Structures in Tropes Networks: Toward a Formal Story Grammar

Jean-Peïc Chou
École polytechnique
Palaiseau FRANCE
jean-peic.chou@polytechnique.edu

Marc Christie
Univ Rennes, Inria, CNRS, IRISA, M2S
Rennes FRANCE
marc.christie@irisa.fr

Abstract

Tropes are cultural narrative conventions that shape our expectations of stories. This paper proposes a new approach to examine movie content relying on tropes. It first presents the architecture of tropes ontology extracted from tvtropes.org, then studies the link prediction problem in trope bipartite networks and discusses the next challenges and numerous applications that ensue from this approach. In addition, we propose to assess the potential of tropes to be the lexicon units of a formal story grammar.

Introduction

Stories are the pillars of all human cultures. They shape our imagination, forge our understanding of the world, and ensure our cultural heritage. Stories have been particularly broken down and analyzed since early XXth century Russian formalism. Propp observed Russian tales regularity and identified 31 narrative functions and 7 types of characters (Propp 1968). Greimas later proposed the first role-based structure for tales with the actantial model and introduced the *isotopy* as the redundancy of semantical elements in a text ensuring the coherence of a story (A.J. Greimas 1966). Spearhead of the structuralism, Barthes proposed an exhaustive analysis model by decomposing stories into complex hierarchical structures of discontinuous and heterogeneous elements with interacting isotopies, understood as meaning units. He distinguished the Function (causality, chronology…), the Action (communication, desire, hardship), and the Narration (R. Barthes 1966), and suggested five conventional ways to interpret a text segment that he named *codes (Barthes and Balzac 1970)*. This narratological heritage has largely been expanded by derived models since then.

Propp's works notably inspired Lakoff who developed the first story grammar (G. Lakoff 1972). This milestone was followed by Colby's specialized grammar of Eskimos folktales (Colby 2009) and the first general grammar by Rumelhart (David E. Rumelhart 1975) which could be applied to wider sets of stories. These grammars define rules that for instance break down stories into setting, theme, plot, and resolution, each of them being decomposed into other subelements such as characters, goal, or episode. However, these grammars have not been able to capture all stories specificities. Indeed, stories evolve with their tellers' fantasies, as their structures, their forms, and their content do. In their evaluation of story grammars, Black and Wilensky pointed out the limits of such structures for story comprehension and urged instead to study the knowledge enabling to understand story content (John B. Black and Robert Wilensky 1979). Garnham also discredited story grammars invoking the finitude of lexicon story elements (Alan Garnham 1983). Eventually, although such studies have been exploited in a few computational contexts such as story generation, the high abstraction of these grammars or their lack of data have made them unsuited to computational uses.

To overcome these deficiencies, we propose to use tropes as interconnected elements of the lexicon of stories, structured in semantic networks. As part of this work, we present a thorough analysis of tropes in five main sections. Our first inquiries will be to introduce tropes definition and state-of-the-art in the first two sections. In the third section, we will dwell on our data itself, examine its structure, and propose a terminology to concepts revolving around. These grounds will enable us to study trope networks to extract statistical patterns suggesting underlying grammar rules in the fourth section. We will eventually conclude on the perspectives raised by this work.

The work presented here is part of a larger research effort to assess the potential of tropes to be the semantic unit of a formal language of stories.

Tropes Definition

Tropes are storytelling devices or conventions. They are recognizable patterns found in all kinds of media. Authors and creators use them as narrative tools to make a specific impression on the audience. Tropes can describe every level of a work: the story and its discourse, characters and their interaction, location, time… The *Save the Princess* trope depicts the universal story plot in which a character, often portrayed by the *Damsel in Distress* trope, is kidnapped, and later rescued by the *Hero* for instance. At a finer scale, tropes can also simply correspond to a way a scene is filmed by a camera. The *Revealing Hug* trope depicts how the camera sometimes focuses on a character whose eyes are open while hugging to suggest that something is on his/her mind. As tropes describe both stories and discourses, we can fully

ISBN: 978-989-54160-3-5

break down and analyze narratives through the prism of tropes.

Tropes are inventoried on a website named *tvtropes* taking the form of a wiki. An active community of contributors have been adding tropes and artworks of all kinds (from advertisements to theater pieces) on the website since 2011. These data are obviously subjective and are therefore regularly modified. Some tropes are redefined, renamed, removed, while others have not been listed yet or emerge from the latest original movies. As such, tropes library is organic, dynamic, and extensive (García-Ortega et al. 2020b). It adapts to current events and will never be complete, evolving over the course of the debates on the forum. Tropes library is for this reason the most extensive story elements lexicon created.

Besides, tropes find their essence in people's story imagination and are thus deeply rooted in one's culture. Many branches of the wiki exist in different languages, containing culturally specific tropes. We will use the English version of the database for it is the most complete one and for the British-American culture is widely spread in the occidental world.

Related Work

As of today, only a few publications focus on or mention tropes. Among them, García-Ortega et al. propose an initial analysis of tropes in films highlighting key figures about *tvtropes* data (García-Ortega et al. 2020a). In his work on *tvtropes* website pages network, Meeks studies tropes main topics and manages to detect six weak communities that tend to settle along six themes (Meeks 2011). He then uses topic modelling to depict works from this classification. The nature and extend of these themes are however questionable as they seem vague and are centered around overly used, and not necessarily significant, tropes.

Tropes have otherwise mainly been used for story generation. Thompson assesses that tropes are well-suited for being story components of advanced grammar in such a task (Thompson 2018) as they are reusable, present a natural semantic hierarchy and can be combined to form more complex and nuanced components than plain *Action Hero*. He also states that a major advantage of tropes is their popular and cultural topicality, unlike unfamiliar Propp's concepts. He created *TropICAL*, a trope-based programming language for story authoring which enables to combine and create tropes as building blocks of simple stories. Guarneri et al. (Andrea Guarneri et al. 2017) selected 94 tropes that they classified into narrative types and used them as story components for their game plot creation tool *GHOST*. Tropes library was completed with cards from the creative card game *Once Upon a Time*. Both previous approaches rely on artist James Harris' way to picture tropes roles in stories structure (Harris 2017). In his artwork named *The Periodic Table of Storytelling*, he classifies tropes based on their Narrative category to build a periodic table made of story atoms. Eventually, García-Ortega et al. built an evolutionary algorithm that generates original sequences of tropes coupled with a neural network predicting movies success from their tropes (García-Ortega et al. 2020b). However, the generated lists lack chronological order and more generally interpretability. Eger and Mathewson help machines and humans collaborate by using tropes as random narrative beats for live improvisational performances (Eger and Mathewson 2018).

Tropes have also been mentioned in movie genre classification tasks. In their study in trailer generation, Smith et al mention *weak* tropes as visual elements characterizing genres (Smith et al. 2017). They note statistical patterns between colors, objects or places and genres: astronauts are present in drama movies while waitresses are found in romances or comedies. Äijälä looks for tropes similarities between movies and assesses the results with a genre analysis (Äijälä 2020). Results on most popular movies are encouraging at a small scale but the hierarchical clustering fails at forming coherent larger communities.

Eventually, *tvtropes* data has been used to conduct diverse cultural studies such as Mellina and Svetlichnaya's analysis of the evolution of the bipartite network with widely used emerging and dying tropes (Mellina and Svetlichnaya, 2011). Otherwise, gender bias in tropes is studied by Gala et al. (Gala et al. 2020), and by Assogba et al. from Bocoup company with an interactive visualization of tropes (Assogba etal.).

Tvtropes Data Networks

In traditional grammar, nature and function are used to define the type and the role of a word in a sentence. Knowing them notably allows us to apply the correct grammar rules. In this section, we will aim at structuring our heterogeneous data to have a sense of the nature and the function of tropes. We will study the *tvtropes* resource as an ontology by extracting and defining underlying networks from the website and more concretely, explore the themes covered by the data.

Tropes Data

Tvtropes is divided into two main parts. The first main part of *tvtropes* includes tropes information. Each of the tropes is always provided with a thorough description which can include its definition, a stereotypical example, the context in which the trope often appears, its supposed origins and its evolution, or the tropes it is close or antinomic to with the mention *See also*. Below, we can generally find an unexhaustive list of examples and their context in various media. The description and the examples can also be accompanied with:

- A *Laconic* page which gives a short definition of the trope,
- A *Quotes* page presenting extracts from works that deal with the trope or give an example of it,
- A Playing With page listing several examples of variations as described above,
- An image illustrating the point,
- A video example.

More than 27,000 tropes have been identified, some having been used in more than 2000 different works. A case in point is the most used trope *Shout Out*, which describes a work referring a person or another work. Figure 1 shows the repartition of the number of listed uses in movies by trope.

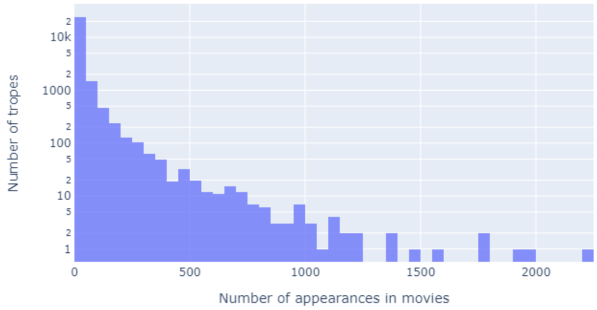

Figure 1: Repartition of tropes based on their appearance

Media and the Bipartite Networks

The second main part of the wiki focuses on *Media*, which correspond to artworks of all sorts. Each work is described similarly to Wikipedia pages and can include links to other pages giving anecdotes, iconic moments, common thoughts, summaries, other information. More interestingly, a list of characters with tropes describing them is sometimes available.

Tropes have been listed for each work and each character. The more fans the movie has, the more complete the inventory of tropes is (García-Ortega et al. 2020b). Figure 2 presents the repartition of the number of listed tropes by movie. We observe that some films include more than 600 tropes which corresponds to 1 trope every 12 seconds on average for a 2-hour-long movie. Most described movies are superhero blockbuster movies (*Thor Ragnarok*, *Thor*, *X-Men Apocalypse*…).

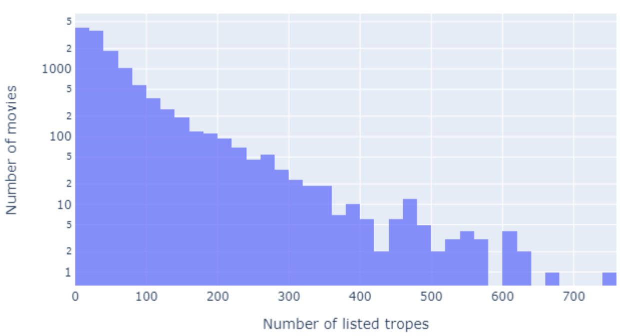

Figure 2: Repartition of movies based on their number of listed tropes

The website structure naturally forms a bipartite graph linking tropes to artworks or characters. Whether focusing on works or characters, we will call this structure the *bipartite* network. The projection of the *bipartite* network results in a co-occurrence graph. These co-occurrence graphs are particularly interesting since they allow to link movies with similar tropes, i.e. similar plot events, for instance.

The Structural Networks

Tvtropes Classification

Tvtropes divides tropes into multiple hierarchical categories and subcategories as on Wikipedia. Tropes are mainly classified into four overlapping folders. The names of the categories explicitly describe the kind of classification inside each of them:

- Genre (Action/Adventure, Comedy, Horror…)
- Media (Film, Game, Theatre…)
- Narrative (*Characters*, *Plots*, *Symbols* …)
- Topical (Alien, Betrayal, Combat …)

Most of the time, tropes belong to several categories or sub-categories. There are no clear levels of folders: a folder can be at two levels below and one level below a category at the same time as shown in Figure 3 for instance. There can namely be multiple paths of varying distances from a folder to another in this data structure. We will call the network resulting from this hierarchy and categorization, the *structural* network.

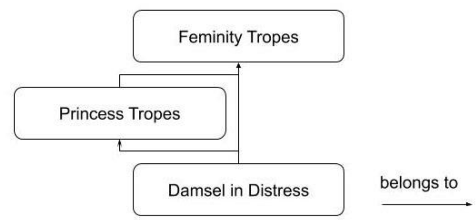

Figure 3: Example of a hierarchical overlapping found on tvtropes.org

Narrative subcategories and Abstraction Scale

We observe that the narrative subcategories describe functions of tropes in a story, some of them having a more or less significant impact on it. For example, *plots* tropes will naturally be more relevant to the story than *Symbol* ones. From this intuition, we propose to determine an abstraction scale of tropes, namely defining levels of description. For this, we exploit the proposed Narrative subcategories which include:

- *Applied Phlebotinum*: Unexplainable plot fuel, something that enables scenario shortcuts (e.g. futuristic technology, magic and powers, tools…)
- *Characterization*: Describe characters, his/her origins, appearance, moral values…
- *Characters*
- *Characters & Casting*: Impact of human actors playing roles
- *Characters as Device*: How a character serves a plot
- *Character Introduction Index*: How characters are introduced
- *Conflict*: Problems that drive a story
- *Dialogue*
- *Genre* (main folder)
- *Motifs*: Symbols that recurrently appear in a work
- *Narrative Devices*: What moves the story forward or organize a scene or sequence
- *Narrator Tropes*

- *Paratext*: Not directly contained in the work content (DVD box, preface…)
- *The Plot Demanded This Index*: Something happening for no other reason than 'the plot requires it to'
- *Plots*: Organize the action of an entire script or episode
- *Settings*: Time, location, and circumstances of the narrative
- *Show, Don't Tell*: Demonstrative techniques rather than informative ones
- *Spectacle*: Techniques to impress the audience (lighting, props, camera tricks…)
- *Symbolism*: Common representation of an idea, belief, event…

Some of these folders contain or belong to each other. We arbitrarily decide not to consider *Characters & Casting*, *Characters as Device*, *Conflict*, and *Genre* to limit the overlapping and to work with heterogeneous categories. These categories correspond to tropes functions in stories.

We pinpoint some strong correlations between categories which have tropes in common. We first consider *Plots* tropes as the highest level descriptors of stories, and thus place them at the top of our abstraction scale. We then naturally build a scale by placing alongside correlated categories and obtain a result observable in Figure 4. We note that:

- *Paratext*, *Settings* and *Applied Phlebotinum* are put aside as they are independent categories, uncorrelated to other ones.
- The *Dialogue* tropes can both be placed at a high and low level. Dialogues can reveal plot-twist and be at the center of a story while many of them only fill a work.

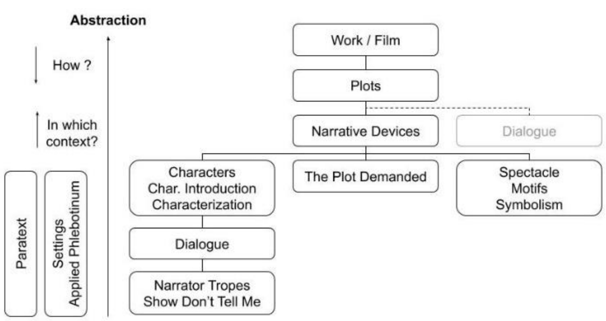

Figure 4: Proposed abstraction scale built from narrative categories correlations

Semantic networks

From the wiki, we can extract three semantic networks that connect tropes with semantic relations:

- Synonymy: The *Revealing Hug* trope is very similar to the *Traitor Shot* trope as they both describe ways to film the suspect behavior of a character.
- Antonymy: *Heroes* versus *Villains*.
- Metonymy: *Revealing Hug* is considered as a sub-trope of *Meaningful Look*.
- Lexical field: *Save the Princess Plots* is most of the time associated with the *Damsel Distress Characters* trope for instance.
- Causality and dependency

We will call these structures as follows:

- *Description* network: Tropes' descriptions mention other tropes, pointing to each other via hyperlinks.
- *Related* network: Connections are also extracted from the *Related* page of each trope. It lists all the pages mentioning at least once the trope. This method completes the previous one. In addition to the descriptions, links can be found in the examples, captions… Both networks correspond to subparts of the *tvtropes* pages network analyzed by Elijah Meeks in 2011 (Meeks 2011).
- *Example* network: We note that examples of use of tropes on movie pages can refer to other tropes when providing the context. If the explication needs to include another trope to tell the context, it means that they share some semantic relation.

Datasets

How was this data collected? Tropes data were first extracted and made available for the scientific community by Kiesel and Grimmes with an RDF-dataset (M. Kiesel, G. A. Grimmes 2010). The last downloadable dump was generated 1st of July 2016 and contains over 20 million RDF statements for a total size of 4.7 GB. However, the size of *tvtropes* database has since then tripled. García-Ortega et al. released *PicTropes*, a more recent dataset including 5,925 films and 18,270 tropes in 2018 (García-Ortega et al. 2018). In this regard, they designed a downloadable web crawler named *tropescraper*. It first extracts all the movies from *tvtropes* then seeks for all the listed tropes on each film page.

We modified and made use of *tropescraper* to extract other available data previously mentioned (tropes categories, tropes per character, semantic networks links).

Exploring Topics in Tropes

These structures have given us a sense of the type of information tropes describe. But in more concrete terms, what are the topics covered by tropes and what are they about? Following Meeks' study (Meeks 2011), we worked on the network of tropes pages. Our attempts to obtain similar themes with community detection were not conclusive. This can be explained by the considerable growth of tvtropes website since 2011. Instead, we extracted tropes topics by studying tropes descriptions and examples with Latent Dirichlet Allocation (LDA) model and topic modelling. We used Gensim library implementation of LDA and define the number of topics by computing the topic coherence score for several numbers.

An initial overview of tropes descriptions presents overused words (e.g. "character", "one", "compare", "work".). We chose to filter words appearing in more than one document out of two. We ran a first analysis on tropes descriptions for which we find the best coherence with 18 topics. The intertopic distance map and three topic descriptions are shown Figure 7. Topic 1 clearly relates to the work as an entity and topic 3 to the characters used to describe a situation. Topic 4 refers to abilities and is close to video games lingo. Topic 3 is harder to semantically distinguish. It mixes notions of family, love, evil, or fuzzy words like car.

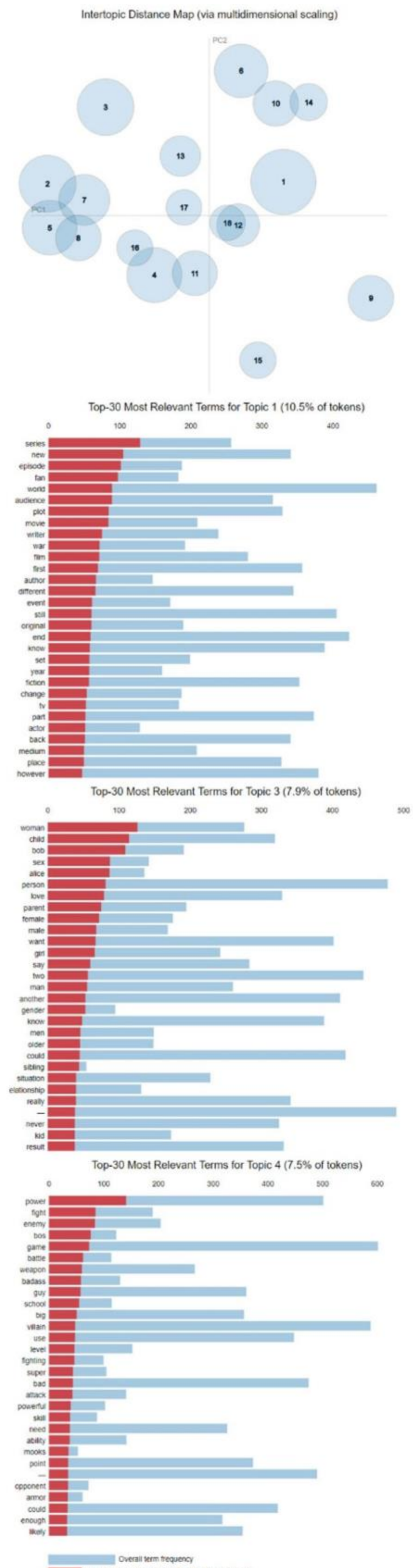

Figure 7: Intertopic Distance Map and descriptions of topics 1, 3, 4 from tropes descriptions with LDA model

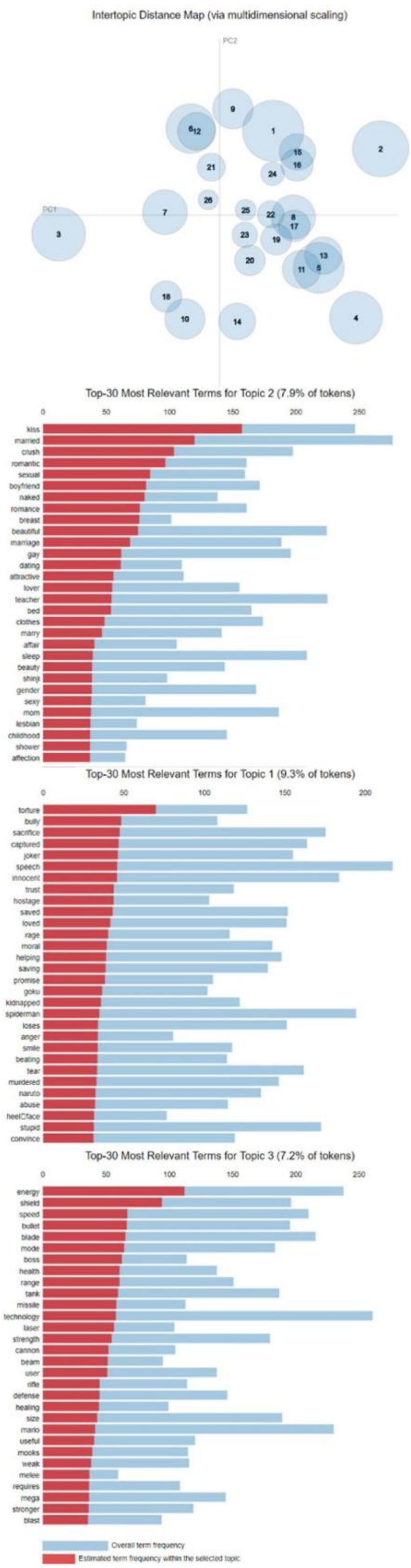

Figure 8: Intertopic Distance Map and descriptions of topics 1, 2 and 3 from tropes descriptions and examples with LDA model

We conducted the same study with tropes descriptions and their examples. We supposed here that the vocabulary employed is representative of each trope. Coherence graph indicated a local maximum around 26 topics. Extracted topics found in Figure 8 illustrate well the diversity of the themes. From love in topic 2 to weapons and video games in topic 3, or topic 1 and suffering.

This study on tropes descriptions shows the diversity of addressed topics and confirms the plurality of the abstraction levels scanned. In addition, it demonstrates the existence of main themes for future topic modelling analysis.

Link Prediction in the Bipartite Network

We will now study tropes within their networks. Predicting tropes presence in a movie from given tropes and information about the movie would suggest the existence of statistical patterns and a semantic logic in tropes interactions. In other words, this would confirm that there is an ensemble of latent rules to follow in order to understand tropes language. This task is equivalent to the well-known link prediction problem.

Problem Definition & Initial Datasets

Our first objective is thus to build an algorithm inferring removed links between movies and tropes of the *bipartite* network. Removed connections depend on our goal. We create and name initial datasets as follows:

- 80-20: For each movie, we remove 20% of the links with tropes to appraise the overall performance.
- N-random: We separate the movies in two parts. The first group (80%) remains intact and serves as training, while the second is connected to a chosen batch of N randomly chosen tropes (20%), serving as test. This situation would happen if we concentrated our efforts in conceiving algorithms dedicated to detecting N tropes.
- Narrative category: Following the previous method, we prepare train-test datasets by removing all links to tropes not in the studied narrative category.

We arbitrarily restrain our study to movies containing more than 5 tropes. This corresponds to 95% of the dataset. Tropes that have been used less than 5 times are also removed. The graph is an undirected bipartite multi-layer graph which can be weighted. Its main structure is given by the movies-tropes *bipartite* network. The other layers contain information about tropes from the previously built ontology, and information about movies content or metadata.

Method

We separately work with each layer of information to assess their efficiency in this task. We only make use of information that a naive algorithm detecting tropes would be able to obtain. Movie tags or collection are for instance not included. The graph is presented in Figure 9 for more clarity. The available information is used as follows:

- Movies *bipartite* network: We draw inspiration from Kunegis et al. who adapted link prediction functions based on probabilities of paths to bipartite networks by keeping the odd components (Kunegis et al.). We use this way odd polynomial link prediction functions, namely the Hyperbolic Sine and the Odd von Neumann Pseudokernel of the transition probability matrix with an empirical $\alpha = 0.1$.

$$\sinh(\alpha A) = \sum_{i=0}^{\infty} \frac{\alpha^{1+2i}}{(1+2i)!} A^{1+2i}$$

$$K_{\mathrm{NEU}}^{\mathrm{odd}}(A) = \alpha A(I - \alpha^2 A^2)^{-1} = \sum_{i=0}^{\infty} \alpha^{1+2i} A^{1+2i}$$

- Characters *bipartite* network: We project the characters bipartite network on tropes nodes with edges weights given by the number of common neighbors. We compute the probabilities of paths between movies and tropes of a random walker starting on a movie, taking a first step to reach the tropes side, then taking a second on this resulting projected graph.
- Semantic and structural networks: For each network, we apply a similar method (the second step is made in the considered network). The networks are not weighted.
- Movies metadata (genres, director, actors, released year): We use a similar method. This time, the random walker takes its first step from a movie to another one with probabilities relying on metadata similarities. Similarities are computed with the Jaccard method.

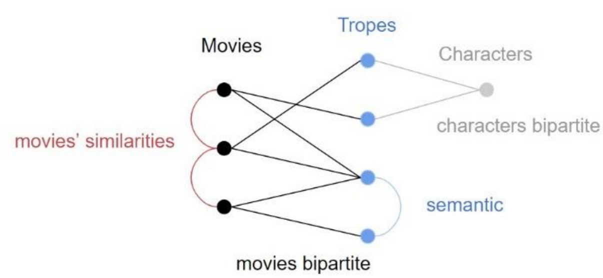

Figure 9: Simplified representation of the multi-layer bipartite network

For each dataset, we compute the probabilities of all tropes to be in a movie and rank them by order of likelihood. To evaluate the performance, we use area under precision-recall curve (AUPR) and area under ROC curve (AUC) as the classes are highly imbalanced (ratio of true edges in the order of 0.001). The metrics can be understood as follows: picking a random true removed link and a random false link, it is the probability that the computed likelihood of the true connection is higher than the false one.

Results

80-20 Dataset We report the results for the 80-20 dataset in Figure 10. Observations and analysis on this figure are listed as follows:

- For every dataset, best results are obtained from movies *bipartite* network and movies genres with respective AUC of 0.855 and 0.847. Both information sources rely on movies content similarities. It is reasonable to state that two movies of the same genres or with similar tropes will have relatively close list of tropes. For further analysis on movie content, we could compare subtitles with LDA as Bougiatiotis and Giannakopoulos do (Bougiatiotis and Giannakopoulos 2016).

- *Characters* tropes also show helpful statistical patterns leading to an AUC of 0.745.
- Tropes categorization into genres and topics does not seem to be efficient in this task as we obtain 0.601 and 0.607 of AUC for the 80-20 dataset.
- Semantic networks are more useful. Especially the *related*-based one which reaches a score of 0.775 compared to a score of 0.653 for the description-based one. This result seems fair as the *related* network is denser.
- With AUC of 0.528 and 0.589, movies director and cast are not valuable, while surprisingly, movies released year gives a relatively better AUC of 0.648.

To assess the potential of a fine-tuned link prediction algorithm, we sum the computed probabilities of each of these layers by attributing to them an empirical weight. The weights reflect the previous results. They can be found in Figure 10. We obtain an AUC of 0.88 for the 80-20 dataset which corresponds to an improvement of 2.9% compared to the previous best score.

	AUC	Weight
Bipartite (sinh)	0,855	1
Bipartite (neumann)	0,855	0
Characters	0,745	0
Description Semantic	0,653	0,1
Related Semantic	0,775	0,1
Tropes Genres	0,601	0,01
Tropes Topics	0,607	0
Movies Genres	0,847	0,1
Movies Year	0,648	0
Movies Directors	0,528	0
Movies Cast	0,589	0
Weighted Sum	0,88	

Figure 10: Results of the link prediction for the 80-20 dataset and empirical weights of the sources of information for link prediction

N-random & Narrative category The study by narrative category and with N-random datasets is shown in Figures 11 and 12. Best score for each source of information is shown in bold. This analysis is not meant to prove the absolute efficiency of a category for this task but discriminates between the best and worst performing categories:

- *Characterization*, *characters*, and *narrative devices* tropes lead to the best results. However, we cannot conclude on the high informative value of these categories as these categories include the most tropes with 7146, 4739 and 5461 ones.
- It can still be noted that we obtain better results with *characters* (4739) than *dialogue* tropes (4739), and with *plots* (2412) than *settings* (2210) as these categories include close number of tropes. This result confirms our initial intuition about high-level tropes.
- N-random datasets present higher AUC scores than their narrative counterparts with similar number of tropes. This indicates that heterogeneous information is more beneficial to predict links.

'Small' Movies and Characters

These results must be examined cautiously. Sine *tvtropes* dataset is inconsistent and biased, even a human would not be able to guess which tropes have been inventoried or not. The community might have based their list on a single scene or an insignificant detail. Besides, fully described movies such as *Star Wars* include more than 600 tropes which cannot make sense without indication of context.

However, we conducted the same work on characters and movies containing between 5 and 30 tropes. With AUCs of 94 and 91, we note a clear improvement that indicates that these works are more easily understandable with our data and make more sense together than huge jumble of tropes.

Perspectives and Future Work

In this study, we have defined tropes as well as their structures and have justified the existence of latent and cultural grammar rules for stories. What kind of outcomes can we hope from these results? What are the next challenges in this field?

Information source	**Applied Phlebotinum**	**Characteri-zation**	**Characters**	**Character Introduction**	**Dialogue**	**Motifs**	**Narrative Devices**	**Narrator**	**Paratext**	**The Plot Demanded This Index**	**Plots**	**Settings**	**Show Don't Tell**	**Spectacle**	**Symbolism**
Number of tropes	1066	7146	4739	53	4739	68	5461	276	303	334	2412	2210	71	2860	55
Bipartite (sinh)	0,607	0,803	**0,811**	0,559	0,759	0,532	0,806	0,609	0,642	0,668	0,799	0,735	0,566	0,770	0,533
Bipartite (neumann)	0,607	0,803	**0,812**	0,559	0,759	0,532	0,806	0,609	0,643	0,668	0,800	0,735	0,566	0,771	0,533
Characters	0,535	0,671	0,685	0,530	0,657	0,514	**0,704**	0,540	0,532	0,590	0,678	0,591	0,521	0,646	0,513
***Description* Semantic**	0,497	**0,560**	0,543	0,498	0,513	0,495	0,560	0,499	0,497	0,502	0,524	0,507	0,496	0,506	0,497
***Related* Semantic**	0,513	0,656	0,636	0,503	0,568	0,496	**0,663**	0,510	0,503	0,526	0,593	0,546	0,501	0,559	0,499
Tropes Genres	0,508	0,564	0,566	0,503	0,540	0,496	**0,575**	0,507	0,499	0,528	0,563	0,533	0,500	0,545	0,498
Tropes Topics	0,500	0,496	0,544	0,498	0,485	0,495	**0,559**	0,503	0,496	0,510	0,536	0,513	0,496	0,512	0,497

Figure 11: Results of the link prediction by narrative category, best score for each source of information is shown in bold

Dataset	100-Random	500-Random	1000-Random	2000-Random	5000-Random
Number of tropes	100	500	1000	2000	5000
Bipartite (sinh)	0,580	0,690	0,758	0,801	**0,821**
Bipartite (neumann)	0,580	0,690	0,758	0,802	**0,821**
Characters	0,536	0,587	0,648	0,693	**0,726**
Description Semantic	0,496	0,504	0,514	0,540	**0,572**
Related Semantic	0,507	0,531	0,561	0,617	**0,679**
Tropes Genres	0,509	0,526	0,542	0,575	**0,584**
Tropes Topics	0,502	0,512	0,525	0,553	**0,570**

Figure 12: Results of the link prediction for N-random datasets, best score for each source of information is shown in bold

A Look at Film Signature Extraction

This subsection is a short presentation of some results from our work on film signatures (Chou and Christie 2021) which was based on the results of this paper. The objective was to determine extractable features that would constitute the gist, namely the signature of a movie. This signature was then exploited for several applications.

In this regard, we started by picturing films as networks of tropes. For each movie, the nodes corresponded to the tropes, and the links were determined by the relationships between tropes in *tvtropes* ontology defined in this paper.

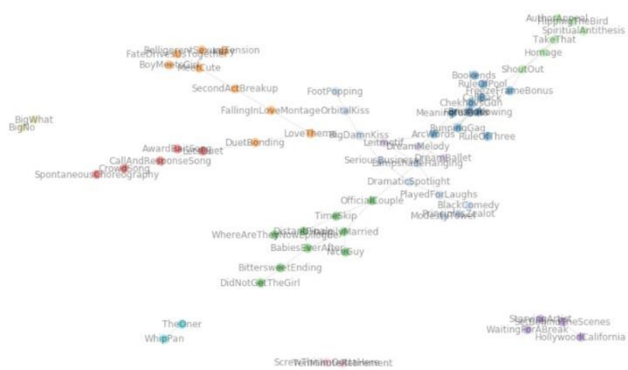

Figure 13: An example of *La La Land* tropes network

The analysis of these graphs of tropes led to promising results. Most central tropes were kept and constituted the signature. We proposed a new way to recommend movies by comparing signatures. When browsing the list of proposed recommendations, the results seemed conclusive. An example is presented Figure 14 for *La La Land* and *Inception*. The first movie is compared to other musicals and romances, and the second with other movies with complex stories or levels of reality.

Another example is the visualization of movies based on their tropes signature. The result shown in Figure 15 is informative since movies are clustered according to their genres and the similarity of their story.

Film comparé	*LaLaLand*	*Inception*
1	*Cinderella1997*	*JacobsLadder*
2	*MyFairLady*	*InTheMouthOfMadness*
3	*FromJustinToKelly*	*EventHorizon*
4	*IsntItRomantic*	*TwelveMonkeys*
5	*ShockTreatment*	*DonnieDarko*
6	*WestSideStory*	*TheFountain*
7	*SinginInTheRain*	*Mirrors*
8	*Enchanted*	*TheShining*
9	*Zombies2018*	*ParanormalActivity*
10	*MoulinRouge*	*DarkCity*

Figure 14: The 10 most similar movies to *La La Land* and *Inception* among works with more than 100 tropes

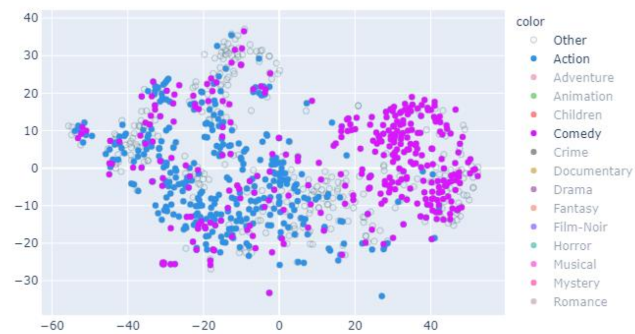

Figure 15: A map of movies based on the similarity of their signature with t-SNE visualization. Action in blue, Comedy in pink

Tropes Prospects for Applications

More generally, the analysis of movie content is essential for many applications ranging from information retrieval to computational creativity. In this regard, tropes vocabulary is very promising as it would enable to fully break down movies or other works. For all such tasks, our previous study has shown that we will need to find the best-suited way of harnessing tropes heterogeneity. Here are two examples of prospects.

Recommendation Systems Film signature extraction briefly presented in the previous subsection is an innovative way to tackle recommendation systems task (Chou and Christie 2021). As of today, previous studies have used tropes as homogenous information to compare movies. Exploiting tropes informative and heterogeneous richness would enable to better understand users' taste by detecting more detailed content similarities - whether they are about a plot twist, a personality trait, or a location for instance. *Iron Man* is a superhero action movie while *Steve Jobs* is a biographical drama. They share at first sight no common points, but their tropes indicate that they both mainly focus on a techno-visionary philanthropist character.

We believe that this level of description could help tackle some of the next challenges of recommendation systems defined by Shi et al (Shi et al. 2014). For instance, long-tail items are items that have low popularity and are therefore harder to compare with collaborative filtering methods. Content analysis with tropes would naturally overcome this problem. Cross-domain recommendation is another challenge that aims at predicting a user's tastes for any media. Our study focuses on movies, but most tropes are common to various domains as the numerous media categories of tvtropes.org indicate. With tropes, similarities between works from different domains would thus easily be assessed.

Story Generation Thompson mentions tropes abstraction as a powerful and useful concept since it allows to break down stories into smaller and smaller components (Thompson 2018) and Guarneri uses James Harris' partial narrative classification (Guarneri et al. 2017). None of them make full use of these abstraction layers, which could be a good starting point to structure and build a story from the general plot to the characters introduction. Such methods are useful creative tools as they offer a rich story vocabulary. But composition rules are still lacking to form a proper grammar and achieve story generation. Again, the analysis of films construction from their signatures might be a lead in this task.

Future challenges

Our analysis of tropes obviously presents significant lacks that must be addressed.

Analysis of Levels, Characters and Media First of all, our work mainly focused on tropes describing a movie as a whole. But such an approach could be applied at other levels of analysis. Working at scene-level or even frame-level sounds promising but the dataset lacks details for the moment. Characters could also be examined since *tvtropes* database includes tropes for more than 30,000 live action movie characters.

Besides, this study has only dealt with storytelling in live-action movies, but such an analysis could be performed on other media involving storytelling (animation, games, television, theatre, literature…). Including other media works would provide more examples to learn from and data to leverage on. Overall, the scope of our work could be extended in many ways.

Temporal Analysis One of our main lacks concerns temporal data. Knowing when a trope appears in an artwork absolutely but also relatively to another would open exciting new perspectives of analysis. Likewise, the duration of a trope is another central temporal feature yet to define and examine.

Lack of data These challenges are primarily a matter of the amount of data available. Therefore, the first stake will be to encourage *tvtropes* community to keep supplying the website database. The second stake will be to allow the community to add temporal information about the appearance of tropes. To this end, we could conceive a cross-platform application facilitating information entry.

Tropes and Originality

Tropes are often accused of being demystifying and cynical conventions ruining the abstract concept of art as they are often used by internet users to criticize works or to point out their impersonality. Besides, they seem to restrain creativity by simplistically breaking it down. To that we argue that tropes breadth and the countless ways to play with them offer a tremendous creativity. Some tropes can seem incompatible until they are exploited in the same plot or mixed to become a new trope. The website provides a tool generating random tropes of multiple narrative categories from which the user is challenged to build a story. *Tvtropes* community states that it makes no sense trying to build a story without any trope. Either the created story is of no interest or the latent processes are immediately inventoried, becoming tropes.

Tropes vocabulary is fired up by our imagination. It keeps growing as our cultures evolve and expands the boundaries of creativity. Tropes do not impose rigid patterns and methods. They should rather be considered as tools helping creators to grasp the content of our imagination, play with it, and reinvent it.

References

Greimas A.J. (1966) Sémantique structurale. Presse Universitaire de France.

Äijälä C. (2020) Using Film-Trope Connections for Clustering Similar Movies. Technical Report.

Garnham A. (1983) What's wrong with story grammars. Cognition 15(1):145–154.

Guarneri A., Ripamonti L.A., Tissoni F., Trubian M., Maggiorini D., Gadia D. (2017) GHOST: a GHOst STory-writer. In: Proceedings of Italian SIGCHI Chapter.

Barthes R, Balzac H de (1970) SZ. Éditions du Seuil, Paris.

Bougiatiotis K, Giannakopoulos T (2016) Content Representation and Similarity of Movies Based on Topic Extraction from Subtitles. In: Proceedings of the 9th Hellenic Conference on Artificial Intelligence.ACM.

Mellina C., Svetlichnaya S., (2011) Trope Propagation in the Cultural Space. Technical Report. Stanford University

Colby B (2009) A Partial Grammar of Eskimo Folktales. American Anthropologist 75:645–662.

Rumelhart D.E., (1975) Notes on a schema for stories. In Representation and Understanding. Morgan Kaufmann, pp 211–236.

Eger M., Mathewson K., (2018) dAIrector: Automatic Story Beat Generation through Knowledge Synthesis. ArXiv preprint:1811.03423

Lakoff G., (1972) Structural Complexity in Fairy Tales. In: The Study of Man, Vol 1.

Gala D., Khursheed M.O., Lerner H., O'Connor B., Iyyer M., (2020) Analyzing Gender Bias within Narrative Tropes.

In: Proceedings of the Fourth Workshop on Natural Language Processing and Computational Social Science. Online, pp 212–217.

García-Ortega R., García-Sánchez P., Guervós J. (2020a) Tropes in films: an initial analysis, ArXiv preprint: 2006.05380.

García-Ortega R., García-Sánchez P., Merelo-Guervós J.J. (2020b) StarTroper, a film trope rating optimizer using machine learning and evolutionary algorithms. Expert Systems 37(6):e12525.

García-Ortega R., García-Sánchez P., Merelo-Guervós J.J., Pitaru G., (2018) Overview of PicTropes, a film trope dataset. ArXiv preprint: 1809.10959

Harris J., (2017) The Periodic Table of Storytelling a visualisation of story tropes. https://jamesharris.design/periodic/.

Black J.B., Wilensky R., (1979) An evaluation of story grammars. Cognitive Science 3(3):213–229.

Kunegis J., Luca E. de, Albayrak S., (2010) The Link Prediction Problem in Bipartite Networks. ArXiv preprint 1006.5367

Kiesel M., Grimmes G. A., (2010) DBTropes—a linked data wrapper approach incorporating community feedback. CEUR Workshop Proceedings 674.

Meeks E., (2011) TVTropes. https://dhs.stanford.edu/social-media-literacy/tvtropes-pt-1-the-weird-geometry-of-the-internet/.

Propp V., (1968) Morphology of the folktale. University of Texas Press, Austin.

Barthes R., (1966) Introduction à l'analyse structurale des récits. Communications.

Shi Y., Larson M., Hanjalic A. (2014) Collaborative Filtering beyond the User-Item Matrix: A Survey of the State of the Art and Future Challenges. ACM Comput. Surv. 47(1).

Smith J.R., Joshi D., Huet B., Hsu W., Cota J. (2017) Harnessing A.I. for Augmenting Creativity: Application to Movie Trailer Creation. In: Proceedings of the 25th ACM Multimedia, pp 1799–1808.

Thompson M., (2018) Building Abstractable Story Components with Institutions and Tropes, University of Bath.

Assogba Y., Ros I., Vallandingham J., Stereotropes. http://stereotropes.bocoup.com/.

Chou J.P., Christie M. (2021). Designing Film Signatures by Using Tropes, ArXiv preprint 3821242.