally serially published, and therefore split in distinct episodes. The authors propose a method to learn a multimodal representation of the narrative based not only on the character network, but also on its textual, visual and audio content. They extract a dynamic character network by considering each episode as a time slice. In each episode, they manually identify the characters and their verbal interactions. They assume that all occurring characters address each other, and consequently connect them all in the considered time slice. The numbers of utterances are used as edge weights. They do not directly use the network, but rather project it into an embedding space using their own method [29]. They also embed the textual content using a Long Short-Term Memory neural network (LSTM), and visual and audio features using a Convolutional Neural Network (CNN). They combine these representations and use the result to assess user preference over a corpus of 35 webtoons.

We conclude this review with two general observations. First, only one work deals with European bandes dessinées, and it does not provide a thorough analysis of the extracted networks, as those are used as part of a larger collection of narratives. Second, all authors extract their character networks through co-occurrence based approaches, differing only in the nature of the selected window: whole comic books for Alberich *et al.* [1] and Gleiser [20], pairs of panels for Murakami *et al.* [39], pairs of pages for Rochat & Triclot [49], and episodes for Lee & Kim [30]. By comparison, we adopt a scene-based approach, which we deem more appropriate to the series studied in this article.

### 3. Network Extraction Method

A comic *page* is composed of a set of chronologically ordered *panels*, generally separated by an empty space called *gutter*. A panel itself can contain characters, background, but also text that can be contained in speech balloons or directly integrated in the drawings, and even iconography. As such, the automatic processing of comics is related to document image understanding, but with additional constraints regarding organization and structure [24]. The problem of extracting information from comics is very difficult, as information can appear under many forms, and is subject to extreme variations from one artist to the other (or even within the same artist's work). It requires solving a number of lower level problems [3], which are themselves hard, and for the most part not satisfactorily treated by state-of-the-art methods. These include identifying panel boundaries, ordering panels, identifying speech bubbles and captions, extracting their textual content, detecting out-of-bubble text, identifying characters and recognizing their poses, attributing speech to characters, detecting and interpreting iconography and graphical conventions. For this reason, a part of the character network extraction process is performed manually in all the works described in Section 2, and this is also the case for us.

In this article, we adopt the general method described in [27] to extract character networks from any type of narrative. It is three-stepped: 1) identify character occurrences throughout the narrative; 2) detect interactions between these characters; and 3) perform some form of temporal integration to obtain a graph. We perform the first and second steps manually (Section 3.1) whereas the third one is automated (Section 3.2).

#### 3.1. Manual Annotation

The annotation step is generally not described in a very detailed way in the literature, if mentioned at all (cf. Section 2). Yet, the method and nature of the annotation have a direct effect on the constituted dataset, and therefore the results obtained later when processing it. Therefore, this step should not be overlooked. Moreover, the task is not as straightforward as it may seem at first glance, as it requires dealing with various specific situations. These are the reasons why we summarize our annotation approach in this section, and review the most typical of these situations. We illustrate them using excerpts from the *Thorgal* series, which is studied later in this article.

We define the notion of *character* as any sentient entity able to have a meaningful interaction. This includes humans of course, but also supernatural creatures such as ghosts, as well as certain animals (e.g. Figure 1a), magical artifacts (e.g. talking swords), or machines (e.g. robots). We ignore dead entities, unless they are supernatural and able to interact with some character (e.g. zombies, vampires, ghosts). We mean *interaction* in a very broad sense here, including active perception (watching, listening), even if unilateral (e.g. Figure 1b). Sometimes, an event involves a group of undetermined or undifferentiated people, akin to extras in a movie: we consider them as a *collective character* and represent them using a single vertex in the graph (e.g. the audience in Figure 1c). When a character impersonates another person, we use his real identity (e.g. Figure 1d). We refer to the characters using their full names when they are mentioned in the graphic novel, otherwise we

## 6 Vincent Labatut



Fig. 1: (a) Muff, the dog of Jolan (Thorgal's son), has a significant impact on the story, which illustrates how animals can be considered as characters. (b) Two thieves watch Aaricia and Vigrid from the distance, thereby performing a remote and unilateral action. (c) Raider Erik tells his misadventure (blueish panels) to the inhabitants of his village, which we consider here as a collective character. (d) Volsung of Nichor is exposed as Thorgal's impersonator before the village's Thing; we represent him as Volsung even when he is mistaken as Thorgal by the other characters.

use a brief description and a unique id to avoid any ambiguity.

We manually determine which characters are interacting based on our understanding of the situation, and use this information to identify scene boundaries. We define a *scene* as a sequence of consecutive panels describing some action involving the same group of characters for an uninterrupted period of time. Consequently, if a character involved in a scene leaves the group (e.g. Figure 2a), or if a new character joins the action (e.g. Figure 2b), then a new scene begins. Similarly, an ellipsis (i.e. omitting a part of the story to suggest the passage of time) also starts a new scene (e.g. Figure 2c). We consider unconscious or asleep characters to be involved



Fig. 2: (a) A group composed of Aaricia (Thorgal's wife) and her children is joined by Solveig, which starts a new scene according to our definition. (b) Aaricia serves dinner to a group, and her departure from the room also starts a new scene. (c) Thorgal goes to sleep in one panel and it is already morning in the next one, which illustrates the concept of ellipsis. (d) Two soldiers carry unconscious Thorgal, believing him dead: we still consider that there is an interaction between these three characters.

in the scene, as long as they are the object of some other character's action (e.g. Figure 2d). Note that a character can still be a part of the interacting group even if not explicitly shown in some panel, provided her involvement can be inferred from the surrounding panels (e.g. Figure 3a).

One could find it surprising that we use the drama concept of *unity of time* to define the notion of scene, but not that of *unity of space*. This is because we want to include, in our broad definition of an action, remote acts such as watching someone from the distance unbeknownst to them (e.g. Figure 2a), or throwing something at someone while remaining hidden (e.g. Figure 3b). Note that some authors take even more remote actions into account, e.g. Alexander considers postal correspondence, in novels [2]. It is worth stressing that the unilateral nature of certain actions (e.g.





Fig. 3: (a) Kriss of Valnor explicitly appears only in the first panel, but we can infer that she participates in the rest of the scene. (b) Example of remote interaction between an off-panel assailant and a group of characters. (c) Scene with two independent groups of characters: Thorgal & Kriss of Valnor (shown in the first panel), and Shenko & Floriana (last panel). (d) Scene involving twice the same character, Torric, at different ages.

*speaking to*, by opposition to *chatting with*) could be used to extract *directed* links. However, for the sake of simplicity, we discard this information in this study, and consequently extract undirected links.

Our approach makes it possible to have several groups of characters involved in separate actions at the same time, resulting in parallel and possibly overlapping scenes, as illustrated by Figure 3c. We treat flashbacks, hallucinations, and dreams, which are usually represented using different graphical codes than the rest of the story (e.g. blue tone in Figure 1c), just as regular scenes. There is an exception though, when the action is told by a visible narrator. We then consider that there are two separate interacting groups (and therefore two overlapping scenes): the narrator and his audience vs. the characters involved in the story told. Note that

some characters can belong to both groups (e.g. a narrator telling a story involving himself). More generally, it is possible that several instances of the same character belong to the same interaction group (e.g. through time travel as in Figure 3d).

### 3.2. Graph Building

We perform the third step of the extraction process, temporal integration, in an automatic way, resulting in a static graph representing the whole narrative. We define edges between the characters that interact during the same scene, and integrate over the narrative.

Our method compares to those described in Section 2 in the following ways. First, like them, we detect characters manually. However, they all infer character interactions based on co-occurrences, whereas we do not make this assumption and identify proper interactions manually. Second, by leveraging the scene to break down the narrative, we use a different temporal granularity. Alberich *et al.* [1] and Gleiser [20] consider character co-occurrences at the level of the whole comic book, and Lee & Kim [30] at the level of the whole webtoon episode. This would not be relevant considering our dataset, because the number of characters appearing in a single volume (34 in average) is large enough to make the extracted graph too dense to be useful. The approach adopted by Murakami *et al.* [39] amounts to using a two-panel sliding window with a single-panel overlap, which generally constitutes a finer granularity than ours, as scenes are often longer than two panels. However, such a window is not narratively sound, as two consecutive panels can belong to completely different subplots. Rochat & Triclot [49] prefer to use a two-page sliding window with a single-page overlap, which generally corresponds to a rougher granularity than ours. Using such a page-based window is questionable when applied to novels, since writers do not control the pagination of the end product, as noted in [27]. But it could be relevant for graphic novels, as the elaboration of the page layout is an important part of the story telling. For instance, scenes often start and end at the first and last panels of a page (not necessarily the same page). Authors also frequently use the last panel of a page to build suspense, as the reader has to turn the page to know what happens next (see for instance Figure SM1 in the Supplementary Material).

## 4. Dataset Description

As mentioned before, in this article we study the character network extracted from the *Thorgal* series of graphic novels. In this section, we first introduce the series itself (Section 4.1), and list the research questions that arose from its study in the literature. We then describe the main characteristics of the dataset resulting from our annotation process, and discuss the resulting statistics (Section 4.2). Next, we explain how we filter the character network to retain only meaningful characters (Section 4.3). In addition, we also compare our dataset to other graphic novel datasets in the Supplementary Material (Section SM1.3).

### 4.1. The *Thorgal* Series

Our dataset is based on the Franco-Belgian series *Thorgal*, which was originally created by writer Jean Van Hamme and artist Grzegorz Rosiński in 1977. It takes

place during a period reminiscent of the European medieval age, and mixes elements of fantasy, mythology, and science-fiction. It mainly follows the adventures of Thorgal (shown in Figure 2c), originally a baby adopted by a Viking chief, and his family. The main secondary characters are his wife Aaricia (Figure 1b), his elder son Jolan (Figure 1a), his daughter Louve (Figure 2a), and his antagonist Kriss of Valnor (Figure 3c).

This graphic novel is still going on nowadays with several other authors and artists, as well as three spinoff series. Two of these are centered around secondary characters (*Kriss of Valnor* and *Louve*), whereas the third focuses on the youth of Thorgal and Aaricia youth (*Young Thorgal*). As of November 2021, a total of 63 volumes were published (39 in the main series and 24 in the spinoff ones), resulting in more than 15 million books sold. This makes *Thorgal* a bestseller on the French-speaking market [46]. In the Supplementary Material, Tables SM1 and SM2 list all the volumes and the 23 narrative arcs constituting the series, respectively. *Thorgal* is also the object of academic work in human sciences [15, 54], was adapted into a video game by Cryo Interactive [45] and there is an ongoing project of TV series adaptation [37] by Academy Award winner Florian Henckel von Donnersmarck.

We chose to study *Thorgal* not only because of its cultural and commercial significance, but also because only two bandes dessinées series have been studied in the complex network literature as of yet (cf. Table 1), and this work is the occasion to expand this corpus. Moreover, both of these series are relatively short, whereas *Thorgal* develops a story involving hundreds of characters over tens of volumes. Another interesting property of the series is the way it handles time. In *Thorgal*, characters get old as the story unrolls, their relationships evolve, some important characters die while others are born. The eponymous character is a teenager in the first volume, whereas in the latest one he is over forty and a father of three, with an elder son that left the parental home.

Based on the literature, we identified two research questions, that we want to answer by analyzing the character network extracted from the series.

**RQ1** *How similar is Thorgal's story to Icelandic sagas?* Thiry [54] stresses that the *Thorgal* series is deeply rooted in the *Íslendingasögur* (*Sagas of Icelanders*), hybridizing it with more modern narrative elements. Such sagas have been thoroughly studied in the complex network literature [34], therefore we can leverage these results to answer this question.

**RQ2** *What is the position of women in Thorgal?* Desfontaine *et al.* [15] argue that *Thorgal* can be seen as a model of soft masculinity, involving a more balanced relationship between men and women. To explore this point, we propose to characterize the position of female characters in the character network of the series.

#### 4.2. Dataset Properties

A single annotator identified the characters, interactions, and scene boundaries for the whole series, as explained in Section 3.1. In order to assess the reliability of this annotation, three independent annotators did the same work on 4 randomly picked volumes. The agreement between these annotators and the main one is strong, with micro-average *F*-measure scores of 0.97 for scene boundaries and of 0.94 for

character identification (cf. Section SM1.2 of the Supplementary Material for more technical details).

In addition, we also annotated the biological sex of the characters: *Female*, *Male*, *Mixed* (for groups of characters), or *Unknown*. Each volume follows the *48CC* bandes dessinées standard, i.e. a 48-page hardcover color book (of which 46 pages are actually drawn). It takes approximately one hour to annotate such a book. Our dataset<sup>b</sup>, including the raw data and the extracted networks, as well as our source code<sup>c</sup> are both available online.

The 63 volumes constituting our dataset contain 21,259 panels distributed over 2,925 pages. They involve a total of 1,480 characters interacting in 4,622 scenes. The average volume and scene include 34.2 and 3.1 characters, respectively. One could assume that longer scenes tend to involve more characters, but this is not the case: the correlation between both quantities is rather weak ( $\rho_k = 0.18$ ).

Table 2: Most frequent characters of the corpus, with their number of occurrences expressed in volumes, pages, scenes and panels. The last row corresponds to the average values over all characters.

Character	Volumes	Pages	Scenes	Panels
Thorgal	54	1,552	1,752	9,852
Aaricia	47	733	816	4,375
Jolan	32	672	805	4,147
Kriss	24	573	683	3,518
Louve	26	481	561	3,010
Muff	15	137	152	761
Average (All)	1.5	9.2	9.8	51.7

The story is clearly dominated by Thorgal and his family, as illustrated by Table 2, that shows the most frequent characters. Besides the main protagonist, Thorgal himself, which is present in 38% of the scenes, these include his wife Aaricia (18%); his eldest son Jolan (17%, introduced in vol.6); the antagonist Kriss of Valnor (15%, introduced in vol.9); his daughter Louve (12%, introduced in vol.16) and Jolan’s dog Muff (3%, shown in Figure 1a, introduced in vol.8 and retired in vol.25). Note that the spinoff series focus on some of these characters (Louve, Kriss, Aaricia), otherwise the imbalance would be even more marked in favor of Thorgal. By comparison, the average character appears in only 0.2% of the scenes. The distribution of scenes by character, showed in red in Figure 4a, has a tail best fit by a power law ( $x_{\min} = 15$ ), according to Clauset *et al.*’s procedure [11]<sup>d</sup>. This means that this strong imbalance between characters, regarding their occurrence over the narrative, is observed also for less-frequent characters. The same observa-

<sup>b</sup>DOI: [10.5281/zenodo.6395875](https://doi.org/10.5281/zenodo.6395875)

<sup>c</sup><https://github.com/CompNet/NaNet/releases/tag/v1.0.1>

<sup>d</sup>We always use Clauset *et al.*’s [11] method to identify distributions in the rest of this article.

tion holds for other narrative units (pages, panels, and to a lesser extent, volumes): cf. Figure SM2 in the Supplementary Material.

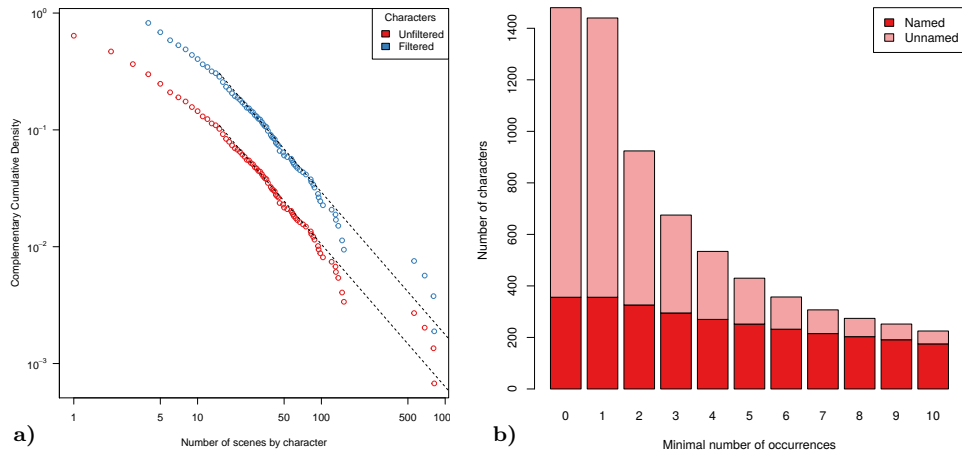


Fig. 4: (a) Complementary cumulative distribution of the number of scenes by character, for all characters (red) and after filtering them (blue). The dotted lines are power laws corresponding to the best fit. Both axes use a logarithmic scale. (b) Number of characters as a function of the minimal threshold used to filter them based on their number of occurrences. The red part represents characters whose proper name is explicitly mentioned in the narrative, whereas the blue part represents the others. Figure available at [10.5281/zenodo.6573491](https://doi.org/10.5281/zenodo.6573491) under CC-BY license.

Finally, 67% of the characters are males, 13% are females, and the rest either correspond to collective characters mixing both sexes (15%) or to entities whose sex could not be determined (5%). It is worth noting that the last three groups are more frequent than males when considering characters appearing only once; however males are prevalent among characters appearing several times. This suggests that the recurrent characters of the series are essentially males, and that other character types make only quick appearances, in general. This constitutes the beginning of an answer to RQ2, but we will come back to this point when studying the network itself.

### 4.3. Character Filtering

When extracting character networks from works of fiction, many authors apply a filtering step in order to remove very minor characters, namely extras, whose contribution to the story is not deemed sufficient to be taken into account during the subsequent analysis [27].

**Criteria** For this purpose, they use mainly three criteria. First, the most widespread is a minimal threshold over the number of occurrences. For instance, Elson *et al.* [16] remove characters that appear three times or fewer in a novel,

or that amount to 1% or less of all occurrences; Suen *et al.* [52] ignore characters that speak fewer than five times over a play or movie; and the *Marvel Chronology Project* dataset used by Alberich *et al.* [1] and Gleiser [20] is built by retaining only super-heroes and super-villains, provided they appear in at least two issues.

The second criterion is to keep only characters interacting with a sufficient number of others characters, which amounts to setting a minimal threshold over the vertex degree. For instance, when dealing with novels, Zhang *et al.* [59] retain only characters interacting with at least six other persons. This criterion is more questionable though, as the literature exhibits examples of low-degree characters that are central according to other topological measures [48], and therefore of great importance for the story. Imagine, for instance, a character interacting with only two persons, each one belonging to an otherwise separate part of the narrative.

Finally, the third criterion is the explicit mention of the character’s proper name. All approaches relying only on Named Entity Recognition to extract character networks from novels implicitly assume that unnamed characters should be ignored. The *Marvel Chronology Project* considers that an entity is a character if it has a proper super-hero or super-villain name, or a full regular name (both first and last names). However, as mentioned by Grener *et al.* [21], even unnamed characters can hold functionally and structurally significant positions in the narrative.

Table 3: Effect of both criteria used to filter characters: number of occurrences (*Occurrences*) and number of distinct interaction partners (*Interactions*).

	<b>Occurrences <math>\leq 3</math></b>	<b>Occurrences <math>&gt; 3</math></b>	<b>Total</b>
<b>Interactions <math>\leq 1</math></b>	221	1	222
<b>Interactions <math>&gt; 1</math></b>	724	534	1,258
<b>Total</b>	945	535	1,480

**Our Approach** In this work, in order to avoid discarding important characters, we use the number of interaction partners (i.e. degree) only to remove leaves and isolates from our networks. Put differently, we keep vertices with a degree of at least two. We do not use the presence of name as a criterion, as certain important characters of *Thorgal* do not have a proper name. For instance, the so-called Guardian of the Keys appears in 10 volumes, each time in a critical role, and volume 17 is even titled *The Guardian of the Keys*. On the contrary, the names of certain extras are revealed incidentally (e.g. a random soldier hailed by his boss during an action scene). We leverage this information though, in order to set up the most appropriate threshold over the minimal number of occurrences, through the elbow method (cf. Figure 4b).

According to our observations, it seems reasonable to remove all characters with three occurrences or fewer. Table 3 summarizes how both criteria affect the filtered characters. It is worth observing that, for the selected threshold values, the occurrence criterion almost completely subsumes the degree one. Put differently, apart from one individual, all characters interacting with a single person (or less)

also participate in three scenes or fewer. There are 534 characters remaining after this process, which amounts to 37% of the initial character set. Among them, 270 are named, which represents 77% of the named characters of the initial character set. They constitute a giant component of 530 characters, whereas the rest are distributed over two very small components. These are constituted of secondary characters that are never showed interacting with others, so we discard them to obtain a connected graph. Note that, in order to assess the effect of this filtering step, we perform our network analysis on both unfiltered and filtered networks. Like its unfiltered counterpart, the distribution of scenes by filtered characters, shown in blue in Figure 4a, has a tail best fit by a power law ( $x_{\min} = 15$ ), according to Clauset *et al.*'s method [11].

## 5. Network Analysis

In this section, we proceed with the descriptive analysis of our character networks. Figure 5 shows the filtered network extracted from the whole *Thorgal* series; whereas the unfiltered version is shown in Figure SM9 (Supplementary Material). The top part of Table 4 contains the main topological properties of both networks. As expected, since it focuses on the most important characters, the filtered network is much denser (see  $\delta$ ,  $\langle k \rangle$ ) and has smaller hubs ( $k_{\max}$ ). Its local transitivity ( $C$ , a.k.a. clustering coefficient) is higher, which indicates that main characters tend to form more triangles. Indeed, many of the removed characters are leaves and therefore not part of triangles, which can also explain the lower diameter ( $d_{\max}$ ) and average distance ( $\langle d \rangle$ ). Interestingly, filtering does not affect seriously degree assortativity ( $\rho_k$ ); however, as we will see later, it affects degree distribution.

Table 4: Main topological properties of the considered networks: numbers of vertices ( $n$ ) and edges ( $m$ ), density ( $\delta$ ), average degree ( $\langle k \rangle$ ), maximal degree ( $k_{\max}$ ), degree assortativity ( $\rho_k$ ), average distance ( $\langle d \rangle$ ), diameter or maximal distance ( $d_{\max}$ ), and average local transitivity ( $\langle C \rangle$ ). The top part corresponds to the unfiltered and filtered networks extracted from *Thorgal*, and the bottom part to two *Íslendingasögur* networks discussed in Section 6.1.

Network	$n$	$m$	$\delta$	$\langle k \rangle$	$k_{\max}$	$\rho_k$	$\langle d \rangle$	$d_{\max}$	$\langle C \rangle$
Unfiltered	1,453	6,579	0.006	9.06	676	-0.17	2.78	7	0.74
Filtered	524	3,702	0.027	14.13	358	-0.18	2.32	5	0.83
<i>18 Sagas</i> [34]	1,549	4,266	0.004	5.51	83	0.07	4.30	19	0.50
<i>Egils Saga</i> [34]	293	769	0.018	5.25	59	-0.07	4.20	12	0.60

In the following, we discuss in more detail the main topological properties of our networks, successively focusing on aspects related to degree (Section 5.1), distance and transitivity (Section 5.2), and centrality (Section 5.3). Additionally, in the Supplementary Material (Section SM2.4), we also leverage Rochat & Triclot's typology [49] (mentioned in Section 2) to study the structure of *Thorgal*'s narrative arc networks.

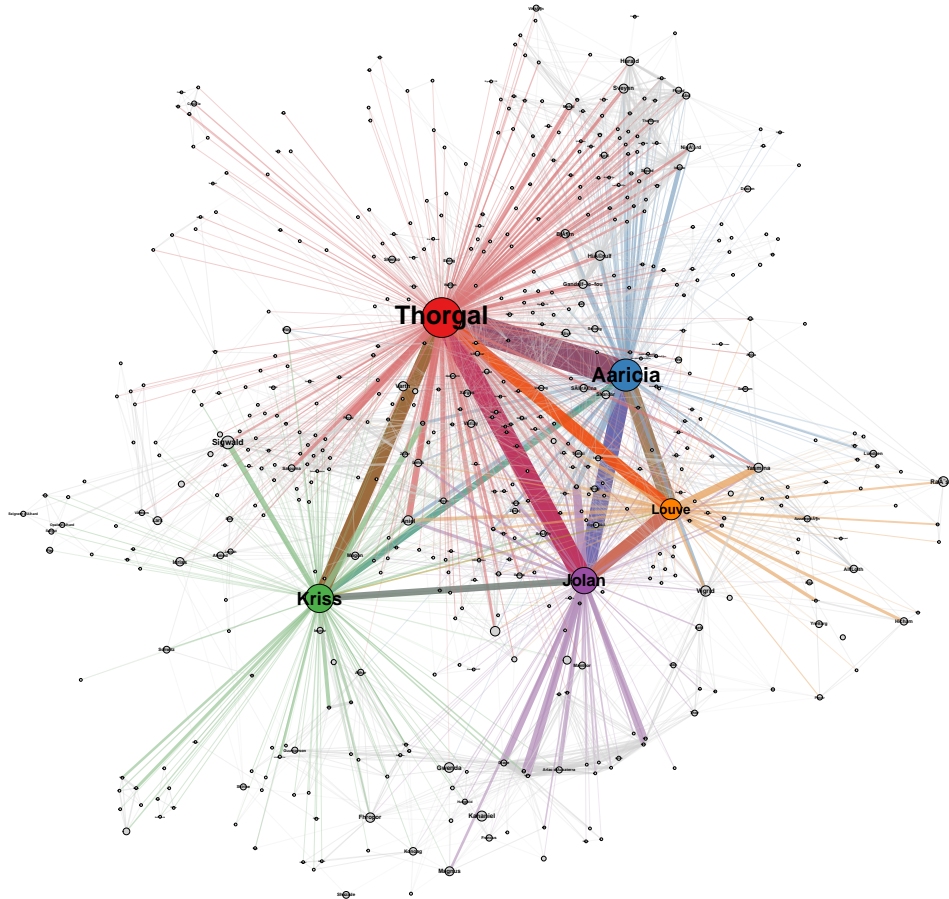


Fig. 5: Filtered network extracted from the whole *Thorgal* series. Vertex size is a function of betweenness, and edge width is a function of the number of co-occurrences. The five most frequent characters (in terms of scenes) are shown in a specific color: Thorgal (red), his wife Aaricia (blue), their elder son Joan (purple), their daughter Louve (Orange), and the antagonist Kriss of Valnor (green). The other characters are shown in gray. The unfiltered network is shown in Figure SM9 (Supplementary Material). Figure available at [10.5281/zenodo.6573491](https://zenodo.org/record/6573491) under CC-BY license.

### 5.1. Degree and Scale-Free Property

**Average Degree** One way to assess the average degree is to compare it with the value expected for a random model of comparable size. Alberich *et al.* [1] argue that their comic book network is a collaboration network, obtained through the projection of a bipartite graph containing two types of vertices (characters vs. comic book issues), and that the selected random model used to study the resulting unipartite network must therefore take this property into account. Like them, we



adopt the model proposed by Newman *et al.* [44] for this purpose. We first extract a bipartite network from our data, which contains character and scene vertices. We then compute the degree of both types of vertices, and use it for the generation of random bipartite networks possessing the same size and degree distribution. We project these bipartite networks into their character dimension, and use the resulting unipartite network as a random model for comparison with our character network.

The expected average degree is 19.57 for the unfiltered network and 37.19 for the filtered one (cf. Table 5). In both cases, this is largely above the values measured for our networks, by factors of 2.16 and 2.63, respectively. This corresponds to what is observed by Newman *et al.* [44] on real-world collaboration networks. By comparison, Alberich *et al.* [1] get an even larger factor of 3.4 on their *Marvel* data, and see this as the result of writers forcing *Marvel* characters to collaborate more often with the same characters than people would in real life. They consider this as a mark of the artificiality of the *Marvel* network. By comparison, in this aspect, the *Thorgal* network is closer to a real-world social network. This could be due to the fact that *Thorgal* contain much fewer recurrent characters: only 13% of the characters appear in several volumes. The main characters obviously interact a lot with one another, but they also frequently go on solo adventures, during which they interact with a number of non-recurrent characters.

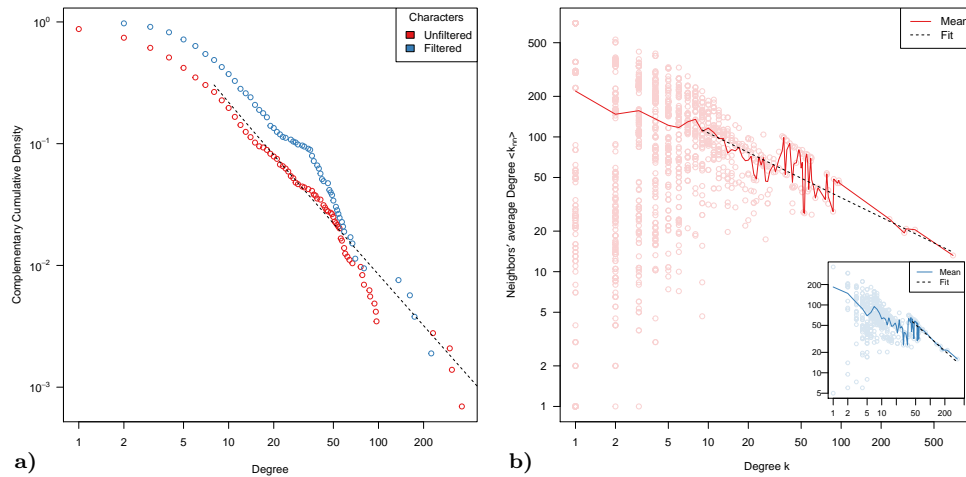


Fig. 6: (a) Degree distribution in the unfiltered (red) and filtered (blue) networks. The dotted line shows the power law best fitting the former. (b) Average degree of the neighbors as a function of vertex degree, for the unfiltered (red) and filtered (blue) networks. The dotted line represents the power law best fitting the function. In both plots, the axes use a logarithmic scale. Figure available at [10.5281/zenodo.6573491](https://zenodo.org/record/6573491) under CC-BY license.

**Scale-Free Property** Figure 6a shows the degree distribution observed in the unfiltered and filtered networks. The tail of the former is best fit by a power law ( $x_{\min} = 8$ ) with exponent  $\alpha = 2.45$ , which is in the [2; 3] range identified by Barabási [4] as characteristic of *scale-free* networks. This is consistent with the presence of many lesser characters acquainted to only a few other persons, and of a few hubs, i.e. very important characters whose number of interacting partners is larger by several order of magnitude. Many real-world networks exhibit the same scale-free property, including the Web, or protein interaction networks, and more interestingly collaboration networks [43]. Such networks are expected to contain large hubs, which in turn affect the average distance [4] (cf. Section 5.2). By comparison, the exponent estimated by Alberich *et al.* [1] for the *Marvel* network is smaller than 2, which is uncommon in real-world networks. Barabási calls this the *anomalous regime*: the hubs of such networks grow faster than the network itself, and  $\langle d \rangle$  does not depend on  $n$ . Alberich *et al.* interprets this as an excessive importance given to the main *Marvel* characters, relative to real-world networks. A power law-distributed degree is often the result of a *preferential attachment* process, as described by Barabási & Albert [5]. In Section SM2.1, we discuss this aspect by considering the *preferential attachment rate* of a vertex as a function of its degree  $k$ .

The *Thorgal* network *seems* realistic with respect to degree distribution. However, it is worth stressing that, for a real-world social network, being scale-free reflects a high heterogeneity in the way social interactions are distributed over the considered population. In the case of a fictional network though, this heterogeneity concerns interactions that are *shown in the narrative*. Maybe certain extras or minor characters have a rich social life that the author chose not to show because this does not help to move the story forward: these characters can be considered as possibly incompletely described. Compared to the unfiltered network, the filtered one is narrowed to the characters whose social interactions are the focus of the story, and can be assumed to be more completely described. Its degree is not power law-distributed, and none of the tested alternative distributions are good fits either. From this perspective, the *Thorgal* network does not appear to be socially very realistic. Moreover, it appears that filtering the lesser characters affects the degree distribution to such an extent that it loses its heavy tail. Interestingly, a variety of degree distributions is observed in the character network literature, besides the power law: truncated power law [34], exponential law [33], log-normal law [58]. As the filtering procedures are also quite different from one author to the other, we hypothesize that the nature of this procedure may affect the degree distribution of the filtered network.

One could assume that the hubs correspond to the most frequent characters discussed in Section 4.2, and more generally that the degree of a vertex depends on the number of occurrences of the corresponding character. This is partly correct indeed, as Spearman's rank correlation coefficient between degree and number of occurrences (expressed in scenes) returns a high value of  $r_s = 0.75$ . Figure SM4a shows that the degree is a non-linear increasing function of the number of scenes, an observation also made by Rochat regarding Rousseau's autobiographical novel *Les confessions* [48]. The match is not perfect, though. On the one hand, it is possible for a character to appear frequently in the narrative while interacting always with

the same few characters, in which case his degree will be small. This is the case of Muff, Jolan’s dog: he is the sixth most frequent character, but he always follows the family. His degree is consequently much lower than for the other members of the family, which often leave for their own adventures. On the other hand, one character could appear in a few scenes in which he interacts with many characters. His degree would be high despite his low frequency. This is the case of the three Elders of Aran, a collective character that essentially appears in a single volume, and consequently has the 170<sup>th</sup> frequency among all characters. These Elders are the organizers of a tournament involving many characters, though, and thus get the 22<sup>nd</sup> higher degree.

**Disassortativity** According to Table 4, both networks are slightly disassortative ( $\rho_k$ ), meaning that hubs tend to connect to low degree vertices. Figure 6b shows how the degree averaged over the neighbors of a vertex evolves as a decreasing function of this vertex degree. In the unfiltered network, the tail of this function is best fit by a power law with exponent  $\gamma = 0.31$  ( $x_{\min} = 9$ ), represented by the dotted line in the figure. The same computation over a degree-preserving randomized version of the network leads to the same result, which means that the observed disassortativity is *structural* [4]. It is caused by the network being scale-free: its hubs are interconnected by fewer edges than required for the network to be neutral (i.e. neither assortative or disassortative). The same observations hold for the filtered network (inset), with  $\gamma = 0.37$ .

By comparison, real-world collaboration networks are generally assortative [43, 4], i.e. vertices tend to connect to other vertices of similar degree. We can therefore conclude that this aspect of the *Thorgal* network is not realistic, reflecting its artificial nature. As shown by Gleiser [20], the *Marvel* network exhibits an even higher level of disassortativity ( $\gamma = 0.52$ ). This could be a feature of fictional character networks, as the same observation was made for other works of fiction, spanning different types of media, e.g. myths [33], novels [60, 48] and TV series [32]. This could be related to the principle of *Chekhov’s Gun*, which states that the narrative should only mention elements necessary to tell the story. As a consequence, narratives generally overlook relationships between minor characters.

## 5.2. Distance, Transitivity, and Small-World Property

The unfiltered network contains a giant component that includes 94% of the nodes, as in most real-world collaboration networks [44]. By comparison, the filtered network is connected by construction. As mentioned before, the filtered network is slightly more compact (see  $\langle d \rangle$ ,  $d_{\max}$  in Table 4), and its transitivity (a.k.a. clustering coefficient,  $\langle C \rangle$ ) is higher. Both observations can be explained by the removal of many leaves (222 out of 1,480 vertices, according to Table 3) during filtering, resulting in a denser and more tightly connected network.

**Small-World Property.** The small-world property was originally defined by Watts & Strogatz [55] by comparing the average distance and the average local transitivity of the considered network to those of two reference networks: a random model and a lattice, both of comparable size and density. A small-world network possesses 1) a small average distance, i.e. of the same order as the random network’s

and much smaller than the lattice’s; and 2) a large average local transitivity, i.e. close to the lattice’s and much larger than the random network’s. In their seminal work, the random model selected by Watts & Strogatz is the well-known Erdős–Rényi model [18], and they use a regular ring lattice. The Erdős–Rényi does not preserve the degree distribution of the original network, which is an essential property of the network, as discussed in Section 5.1. For this reason, other authors [53] propose to use the Configuration model [38] instead. For the same reason, we do not use a regular ring lattice, but a ring lattice with the same degree distribution as in the original graph (cf. Section SM2.2 of the Supplementary Material). Finally, we also consider the bipartite model described in Section 5.1.

Table 5: Average degree ( $\langle k \rangle$ ), maximal degree ( $k_{\max}$ ), degree assortativity ( $\rho_k$ ), average distance ( $\langle d \rangle$ ), diameter ( $d_{\max}$ ) and average local transitivity ( $\langle C \rangle$ ) obtained for both versions of the *Thorgal* network and for the selected models.

	Model	$\langle k \rangle$	$k_{\max}$	$\rho_k$	$\langle d \rangle$	$d_{\max}$	$\langle C \rangle$
<b>Unfiltered</b>	<i>Thorgal</i> Network	9.06	676.00	-0.17	2.78	7.00	0.737
	Bipartite Model	19.57	932.51	-0.23	2.35	5.33	0.641
	Erdős–Rényi Model	9.06	20.33	0.00	3.56	6.07	0.006
	Configuration Model	9.07	697.00	0.00	3.07	7.33	0.113
	Ring Lattice	9.08	10.00	1.00	136.87	275.00	0.577
<b>Filtered</b>	<i>Thorgal</i> Network	14.13	358.00	-0.18	2.32	5.00	0.829
	Bipartite Model	37.20	495.68	-0.25	1.94	3.02	0.561
	Erdős–Rényi Model	14.13	26.57	0.00	2.66	4.03	0.027
	Configuration Model	14.12	364.00	0.00	2.57	4.96	0.159
	Ring Lattice	14.10	18.00	0.90	33.39	66.00	0.643

Table 5 summarizes the statistics obtained for the original networks and the considered models. For both the unfiltered and filtered networks, the average distance is of the same order of magnitude as for all considered random models, and much smaller than for the lattice. On the contrary, their transitivity is much larger than the random models’, and even than the lattice’s. Based on Watts & Strogatz’s definition [55], we can conclude that the network is small-world. Newman *et al.* [44] observe that real-world collaborative networks tend to have a transitivity roughly twice as large as for their bipartite model. For the *Thorgal* networks, this factor is 1.15 (unfiltered) and 1.48 (filtered), which puts them somewhere between the real-world and random cases. By comparison, the transitivity of the *Marvel* network is twice as large as the bipartite model’s. However Alberich *et al.* do not consider it as small-world, because its transitivity is small in absolute terms [1].

The small-world property can alternatively be defined as a logarithmic dependence between the average distance and the network size and average degree [55, 4]:  $\langle d \rangle \sim \ln n / \ln \langle k \rangle$ . The scale-free nature of the unfiltered network even indicates that it should be ultra-small-world, i.e.  $\langle d \rangle \sim \ln \ln n$  [12, 4]. As a way of testing this assumption, we propose to extract a series of *cumulative* networks, using one scene as the time step. Each network in this series models the characters and interactions

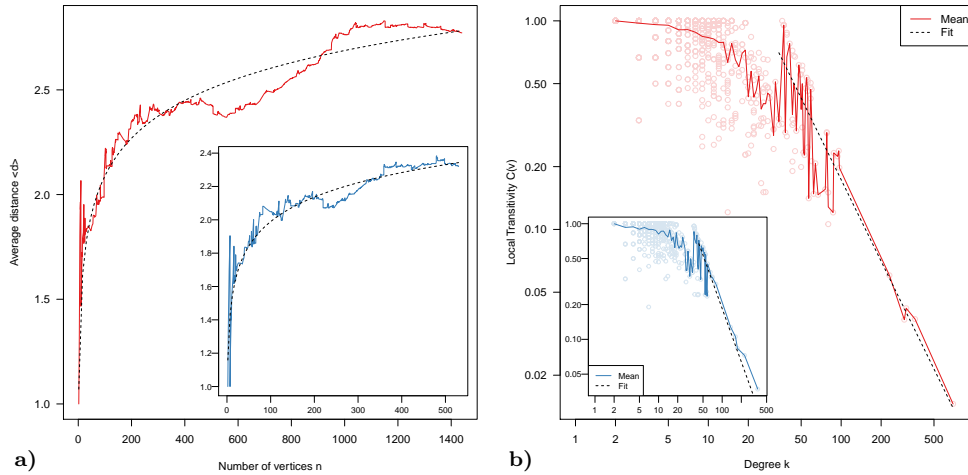


Fig. 7: (a) Evolution of the average distance as a function of the number of vertices in the unfiltered (red) and filtered (blue) networks. Each time step corresponds to one scene. The dotted lines represent the best logarithm fits. (b) Local transitivity as a function of the degree, in the unfiltered (red) and filtered (blue) networks, using a logarithmic scale on both axes. The dotted lines represent the best power law fits. Figure available at [10.5281/zenodo.6573491](https://doi.org/10.5281/zenodo.6573491) under CC-BY license.

occurring in the scene of interest, as well as those of the previous scenes. Figure 7a shows how their average distance evolves over the sequence, as a function of the number of vertices. This function is well fit by a logarithm (represented by a dotted line in the figure), thereby confirming the small-world property. This holds for both unfiltered and filtered versions of the network.

The average distance and transitivity of the *Thorgal* networks are revealing of the story told in the series. *Thorgal* is a rather episodic series, meaning it can be broken down into a few loosely connected narrative arcs (cf. Table SM2, Supplementary Material). The large average transitivity reflects the presence of relatively separated cliquish subgroups matching the social groups involved in these episodes. The small average distance indicates that the story is compact [27], which means that there is a certain narrative proximity between all the characters, even if these are very different and appear in distant moments of the story. This shows that, despite its episodic nature, the series is so centralized around Thorgal and its family and friends, that these efficiently connect very remote parts of the character space.

**Hierarchical Structure** Figure 7b shows the average transitivity as a function of the degree. In both the unfiltered and filtered networks, the tail of this decreasing function is best fit by a power law with exponent 0.42 ( $x_{\min} = 39$ ) and 1.53 ( $x_{\min} = 44$ ), respectively. This behavior is similar to various real-world networks, including collaboration networks [47, 4], and indicates the presence of a hierarchical structure. Gleiser shows that it is also the case of the *Marvel* network [20]. Vertices with small degree tend to have a higher transitivity, forming tightly-knit clusters, whereas hubs

tend to link these small structures and make the network connected at a higher level.

However, it is worth stressing that the dispersion observed in terms of transitivity is very different for *Thorgal* and *Marvel*, compared to other networks. Indeed, it mainly concerns small degree nodes, whereas in the cases shown in the literature, hubs tend to exhibit the wider range of transitivity [47, 4]. This consistently low transitivity among the main characters suggests that the structure is more pronounced at the top of the hierarchy. By comparison, in real-world networks, top vertices tend to collaborate with individuals which are not as consistently disconnected from each other.

In the *Marvel* network, Gleiser [20] observes that the small clusters are organized around a few super-heroes surrounded by super-villains. Central super-heroes from different clusters are connected, forming higher level bonds, whereas this is not true for super-villains. Gleiser assumes that this is due to the *Comics Authority Code* (CAC), a set of rules applying to comics. It prevents authors from having villains playing leading roles, and requires that good triumphs. Consequently, villains are not likely to band together, whereas hero collaboration allows making the story more interesting, even though the end is known in advance. In *Thorgal*, we observe clusters of 5–20 characters roughly matching specific narrative arcs or one-shot volumes. An interesting difference with the *Marvel* network is that some of these groups are connected through antagonists (e.g. Kriss of Valnor, Varth). This can be explained by the fact that *Thorgal* is not constrained by the CAC, and that its characters are much less Manichean: some of them hold an ambiguous position (e.g. Kriss of Valnor, who is both Thorgal’s antagonist and lover).

### 5.3. Vertex Centrality

We consider four standard centrality measures to characterize the position of individual vertices in the network, as these are widely used in the character network literature: degree, Eigenvector centrality [6], betweenness [19], and closeness [19].

**Comparison Between Measures** As mentioned before, high degree vertices represent characters interacting with many partners. A high betweenness typically denotes a broker, i.e. a character connecting otherwise well separated parts of the network, which in our case correspond to clusters matching distinct narrative arcs. The closeness allows describing the position of a character in terms of core vs. periphery, at the global and/or local levels. The Eigenvector centrality highlights hubs located in dense parts of the networks, containing other hubs. In addition to the power law distribution already described for the degree in Section 5.1, and similarly to Rochat’s results for Rousseau’s *Les Confessions* [48], we observe a long-tailed distribution for the Eigenvector centrality and betweenness, whereas the closeness has a more homogeneous distribution (i.e. normal-like).

Plotting the measures against each other shows a general positive correlation between them, as shown by Figure SM5 (Supplementary Material). This is confirmed when computing Spearman’s rank correlation: closeness vs. Eigencentrality (0.94), degree vs. Eigencentrality (0.78), betweenness vs. degree (0.70), closeness vs. degree (0.63), betweenness vs. Eigencentrality (0.50), betweenness vs. closeness (0.35). Overall, our results exhibit a similar level of correlation compared to those obtained by Rochat for *Les confessions* [48]. This relatively high correlation re-

flects the fact that most vertices get a similar (relative) centrality according to all measures [48]. But what interests us here are the vertices who do *not*.

Table 6: Ranks of the most central characters, according to the four considered centrality measures. All characters among the 12 most frequent or central characters (according to at least one measure) are represented.

Character	Scene Rank	Degree Rank	Eigencentality Rank	Betweenness Rank	Closeness Rank
Thorgal	1	1	1	1	1
Aaricia	2	2	2	3	2
Jolan	3	3	3	4	3
Kriss	4	4	5	2	4
Louve	5	5	4	5	5
Muff	6	8	7	28	8
Sigwald	7	10	41	6	60
Petrov	7	11	44	22	48
Yasmina	9	24	47	10	51
Aniel	10	16	43	14	24
Tjall	11	7	6	39	6
Lehla	12	9	9	16	9
Vigrid	13	18	46	7	49
Argun	15	13	12	18	10
Gandalf	16	6	8	8	7
Hierulf	21	11	38	11	40
Darek	21	13	10	30	14
Variay	45	21	13	36	11
Gwenda	67	79	248	9	55
Veronar	125	19	14	53	12
Elders	169	21	11	96	13
Kahaniel	184	164	864	12	782

**Typology of secondary characters** Table 6 shows the *ranks* of the most frequent and central characters, according to their centrality. It includes all characters which are among the 12 most central characters according to at least one measure. The 5 most central characters are always the same, almost in the same order, whatever the considered centrality measure. In addition, these correspond to the most frequent characters (identified using specific colors in Figure 5): Thorgal and his family. However, the measures disagree regarding the other characters, allowing us to distinguish four different cases.

First, six characters have a much lower betweenness rank, compared to the other measures. Among them, Muff, Tjall, Darek and Lehla are recurrent sidekicks, that support the main protagonists for extended periods of time; whereas Veronar and the Elders are antagonists that appear only in a single volume. In both cases, they are well-connected locally, but at least one of their neighbors has a better position (generally: one of the main characters) and claims most of the shortest paths, hence the low betweenness.

Second, quite similarly to the main protagonists, two characters have approxi-

mately the same rank for all measures. Argun is a friend of Thorgal's family, that follows them in their adventures, and Gandalf is one of their main antagonists. The main difference with our first category is a higher betweenness rank, which can be explained by a stronger recurrence of these characters: they appear in more narrative arcs, which makes them better connected to more distinct parts of the network.

Third, and contrarily to the first category, four characters have a much better betweenness rank, compared to the other measures. They connect some of the main characters to relatively remote parts of the network, corresponding to certain narrative arcs, but are not very well connected locally. Vigrid is a lesser god that interacts a lot with Thorgal and Aaricia in separate adventures, connecting them to major gods. Yasmina is a friend of Louve and has an important broker role in her spinoff. Gwenda and Kahaniel are recent antagonists connecting the main series to some narrative arcs developed in the spinoffs. They are both ranked poorly for all measure except betweenness.

Fourth, in addition to a better betweenness rank, a few characters also have a better degree rank. This means that they are at the same time secondary hubs and brokers. Their low Eigencentality and closeness ranks indicate that their surroundings are not tightly connected, though. The remaining characters lie in between these somewhat extreme four categories.

## 6. Research Questions

In this section, we tackle the research questions described in Section 4.1. In Section 6.1, we compare the structure of *Thorgal* and *The Icelanders Saga*. In Section 6.2, we study the position of women in *Thorgal*.

### 6.1. Comparison with the *Íslendingasögur*

In this section, we focus on RQ1, i.e. the hypothesized similarity between *Thorgal* and the *Íslendingasögur*, or *Sagas of Icelanders*. It is a corpus of medieval Icelandic literature telling the story of the early settlers of Iceland. It is based on events that took place between the 9<sup>th</sup> and 11<sup>th</sup> centuries, but also includes supernatural elements. Some of these stories focus on a single adventurer, while others follow whole families over several generations.

As mentioned in Section 4.1, Thiry argues that *Thorgal* can be seen as a hybridization between the *Íslendingasögur* and more modern narrative elements [54]. On the one hand, like this Icelandic corpus, the *Thorgal* series is a collection of adventures that generally focus on the main character, but also tell the story of his family. Moreover, they combine realistic and supernatural elements derived from the Norse mythology. On the other hand, Thiry notes that *Thorgal* is an anti-mythological hero fighting to take control over his destiny, and distinguishing himself from the traditional Viking figure [54]. Moreover, the series references other mythological sources (e.g. Greek tragedies), as well as fantasy, science-fiction and surrealist tropes.

In [34], Mac Carron & Kenna extract several character networks based on 18 sagas of the *Íslendingasögur*, and study their structures in order to assess the level of historicity of the narrative. For this purpose, they use standard topological mea-



tures to describe the networks and compare them to values obtained in the literature for fictional and real-world social networks. Their assumption is that if the *Íslendingasögur* networks reflect a realistic societal structure, they should be similar to the latter. We proceed similarly to compare *Thorgal* to the *Íslendingasögur*.

**Aggregation of 18 Sagas** We first compare the *Íslendingasögur* network built upon the whole selection of 18 texts, and the unfiltered version of the *Thorgal* network. As shown in Table 4, they roughly contain the same numbers of characters, which allows comparing their characteristics directly. This reveals several similarities. Both are built around a giant component, corresponding to 99% of the characters for the former and 94% for the latter. They both are small-world, exhibiting a high average local transitivity and a small average distance (cf. Section 5.2). They both are scale-free, following power law distributions with exponents in the range [2; 3] ( $\alpha = 2.99$  and 2.45, respectively) (cf. Section 5.1). Both are also robust to random attacks but fragile to targeted attacks (cf. Section SM2.3, in the Supplementary Material).

But there are also noticeable differences. The *Thorgal* network is 50% denser. Its hubs are much larger, *Thorgal* possessing a degree more than 8 times that of the largest Icelandic hub. It is clearly disassortative, whereas the *Íslendingasögur* is slightly assortative ( $\rho_k = -0.17$  vs. 0.07). The *Thorgal* network is more compact, with substantially smaller average distance ( $\langle d \rangle = 2.78$  vs. 4.30) and diameter ( $d_{\max} = 7$  vs. 19), and higher transitivity ( $\langle C \rangle = 0.74$  vs. 0.50). Overall, this depicts a network that is not as centered on a tight group of characters as *Thorgal*, and whose edge distribution better matches real-world social networks. This realism is precisely what makes the *Íslendingasögur* special [34, 54], and we can only conclude that *Thorgal* differs significantly from this Icelandic saga, at least from the perspective of the character network.

**Egils Saga** However, this difference could be explained by the fact that the considered Icelandic network is the aggregation of 18 stories which, although sharing many characters, focus on different individuals or families. By comparison, *Thorgal* has a single focus, on the main character and his direct family (half of the 20 most frequent characters are direct relatives). Therefore, we now turn to the network of a single Icelandic story. Among those considered by Mac Carron & Kenna [34], we select *Egils Saga*, because it is the most similar to the story told in *Thorgal*. First, it is at the same time a *family saga* (telling the story of a family) and a *poet's saga* (focusing on the adventures of a protagonist). Second, it is not set in a single location, as the protagonist travels a lot. Third, it contains more supernatural elements than the other 17 sagas. We compare it with the filtered *Thorgal* network, as they contain a similar number of vertices.

As before, both networks have a giant component and are small-world. However, the filtered *Thorgal* network is not scale-free, whereas *Egils Saga* degree follows a log-normal distribution [35], and is therefore heavy-tailed. As before, the *Thorgal* network is 50% denser and has larger hubs, it is more compact in terms of average distance and diameter, and has a higher transitivity. However, unlike the aggregated Icelandic network, *Egils Saga* is slightly disassortative ( $\rho_k = -0.07$ ), which makes it similar to *Thorgal*, on this point. According to Mac Carron & Kenna, this is

characteristic of sagas centering on one or a few characters, by opposition to a larger society [34]. Finally, and as before, both networks are robust to random attacks but fragile to targeted attacks.

**Concluding Remarks** According to Thiry, *Thorgal* retains mainly three aspects of the Icelandic saga: 1) the story follows an extraordinary man and his family (also extraordinary), including his ancestors and children; 2) it depicts the Viking daily life, customs, and system of belief with a certain historical accuracy; and 3) it uses the Norse mythology as a background, but also as a driving narrative force. Certain structural similarities exhibited by both *Thorgal* networks with respect to the *Íslendingasögur* seem to support this observation: similar numbers of characters, small-world and scale-free structures suggesting realistic social interactions, strong focus on Thorgal and his family, high transitivity revealing an episodic story, supernatural entities with central positions (e.g. Vigrid, Kahaniel). However, there are also clear differences, in particular the *Thorgal* networks are denser and more compact, with a stronger degree disassortativity, which reveals a higher focus on the main characters. In [33], Mac Carron & Kenna make similar observations when comparing networks extracted from mythological narratives and from other works of fiction, including novels, plays, and the *Marvel* corpus. These differences could illustrate the hybridization noted by Thiry: recycling of foreign myths (Greek, Abrahamic, but also contemporary), more frequent occurrence of supernatural events through fantasy and science-fiction subplots, and Thorgal acting as a (post)modern hero by refusing the fate decided for him by the Norse gods.

## 6.2. Position of Women

One characteristic of the *Thorgal* series is that its eponymous protagonist does not comply with the Viking value system of his adoptive clan. Instead, he goes by his own rules, which are at odds with those of his time, but consistent with the modern standards of the series authors' [15]. As a teenager, he is an outcast in his village, and is consequently not trained as a warrior but as a skald (a Scandinavian poet): his male identity is not built upon brute force and hegemony, but rather on empathy and humanism, and on more balanced social gender relations. This is not to say that Thorgal does not exhibit traditional virile features that characterize mythological heroes (athleticism, courage, pugnacity, righteousness, etc.), but rather that those are mitigated by his personal values, resulting in what Desfontaine *et al.* [15] call a *soft masculinity*. This aspect is highlighted by the fact that the main female characters are strong and independent women and girls. In particular, his main antagonist is Kriss of Valnor, which challenges him on most levels, and is treated on an equal foot not only by Thorgal himself, but more generally by the other characters.

**Individuals** This section focuses on RQ2, studying these aspects of *Thorgal* by analyzing the position of female characters in the network. Figure 8 shows the filtered network using colors to represent the characters' sex. It appears clearly that males are prevalent. Among the unfiltered characters whose sex could be identified, 84% are males and 16% are females. After filtering, the proportion of males decreases to 79% for 21% females, which indicates that the proportion of women

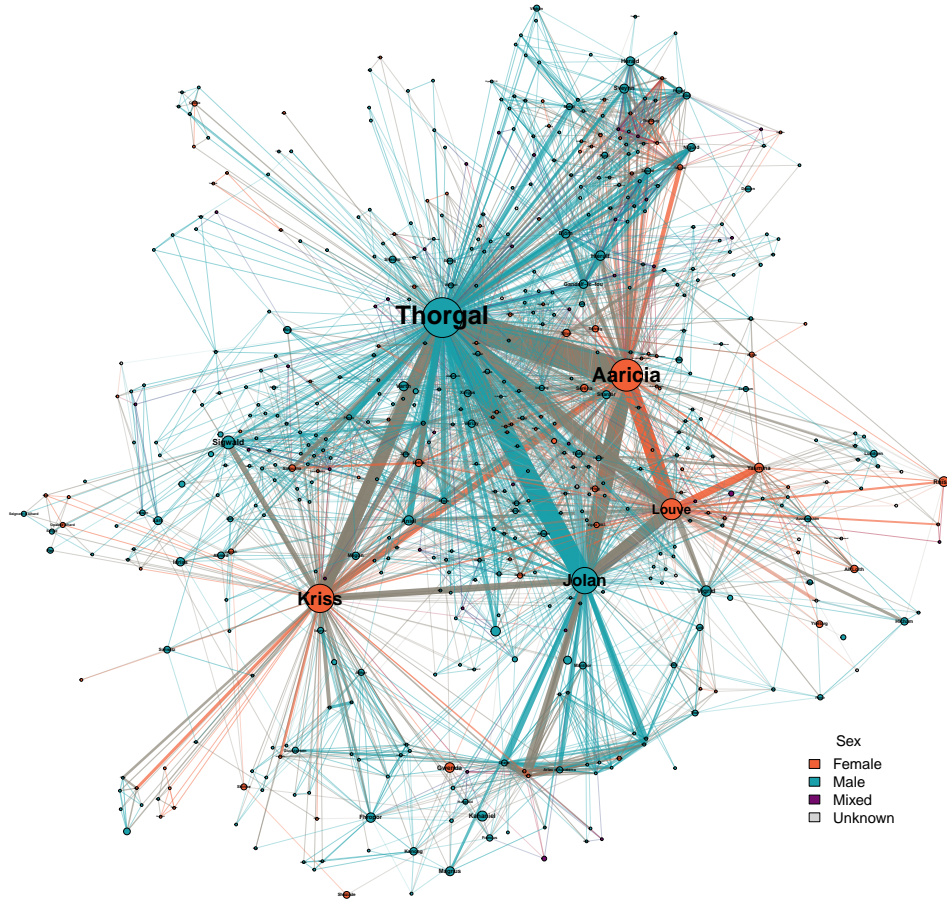


Fig. 8: Sex of the characters in the filtered network, using the layout of Figure 5. Figure available at [10.5281/zenodo.6573491](https://zenodo.org/record/6573491) under CC-BY license.

is larger when focusing on important characters. By comparison, there are 22% women among *Marvel* characters, with also an improvement when focusing on important characters: 31% [23]. However, it is worth stressing that only super-heroes and super-villains are taken into account in these statistics, whereas our *Thorgal* dataset contains all characters appearing in the narrative.

When focusing on the 10 most central characters, we find 4 or 5 females, depending on the considered centrality measure. In [26], Kagan *et al.* study gender bias in movies spanning a period of one century, through the analysis of their character networks. By comparison with *Thorgal*, they find a lower average of 33% women among the 10 most central roles. Note that their dataset includes only *named* characters. They get a slightly lower proportion when considering the 5 most central roles. However, in the case of *Thorgal*, 3 out of the 5 most central characters are always female, whatever the considered measure. Moreover, as explained in Section 5.3,

these top 5 characters are clearly more central than the rest of the vertices. Thorgal is unambiguously the main protagonist of the series, but 60% of the core characters are females. In his study of *Marvel* comics, Gleiser [20] observes that *all* the central characters are males. Moreover, women do not act as bridges between parts of the network, as also observed for super-villains. In *Thorgal*, the high betweenness of certain female characters shows that some women hold this intermediary position. On the same note, if in *Marvel* comics women tend to be heroes and not villains [23], in *Thorgal* the main antagonist (and fourth most frequent character) is a woman, Kriss of Valnor.

We compare the full distributions of the four centrality measures studied in Section 5.3, relative to character sex (cf. Figure SM8). We use the permutation test for integer-valued measures and the Kolmogorov–Smirnov test for real-valued measures. We find significant differences ( $p < 0.05$ ) for the degree and Eigencentrality, but not for the betweenness and closeness. This indicates that despite the strong male prevalence, female characters manage to retain important positions in terms of long-range centrality. In short, women tend to be less central than men for a separate episode, but they are as important when considering the whole narrative.

**Relationships** Regarding the relationships, in the unfiltered network, the majority of edges model full male interactions (51%), followed by mixed (43%) and full female (6%) interactions. The proportions are very similar for the filtered network, with slightly more female and mixed edges (7%). This is of course due to the large sex imbalance observed before between the characters. However, if the edges were randomly distributed, we would get 70% of male, 27% of mixed and 3% of female edges: we can thus conclude that the network structure favors interactions involving females, more than expected by chance. We compute the density of the male- and female-induced subgraphs. The unfiltered and filtered female subgraphs are respectively three and two times denser than their male counterparts. This confirms that the narrative develops proportionally more connections between female than male characters. The sex-based assortativity measure is almost zero in both cases though, which shows that this property is hidden by the wide prevalence of male characters.

Following the method used by Kagan *et al.* [26], we enumerate the triangles constituting the network, distinguishing the vertices only in terms of sex. Interestingly, in both the unfiltered and filtered networks, the most frequent type of triangle is FMM (one female and two male characters, 44%). Triangles of type MMM are only second (33%), followed by FFM (20%) and FFF (3%) women. By comparison, in their corpus of movies, Kagan *et al.* obtain very similar proportions, with an important difference: MMM triangles are the most frequent (41%), and FMM ones only come second (36%). In both cases, 77% of the triangles contain more men than women. Kagan *et al.* notice some differences between movie genres: romantic movies tend to contain an increased number of triangles with a majority of women (29% FFM and 6% FFF), whereas it is the opposite for war movies (25% FMM and 64% MMM). This is interesting, because *Thorgal* contains a number of fight scenes, and some characters participate in battles and even lead a war. One would therefore expect a much larger number of MMM triangles. There are love stories in *Thorgal*, but only a few, especially between the protagonist and his wife: this does

not seem to be enough to compensate for the war scenes. The Viking background of the narrative implies that women participate in warfare, which could be a better explanation of this peculiarity.

**Gender Equality Testing** The Bechdel test is a popular method to test whether a narrative represents women fairly. It has three criteria: 1) there must be at least two named female characters; 2) they must talk to each other; and 3) about another topic than men. Out of the 4,622 scenes constituting our dataset, 441 (10%) involve only women. Among them, 200 have more than one female character, which talk to each other in 172 cases. After a manual assessment of these scenes, it turns out that 87 (51%) comply with the third criterion of the Bechdel test.

However, as noted by Kagan *et al.* [26], the Bechdel test can be considered as too lenient, as a single interaction is enough for a narrative to pass, independently of the rest of the story. Instead, they propose a quantitative approach relying on a gender-based comparison of the numbers of interactions. They use the notion of *volume* of a subgraph, which is the sum of the degrees over the vertices constituting this subgraph. Their *Gender Degree Ratio* is the ratio of the male to female volumes. They consider that a situation of gender equality is described by a ratio between 0.8 and 1.2. In the case of *Thorgal*, the ratio is 0.32, so the series is clearly out of this interval. By comparison, Kagan *et al.* get an average ratio of 0.6 for their corpus of movies, of which only 12% pass this test.

One could argue that the degree-based volume does not represent the number of *punctual* interactions though, but rather the number of interaction *partners*. This is likely to hide situations where recurrent female characters have many interactions over time. We propose to compute the *strength*-based volume instead. Vertex strength is the total weight of the edges attached to the vertex of interest. In our case, we use the number of distinct interaction between two characters as weight. The ratio obtained for *Thorgal* is 0.40: this is higher than before, which indicates that this series matches the case mentioned above, but still far from Kagan *et al.*'s minimal threshold of 0.8.

**Concluding Remarks** The soft masculinity of *Thorgal* is not obvious at first sight in the network structure, when only considering sex distribution among the characters. There is a large male prevalence, on par with other works of fiction such as the *Marvel* comics and popular movies. However, focusing on the most important characters reveals that women indeed hold a different position in this network. They are prevalent in the top 5 characters, and hold central roles, unlike in the other considered works of fiction. In terms of relationships, the *Thorgal* network similarly exhibits much more full male edges, but full female edges are more frequent than explained by the male vertex prevalence. Its triangles tend to involve more females than in other works of fiction. However, if *Thorgal* passes the Bechdel test, its gender degree ratio does not meet the criteria proposed by Kagan *et al.* [26], which are fulfilled by many of the movies they study. This could be due to them discarding unnamed characters, though.

## 7. Conclusion

In this article, we proposed a new corpus describing the characters of the graphic novel *Thorgal*, and the interactions occurring between them in the course of the narrative. We then leveraged this dataset to extract the series character network. We conducted a descriptive analysis which showed that it exhibits certain realistic properties (i.e. features also present in real-world social networks): small-worldness, scale-freeness, hierarchical structure, and sensitivity to attacks. By comparison, the comics-based *Marvel* network studied by Alberich *et al.* [1] is less realistic (less transitive, not scale-free, robust to targeted attacks). However, the *Thorgal* network is also disassortative, a trait generally found in fictional character networks. Moreover, we argue that it exhibits only the *appearance* of realism: the structure of the network does not necessarily represent realistic social interactions, but rather the outcome of the creative process behind the unfolding of the story, which happens to possess some of the properties observed when modeling real-world relationships.

During our descriptive analysis, we compared two versions of our network: unfiltered (all characters) vs. filtered (important characters only). Filtering is generally performed quite systematically (and implicitly) in the character networks literature, and we wanted to assess its effect. It turns out it essentially affects the degree distribution, resulting in a loss of the scale free property. Interestingly, fictional character networks are often *not* scale free [33]: according to our results, this could be explained not by the intrinsic nature of works of fiction, but rather by the way the networks are extracted and, more importantly, filtered.

In addition, we leveraged our character network to answer two research questions related to the series. In summary, we first showed that the similarity between *Thorgal* and the *Sagas of Icelanders* observed in [54] does not transpire in the network structure in a meaningful way, especially when considering the characteristics of other fictional narratives. Second, we showed that the notion of soft masculinity discussed in [15] seems to affect the network structure, both intrinsically and when compared to other works of fiction.

Our work can be extended in several ways. We plan to expand the dataset by annotating the character interactions in order to include directionality and polarity: this will allow us to extract directed and signed networks, and therefore to apply additional tools to complement our analysis. Another direct extension is to use the narrative smoothing method presented in [8] to extract a dynamic version of our network and study its evolution. We also want to conduct a comparative study of character networks extracted from works belonging to the three main graphic novel traditions, namely North American, European, and East Asian.

## Acknowledgments

I first thank Christian Labatut, who introduced me to *Thorgal* and comics in general, more than 30 years ago. I also thank Arthur Amalvy, Noé Cécillon, and Elise Labatut, whose annotation work helped assess the quality of the main corpus.

## Credits

The French version of *Thorgal* is published by Le Lombard, and the English version by Cinebook Ltd. The panels shown in this article are taken from the following

books, all authored by Jean Van Hamme (writer) and Grzegorz Rosiński (artist):

- Figure 1a: *Thorgal* vol.16 “Wolfcub”, p.18, 1990.
- Figure 1b: *Thorgal* vol.14 “Aaricia”, p.41, 1989.
- Figure 1c: *Thorgal* vol.20 “The Brand of the Exiles”, p.9, 1995.
- Figure 1d: *Thorgal* vol.17 “The Guardian of the Keys”, p.28, 1991.
- Figure 2a: *Thorgal* vol.20 “The Brand of the Exiles”, p.16, 1995.
- Figure 2b: *Thorgal* vol.21 “Ogotai’s Crown”, p.33, 1995.
- Figure 2c: *Thorgal* vol.18 “The Sun Sword”, p.8, 1992.
- Figure 2d: *Thorgal* vol.27 “The Barbarian”, p.48, 2002.
- Figure 3a: *Thorgal* vol.28 “Kriss of Valnor”, p.19, 2004.
- Figure 3b: *Thorgal* vol.18 “The Sun Sword”, p.40, 1992.
- Figure 3c: *Thorgal* vol.18 “The Sun Sword”, p.48, 1992.
- Figure 3d: *Thorgal* vol.15 “The Master of the Mountains”, p.36, 1989.
- Figure SM1: *Thorgal* vol.11 “The Eyes of Tanatloc”, p.11–12, 1986.

## References

- [1] Alberich, R., Miro-Julia, J., and Rossello, F., Marvel Universe looks almost like a real social network, *arXiv cond-mat.dis-nn* (2002) cond-mat/0202174.
- [2] Alexander, S., Population thinking and narrative networks: Dickens, joyce, and the wire, *Modern Language Quarterly* **82** (2021) 315–343.
- [3] Augereau, O., Iwata, M., and Kise, K., A survey of comics research in computer science, *Journal of Imaging* **4** (2018) 87.
- [4] Barabási, A.-L., *Network Science* (Cambridge University Press, 2015).
- [5] Barabási, A.-L. and Albert, R., Emergence of scaling in random networks, *Science* **286** (1999) 509.
- [6] Bonacich, P. F., Power and centrality: A family of measures, *American Journal of Sociology* **92** (1987) 1170–1182.
- [7] Bost, X., Gueye, S., Labatut, V., Larson, M., Linares, G., Malinas, D., and Roth, R., Remembering winter was coming: Character-oriented video summaries of TV series, *Multimedia Tools and Applications* **78** (2019) 35373–35399.
- [8] Bost, X., Labatut, V., Gueye, S., and Linares, G., Narrative smoothing: dynamic conversational network for the analysis of TV series plots, in *Advances in Social Networks Analysis and Mining - 2nd International Workshop on Dynamics in Networks* (2016), pp. 1111–1118, doi:[10.1109/ASONAM.2016.7752379](https://doi.org/10.1109/ASONAM.2016.7752379).
- [9] Bost, X., Labatut, V., Gueye, S., and Linares, G., Extraction and analysis of dynamic conversational networks from TV series, in *Social Network Based Big Data Analysis and Applications*, Lecture Notes in Social Networks, chapter 3 (Springer, 2018), pp. 55–84.
- [10] Chute, H., Comics as literature? reading graphic narrative, *Publications of the Modern Language Association* **123** (2008) 452–465.
- [11] Clauset, A., Shalizi, C. R., and Newman, M. E. J., Power-law distributions in empirical data, *SIAM review* **51** (2009) 661–703.
- [12] Cohen, R. and Havlin, S., Scale-free networks are ultrasmall, *Physical Review Letters* **90** (2003) 058701.
- [13] Couch, C., The publication and formats of comics, graphic novels, and tankobon, *Image & Narrative* **1** (2000).
- [14] Cronin, B., How many pages long have marvel comic stories been over the years?, <https://www.cbr.com/marvel-comics-how-many-pages-long/> (2017), <https://www.cbr.com/marvel-comics-how-many-pages-long/>.
- [15] Desfontaine, P., Mesyngier, P., and Renaud, J.-N., Thorgal, “soft” masculinity as a model for young people?, *Agora débats/jeunesses* **78** (2018) 67–85.

- [16] Elson, D. K., Dames, N., and McKeown, K. R., Extracting social networks from literary fiction, in *48th annual meeting of the association for computational linguistics* (2010), pp. 138–147, <http://dl.acm.org/citation.cfm?id=1858696>.
- [17] Ercolessi, P., Bredin, H., Sénac, C., and Joly, P., Segmenting TV series into scenes using speaker diarization, in *Workshop on Image Analysis for Multimedia Interactive Services* (2011), pp. 13–15, <http://herve.niderb.fr/download/pdfs/Ercolessi2011.pdf>.
- [18] Erdős, P. and Rényi, A., On the evolution of random graphs, *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* **5** (1960) 17–61.
- [19] Freeman, L. C., Centrality in social networks i: Conceptual clarification, *Social Networks* **1** (1978) 215–239.
- [20] Gleiser, P. M., How to become a superhero, *Journal of Statistical Mechanics* **2007** (2007) P09020.
- [21] Grener, A., Luczak-Roesch, M., Fenton, E., and Goldfinch, T., Towards a computational literary science: A computational approach to dickens’ dynamic character networks, Technical report, Zenodo (2017), doi:10.5281/zenodo.259499.
- [22] Hettinger, L., Becker, M., Reger, I., Jannidis, F., and Hotho, A., Genre classification on german novels, in *26th International Workshop on Database and Expert Systems Applications* (2015), doi:10.1109/DEXA.2015.62.
- [23] Hickey, W., Comic books are still made by men, for men and about men, <https://fivethirtyeight.com/features/women-in-comic-books/> (2014), <https://fivethirtyeight.com/features/women-in-comic-books/>.
- [24] Hirata, N. S. T., Montagner, I. S., and Hirata, R., Comics image processing: learning to segment text, in *1st International Workshop on coMics ANalysis, Processing and Understanding* (2016), p. 11, doi:10.1145/3011549.3011560.
- [25] Jeong, H., Neda, Z., and Barabási, A.-L., Measuring preferential attachment for evolving networks, *Europhysics Letters* **61** (2003) 567–572.
- [26] Kagan, D., Chesney, T., and Fire, M., Using data science to understand the film industry’s gender gap, *Palgrave Communications* **6** (2020).
- [27] Labatut, V. and Bost, X., Extraction and analysis of fictional character networks: A survey, *ACM Computing Surveys* **52** (2019) 89.
- [28] Lee, O.-J. and Jung, J. J., Affective character network for understanding plots of narrative contents, in *Affective Computing and Context Awareness in Ambient Intelligence* (2016), <http://ceur-ws.org/Vol-1794/afcai16-paper3.pdf>.
- [29] Lee, O.-J. and Jung, J. J., Story embedding: Learning distributed representations of stories based on character networks, *Artificial Intelligence* **281** (2020) 103235.
- [30] Lee, O.-J. and Kim, J.-T., Learning multi-modal representations of narrative multimedia: a case study of webtoons, in *ACM International Conference on Research in Adaptive and Convergent Systems* (2020), doi:10.1145/3400286.3418216.
- [31] Lefevre, P., The importance of being ‘published’. a comparative study of different comics formats, in *Comics & Culture: Analytical and Theoretical Approaches to Comics* (Museum Tusulanum Press, 2000), pp. 91–105.
- [32] Liu, D. and Albergante, L., Balance of thrones: a network study on ‘Game of Thrones’, *arXiv cs.SI* (2017) 1707.05213.
- [33] Mac Carron, P. and Kenna, R., Universal properties of mythological networks, *Europhysics Letters* **99** (2012) 28002.
- [34] Mac Carron, P. and Kenna, R., Network analysis of the Íslendinga sögur - the sagas of icelanders, *European Physical Journal B* **86** (2013) 407.
- [35] MacCarron, P., *A Network Theoretic Approach to Comparative Mythology*, Phd thesis (2014), <https://curve.coventry.ac.uk/open/file/9bfc043d-497e-4217-82ab-e19963b53290/1/maccarroncomb.pdf>.
- [36] Markovic, R., Gosak, M., Perc, M., Marhl, M., and Grubelnik, V., Applying network theory to fables: complexity in slovene belles-lettres for different age groups, *Journal of Complex Networks* **7** (2018) 114–127.
- [37] Meza, E., Henckel von Donnersmarck puts ‘Thorgal’ TV series at top of his list



- (2018), <https://variety.com/2018/film/news/henckel-von-donnensmarck-puts-thorgal-tv-series-at-top-of-his-list-1202923350/>.
- [38] Molloy, M. and Reed, B., A critical point for random graphs with a given degree sequence, *Random Structures and Algorithms* **6** (1995) 161–179.
- [39] Murakami, H., Kyogoku, R., and Ueda, H., Creating character connections from manga, in *3rd International Conference on Agents and Artificial Intelligence* (2011), pp. 677–680, <https://www.scitepress.org/Papers/2011/31965/31965.pdf>.
- [40] Murakami, H., Nagaoka, Y., and Kyogoku, R., Creating character networks from comics using frames and words in balloons, in *7th International Congress on Advanced Applied Informatics* (2018), pp. 1–6, doi:10.1109/iaai-aai.2018.00012.
- [41] Murakami, H., Nagaoka, Y., and Kyogoku, R., Creating character networks with kinship relations from comics, *International Journal of Service and Knowledge Management* **4** (2020).
- [42] Newman, M. E. J., Clustering and preferential attachment in growing networks, *Physical Review E* **64** (2001) 025102.
- [43] Newman, M. E. J., The structure and function of complex networks, *SIAM Review* **45** (2003) 167–256.
- [44] Newman, M. E. J., Strogatz, S. H., and Watts, D. J., Random graphs with arbitrary degree distributions and their applications, *Physical Review E* **6402** (2001) 026118.
- [45] Parker, S., Curse of Atlantis: Thorgal’s Quest ships (2003), <https://www.gamespot.com/articles/curse-of-atlantis-thorgals-quest-ships/1100-6028714/>.
- [46] Potet, F., Bd “Thorgal”, la vie sans Rosinski (2018), [https://www.lemonde.fr/bande-dessinee/article/2018/11/23/bd-thorgal-la-vie-sans-rosinski\\_5387303\\_4420272.html](https://www.lemonde.fr/bande-dessinee/article/2018/11/23/bd-thorgal-la-vie-sans-rosinski_5387303_4420272.html).
- [47] Ravasz, E. and Barabási, A.-L., Hierarchical organization in complex networks, *Physical Review E* **67** (2003) 026112.
- [48] Rochat, Y., *Character Networks and Centrality*, Ph.D. thesis, Université de Lausanne (2014), [http://infoscience.epfl.ch/record/203889/files/yrochat\\_thesis\\_infoscience.pdf](http://infoscience.epfl.ch/record/203889/files/yrochat_thesis_infoscience.pdf).
- [49] Rochat, Y. and Triclot, M., Les réseaux de personnages de science-fiction : échantillons de lectures intermédiaires, *ReS Futuræ* **10** (2017).
- [50] Schmidt, D., Zehe, A., Lorenzen, J., Sergel, L., Düker, S., Krug, M., and Puppe, F., The fairynet corpus - character networks for german fairy tales, in *5th Joint SIGHUM Workshop on Computational Linguistics for Cultural Heritage, Social Sciences, Humanities and Literature* (2021), pp. 49–56, <https://aclanthology.org/2021.latechclfl-1.6/>.
- [51] Stiller, J., Nettle, D., and Dunbar, R. I. M., The small world of Shakespeare’s plays, *Human Nature* **14** (2003) 397–408.
- [52] Suen, C., Kuenzel, L., and Gil, S., Extraction and analysis of character interaction networks from plays and movies, in *Digital Humanities* (2013), [http://snap.stanford.edu/class/cs224w-2011/proj/laneyk\\_Finalwriteup\\_v1.pdf](http://snap.stanford.edu/class/cs224w-2011/proj/laneyk_Finalwriteup_v1.pdf).
- [53] Telesford, Q. K., Joyce, K. E., Hayasaka, S., Burdette, J. H., and Laurienti, P. J., The ubiquity of small-world networks, *Brain Connectivity* **1** (2011) 367–375.
- [54] Thiry, M., Reshaping the scandinavian saga through hybridity: “Thorgal”, an anti-mythological hero, in *International Congress of Medieval Studies* (2019), [https://dial.uclouvain.be/pr/boreal/object/boreal:219470/datastream/PDF\\_01/view](https://dial.uclouvain.be/pr/boreal/object/boreal:219470/datastream/PDF_01/view).
- [55] Watts, D. J. and Strogatz, S. H., Collective dynamics of ‘small-world’ networks, *Nature* **393** (1998) 440–442.
- [56] Waumans, M. C., Nicodème, T., and Bersini, H., Topology analysis of social networks extracted from literature, *PLoS ONE* **10** (2015) e0126470.
- [57] Weng, C.-Y., Chu, W.-T., and Wu, J.-L., RoleNet: Movie analysis from the perspective of social networks, *IEEE Transactions on Multimedia* **11** (2009) 256–271.
- [58] Yose, J., Kenna, R., MacCarron, P., Platini, T., and Tonra, J., A networks-science investigation into the epic poems of Ossian, *Advances in Complex Systems* **19** (2016) 1650008.
- [59] Zhang, C., Zhang, Q., Yu, S., Yu, J. J. Q., and Song, X., Complicating the social

- networks for better storytelling: An empirical study of chinese historical text and novel, *IEEE Transactions on Computational Social Systems* **8** (2021) 754–767.
- [60] Zhang, J., Zhao, H., Xu, J.-Q., and Wang, J.-F., Small-world and scale-free features in harry potter, *TELKOMNIKA Indonesian Journal of Electrical Engineering* **12** (2014) 6411–6416.

# Supplementary Material

## Complex Network Analysis of a Graphic Novel: The Case of the Bande Dessinée *Thorgal*

Vincent Labatut

This document contains additional resources and results completing those already shown in the main article. It also contains some results that could not be placed in the main text due to lack of space.

### SM1. Corpus

This section provides some information regarding the titles and characteristics of the volumes and narrative arcs of *Thorgal* (Section SM1.1). It also compares the *Thorgal* corpus with other comic novel-based datasets (Section SM1.3).

#### SM1.1. *Thorgal* Series

Table SM1 lists the volumes constituting the *Thorgal* series, including the main series and its spinoffs. They are shown by order of publication date. The volumes of the main series are simply numbered, whereas the spinoffs are prefixed with a letter: *L* (*Lowve*), *K* (*Kriss of Valnor*), and *J* (*Young Thorgal*, or *La jeunesse de Thorgal* in French).

Table SM2 lists the narrative arcs of the *Thorgal* series, according to the official website<sup>e</sup>, also by order of publication. Each arc is typically constituted of 2–3 volumes.

Figure SM1 complements our explanations from Section 3.2 by illustrating how artists use page layout to control the pace of the narrative.

As explained in Section 4.2 and shown in Figure 4a, the distribution of scenes by character in *Thorgal* is best fit by a power law, both for unfiltered and filtered characters. For the sake of exhaustiveness, Figure SM2 shows the distributions of volumes, pages and panels by character, for both the unfiltered and filtered versions of our dataset. The volume distribution (Figure SM2a) is best fit by a truncated power law for filtered characters, whereas no good fit is detected for unfiltered characters. The other distributions (Figures SM2b and SM2c) all follow power laws.

#### SM1.2. Annotation

As explained in Section 4.2, we assess the quality of the main annotation by comparing it with the annotations produced by three additional annotators, on four randomly chosen volumes: *Thorgal* vol.6, 16 and 20, and *Kriss of Valnor* vol.1. In order to conduct this comparison, we consider the main annotation as our ground

<sup>e</sup><http://www.thorgal.com/>

Table SM1: List of Thorgal volumes, by order of publication date.

Nbr	Volume title	Nbr	Volume title
01	The Betrayed Sorceress	K1	I Forget Nothing
02	The Island of Frozen Seas	33	The Blade Ship
03	The Three Elders of Aran	L1	Raissa
04	The Black Galley	K2	The Sentence of the Valkyries
05	Beyond the Shadows	K3	Fit for a Queen
06	The Fall of Brek Zarith	L2	The Severed Hand of the God Tyr
07	Child of the Stars	J1	The Three Minkelsonn Sisters
08	Alinoe	L3	The Kingdom of Chaos
09	The Archers	K4	Alliances
10	The Land of Qa	34	Kah-Aniel
11	The Eyes of Tanatloc	J2	The Eye of Odin
12	City of the Lost God	L4	Crow
13	Between Earth and Sun	K5	Red like Raheborg
14	Aaricia	L5	Skald
15	The Master of the Mountains	J3	Runa
16	Wolf Cub	K6	The Island of Lost Children
17	The Guardian of the Keys	J4	Berserkers
18	The Sun Sword	L6	Queen of the Black Elves
19	The Invisible Fortress	35	The Scarlet Fire
20	The Brand of the Exiles	L7	Nidhogg
21	Ogotai's Crown	J5	Slivia
22	Giants	K7	The Time Mountain
23	The Cage	J6	The Ice Drakkar
24	Arachnea	K8	The Master of Justice
25	The Blue Plague	36	Aniel
26	The Kingdom Beneath the Sand	J7	The Blue Tooth
27	The Barbarian	37	The Hermit of Skellingar
28	Kriss of Valnor	J8	The Two Bastards
29	The Sacrifice	38	The Selkie
30	I, Jolan	J9	The Tears of Hel
31	Thor's Shield	39	Neokora
32	The Battle of Asgard		

truth, and leverage the Precision, Recall, and  $F$ -measure separately for scene boundary detection and for character identification. Note that for the former task, these measures are very conservative, as they do not involve any tolerance regarding the position of the boundaries [17].

A scene boundary correctly located is a true positive (TP). A scene boundary incorrectly defined by a secondary annotator is a false positive (FP). A scene boundary missed by a second annotator is a false negative). For characters identification, we first match each scene detected by a secondary annotator to a scene of the main annotation. For this purpose, we select the ground truth scene with the largest intersection, in terms of panels. Then, a character detected by a secondary annotator is a true positive if it is also present in the ground truth. It is a false positive if it is not present in the ground truth. On the contrary, a character that is not detected by the secondary annotator while it appears in the ground truth is a false negative.

Based on the numbers of TPs, FPs, and FNs, we can compute the Precision and

Table SM2: Narrative arcs of the *Thoragal* series, according to the official website. The last column indicates the type of network according to Rochat & Triclot’s nomenclature [49]: *kernel* (central group of characters), *unicentric* (a single main character), *polycentric* (several distinct centers), *acentric* (no clear center).

Number	Arc title	Volumes	Type
1	The Queen of Frozen Seas	01–02	Unicentric
2	The Cursed Village	03	Acentric
3	The Masters of Brek Zarith	04–06	Kernel
4	The Origins	07, 14	Polycentric
5	The Dreamed Child	08	Acentric
6	The Great Country	09–13	Polycentric
7	To the North	15–17	Polycentric
8	Shaigan-the-Merciless	18–23	Polycentric
9	The Wandering Viking	24–26	Kernel
10	The Last Trip	27–29	Polycentric
11	The Successor	30–32	Polycentric
12	The Valkyries’ Court	K1–K2	Unicentric
13	The Red Mages	33–36	Polycentric
14	The Fallen Mage	L1–L3	Unicentric
15	North-Levant	K3–K5	Polycentric
16	The Sald’s Song	J1–J2	Polycentric
17	The Black She-Wolf	L4–L5	Polycentric
18	The Beast-Warriors	J3–J4	Acentric
19	Aniel	K6–K8	Polycentric
20	The Underground Kingdoms	L6–L7	Kernel
21	The Betrayed Sorceress	J5–J6	Polycentric
22	The Kingdom of the Danes	J7–J9	Unicentric
23	New Horizons	37–39	Polycentric

Recall, using their standard definitions:

$$Pre = TP / (TP + FP) \quad (1)$$

$$Rec = TP / (TP + FN). \quad (2)$$

The  $F$ -measure is the harmonic mean of the Precision and Recall, i.e.

$$F = 2 \frac{Pre \cdot Rec}{Pre + Rec}. \quad (3)$$

These three definitions hold for both tasks (scene boundary detection and character identification).

Here, we consider the *micro-average* versions of these measures, which means that the TP, FP and FN numbers are computed over all annotators. Tables SM3 and SM4 show the scores obtained for each considered volume separately, as well as overall (for all four volumes at once). Table SM3 focuses on the scene detection task, and Table SM4 on character identification.

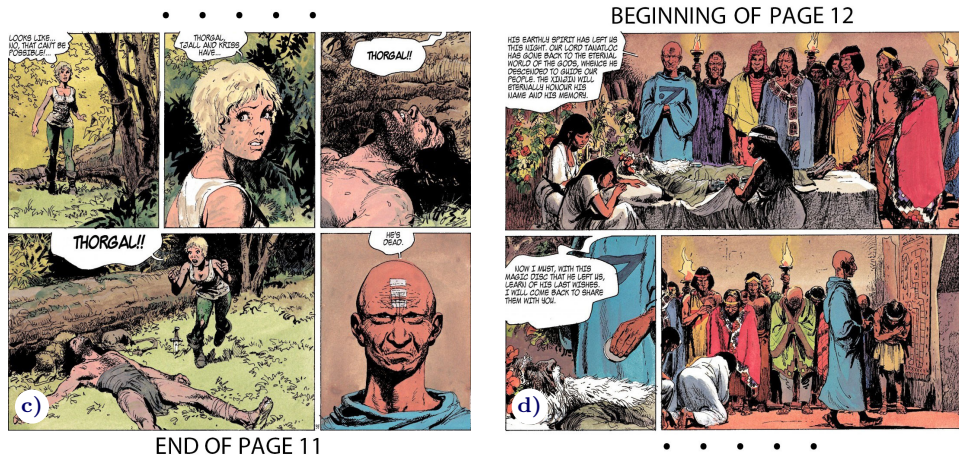


Fig. SM1: Example of the way artists use page layout to build up suspense. (a) Ariacia finds Thorgal unconscious, and as he was previously very sick, the reader assumes that he may be dead. The last panel is part of another scene, happening in a completely different place. (b) The next page reveals who is actually dead: Variay announces the death of Tanatloc.

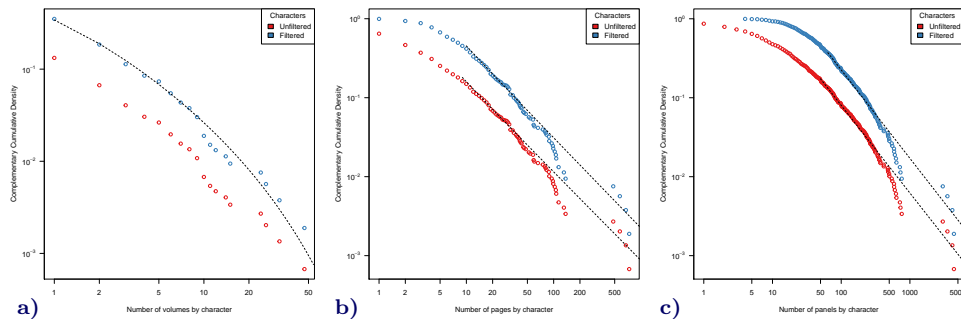


Fig. SM2: Complementary cumulative distribution of the number of (a) volumes, (b) pages, and (c) panels by character, for all characters (red) and after filtering them (blue), in *Thorgal*. The dotted lines represent the best fits, when available (see text). Figure available at [10.5281/zenodo.6573491](https://doi.org/10.5281/zenodo.6573491) under CC-BY license.

The agreement level is good, with an overall  $F$ -measure of 0.97 and 0.94, respectively. A manual inspection reveals that the disagreement between annotators are generally not due to errors, but rather to differences in the way they understand a scene. As explained in Section 3.1, our annotation process requires identifying the characters involved in a scene, even when they are not shown explicitly in certain panels. This necessitates a good understanding of the story, and also to make some inference and assumptions. Sometimes, the annotators do not reach the same conclusion on these aspects. Another source of disagreement is the identification

Table SM3: Precision, Recall and  $F$ -measure (all three micro-average) obtained by the three secondary annotators on the scene boundary detection task, for each one of the 4 considered volumes, and overall.

Volume	$TP$	$FP$	$FN$	$Pre$	$Rec$	$F$
<i>Thorgal</i> vol.6	189	8	7	0.959	0.964	0.962
<i>Thorgal</i> vol.16	221	8	16	0.965	0.932	0.948
<i>Thorgal</i> vol.20	319	5	8	0.985	0.976	0.980
<i>Kriss of Valnor</i> vol.1	313	8	12	0.975	0.963	0.969
<b>Overall</b>	1,042	29	43	0.973	0.960	0.967

of extras and minor characters, which are sometimes mistaken with other similar characters.

Table SM4: Precision, Recall and  $F$ -measure (all three micro-average) obtained by the three secondary annotators on the character identification task, for each one of the 4 considered volumes, and overall.

Volume	$TP$	$FP$	$FN$	$Pre$	$Rec$	$F$
<i>Thorgal</i> vol.6	492	31	30	0.941	0.943	0.942
<i>Thorgal</i> vol.16	595	34	31	0.946	0.950	0.948
<i>Thorgal</i> vol.20	1,181	85	96	0.933	0.924	0.929
<i>Kriss of Valnor</i> vol.1	779	46	26	0.944	0.968	0.956
<b>Overall</b>	3,047	196	183	0.940	0.943	0.941

### SM1.3. Comparison

It is worth comparing some characteristics of our dataset with those listed in Section 2, whenever possible. For the European and Asian graphic novels studied by Rochat & Triclot [49], we use the dataset provided in the article<sup>f</sup> to compute the statistics ourselves. For the *Marvel* comics, we leverage the information provided in Alberich *et al.*'s article [1, 20], and also use the dataset available online<sup>g</sup>. Unfortunately, the Manga- and Webtoon-based datasets respectively constituted by Murakami *et al.* [39, 40, 41] and Lee & Kim [30] are not available and consequently could not be included in this comparison. Note that we focus on the filtered version of *Thorgal*, as only the main characters are considered in the other datasets.

Table SM5 summarizes the main statistics of the mentioned graphic novels. It reflects some differences in the way these are published in America, Europe and Japan, from the perspective of their format and production pace. These differences are important in the context of our study, as they are known to affect various aspects of the graphic novels, in particular their narrative structure [31]. In the United

<sup>f</sup><https://github.com/mtriclot/Belfort>

<sup>g</sup><http://bioinfo.uib.es/~joemiro/marvel.html>

Table SM5: Statistics related to characters, volumes, and chapters, for *Thorgal* (filtered version) and certain graphic novels listed in Section 2.

Statistic	Thorgal	Marvel	Metabarons	Aldebaran	Akira	Gunnm
Characters	530	6,486	102	30	42	37
Volumes	63	12,942	8	5	3	6
Chapters	–	–	–	–	28	32
Characters/Volume	15.25	14.90	27.00	21.00	21.33	12.50
Characters/Chapter	–	–	–	–	9.00	6.84
Volumes/Character	2.25	7.47	2.20	3.50	1.56	2.50
Chapters/Character	–	–	–	–	6.15	7.30

States, the primary medium of dissemination is the publication of monthly issues of approximately 30 pages, which are generally not systematically collected in omnibus books [13]. Recent *Marvel* issues tend to be 20 pages long [14]. In France and Belgium, bandes dessinées used to be pre-published in specialized weekly magazines collecting various distinct series, before being released as hardcover albums. Nowadays, such pre-publication still exist, but is less frequent, and the main means of dissemination is the album, typically 46 pages long. *Thorgal* and *Aldebaran* follow this format, whereas *Metabarons* is a bit longer, with 62-page volumes. In Japan, mangas are also pre-published in similar weekly or monthly magazines, under the form of 20-page long chapters, then collected in so-called tankobons [31]. The number of pages produced is larger than in Europe though, as mangas are generally produced by studios and not by individual authors. The *Akira* and *Gunnm* volumes contain approximately 175 pages, and their chapter 20–30 pages. Based on these lengths, manga chapters seem to better fit a comparison with American and European volumes, *a priori*.

The compared datasets vastly differ in their numbers of volumes. The *Marvel* dataset [1] is the largest, as it describes the work of many people over forty years. By comparison, the main *Thorgal* series was produced by two persons; this dataset is second in number of volumes only because it is complete. It is not the case of the datasets produced by Rochat & Triclot [49]: *Metabarons* is part of the *Incal* universe, *Aldebaran* focuses only on the second arc of the larger series, and *Akira* and *Gunnm* contain only the first few volumes of their respective series. Despite these differences, the average number of volumes by character is rather similar for all datasets (2–4) except *Marvel*. The average number of characters by volume is also relatively close for all series (15–20), except for *Metabarons* (27). Finally, contrary to our initial assumption, the statistics obtained for the mangas when using chapters instead of volumes are the most different from the other datasets. This could be due to a more action-oriented way of staging the story, involving fewer characters in each chapter.

## SM2. Descriptive Analysis

This section contains part of our descriptive analysis that could not fit in the main article. These include a study of the preferential attachment mechanism in *Thorgal* (Section SM2.1); a description of the lattice-based generative model used in Section 5.2 to assess the small-world property (Section SM2.2); some additional plots



related to the four centrality measures considered in the article (Section SM2.3); and a study of the narrative arcs of *Thorgal* based on the graph typology proposed by Rochat & Triclot in [49] (Section SM2.4).

### SM2.1. Preferential Attachment

A power law-distributed degree is often the result of a *preferential attachment* process, as described by Barabási & Albert [5]. It is possible to assess this property by estimating  $\Pi(k)$ , the preferential attachment rate of a vertex as a function of its degree  $k$ . In case of preferential attachment, this function is supposed to be linear in  $k$ , whereas if there is no preferential attachment, it should be independent of  $k$ . Following the method used in the literature [25, 4], we do not study directly this function, but rather the *cumulative* preferential attachment rate, noted  $\kappa(k)$ . Figure SM3a shows an estimation of  $\kappa(k)$  for both unfiltered and filtered networks. The left tail of the function is best fit by a power law with exponent 0.98 for the unfiltered network and 0.72 for the filtered one. The former displays an almost *linear* preferential attachment, known to produce scale-free networks. The latter exhibits a *sublinear* preferential attachment: this scaling regime is not sufficient to produce a power law degree distribution, which confirms our previous observation. Albeit apparently less frequent, this sublinear regime is also observed in real-world collaboration networks [42, 25, 4], though.

In the context of a character network, preferential attachment means that new characters tend to interact more with existing characters that possess more acquaintances. In the case of a series like *Thorgal*, the narrative typically focuses on a few main characters, and new characters are generally introduced *because* they interact with them. An absence of preferential attachment would require the introduction of new characters related to minor characters only, which is very uncommon. The fact that preferential attachment is sublinear in the filtered network indicates that this observation is not as strong when discarding the most minor characters. There is no point of comparison for the preferential attachment rate in the character network literature, at least regarding other graphic novels. However, in novels, Waumans *et al.* [56] detect a linear dependence in the case of the *Harry Potter* series.

Figures SM3b and SM3c shows the external and internal cumulative preferential attachment rates, noted  $\kappa(k)$  and  $\kappa(k_1k_2)$ , respectively. The former (Figure SM3b) reflects the probability for a new vertex to attach to an existing vertex, depending on its degree. The latter (Figure SM3c) is related to the probability for a new edge to attach to two existing vertices, depending on the product of their degrees. Both functions are supposed to be linear in presence of preferential attachment [25, 4]. Compared to Figure SM3a, these representations allow distinguishing between the effects of new edges connecting new vertices to existing ones (so-called *external* edges) vs. new edges connecting two existing vertices (*internal* edges). See [25, 4] for more methodological details.

All functions are well fit by a power law, with approximately the same exponents for the unfiltered and filtered networks. The external rate is related linearly to the degree, with an exponent of 0.99. The exponent is 0.44 for the internal rate, which corresponds to a sublinear regime. Note that there are approximately 10 times as many external than internal edges, on the considered period. Therefore, preferential attachment is stronger for edges connecting a new vertex to an existing one than

for edges connecting two existing vertices. This indicates that new characters tend to interact first with important characters rather than minor ones, and then only develop connections with less important characters.

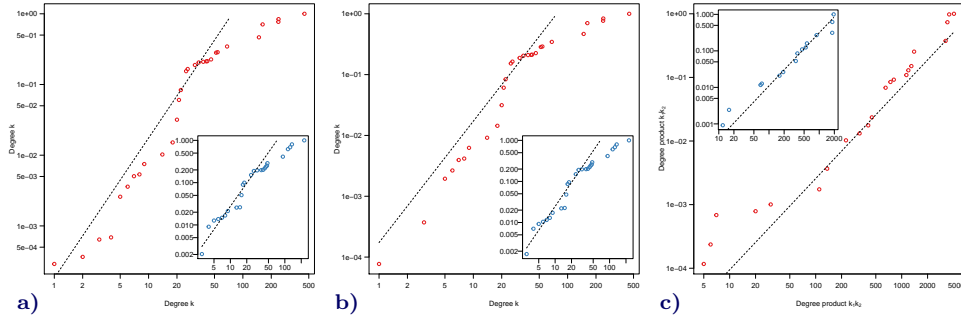


Fig. SM3: (a) Cumulative preferential attachment of new vertices, as a function of the degree. (b) Cumulative preferential attachment of new vertices, as a function of the degree. (b) Cumulative preferential attachment of new internal edges, as a function of the degree product. The dotted lines show the power law best fitting the data. The unfiltered network is represented in red, and the filtered one in blue. In all three plots, the axes use a logarithmic scale. Figure available at [10.5281/zenodo.6573491](https://doi.org/10.5281/zenodo.6573491) under CC-BY license.

### SM2.2. Lattice-Based Model

To obtain a lattice-based graph following a specified degree sequence, we proceed as follows. We start from a regular ring lattice, whose degree corresponds to the minimal value specified in the degree sequence. By construction, at this moment, all vertices have the same degree.

Second, we randomly pick one of its vertex, which we note  $v_1$ , to become the largest hub of the final network. Its first left- and right-hand neighbors, spatially speaking, will consequently become the second and third largest hubs. Similarly, its second neighbors will become the fourth and fifth largest hubs, and so on.

Third, we add edges between  $v_1$  and its neighbors, starting from the spatially closest ones, until  $v_1$  reaches the targeted degree. This will also increase the degree of the concerned neighbors. We proceed similarly with the second largest hub to be, and go on until all vertices have the desired degree.

### SM2.3. Centrality

**Correlation Study** Figure SM4 shows character centrality as a function of their number of occurrences expressed in scenes. Note the logarithmic scales. It appears that centrality is an increasing function of the number of occurrences, whatever the considered centrality measure. Spearman's rank correlation coefficient is relatively high for all four measures vs. scenes. For the unfiltered network, we get 0.75 (degree),

0.58 (Eigencentrality), 0.73 (betweenness), and 0.49 (closeness). The same observation and similar correlation values hold when considering occurrences expressed in panels, pages and volumes (not shown). The dispersion is much larger for infrequent characters, whereas the 5 most frequent characters are the most central according to all four considered measures.

The Eigenvector centrality computed for the unfiltered network exhibits a clear separation between the 85 least central characters, whose centrality score is almost zero, and the rest of the population. They all correspond to filtered characters, which hints at using the Eigenvector centrality as the filtering criterion. However, the first non-minor character (Shazade) is the 113<sup>th</sup> by order of increasing Eigencentrality, whereas we filtered around 1,000 characters with the method described in Section 4.3. We can conclude that this centrality measure alone is not sufficient to perform the filtering.

Figure SM5 allows assessing the association for each pair of considered centrality measures, in the unfiltered (red) and filtered (blue) networks. Visual assessment reveals that there is some form of non-linear relationship for all of them. Of course, that is a general observation, since some vertices of interest differ in how they are ranked by centrality measures, as discussed in Section 5.3.

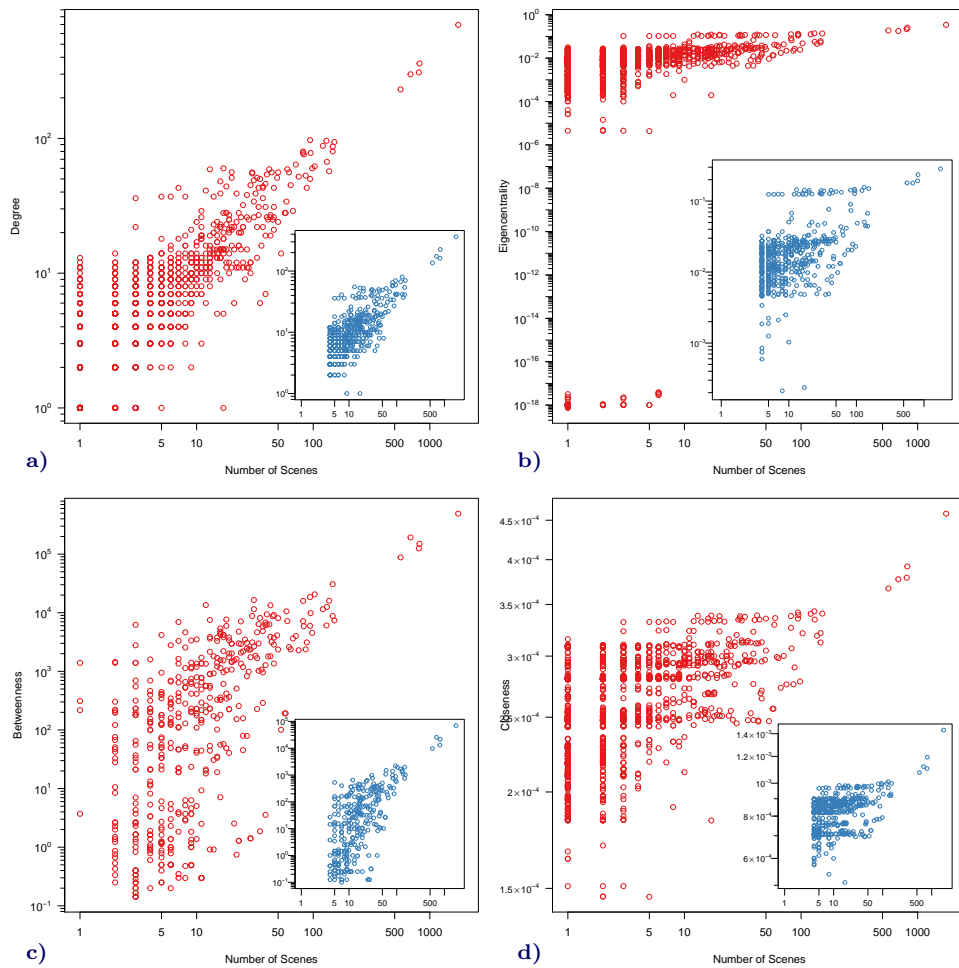


Fig. SM4: Centrality of characters as a function of their number of occurrences expressed in scenes, for all characters (red) and after filtering them (blue). (a) Degree. (b) Eigenvector centrality. (c) Betweenness. (d) Closeness. All axes use a logarithmic scale. Figure available at [10.5281/zenodo.6573491](https://zenodo.org/record/6573491) under CC-BY license.

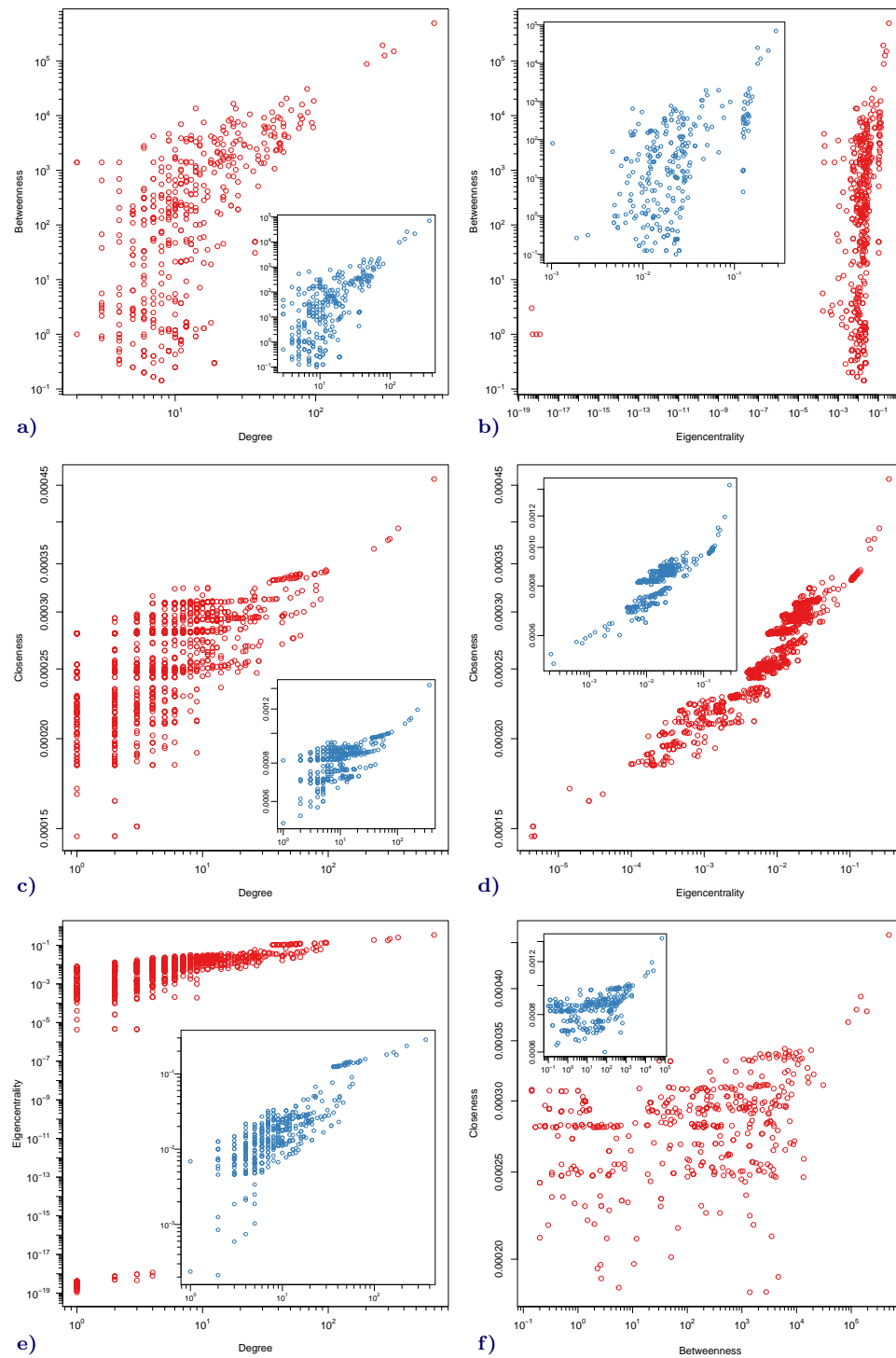


Fig. SM5: Comparison of four centrality measures (degree, Eigenvector centrality, betweenness and closeness), for all characters (red) and after filtering them (blue). All axes use a logarithmic scale. Figure available at [10.5281/zenodo.6573491](https://zenodo.org/record/6573491) under CC-BY license.

**Robustness to Attacks** One way to study whether the network structure relies on one or a few important vertices, is to study its robustness to targeted vs. random attacks. A targeted attack consists in removing the most central vertices [34] or edges [20] of the network, by opposition to randomly picking them. One assesses the network robustness by monitoring how the size of its largest component evolves as a function of the number of removed vertices or edges. In [33], Mac Carron & Kenna apply this approach while studying character networks extracted from myths, and comparing them to real-world and fictional social networks. The latter type corresponds to a selection of novels and plays, and also includes the *Marvel* dataset of Alberich *et al.* [1]. Mac Carron & Kenna identify robustness to targeted attacks as a sign of the network artificiality, in the sense that its structure relies too much on a few important characters to be realistic. All three considered types of social networks tend to be robust to random attacks, however real-world ones are vulnerable to targeted attacks whereas fictional ones are robust to these. The mythological networks studied in [33] tend to be on the realistic side.

We adopt the same method to study *Thorgal*. Both the unfiltered and filtered networks are robust to random attacks: removing 5% of the vertices results in giant components containing 89% and 100% of the remaining vertices, respectively. On the contrary, they are fragile to targeted attacks, for all four considered centrality measures. The effect is noticeably stronger for the unfiltered network, which can be explained by its lower density. When considering degree and Eigencentrality, the largest component retains half the vertices for the unfiltered network, and three quarters for the filtered one. When considering betweenness and closeness, we get 5% and 50% of the vertices remaining in the unfiltered and filtered networks, respectively. We can conclude that the *Thorgal* exhibits the characteristics of real-world social networks regarding these attacks.

#### SM2.4. Graph Typology

As mentioned in Section 2, when discussing the structure of the narratives constituting their science-fiction corpus, Rochat & Triclot [49] identify four main types of networks based on their overall structure. This typology is based on the relative position and interconnection of the main vs. secondary characters.

**Typology** First, the *kernel* network is built around a dense group of characters, dominated by a protagonist, and surrounded by some loosely connected periphery. In this type of story, the protagonist is typically assisted by some sidekicks, that follow him or her everywhere. The periphery can either be single characters or social groups, that are connected to the central group but disconnected from each other.

Second, the *unicentric* network can be seen as an extreme version of the previous type, in which the kernel is reduced to a single vertex constituting its clear center. It represents the story of an isolated character facing separated social groups. The protagonist may have sidekicks, but only punctually.

Third, the *polycentric* network is also built around an interconnected group of vertices, but each of these central characters acts as a gateway, providing exclusive access to one or several social groups. Such structure typically appears in case of *hypodiegesis*, i.e. when a protagonist tells his story, which involves a separate social group, to some other main character; or when the novel is a *fix-up* (a rework of

some originally separate short stories).

Fourth, the *acentric* network has no recognizable center at all. This is caused either by a large number of characters with well distributed interactions, or on the contrary when the focus is put on a small number of characters, without any periphery. The former case is typical of serial, long-term stories, that have the time to develop the relationships between secondary characters. The latter is characteristic of behind-closed-doors narratives, in which all characters are important because they are few.

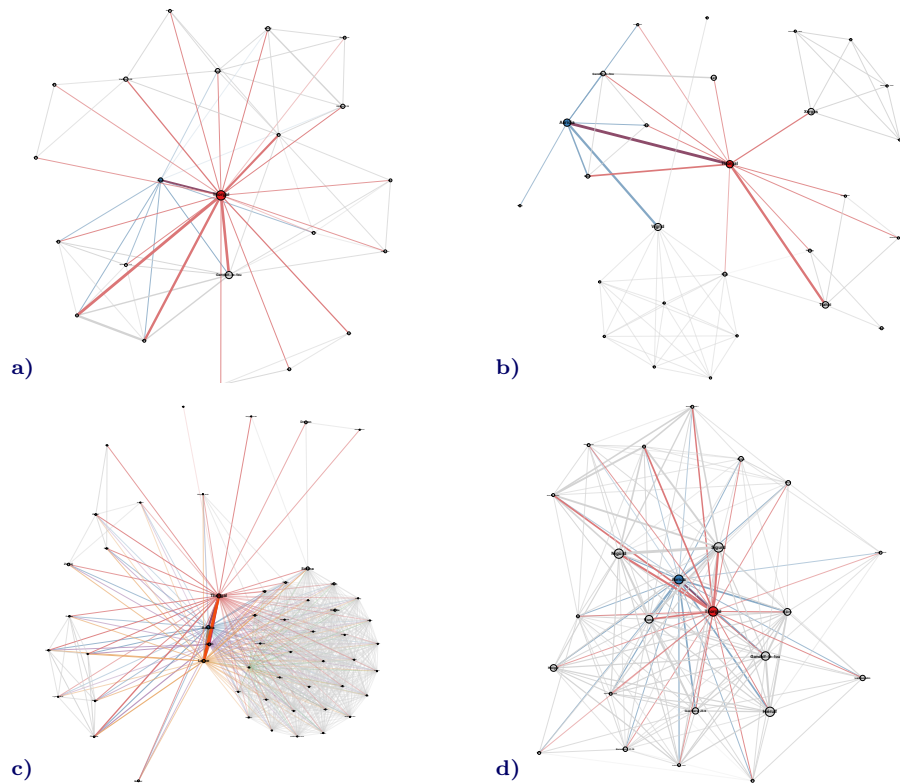


Fig. SM6: Selection of arc networks illustrating Rochat & Triclot's typology. Like in Figure 5, the five most frequent have a specific color; and vertex size and edge width are functions of betweenness and co-occurrences, respectively. The layout is different in each plot, to match the graph structure. (a) Arc #1 is *unicentric*, as it introduces the main character, Thorgal. (b) Arc #4 is *polycentric*, because it is a collection of short stories. (c) Arc #9 exhibits a kernel constituted of Thorgal's close family. (d) Arc #18 is *acentric* due to the uniform distribution of edges. Figure available at [10.5281/zenodo.6573491](https://zenodo.org/record/6573491) under CC-BY license.

**Overall *Thorgal* Network** When defining their typology, Rochat & Triclot experiment with various edge weight thresholds to remove minor characters [49], therefore when analyzing *Thorgal*, we focus on our filtered network. As illustrated by Figure 5, the filtered network extracted from the full series combines aspects of the *kernel* and *polycentric* types: the colored vertices corresponding to the five main characters constitute a kernel, but some of them are connected in an exclusive way to whole peripheral parts of the network. By comparison, Rochat & Triclot describe the network of the *Metabarons* graphic novel as *acentric*, and explain this by its serial nature. But *Thorgal* is also a serial narrative, and furthermore it is much longer than *Metabarons* (63 vs. 8 volumes). The fact that it has a clear center nonetheless can be explained by the spatial (and therefore social) separation between the groups met by the main characters, who go all over the world during their many travels.

**Arc-Focused *Thorgal* Networks** In addition, *Thorgal* can be broken down into 23 narrative arcs, whose respective networks include instances of all four of Rochat & Triclot's classes: kernel (3/23), unicentric (4), polycentric (13) and acentric (3). Table SM2 lists the type of each arc, and Figure SM6 gives 4 examples.

Arc #1 (Figure SM6a) introduces the protagonist, Thorgal, and is completely built around him, constituting a typical case of unicentric network. Incidentally, the same observation holds for the first volumes of the spinoffs: Arc #12 (Kriss of Valnor) and #14 (Louve).

Arc #4 (Figure SM6b) is a collection of short stories, and also has a *story within a story* situation (a.k.a. *hypodiegesis*), resulting in a polycentric structure.

Arc #5 takes place on a small island; it involves a handful of characters, and illustrates the behind-closed-doors acentric narrative. Arc #18 (Figure SM6d) is also acentric, but it corresponds to the case where many characters are interconnected.

Arc #9 (Figure SM6c) exhibits a kernel structure, as it follows the story of Thorgal's family, whose members stay together (for once).

Overall, the large prevalence of the polycentric structure reflects the fact that a majority of volumes tell several parallel stories at once, alternatively focusing on several main characters.

### SM3. Research Questions

This Section treats an additional research question (Section SM3.1), and provides extra figures regarding centrality distribution depending on character sex (Section SM3.2).

#### SM3.1. *Substitution of the Protagonist*

As noted by Desfontaine *et al.* [15], Thorgal and his elder son Jolan start sharing the stage in the volumes of the main series published right before their article. This corresponds to a time when Jolan reaches his late teens in the series. Figure SM7 shows the evolution of the strength of both characters, as a function of the volumes. Only the volumes in which at least one of them appears are represented. It appears indeed that Jolan overtakes Thorgal from volumes 28 to 32 in the main series. This event corresponds to an editorial choice from the authors, which became evident to the readers with volume 30, titled *I, Jolan*.



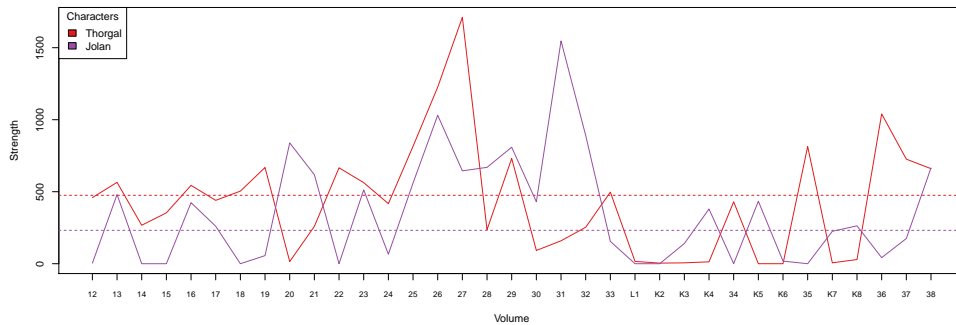


Fig. SM7: Evolution of the strength of Thorgal (red) and Jolan (purple) over the volumes. The dotted lines represent their respective average strength over the whole series, as a reference. Figure available at [10.5281/zenodo.6573491](https://zenodo.org/record/6573491) under CC-BY license.

In the preceding volume (*The Sacrifice*), Thorgal is critically ill. Jolan must accept the help of a mage, Manthor, to save him. In exchange, he swears to serve him for several years. This narrative trick is the occasion to give some independence to Jolan, which starts living his own adventures apart from his family, together with four other teenagers. This change matches the retirement of Jean Van Hamme, the historic writer of the series, and his replacement by new writer Yves Sente. It also corresponds to an effort by the publisher to rejuvenate the readership of the series. Indeed, as mentioned before, characters get older in *Thorgal*, including the protagonist. Providing a teenage protagonist to the main series is a way to attract younger readers. This is also one of the motivations for the introduction of the *Lowe* spinoff, which tells the story of Thorgal's kid daughter.

Interestingly, this situation only lasts a few volumes. After volume 33, Thorgal starts dominating again. This is due to a period of instability in the creative team, which saw several changes of writer in a short period of time. The idea of replacing Thorgal by a younger character as the new main protagonist was finally overturned, as shown in Figure SM7.

### SM3.2. *Centrality vs. Sex*

Figure SM8 shows the distribution of the four centrality measures, while distinguishing male and female characters (mixed and undetermined characters are not represented due to their small number). As before, the main plots show the distributions obtained for the unfiltered network, and the insets show the those obtained for the filtered networks.

## SM4. Network Figures

This section provides additional figures representing the whole network (Section SM4.1) as well as arc-specific networks (Section SM4.2).

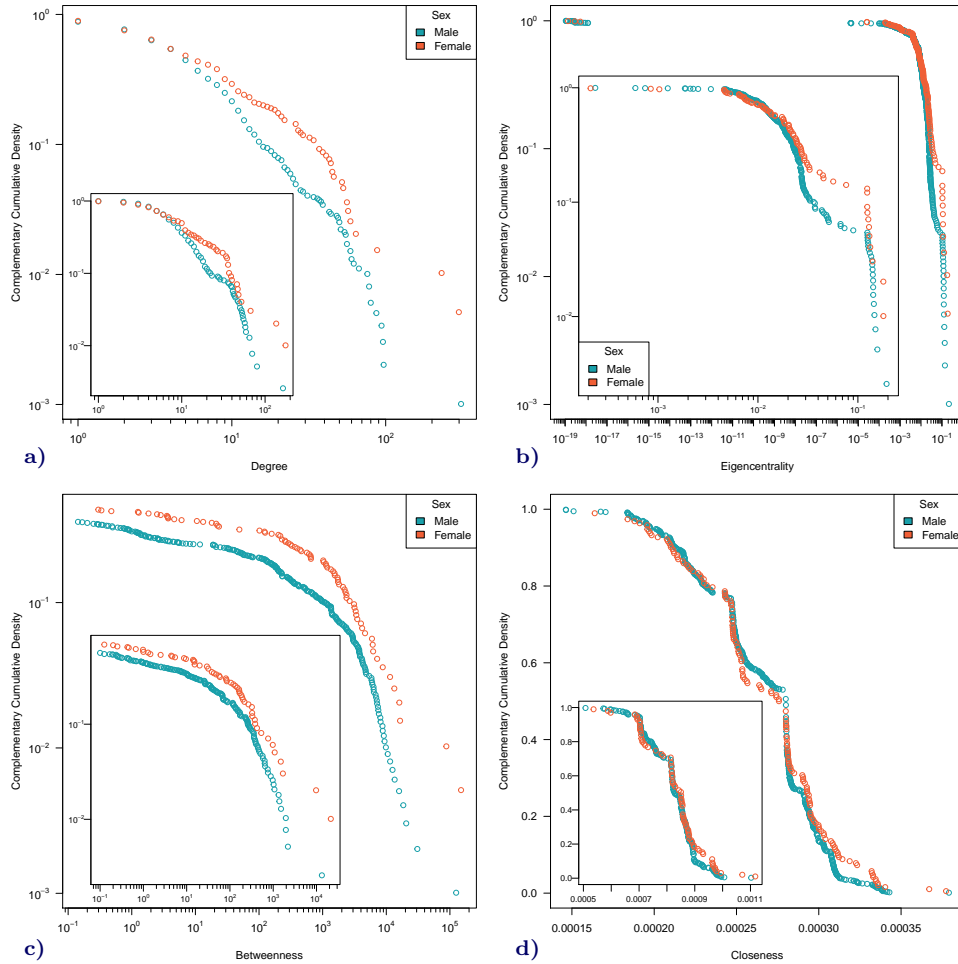


Fig. SM8: Distribution of four centrality measures depending on character sex, for all characters (main plots) and after filtering them (inset). (a) Degree. (b) Eigenvector centrality. (c) Betweenness. (d) Closeness. All axes use a logarithmic scale. Figure available at [10.5281/zenodo.6573491](https://zenodo.org/record/6573491) under CC-BY license.

#### SM4.1. Whole Series

Figure SM9 shows the *unfiltered* version of the Thorgal network, i.e. with all characters, even minor ones. By comparison, Figure 5 represents the *filtered* network, focusing only on primary characters as defined in Section 4.3. Figure SM10 highlights (in red) the vertices that are removed during the filtering step.

#### SM4.2. Arc Networks

Figures SM11 and SM12 show the character networks obtained for each narrative arc of *Thorgal*, as listed in Table SM2. To ease visual comparison, all plots use the

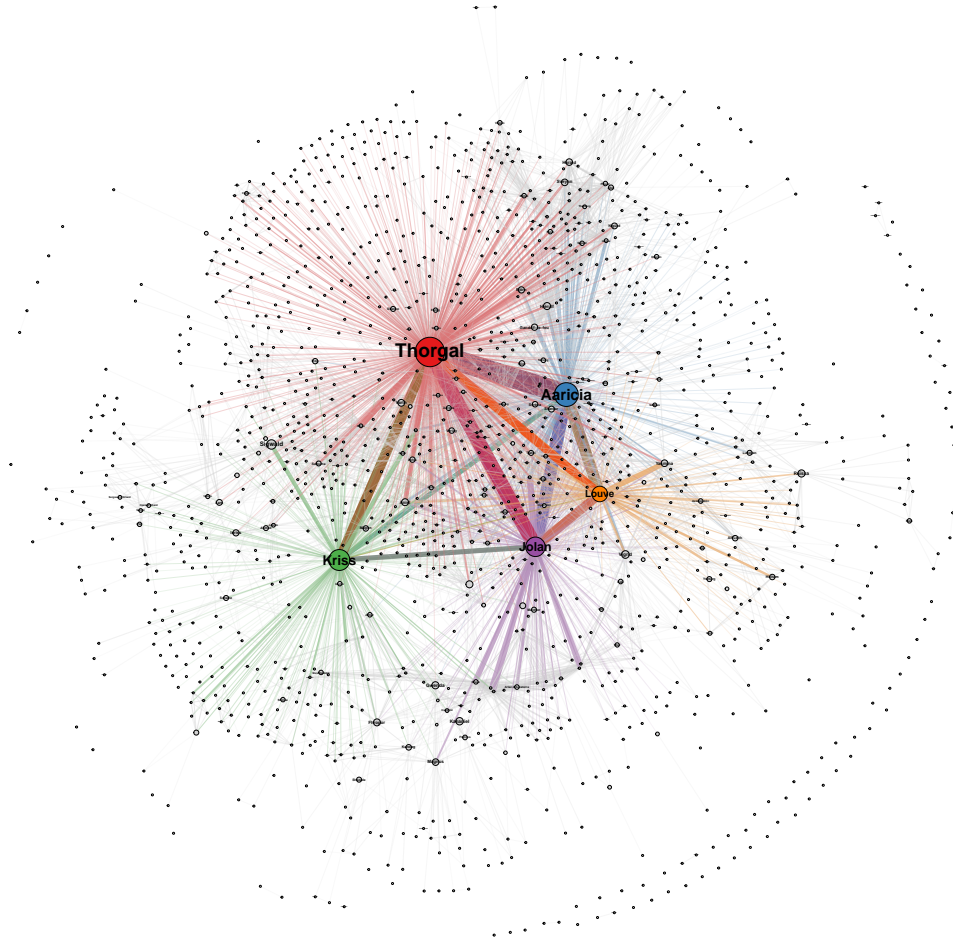


Fig. SM9: *Unfiltered* network extracted from the whole *Thorgal* series. Vertex size is a function of betweenness, and edge width is a function of the number of co-occurrences. The five most frequent characters (in terms of scenes) are shown in a specific color: Thorgal (red), his wife Aaricia (blue), their elder son Jolan (purple), their daughter Louve (Orange), and the antagonist Kriss of Valnor (green). See Figure 5 for the *filtered* version of this plot. Figure available at [10.5281/zenodo.6573491](https://zenodo.org/record/6573491) under CC-BY license.

same layout, so that the same vertex is always located at the same place.

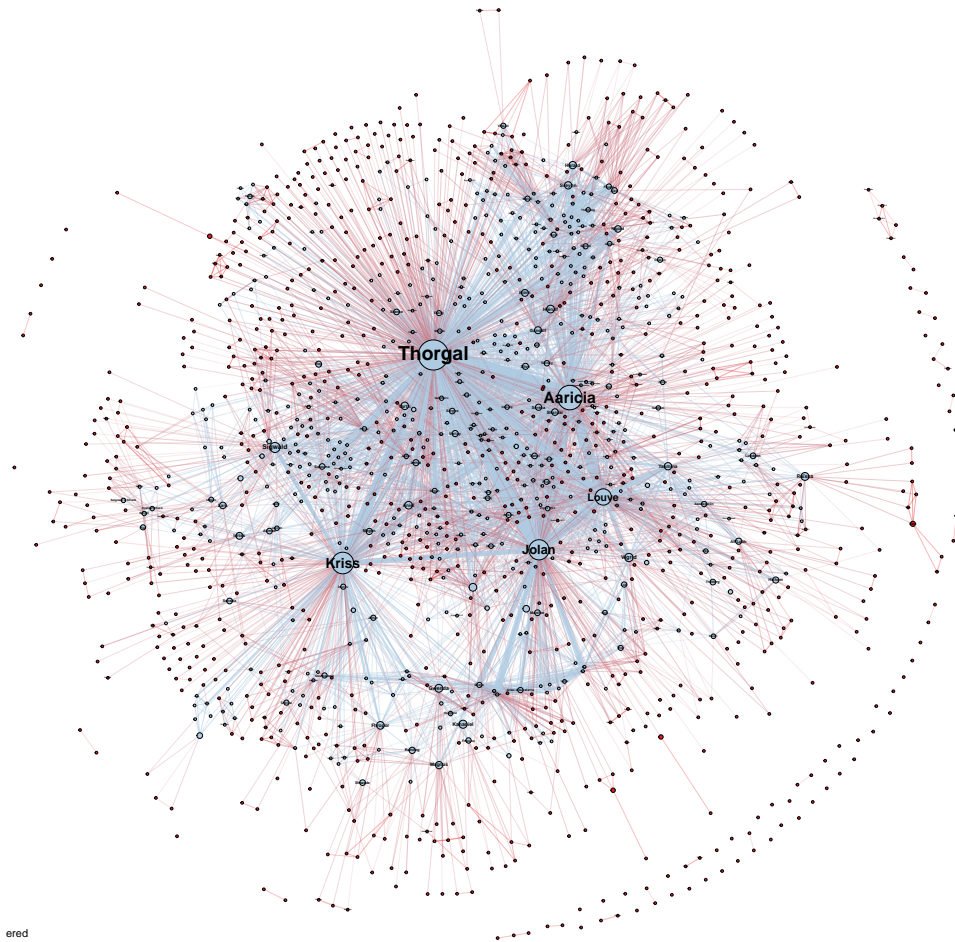


Fig. SM10: *Unfiltered* network showing the filtered vertices (in red). Vertex size and edge width have the same meaning as in Figure SM9. Figure available at [10.5281/zenodo.6573491](https://zenodo.org/record/6573491) under CC-BY license.

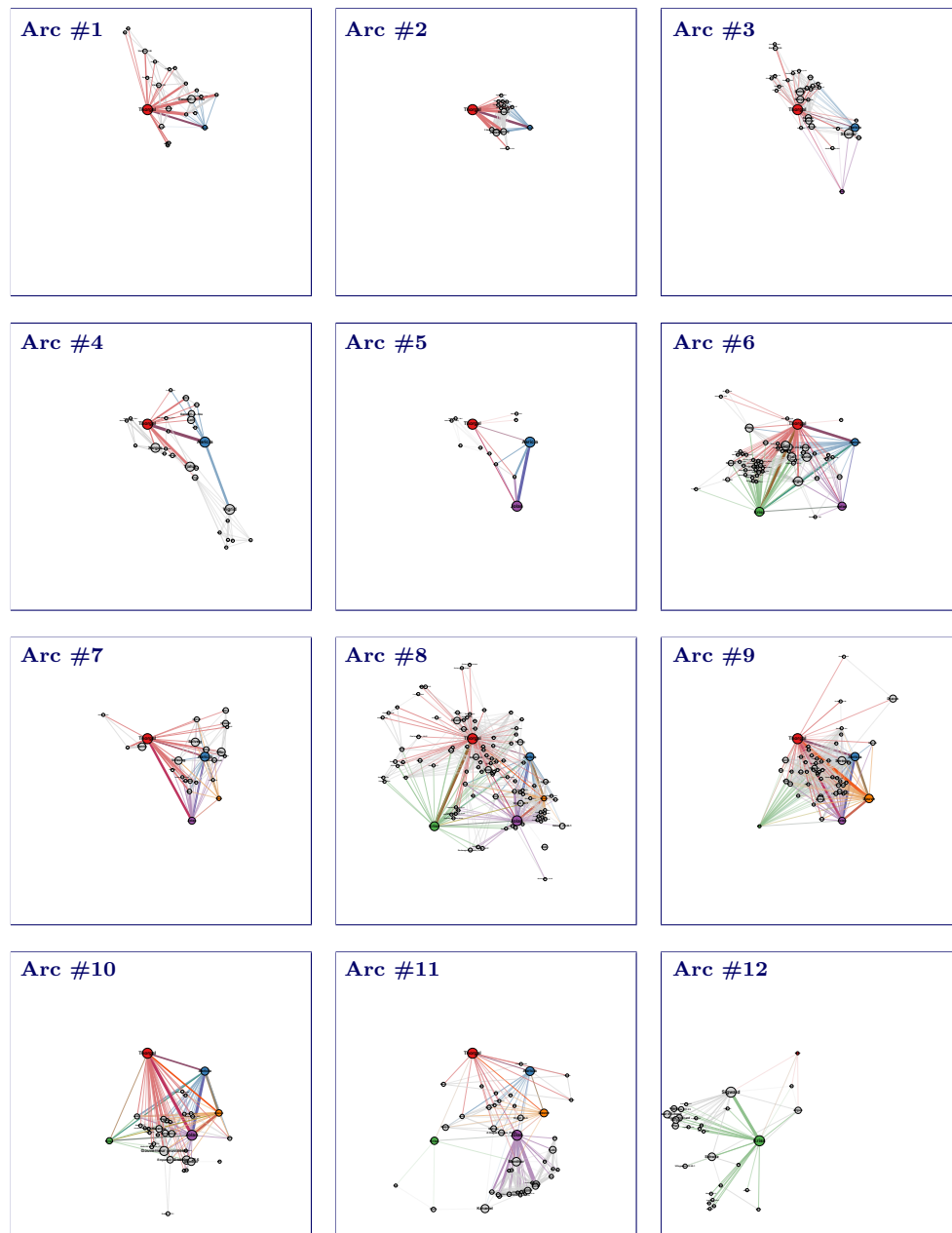


Fig. SM11: Filtered characters involved in narrative arcs #1–12 of the *Thorgal* series. As in Figures 5 & SM9, vertex size is a function of betweenness, edge width is a function of the number of co-occurrences, and the five most frequent characters are shown in a specific color: Thorgal (red), his wife Aaricia (blue), their elder son Jolan (purple), their daughter Louve (Orange), and the antagonist Kriss of Valnor (green). Vertex positions are fixed, to ease visual comparison. See Figure SM12 for narrative arcs #13–23, and Table SM2 for the arc titles. Figure available at [10.5281/zenodo.6573491](https://zenodo.org/record/6573491) under CC-BY license.

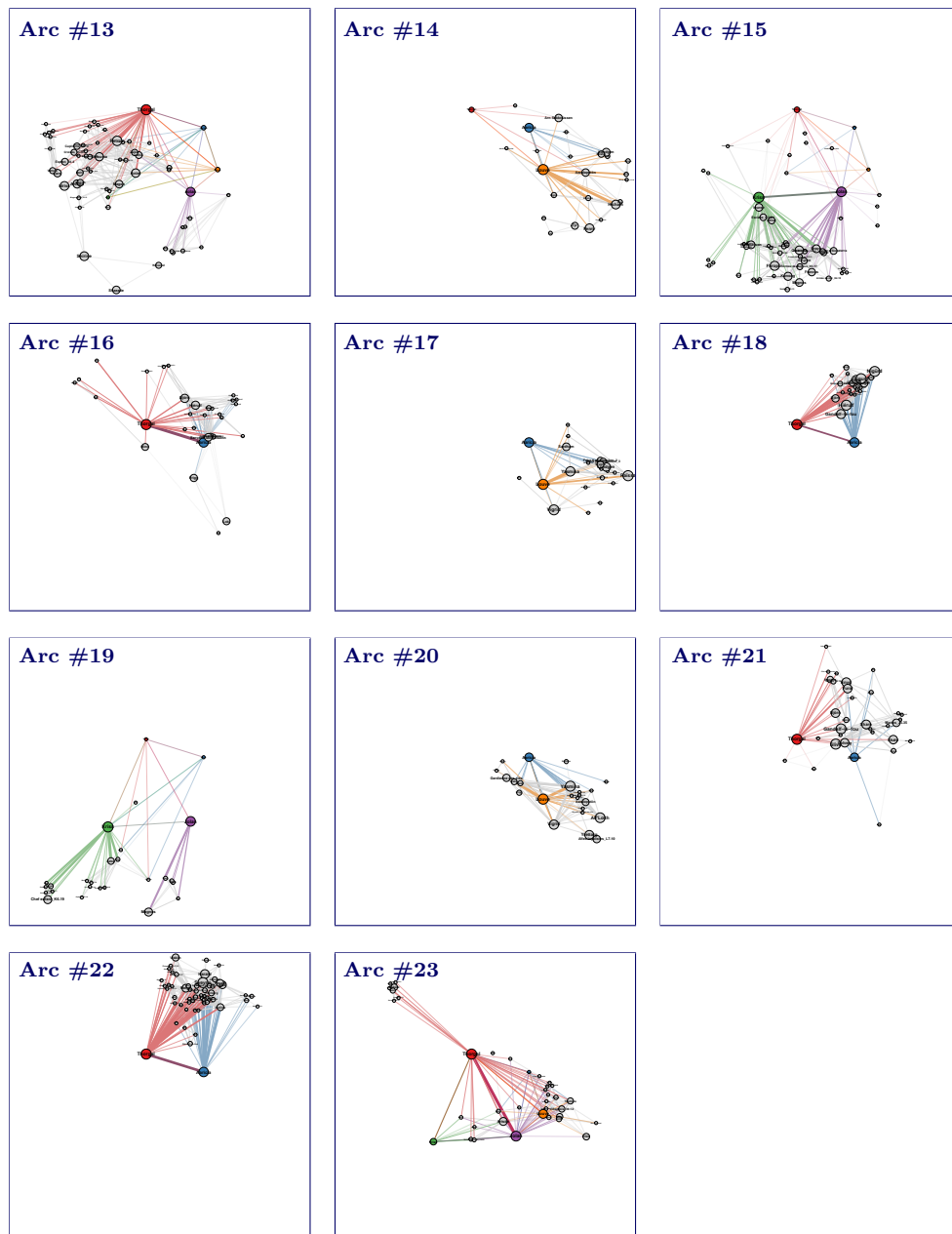


Fig. SM12: Filtered characters involved in narrative arcs #13–23 of the *Thorgal* series. As in Figures 5 & SM9, vertex size is a function of betweenness, edge width is a function of the number of co-occurrences, and the five most frequent characters are shown in a specific color: Thorgal (red), his wife Aaricia (blue), their elder son Jolan (purple), their daughter Louve (Orange), and the antagonist Kriss of Valnor (green). Vertex positions are fixed, to ease visual comparison. See Figure SM11 for narrative arcs #1–12, and Table SM2 for the arc titles. Figure available at [10.5281/zenodo.6573491](https://zenodo.org/record/6573491) under CC-BY license.