Movienet: a movie multilayer network model using visual and textual semantic cues

Youssef Mourchid, Benjamin Renoust, Olivier Roupin, Lê Văn, Hocine Cherifi, Mohammed El Hassouni

Mourchid *et al. Applied Network Science* (2019) 4:121
https://doi.org/10.1007/s41109-019-0226-0

Applied Network Science

RESEARCH Open Access

Movienet: a movie multilayer network model using visual and textual semantic cues

Youssef Mourchid[1,3*], Benjamin Renoust[2], Olivier Roupin[2], Lê Văn[2], Hocine Cherifi[3] and Mohammed El Hassouni[1]

*Correspondence:
youssefmour@gmail.com
[1]LRIT URAC 29, Mohammed V University, Rabat, Morocco
[3]LIB, University of Burgundy, Dijon, France
Full list of author information is available at the end of the article

Abstract

Discovering content and stories in movies is one of the most important concepts in multimedia content research studies. Network models have proven to be an efficient choice for this purpose. When an audience watches a movie, they usually compare the characters and the relationships between them. For this reason, most of the models developed so far are based on social networks analysis. They focus essentially on the characters at play. By analyzing characters interactions, we can obtain a broad picture of the narration's content. Other works have proposed to exploit semantic elements such as scenes, dialogues, *etc*.. However, they are always captured from a single facet. Motivated by these limitations, we introduce in this work a multilayer network model to capture the narration of a movie based on its script, its subtitles, and the movie content. After introducing the model and the extraction process from the raw data, we perform a comparative analysis of the whole 6-movie cycle of the Star Wars saga. Results demonstrate the effectiveness of the proposed framework for video content representation and analysis.

Keywords: Multilayer network, Movie analysis, Movie script, Subtitles, Multimedia analysis

Introduction

Since ancient times, humans have been telling stories, putting on scene different characters in their own rich world. Each story forms a small universe, sometimes intertwining with one another. The creation of a story is a careful recipe that brings together characters, location, and other elements so that it catches a reader, a viewer, or a listener's full attention. To collect these stories, books present and structure these elements such that any reader would assemble them in their mind, building their own vision of the story.

Movies follow the same narrative principles, but stimulate viewers differently by providing a fully constructed visual world that is the product of movie director's and its team's vision. Viewers' perception can be manipulated, motivating in them the elicitation of different emotions, and their progression into some unknown universe, such as it is done is science-fiction movies. The articulation of the story elements can be the hallmarks of a director's fingerprint, characterizing genre and stories or even movie rating prediction.

Network modelling puts into relation different entities, therefore it has naturally become a powerful tool to capture the elements articulation in stories (Rital et al. 2005; Park et al. 2012; Waumans et al. 2015; Tan et al. 2014; Renoust et al. 2016; Renoust et

al. 2016; Mish 2016; Mourchid et al. 2018; Viard and Fournier-S'niehotta 2018; Markovič et al. 2018). Such network models have been applied to many different types of stories, starting with written stories in books (Waumans et al. 2015; Markovič et al. 2018), in news events from news papers and TV (Renoust et al. 2016), in television series (Tan et al. 2014), and eventually in the target medium of this paper: movies (Park et al. 2012; Mourchid et al. 2018). The topology and structure of these networks have been investigated both visually (Renoust et al. 2016; Renoust et al. 2016) and analytically (Waumans et al. 2015; Rital et al. 2005), and may in turn be used for prediction tasks (Viard and Fournier-S'niehotta 2018). These narrative networks built from large scale archives can be automatically created (Waumans et al. 2015; Renoust et al. 2016; Renoust et al. 2016) or use manual annotations (Mish 2016).

Social network analysis is one main focus of video network analysis, so naturally most of the related works put into relation characters at play in a story. But this only reveals one part of the story. In order to investigate an event, journalists use the 5 W-questions (Chen et al. 2009; Kipling 1909; Kurzhals et al. 2016) (which are *Who?*, *What?*, *When?*, *Where?* and *How/Why?*). Answering the most complex question *How/Why?* is the whole focus of analytics at large, often done through the articulation of the other four questions. Social network analysis then mostly focuses on *Who?* and puts it in perspective with other questions such as time (*When?*) for dynamic social networks (Sekara et al. 2016), or with semantics (*What?*) in content analysis (Park et al. 2012; Renoust et al. 2014), location (*Where?*) with additional sensor networks (Bao et al. 2015), and even the multiple combinations of those (*i.e.* streamgraphs) (Latapy et al. 2018; Viard and Fournier-S'niehotta 2018). Our goal is to provide a more holistic analysis over the different story elements by using a multilayer network modeling.

The recommended process of movie creation starts with the writing of the script, which is a text that is usually structured. A movie script assembles all movie elements in a temporal fashion (scenes, dialogues) and highlights specific information such as characters and setting details, so that it supports automatic movie analysis (Jhala 2008; Mourchid et al. 2018). In recent years, image analysis tools have tremendously enhanced our automatic understanding of image content (Guo et al. 2016), and although tasks such as picture localization remain challenging (Demirkesen and Cherifi 2008; Pastrana-Vidal et al. 2006), we may enrich textual approaches with face detection and recognition (Jiang and Learned-Miller 2017; Cao et al. 2018) or with scene description (Johnson et al. 2016; Yang et al. 2017).

In our previous work (Mourchid et al. 2018), we introduced a network analysis that deploys across *Who?*, *What?* and *Where?* extracted from the textual cues contained in the script, articulated around *When?* as the script unfolds. We capture these by proposing a multilayer network model that describes the structure of a movie in a richer way as compared to regular networks. It enriches the single character network analysis, and allows to use new topological analysis tools (Domenico et al. 2014).

In this paper, we extend this approach into multiple direction.

- We extend the original model based only on the script information in order to exploit the multimedia nature of information. It integrates, now, information contained in the movie (through shot segmentation, dense captioning, and face analysis) and in the subtitles.

- We additionally root the model on the multilayer network formalism proposed by Kivelä (Kivelä et al. 2014), to articulate characters, places, and themes across modalities (text and image).
- From single movies, we extend our model analysis to the first six movies of the Star Wars saga.

After discussing the related work in the next section, we introduce the proposed model called Movienet in "Modeling stories with Movienet" section. We describe how we extract the multilayer network in "Extracting the multilayer network" section, before deploying the analysis in "Network analysis" section on the Star Wars saga (Lucas 1977; 1980; 1983; 1999; 2002; 2005). We finally conclude in "Conclusion" section.

Related work

Network-based analysis of stories is widely spread, first for topical analysis (Kadushin 2012; Renoust et al. 2014). But when applied to multimedia data and movies, the analysis first focused on scene graphs (Yeung et al. 1996; Jung et al. 2004; EAC et al. 2019) for their potential for summarization. Character networks then became a natural focus for story analysis which from literature (Knuth 1993; Waumans et al. 2015; Chen et al. 2019) expanded to multimedia content (Weng et al. 2009; Tan et al. 2014; Tran and Jung 2015; Renoust et al. 2016; Mish 2016; He et al. 2018). Particular attention has been paid to dialogue structure (Park et al. 2012; Gorinski and Lapata 2018), which leads to an extension of network modeling to multilayer models (Lv et al. 2018; Ren et al. 2018; Mourchid et al. 2018).

Scene graphs: Some studies have proposed graphs based on scenes segmentation and scenes detection methods to analyze movie stories. Yeung et al. (1996) proposed an analysis method using a graph of shot transitions for movie browsing and navigation, to extract the story units of scenes. Edilson et al. (2019) extends this approach by constructing a narrative structure to documents. They connect a network of sentences based on their semantic similarity, which can be employed to characterize and classify texts. Jung et al. (2004) use a narrative structure graph of scenes for movie summarization, where scenes are connected by editorial relations. Story elements such as major characters and their interactions cannot be retrieved from these networks. Our work contrasts in using additional sources (scripts, subtitles, etc).

Character networks in stories: Character network analysis is a traditional exercise of social network analysis, with the network from *Les Misérables* now being a classic of the discipline (Knuth 1993), and still inspires current research. Waumans et al. (2015) create social networks from the dialogues of the *Harry Potter* series, including sentiment analysis and generating multiple kind of networks, with the goal of defining a story signature based on the topological analysis of its networks. Chen et al. (2019) propose an integrated approach to investigating the social network of literary characters based on their activity patterns in the novel. They use the minimum span clustering (MSC) algorithm for the identification of the character network's community structure, visualizing the community structure of the character networks, as well as to calculate centrality measures for individual characters.

Co-appearance social networks: Co-appearance networks, connecting when co-appearing characters on screen, have been an important subject of research, even reaching the characters of the popular series Game of Thrones (Mish 2016). *RoleNet*

(Weng et al. 2009) identifies automatic leading roles and corresponding communities in movies through a social network analysis approach to analyze movie stories. He et al. (2018) extend co-appearance network construction with a spatio-temporal notion. They analyze social centrality and community structure of the network based on human-based ground truth. Tan et al. (2014) analyze the topology of character networks in TV series based on their scene co-occurrence in scripts. *CoCharNet* (Tran and Jung 2015) uses manually annotated co-appearance social network on the six Star Wars movies, and propose a centrality analysis. Renoust et al. (2016) propose an automatic political social network construction from face detection and tracking data in news broadcast. The network topology and importance of nodes (politicians) is then compared across different time windows to provide political insights. Our work is very inspired by these co-appearance social networks, which give an interesting insight for the roles of characters, but they are still insufficient to fully place the characters in a story, which is why we rely on additional semantic cues.

Dialogue-based social networks: Social networks derived from dialogue interaction in movie scripts have been used for different purposes. *Character-net* (Park et al. 2012) proposes a story-based movie analysis method via social network analysis using movie script. They construct a weighted network of characters from dialogue exchanges in order to rank their role importance. Based on a corpus of movie scripts, Gorinski et al. (2018) proposed an end-to-end machine learning model for movie overview generation, that uses graph-based features extracted from character-dialogue networks built from movie scripts.

Similar to co-appearance networks, these approaches only use a social network for video analysis based on dialogue interaction, which cannot provide a socio-semantic construct of the video narration content. Having a different purpose, the proposed model gives a W-question based semantic overview of the movie story, tapping into the very multimedia nature of movies.

Multilayer network approaches: Recent approaches use multiplex networks to combine both visual and textual semantic cues. StoryRoleNet (Lv et al. 2018) is not properly a multilayer approach, but it well displays the interest of multimodal combination. It provides an automatic character interaction network construction and story segmentation by combining both visual and subtitle features. In the *Visual Clouds* (Ren et al. 2018) networks extracted from TV news videos are used as a backbone support for interactive search refinement on heterogeneous data. However, layers cannot be investigated individually. In a previous work (Mourchid et al. 2018), we introduced a multilayer model to describe the content of a movie based on the movie script content. Keywords, locations, and characters are extracted from the textual information to form the multilayer network. This paper builds on this work by further exploiting additional medium sources, such as subtitles and the image content of the video to enrich the model and to refine the multilayer extraction process. The proposed model is fully multimedia, as it takes into account text-based semantic extraction, and image-based semantic cues from face recognition and scenes captioning, in order to capture a richer structure for the movies.

Modeling stories with Movienet

To describe a complete story, four fundamental questions are investigated (*Who?*, *Where?*, *What?*, *When?* often refered as the four *Ws*) (Flint 1917; Kipling 1909). Inferring

How/Why? can be done while articulating the other *Ws* making them essential bricks of analysis:

Given our context of movie understanding, we may reformulate the four *Ws* as follows:

- *Who?* denotes **characters** and *people appearing* in a movie;
- *Where?* denotes **locations** where actions of a movie take a place;
- *What?* denotes **subjects** which the movie talks about and *other elements that describes* a movie scene.
- *When?* denotes the **time** that guide the succession of events in the movie.

Answering these questions form the entities *characters* (mentioned in the script), *locations* (as depicted by the script), *keywords* (conversation subjects understood from dialogues), *faces* (as people *appear* on screen), and *captions* (that describe a scene) – which ground our study. *Time* is a special case to infer connections, but we do not treat it as an entity in our model.

Our goal is to help formulate movie understanding by articulating these four *Ws*. In a preliminary work, we exploited the information contained in the movie script in order to construct a multilayer network. However, we neglected the complementary information contained in the movie and the subtitle. Using both visual and textual information allows a better understanding of the content and therefore a richer representation.

We propose a multilayer graph model that complete the previous model formulation (Mourchid et al. 2018) by exploiting two additional layers, *faces* and *captions*. The multilayer graph puts these elements together as they form a story by exploiting two new sources that are subtitles and the video content. This model is made of five layers in order to represent each type of entity *characters*, *keywords*, *locations*, *faces*, and *captions*, with multiple relationships between them.

Following Kivelä's definition (Kivelä et al. 2014) of multilayer networks, we model two main classes of relationships: intra-layer relationships, between nodes of a same category, such as two faces appearing in the same scene; and inter-layer relationships which capture the interactions between nodes of different categories, such as when a caption describes a scene where a character is present. Altogether, the multiple families of nodes and edges form a multilayer graph as illustrated in Fig. 1.

We now define our multilayer graph $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ such that:

- $V_C \subseteq \mathbb{V}$ represents the set of characters $c \in V_C$,
- $V_L \subseteq \mathbb{V}$ represents the set of locations $l \in V_L$,

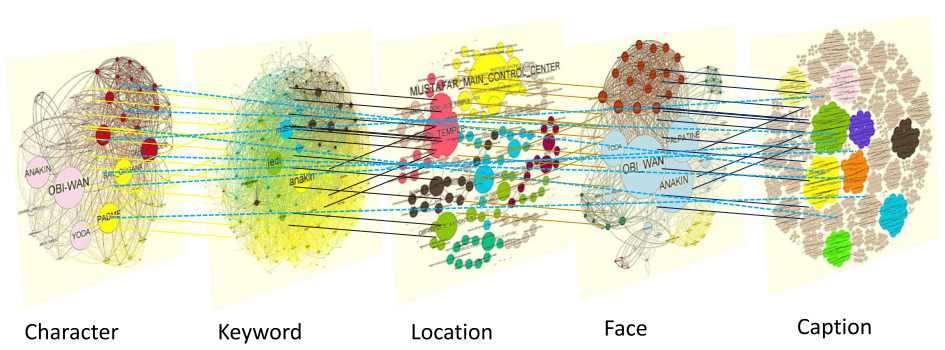


Fig. 1 A conceptual presentation of our multilayer network model: Five main layers of nodes, *Character* G_{CC}, *Keyword* G_{KK}, *Location* G_{LL}, *Face* G_{FF} and *Caption* G_{CaCa} are interacting within and across each layer

- $V_K \subseteq \mathbb{V}$ represents the set of keywords $k \in V_K$.
- $V_F \subseteq \mathbb{V}$ represents the set of faces $f \in V_F$.
- $V_{Ca} \subseteq \mathbb{V}$ represents the set of captions $ca \in V_{Ca}$.

The different families of relationships can then be defined as:
Intra-layer:

- $e \in E_{CC} \subseteq \mathbb{E}$ between two characters such that $e = (c_i, c_j) \in V_C^2$, when a character $c_i \in V_C$ is conversing with another character $c_j \in V_C$.
- $e \in E_{LL} \subseteq \mathbb{E}$ between two locations such that $e = (l_i, l_j) \in V_L^2$, when there is a temporal transition from one location $l_i \in V_L$ to the other $l_j \in V_L$.
- $e \in E_{KK} \subseteq \mathbb{E}$ between two keywords such that $e = (k_i, k_j) \in V_K^2$, when $k_i \in V_K$ and $k_j \in V_K$ belong to the same subject.
- $e \in E_{FF} \subseteq \mathbb{E}$ between two faces such that $e = (f_i, f_j) \in V_F^2$, when $f_i \in V_F$ and $f_j \in V_F$ appear in the same scene.
- $e \in E_{CaCa} \subseteq \mathbb{E}$ between two captions such that $e = (ca_i, ca_j) \in V_{Ca}^2$, when $ca_i \in V_{Ca}$ and $ca_j \in V_{Ca}$ describe the same scene.

Inter-layer:

- $e \in E_{CK} \subseteq \mathbb{E}$ between a character and a keyword such that $e = (c_i, k_j) \in V_C \times V_K$, when the keyword $k_j \in V_K$ is pronounced by the character $c_i \in V_C$.
- $e \in E_{CL} \subseteq \mathbb{E}$ between a character and a location such that $e = (c_i, l_j) \in V_C \times V_L$, when a character $c_i \in V_C$ is present in location $l_j \in V_L$.
- $e \in E_{CF} \subseteq \mathbb{E}$ between a character and a face such that $e = (c_i, f_j) \in V_C \times V_F$, when a character $c_i \in V_C$ appears in the same scene of $f_j \in V_F$.
- $e \in E_{CCa} \subseteq \mathbb{E}$ between a character and a caption such that $e = (c_i, ca_j) \in V_C \times V_{Ca}$, when a character $c_i \in V_C$ appears in the same scene which $ca_j \in V_{Ca}$ describes.
- $e \in E_{KL} \subseteq \mathbb{E}$ between a keyword and a location such that $e = (k_i, l_j) \in V_K \times V_L$, when a keyword $k_i \in V_K$ is mentioned in a conversation taking place in the location $l_j \in V_L$.
- $e \in E_{KF} \subseteq \mathbb{E}$ between a keyword and a face such that $e = (k_i, f_j) \in V_K \times V_F$, when a keyword $k_i \in V_K$ is mentioned in a scene where $f_j \in V_F$ appears.
- $e \in E_{KCa} \subseteq \mathbb{E}$ between a keyword and a caption such that $e = (k_i, ca_j) \in V_K \times V_{Ca}$, when a keyword $k_i \in V_K$ is mentioned in a scene which $ca_j \in V_{Ca}$ describes.
- $e \in E_{LF} \subseteq \mathbb{E}$ between a location and a face such that $e = (l_i, f_j) \in V_L \times V_F$, when a face $f_j \in V_F$ appears in the same scene which contains the location $l_i \in V_L$.
- $e \in E_{LCa} \subseteq \mathbb{E}$ between a location and a caption such that $e = (l_i, ca_j) \in V_L \times V_{Ca}$, when a caption $ca_j \in V_{Ca}$ describe a scene that contains the location $l_i \in V_L$.
- $e \in E_{FCa} \subseteq \mathbb{E}$ between a face and a caption such that $e = (f_i, ca_j) \in V_F \times V_{Ca}$, when a face $f_i \in V_F$ appears in the same scene that $ca_j \in V_{Ca}$ describes.

Edge direction and weight are not considered for the sake of simplicity. Moreover, as we do not intend to study the network dynamics, time is not directly taken into account. However, time supports everything: the existence of a node or an edge is defined upon time, unrolled by the order of movie scenes.

As a shortcut, we can now refer to subgraphs by only considering one layer of links and its induced subgraph:

- $G_{CC} = (V_C, E_{CC}) \subseteq \mathbb{G}$ refers to the subgraph of character interaction;
- $G_{KK} = (V_K, E_{KK}) \subseteq \mathbb{G}$ refers to the subgraph of keyword co-occurrence;
- $G_{LL} = (V_L, E_{LL}) \subseteq \mathbb{G}$ refers to the subgraph of location transitions;
- $G_{FF} = (V_F, E_{FF}) \subseteq \mathbb{G}$ refers to the subgraph of face interaction;
- $G_{CaCa} = (V_{CaCa}, E_{CaCa}) \subseteq \mathbb{G}$ refers to the subgraph of caption co-occurrence;
- $G_{CK} = (V_C \cup V_K, E_{CK}) \subseteq \mathbb{G}$ refers to the subgraph of characters speaking keywords;
- $G_{CL} = (V_C \cup V_L, E_{CL}) \subseteq \mathbb{G}$ refers to the subgraph of characters standing at locations;
- $G_{CF} = (V_C \cup V_F, E_{CF}) \subseteq \mathbb{G}$ refers to the subgraph of characters appearing with faces;
- $G_{CCa} = (V_C \cup V_{Ca}, E_{CCa}) \subseteq \mathbb{G}$ refers to the subgraph of characters described by captions;
- $G_{KL} = (V_K \cup V_L, E_{KL}) \subseteq \mathbb{G}$ refers to the subgraph of keywords mentioned at locations.
- $G_{KF} = (V_K \cup V_F, E_{KF}) \subseteq \mathbb{G}$ refers to the subgraph of keywords said by faces.
- $G_{KCa} = (V_K \cup V_{Ca}, E_{KCa}) \subseteq \mathbb{G}$ refers to the subgraph of keyword said at the same scene which caption describe.
- $G_{LF} = (V_L \cup V_F, E_{LF}) \subseteq \mathbb{G}$ refers to the subgraph of faces appearing at locations.
- $G_{LCa} = (V_L \cup V_{Ca}, E_{LCa}) \subseteq \mathbb{G}$ refers to the subgraph of captions describing locations.
- $G_{FCa} = (V_F \cup V_{Ca}, E_{FCa}) \subseteq \mathbb{G}$ refers to the subgraph of captions describing faces.

Now that we have set the model, we need to extract elements from scripts, subtitles, and movie clips. This allows for the analysis of various topological properties of the network in order to gain a better understanding of the story.

Extracting the multilayer network

We now describe the data and methodology used to build the multilayer network of a movie. Figure 2 illustrates the methodology processing pipeline. Very much inspired by the work from Kurzahls et al. (2016), we align scripts, subtitles and video, from which we extract different entities. After introducing the extraction of the various entities and interactions from each data source, we explain how to build the network based on this information.

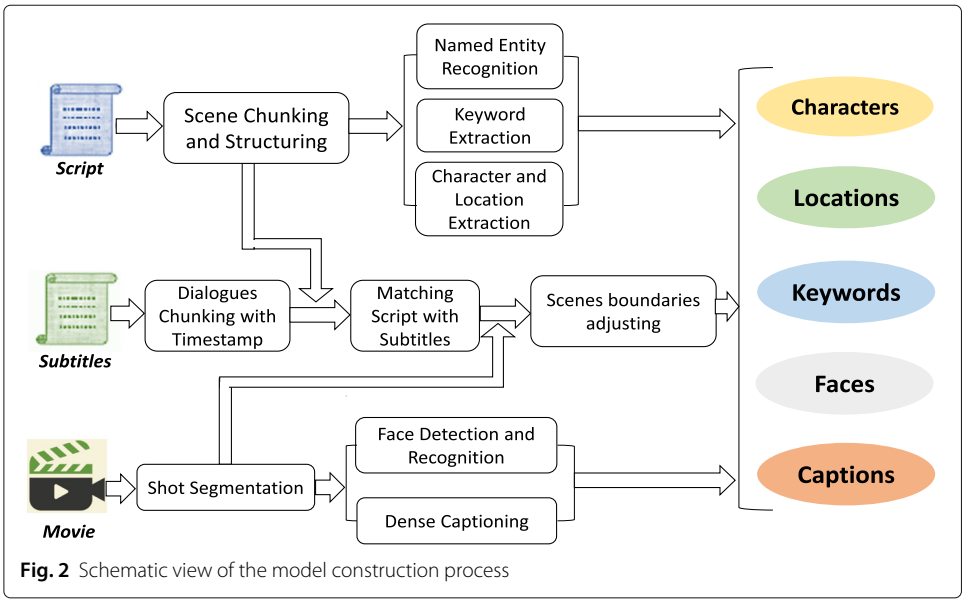


Fig. 2 Schematic view of the model construction process

Data description

Three data sources are used for this task: script, subtitles and video.

Definitions

In order to remove any ambiguity, we first define the following dedicated glossary.

- Script: A text source of the movie which has descriptions about scenes, with setting and dialogues.
- Scene: Chunk of a script, temporal unit of the movie. The collection of all scenes form the movie script.
- Shots: Continuous (uncut) piece of video, a scene is composed of a series of shots.
- Setting: The location a scene takes place in, and its description.
- Character: Denotes a person/animal/creature who is present in a scene, often impersonated by an actor.
- Dialogues: A collection of utterances, what all characters say during a scene.
- Utterance: An uninterrupted block of a dialogue pronounced by one character.
- Conversation: A continuous series of utterances between two characters.
- Speaker: A character who pronounced an utterance.
- Description: A script block which describes the setting.
- Location: Where a scene takes place, or mentioned by a character.
- Keyword: Most relevant information from an utterance, often representative of its topic.
- Time: the time information extracted by aligning the script and subtitles.
- Subtitles: a collection of blocks which have a time information.
- Subtitles block: a block of the collection of utterance that has a start and end time.
- Keyframe: a keyframe is a picture extracted from the movie. Keyframes are extracted at regular intervals (every second) to ease image processing.
- Face: a character's face detected in a keyframe, associated to an image bounding box.
- Caption: a descriptive sentence detected in a keyframe, associated to an image bounding box.

Script

Scripts happen to be very well-structured textual documents (Jhala 2008). A script is composed of many scenes, each scene contains a location, scene description, characters and their dialogues. The actual content of a script often follows a semi-regular format (Jhala 2008) such as depicted in Fig. 3. It usually starts with a heading describing the location and time of the scene. Specific keywords give important setting information (such as inside or outside scene) and character and key objects are often emphasized. The script then follows in a series of dialogues and setting descriptions.

Subtitles

Subtitles are available in a SubRip Text (SRT) format and consist of four basic information (Fig. 4): (1) a number to identify the order of the subtitles; (2) the beginning and ending time (hours, minutes, seconds, milliseconds) in which the subtitle should appear in the movie; (3) the subtitle text itself on one or more lines and (4) typically an empty line to indicate the end of the subtitle block. However, subtitles do not include information about characters, scenes, shots, and actions whereas dialogues in a script do not include time information.

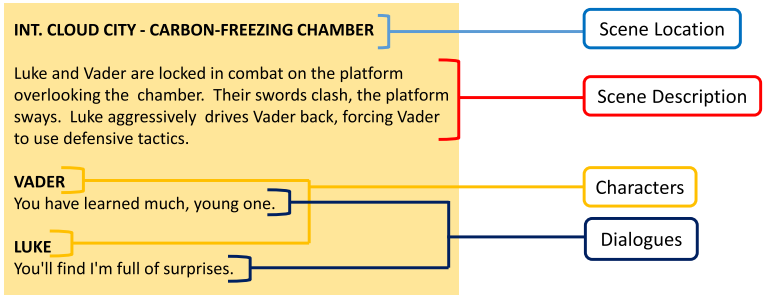


Fig. 3 Snippets of a resource script describing the movie *The Empire Strikes Back*, displaying different elements manipulated (characters, dialogues and locations)

Video

A movie's video can be divided into two components: a soundtrack (that we do not approach in this work) and a collection of images (the motion is then implied from the succession of these images). A movie is composed of scenes which are decomposed in shots. Scenes make up the actual unit of action which composes the movie. Each scene provides visual information about characters, locations, events, *etc.*

Script processing

We now describe each step of the script processing pipeline. This process is language dependent, so we restrict our study to English scripts only. However, note that the framework can be easily adapted to other languages.

Scene chunking and structuring

As we mentioned above, scenes are the main subdivisions of a movie, and consequently our main unit of analysis. During a scene, all the critical elements of a movie (all previously defined entities) interact. Each scene contains information about characters who talk, location where the scene takes place, and actions that occur. Our first goal is then to identify those scenes.

Fortunately scripts are structured and give away this information. We then need to *chunk* the script into scenes. In a script, a scene is composed as follows. First, there is a technical description line written in capital letters for each scene. It establishes the

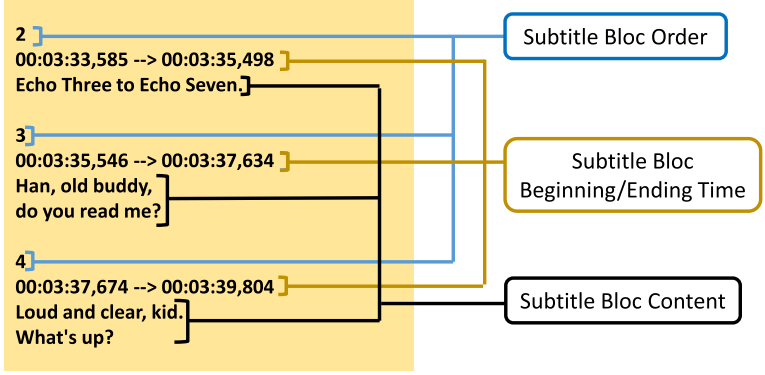


Fig. 4 Snippets of a resource file from *The Empire Strikes Back* subtitles

physical context of the action that follows. The rest of a scene is made of dialogue and description. Each scene starts by a set information, *INT* or *EXT*, which indicates whether a scene takes place inside or outside, the name of the location, and also the time of day (e.g. *DAY* or *NIGHT*).

Within a scene heading description, important people and key objects are usually highlighted in capital letters that we may harvest while analyzing the text. Character names and their actions are always depicted before the actual dialogue lines. A line indent also helps to identify characters and dialogue parts in contrast to scene description. We can harvest scene locations and utterance speakers, by structuring each scene into its set of descriptions and dialogues. Finally, we identify conversations and characters present at a scene. Specific descriptions can then be associated to locations, and dialogues to characters. After chunking, we then obtain a scene structured into the following elements (as illustrated in Fig. 3): a scene location, a description block, and a series of dialogues blocks assigned to characters.

Semantic extraction

The next step is to identify the actual text content that is attributed to locations or to speakers. Fortunately, Named Entity Recognition (NER) (Nadeau and Sekine 2007) is a tool of natural language processing that labels significant words extracted from a text content with categories such as *organizations*, *people*, *locations*, *cities*, *quantities*, *ordinals*, *etc.* We apply NER to each scene description block and discard the irrelevant categories. However, this process is not perfect and many words can end up mislabelled due to the ambiguous context of the movie, especially within the science-fiction genre. In a second pass, we manually curate the resulting list of words and assign them to our fundamental categories: *characters*, *locations*, and *keywords*.

Because ambiguity also includes polymorphism of semantic concepts, we next assign a unique class for synonyms referring to the same concept (*i.e.* $\{LUKE, SKYWALKER\} \rightarrow LUKE$). NER also helps us identifying characters present at a scene who are mentioned in utterances. Many public libraries are available for NER, and we used the spaCy library (Al Omran and Treude 2017) because of its efficiency in our context.

We may now identify *keywords* within dialogues. We investigated three methods to measure the relevance of keywords: *TF-IDF* (Salton et al. 1975; Li et al. 2007), LDA (Blei et al. 2003) and Word2Vec (Yuepeng et al. 2015). Because dialogue texts are made of short sentences (even shorter after stop-words removal), empirical results of Word2Vec and *TF-IDF* rendered either too few words with a high semantic content, or too much words without semantic content. Only LDA, brought the best trade-off, but still included some level of noisy semantic-less words. We manually curated the resulting words by removing the remaining noise (such as *can*, *have*, and so on).

Video processing

Since video information also allows for answering a few of the *W* questions, we introduce two techniques in this paper borrowed from computer vision: face detection and recognition to address *Who*, and dense captioning to address *What*. These are computationally intensive processes, so we first apply a rough shot detection using the PySceneDetect tool (Castellano 2012), then extract for each shot only one keyframe every second, which should maintain a good granularity to match with scenes. This renders an average of $\sim$8k key-frames per movie. Key-frames can then be analyzed in parallel.

Face detection and recognition

Before knowing who appears in a scene, we need to detect if there is a face or not. This is the task of face detection applied in each frame. To extract those faces, we deployed a state-of-the-art face detector based on the faster R-CNN architecture (Jiang and Learned-Miller 2017) that is trained with WIDER (Yang et al. 2016). This algorithm proposes bounding boxes for each detected face (in average obtaining ∼5k detected faces per movie). We then manually remove all false positive detections (around 6.5% in average).

We now need to identify who the faces belong to. We also wish to match the faces that belong to the same people. For each of the valid faces we use another state-of-the-art embedding technique, the ResNet50 architecture (He et al. 2016) trained on the VGGFace2 dataset (Cao et al. 2018). This allows us to obtain a 2048 dimensional vector that corresponds to each detected face. Traditional retrieval approaches are challenged because of the specific characteristics of our dataset (pairwise distances are very close within a shot *and* very far between shots, in addition to other motion blur and lighting effects). Since the number of detected faces is limited for each movie, we only use automated approaches to assist manual annotation. We project the vector space in 2D using *t*-SNE (Gisbrecht et al. 2015) and manually extract obvious clusters within the visualization framework Tulip (Auber et al. 2017). In order to quick-start the cluster creation, we applied a DBScan clustering (Ester et al. 1996), for which we fine tuned parameters on our first manually annotated dataset, reaching a rough 17% accuracy. Based on the detected clusters, and on the movie distribution, we then create face models as collections of pictures to incrementally help retrieving new pictures of the same characters. With the results still containing many errors, we finally manually curated them all to obtain a clean recognition for each character.

Dense captioning

One could wish also to explore what objects and relations could be inferred from the scenes themselves. The dense captioning task (Johnson et al. 2016) attempts to use tools of computer vision and machine learning to describe textually the content of an image. We used an approach with inner joints (Yang et al. 2017) trained with the Visual Genome (Krishna et al. 2017). This computes bounding boxes and sentences for each frame, accompanied with a confidence index $w \in [0, 1]$.

Depending on the rhythm of the movie, frame extraction may still result in very similar consecutive frames. As a consequence, dense captioning of these consecutive frames may be very similar. However, the similar captions may be assigned very different confidence index. In order to extract the most relevant captions in this context, we propose to use this confidence index to rank then filter captions.

We extend the *TF-IDF* definition (Salton et al. 1975) $tfidf = tf * idf$ to one incorporating caption confidence index. The notion of document here corresponds to a scene, and instead of a term, we have a caption. We define $tf(ca_i, s)$ the weighted frequency of caption ca_i in a scene s as follows:

$$tf(ca_i, s) = \frac{\sum_{fr \in s} w_{ca_i, fr}}{\sum_{fr \in s} \sum_{ca \in f} w_{ca, fr}}$$

where ca denotes a caption having a confidence index $w_{ca,fr}$ in a frame fr of a scene s. We then define $idf(ca_i, S)$ the inverse scene frequency such as:

$$idf(ca_i, S) = log\left(\frac{|S|}{|\{s \in S : ca_i \in s\}|}\right)$$

with $\{s \in S : ca_i \in s\}$ denoting the scenes s which contain the caption ca_i in the corpus made of all the scenes in the movie S.

We keep the top 40 captions per scene. Captions are simple sentences, such as *"a white truck parked on the street"*, and their generation process make them resemble a lot one another (due to the limitations of the training vocabulary and relationships). To further extract their semantic content, we compute their n-grams (Cavnar and Trenkle 1994) ($n = 4$, keeping a maximum of one stop word in the n-gram).

Each resulting n-gram is then represented by a bag of unique words that we sort in order to cover permutations and help matching between scenes. The piece of sentences formed may then be used as an additional keyword layer obtained from the visual description of the scene,

Time alignment between script and subtitles

We now need to match the semantic information extracted from the script to the one extracted from the video. This can naturally be done by aligning the script with the time of the movie. The movie is played along time, but the script has no time information. Fortunately dialogues are reported in the script, and they correspond to people speaking in the movie. Subtitles are the written form of these dialogues, and they are time-coded in synchronization with the movie. The idea is to use them as a proxy to assign time-codes of matching dialogues in the script. Hence, we should have rough approximations of when scenes occur through dialogues start/end boundaries.

Unfortunately, the exact matching of scripts and dialogues greatly varies between versions of the script and movie. Sometimes a scene may appear in the script but not in the movie, and vice versa. Additionally, the order and wording may greatly differ between the two.

To deal with these issues, we proceed in multiple steps as introduced by Kurzhals et al. (2016). Scenes are decomposed in blocks, for which each is a character utterance. We then normalize the text on both sides through stemming. The idea is then to assign each of the utterance block to its corresponding counterpart in the subtitles. A first step checks for an absolute equality of subtitles and script dialogue. A second step is for textual inclusion between script and subtitles. This does not work for all utterances but the matching part gives search window constraints for our next step. For the remaining blocks, we compute their *TF-IDF* weighted vectors (Salton et al. 1975) and match with minimal cosine similarity.

Keywords and characters can then precisely be identified. But since a scene compiles a series of utterance, we get as a result a rough approximation of each scene's time boundaries, and each location too. To better align scenes and the video, we further refine the scene boundaries to those of the beginning and ending shot boundaries each scene is falling into, as shown in Fig. 5.

Many scenes however do not contain any dialogue (a battle scene which contains only a description of what's happening in it) and therefore cannot be matched to any subtitle block (these scenes are often used to better pace the narration, and may typically display an action from the outside, for example a moving vehicle). In other cases, scenes cannot be matched with subtitles when the dialogues are too small or have changed too much,

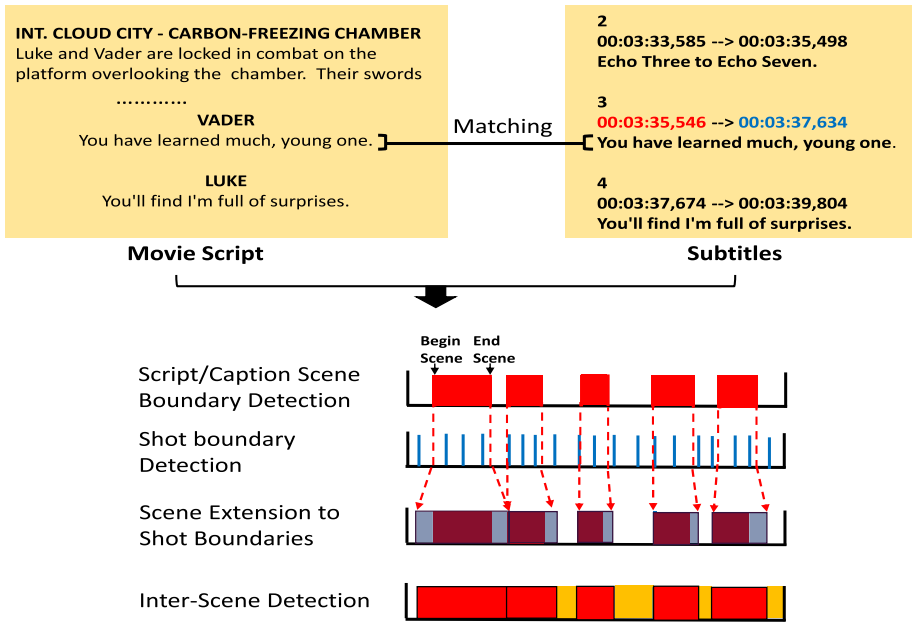


Fig. 5 Data fusion for script/subtitles alignment. First, the script is matched with the subtitles. Then, we refine the scene boundary with the beginning and ending shot boundaries

and many scenes have actually been erased from script to the final movie cut. Table 1 summarizes these statistics.

The placement of some of these scenes may still be inferred from the matching of other scenes. Indeed, a scene that has not been matched can be fitted between its two neighboring scenes if they have been matched previously. When more than one consecutive scenes cannot be matched, we create a meta scene to regroup them. For instance, if we have a gap of consecutive scenes between *Scene 1 (00:02:00–00:02:20)* and *Scene 5 (00:02:46–00:03:52)*, we create the *Meta Scene 2–4 (00:02:20–00:02:46)* which starts from the end of *Scene 1* and ends at the beginning of *Scene 5*.

Network construction

As a result of the previous steps, we now have alignment between scenes, with location, characters, and keywords, and video frames, with faces, and descriptive captions. These form the entities to build the multilayer network made of the individual layers V_L, V_C, V_K, V_F, and V_{Ca}.

Table 1 Number of scenes, matched, retrieved, and missed from the script to caption, for each episode of the Star Wars saga as a pre-processing for use cases in "Network analysis" section

Episode	# Script-caption matching scenes	# Boundary based retrieved scenes (#empty)	# Meta scenes (#empty)	# Total scenes (#empty)
SW1	109	23 (14)	51 (36)	183 (50)
SW2	58	22 (17)	70 (46)	150 (63)
SW3	75	16 (14)	97 (66)	188 (80)
SW4	223	66 (52)	193 (172)	479 (224)
SW5	146	52 (51)	77 (70)	275 (121)
SW6	89	27 (19)	22 (23)	138 (42)

Note that, in the creation process, many scenes were actually removed and changed from their original version to the final cut, explaining the amount of mismatches (empty scenes are usually scene cuts giving a rhythm to the movie

Let us revisit our investigative questions in the context of a scene: *Where does a scene take place?* is identified by the *locations*. *Who is involved in a scene?* may be tackled by *characters*, but also through the other question *Who appears in a scene?* which is identified through *faces*. *What is a scene about?* is identified through keywords, but also partly by answering *What is represented in a scene?*, tackled by captions.

We now wish to infer the relationships we described in "Modeling stories with Movienet" section. Two characters c_i, c_j can be connected when they participate in a same conversation, hence forming an edge $e_{c_i,c_j} \in E_{CC}$. We connect two locations $e_{l_i,l_j} \in E_{LL}$ when there is a temporal transition between the locations l_i and l_j (analogous to geographical proximity), *i.e.* following the succession of two scenes. Keywords k_i, k_j co-occurring in a same conversation create an edge $e_{k_i,k_j} \in E_{KK}$. If two faces f_i and f_j appear in the same scene, an edge $e_{f_i,f_j} \in E_{FF}$. Two captions ca_i and ca_j describing the same scene can also be associated by an edge $e_{ca_i,ca_j} \in E_{CaCa}$.

Using the structure extracted from the script, subtitles, and movie content, we can add additional links between categories. An edge $e_{c_i,l_j} \in E_{CL}$ associates a character c_i with a location l_j when the character c_i appears in a scene taking place at location l_j. When a character c_i speaks an utterance in a conversation, for each keyword k_j that is detected in this utterance, we create an edge $e_{c_i,k_j} \in E_{CK}$. If a character c_i is present in the same scene as the face f_j an edge $e_{c_i,f_j} \in E_{CF}$ is created between them. An edge $e_{c_i,ca_j} \in E_{CCa}$ links a character c_i with a caption ca_j if the caption describes a scene in which the character appears. We can associate the keywords k_i extracted in conversation placed in a location l_j to form the edge $e_{k_i,l_j} \in E_{KL}$. We create an edge $e_{k_i,f_j} \in E_{KF}$ between a keyword k_i and a face f_j if the keyword is mentioned in a scene where the face is present. When a keyword k_i is mentioned in a scene which the caption ca_i describes, we create an edge $e_{k_i,ca_j} \in E_{KCa}$. A link $e_{l_i,f_j} \in E_{LF}$ is created between a location l_i and a face f_j when a location is in the scene where the face appears. We associate an edge $e_{l_i,ca_j} \in E_{LCa}$ between a location l_i and a caption ca_j, if the location is in the scene that the caption describes. Finally, when a face f_i appears in a scene that the caption ca_j describes, an edge $e_{f_i,ca_j} \in E_{FCa}$ is created. A resulting graph combining all layers is visualized in Fig. 1.

Network analysis

We now wish to perform a network analysis of the whole 6-movie Star Wars saga (hereafter SW). With many people to keep track of during the six movies, it can be a challenge to fully understand their dynamics. To demystify the saga, we turn to network science. After turning every episode of the saga into a multilayer network following the proposed model, our first task is to investigate their basic topological properties. We then further investigate node *influence* as proposed by Boglio et al. (2017), on centralities that are defined for single-layer and multilayer cases: the *Influence Score* is computed by the average ranking of three centralities.

The three centrality measures we consider are defined for both single and multilayer cases (Domenico et al. 2013; Ghalmane et al. 2019a). Additionally *Degree*, *Betweenness* and *Eigenvector* centrality are among the most influential measures. *Degree* centrality measures the direct interactions of a story element. The *Betweenness* centrality measures how core to the plot a story element might be. The *Eigenvector* centrality then measures the relative influence of a story element in relation to other influential elements.

As a result, after studying influence score on separated layers, we then study it on our multilayer graphs.

Description of the data

First, a quick introduction to the SW saga: The saga began with *Episode IV – A New Hope (1977)* (Lucas 1977), which was followed by two sequels, *Episode V – The Empire Strikes Back (1980)* (Lucas 1980) and *Episode VI – Return of the Jedi (1983)* (Lucas 1983), often referred to as *the original trilogy*. Then, the prequel trilogy came, composed of *Episode I – The Phantom Menace (1999)* (Lucas 1999), *Episode II – Attack of the Clones (2002)* (Lucas 2002), and *Episode III – Revenge of the Sith (2005)* (Lucas 2005). Movies and subtitles are extracted from DVD copies, and scripts can be acquired from the Internet Movie Script Database (The Internet Movie Script Database (IMSDb)) and Simply Scripts (Simply Scripts) depending on the format.

The SW saga tells the story of a young boy (Anakin), destined to change the fate of the galaxy, who is rescued from slavery and trained by the Jedi (the light side), and groomed by the Sith (the dark side). He falls in love and marries a royalty, who fell pregnant. The death of his mother pushes him to seek revenge, so he gets coerced by the Sith. He is nearly killed by his former friend, but is saved by the Sith Emperor to ultimately stay by his side. His twin children are taken and hidden away, they grow up independently, one becomes a princess (Leia) and the other one becomes a farm hand (Luke). Luke stumbles upon a message from a princess in distress and seeks out an old Jedi who, knowing Luke's heritage, begins training him. To rescue the princess, they hire a mercenary (Han Solo) and save her. She turns out to be Luke's long lost twin sister. Discovering the identity of Luke, the emperor tries, with the help of Anakin, to turn him to the dark side. When that fails, he attempts to execute him, but Anakin, at the sight of his son's suffering, turns against the emperor saving the galaxy.

Topological properties of individual layers

Now that we have set the model, we are able to compute measures characterizing it at a macro level. To do so, we measure the basic topological properties of each layer. The number of nodes, number of edges, the network density, the diameter, the average shortest path length, the clustering coefficient and assortativity measure (degree correlation coefficient) are measured for each layer and reported in Fig. 6.

A first observation is that the character layer G_{CC} contains less nodes than the face layer G_{FF}. The number of nodes of location G_{LL} and keyword G_{KK} layers are rather stable across the movies, but the number of nodes in the caption layer G_{CaCa} is varying a lot, and looks quite different between the original and prequel series.

For all movies, the location G_{LL} layer are made of one single connected component and also for the character G_{CC} layer except for episodes IV and VI. The face layer G_{FF} has a few isolated components, related to extra characters that play no significant role in the story. From the semantic point of view, the keyword layer G_{KK} has a few isolated nodes, and the caption layer G_{CaCa} has a large number of isolated components.

Results show that the character layer G_{CC} is denser in comparison to all other layers. Indeed, we can expect much more connections among characters, since they exchange dialogues. By comparison, the face layer G_{FF} shows a much higher number of edges than the character layer, both having a very high clustering coefficient, suggesting the existence

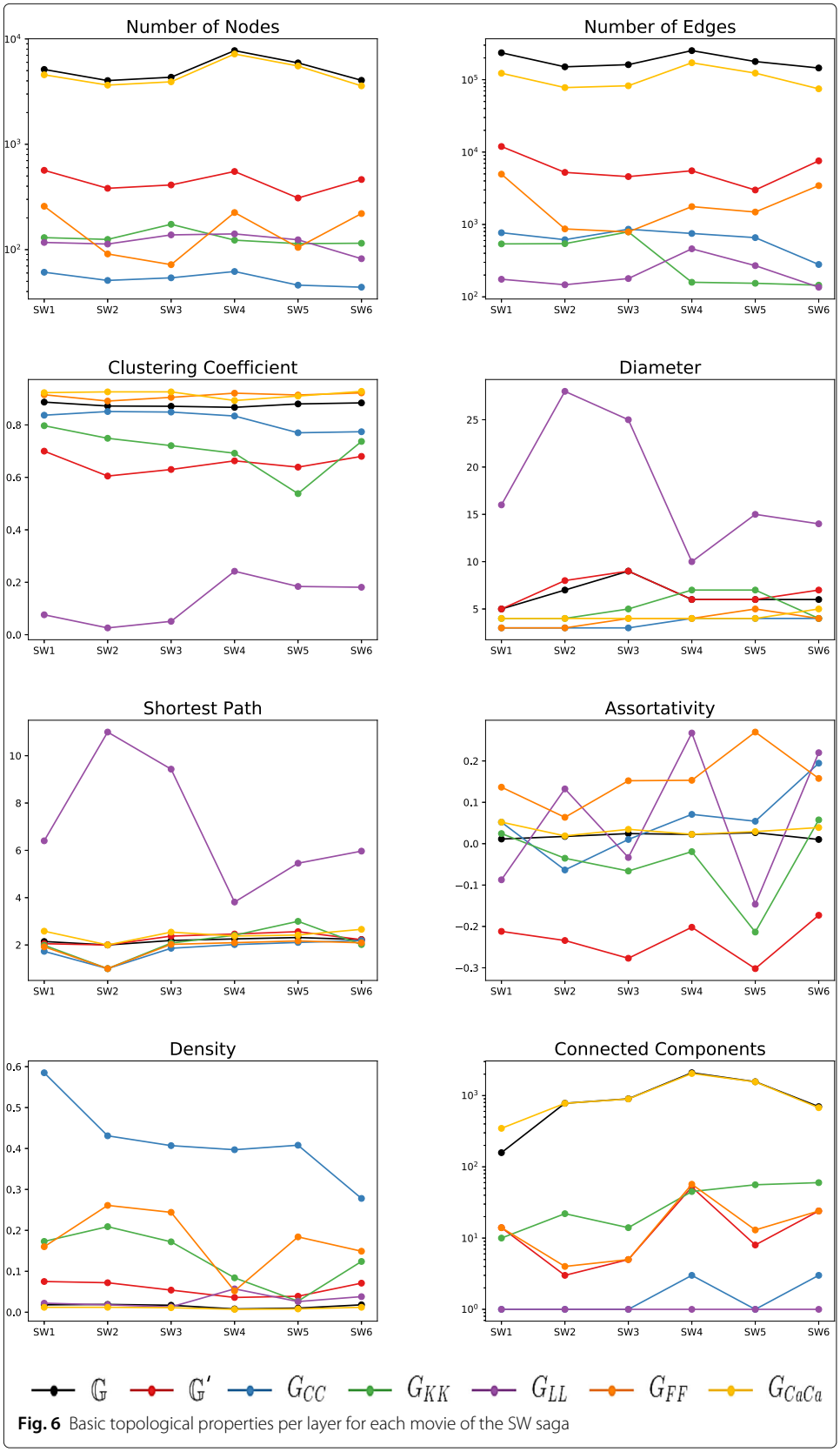


Fig. 6 Basic topological properties per layer for each movie of the SW saga

of social communities. The keyword layer G_{KK} also shows a large clustering coefficient, despite a more limited number of edges.

Location layers G_{LL} display quite a high diameter and the longest average shortest path. This is due to the limited amount of locations and very few temporal transitions between locations that introduce long paths. Only a few sets can be considered hubs. On the opposite, the caption layer G_{CaCa} shows a diameter of 4 and clustering coefficient much closer to 1, because each scene creates a clique of unique captions. The face layer G_{FF} shows the highest assortativity, as we may suspect for main and secondary characters to appear together most often, while tertiary characters (*i.e.* extras) often appear in group.

Caption layers G_{CaCa} show the largest number of nodes and edges with the lowest density. This is due to their generation and construction which creates cliques of many captions for each scenes, which are connected only later on through a few number of captions. As a consequence, captions have many connected components, and display a very short diameter and average shortest past with a high clustering coefficient and an almost null assortativity.

Another consequence is that global characteristics of the multilayer graphs follow mostly those of the caption layers because of their overwhelming number of nodes and clique edges in comparison to all other layers.

We now compare the prequel series (SW1–3) with the original movies (SW4–6). While the average number of nodes in the character layer is comparable, the number of nodes in faces are very different, with much more faces in the original series and the first episode of the prequel. This may be due to the increase use of storm trooper faces during the prequel trilogy, which are not properly detected with our face detector due to their mask. SW1 displays an extremely large amount of face co-occurrence. This is probably due to the scenes putting in action large crowds like during the pod race and other ceremonies. The original trilogy shows on average a high number of face links, with a peak at the last episode, due to the presence of the many Ewoks.

With the exception of SW6, which displays the lowest number of location nodes, the average number of locations are rather similar between the movies, but SW4, the original movie, contains the highest number of transitions between locations. However, this episode does not exhibit a high diameter in comparison to the prequel series, and it displays, together with the original trilogy, the highest clustering coefficient and lowest average shortest path length, suggesting that clusters of locations may occur. This may be the mark of a different style of cuts that depends on the generation of the movie.

The number of keyword nodes is quite comparable between the movies, but the connectivity of those keywords greatly varies between the two trilogies, the prequel trilogy shows a lot more edges in keywords. The number of captions seems, on average, slightly higher in the original series than in the prequel.

As illustrated in Fig. 6, there seems to be a significant difference rather consistent across both trilogies in terms of global metrics, all layers considered. Nonetheless, the clustering coefficients remain stable across movies for their individual layers.

Node influence within individual layers

We first investigate the movies for each individual layer. Due to the large number of movie×layer combinations, we only present the result of the influence score (IS) (Bioglio and Pensa 2017). A full detailed account for each episode may be found in the Additional

file 1 of this paper. For each layer, we report the top 10 nodes sorted by their influence score for each SW episode.

Ranking characters

We first report on the ranking of IS as collected in Table 2. In the prequel trilogy, *Anakin* is always among the top 3 characters. In the original trilogy, his second identity *Vader*, who is first seen in SW3, only appears in the second top tier. *Obi-Wan* gradually gains importance in the prequel trilogy being the top character of the third movie, while his second identity as *Ben* only gets in the last tier of the first movie of the original trilogy.

Focusing on the first trilogy, *Padme/Amidala* is in the second tier in the first movie, then becomes the main character of the second movie, before being overtaken by *Palpatine* in the third movie, who has a steady growth from the first to the third movie (note that his second identity as *the Emperor* does not appear in the top of the original trilogy). We can add that *Qui-Gon* is the main character of the first movie. The main antagonist characters are also well presented in this top 10 ranking. We have *Nute Gunray* in the first episode, *Count Dooku* in the second episode, and *Grievous* in the third episode.

In the original series, *Luke Skywalker* and *Han Solo* are always in the top 3 characters, with the intrusion of *C-3Po* and *princess Leia*. Beyond *Vader*, antagonists like *Tarkin*, *Piett*, *Veers*, and *Jabba* make their appearance in the top 10 characters too. We can notice that *Lando* only appears in the top of the 6th movie. In addition, *Artoo* and *Chewbacca* are also important protagonists who did not appear in this ranking because they were not properly identified as speakers.

Ranking faces

Observing the ranking of faces in Table 3 gives a different side of the story, and some new characters make it to the top, due to the length of some scenes. The main changes we observe happen in the second and last tiers of the rankings.

For example, Padme is a role, that was played by different characters, and since Amidala is also Padme, *Amidala's doppelganger* makes it to top ranking. It seems that she is not playing an important role in the movie, but its presence in almost all scenes makes her in the top of the list. *Shmi* (the mother of *Anakin*) and *Sebulba* (Anakin's main opponent during the pod race) are two important characters for the narration of Anakin's side of the story. *Jango Fett* and *Boba Fett* are two key characters in the construction of the drone army, who appear only from their face occurrence in SW2. In SW3, we may notice the addition of *Chewbacca* first, who happen to be a key character in the following trilogy. We may also underline the appearance of *Mace Windu* who does not play a major role for this episode, but who is played by the very popular actor Samuel Lee Jackson.

In the original trilogy of SW, we may also confirm the characters ranking with *Luke Skywalker*, *Leia*, and *Han Solo* on the top rankings. However in the whole trilogy, we see *Chewbacca* reach the first half of the rankings, and interesting newcomers such as the *Cantina's bartender*, central to the iconic Cantina scene in SW4. Secondary characters as technicians and stormtroopers reach in SW5, which exposes more the military organization of the rebellion. SW5 introduces a lot of new characters such as one of *Jabba's musicians*, *Biggs* (a member of the rebel) and an *Ewok*.

All in all, faces and characters are mostly common when we compare the top protagonists, but interesting changes occur on the secondary characters, and introduces key

Table 2 Top 10 nodes sorted and their influence score of the character layer G_{CC} for each of the 6 SW movies

CHARACTERS G_{CC}											
SW1		SW2		SW3		SW4		SW5		SW6	
QUI-GON	1,00	PADME	1,00	OBI-WAN	1,00	LUKE	1,00	HAN SOLO	1,00	HAN SOLO	1,33
ANAKIN	2,33	ANAKIN	2,00	PALPATINE	2,67	C-3PO	2,00	LUKE	2,33	C-3PO	2,67
JAR JAR	2,67	OBI-WAN	3,00	ANAKIN	3,00	HAN SOLO	3,00	LEIA	3,00	LUKE	3,33
OBI-WAN	4,67	M.WINDU	4,33	PADME	3,33	LEIA	4,00	C-3PO	3,67	LANDO	3,67
PADME	5,00	YODA	4,67	YODA	5,00	VADER	5,33	PIETT	5,00	LEIA	4,00
AMIDALA	5,67	PALPATINE	6,33	B.ORGANA	6,33	BIGGS	6,67	VADER	6,00	VADER	6,00
PANAKA	7,33	C-3PO	6,67	D.VADER	8,00	I.OFFICER	8,67	RIEEKAN	7,33	ACKBAR	8,00
NUTE	8,00	JAR JAR	8,33	N.GUNRAY	8,33	BEN	8,67	ANNOUNCER	8,33	WEDGE	8,67
PALPATINE	9,33	C.DOOKU	9,67	GRIEVOUS	10,00	TARKIN	9,33	WEDGE	9,33	COMMANDER	9,00
R.OLIE	10,67	M.AMEDDA	11,00	M.AMEDDA	10,67	R.LEADER	11,33	VEERS	10,33	JABBA	10,33

Table 3 Top 10 nodes sorted and their influence score of the face layer G_{FF} for each of the 6 SW movies

FACES G_{FF}										
SW1		SW2		SW3		SW4		SW5		SW6	
QUI-GON	1,00	ANAKIN	1,33	ANAKIN	1,00	LUKE	1,00	LEIA	1,00	HAN SOLO	1,67
A.DOPPELGANGER	2,00	OBI-WAN	2,00	OBI-WAN	2,00	HAN SOLO	2,33	HAN	2,67	LUKE	2,00
ANAKIN	3,00	AMIDALA	2,67	PALPATINE	3,00	LEIA	3,33	LUKE	4,00	LEIA	3,00
JAR JAR	4,00	M.WINDU	4,00	YODA	4,00	OBI-WAN	4,33	CHEWBACCA	4,00	C-3PO	4,67
OBI-WAN	5,00	YODA	5,67	B.B.ORGANA	5,33	CHEWBACCA	4,67	L.TECHNICIAN	5,67	CHEWBACCA	5,33
PANAKA	6,33	DOOKU	6,67	GRIEVOUS	7,00	C-3PO	5,33	C-3PO	6,00	LANDO	6,67
SHMI	6,67	J.FETT	8,67	M.WINDU	7,33	DODGE	8,67	PIETT	8,00	P.FOLLOWER	7,67
PADME	8,33	B.FETT	9,67	NUTE_GUARD	8,00	TARKIN	9,00	RIEEKAN	8,33	BIGGS	9,00
SEBULBA	9,67	KI-ADI	10,33	M.AMEDDA	9,00	R.LEADER	9,00	STORMTROOPER	10,33	J.MUSICIAN_1006_0	10,67
PALPATINE	11,67	P.FOLLOWER	11,67	CHEWBACCA	9,33	C.BARTENDER	10,33	R.OFFICER	11,33	EWOK_1	14,33

characters either from the length of scenes (like *Sebulba*), because they would not speak (like *Chewbacca*), or for more commercial reasons (like *Mace Windu*).

Ranking locations

We report the ranking results of locations in Table 4. Note that we made abbreviations to improve the table readability. The table of abbreviations may be found in the Additional file 1. We may first notice that in the prequel series, there is no actual redundancy of locations, whereas the original series has the *Millenium Falcon* as a key location to access most of the others. However the locations are described in a tree manner (e.g. *Hoth - Ice plain - Snow Trench)*, but since it is not consistent across all movies, we only consider them as leaves in this study and keep the hierarchical analysis for a future work.

The top location of the first movie is the *Federation Battleship Bridge* (*FBB*), where the movie starts, and where the two first antagonists are introduced. The ship is wide and contains many different areas hence making a central area in the location layer. In the second movie, there is no one top location but a more evenly distributed top locations, among them *Cockpit Naboo Starship - Sunset* (*CNSS*) in which Anakin and Padme travel to Tatooin, the *Senate Building - Padme's Appartement Bedroom* (*SBPAB*) in which Anakin and Padme start developing their relationship, and *Space* (SP) which is central to battles. In SW3, the *Plaza Jedi Temple-Coruscant* (*PJTC*) is the heart location where all dramatic development happened.

In the original series, from SW4, the main locations are the *Space Craft in Space* (*SIS*) because space battles are central to movie, and even from the first scene, and the final battle from *Luke's XWing Fighter - Cockpit* (*LXFC*), where he destroys the Death Star. These locations are central because these scenes display a lot of cuts between different vessels. The last two movies are really centred on the *Millenium Falcon*, from the *Main Hangar* (*MHMFC*) in SW5 and the *cockpit* (*MFC*) in SW6. The Millenium is iconic of the original series, and the main protagonists travel in this space ship.

Ranking keywords

We now report the ranking of keywords in Table 5, of which we find mentions to some key characters.

In the prequel series, there is mention of the *chancellor* as a key word in all three episodes, and growing to the third episode since the chancellor is the Emperor corrupting Anakin. *Queen* is specific to SW1 which the movie revolves around. *Annie* (Anakin) is mentioned in the second movie, which is interesting since it is his tender name, and the movie develops their relationship with Padme. *Windu* and *Yoda* are mentioned in the third movie, which revolves around the conflict between the Jedi council they represent and Anakin.

Beyond character keywords, the *federation*, *senate*, *republic* are recurring keywords highlighting the political tone of the first series. *Master*, *Jedi* and the *Force* make the relationship with the "religious/magic" part of the series.

In the original series, a lot of main characters enter the top ranking. We can mention that *Han* is on the top of SW4, beyond the main character who is *Luke*. *Artoo* (R2-D2) and *Chewie* (Chewbacca) are also introduced SW4, which is interesting because neither the script characters or the face detection helped reveal Artoo in the main protagonists. From SW5, *father* is by far the top keyword, which is the key revelation of this episode.

Table 4 Top 10 nodes sorted and their influence score of the location layer G_{LL} for each of the 6 SW movies

LOCATIONS G_{LL}											
SW1		SW2		SW3		SW4		SW5		SW6	
FBB	1,00	CNSS	3,00	PJTC	1,33	SIS	3,00	MHMFC	1,00	MFC	1,00
NGP	4,00	SBPAB	3,67	MMCC	2,67	LXFC	3,67	HB	4,67	DSCR	2,67
TCH	4,67	SP	4,00	MCP	11,00	DSCR	6,33	HRBMHD	5,00	RSCB	3,67
TDNS	6,67	TCKLP	5,33	CSCMA	11,33	SATDS	7,00	HRBCC	8,00	ETTR	4,67
FBCR	9,33	CCD	6,00	ASH	12,00	MFC	7,67	BOCCWVD	8,00	SKI	7,00
AHMR	11,33	TCCE	6,33	OBS	13,00	MOWR	9,00	DVSDBMCD	8,33	SRF	7,67
NSMA	11,67	GLA	8,00	PJTCR	13,00	SOTDS	9,00	SIF	9,33	DSMDB	8,33
SCU	12,33	THMF	10,00	ULP	13,33	MFGC	10,33	HIPST	12,00	FGB	9,00
NSC	13,00	GEA	10,67	IDC	15,33	RLC	12,67	MFGAC	13,67	RTJPT	9,33
NPTR	14,67	TC	15,33	LPN	17,00	DSCOR	13,00	LSRLC	16,33	JTR	9,67

Table 5 Top 10 nodes sorted and their influence score of the keyword layer G_{KK} for each of the 6 SW movies

KEYWORDS G_{KK}										
SW1		SW2		SW3		SW4		SW5		SW6	
federation	2,00	jedi	1,00	jedi	1,33	ship	1,00	father	1,00	han	1,00
jedi	3,00	senator	2,00	anakin	1,67	han	2,33	master	4,33	luke	2,00
queen	3,67	master	3,00	chancellor	4,33	imperial	4,00	ship	4,67	artoo	4,00
senate	5,33	great	4,33	master	4,33	hear	4,00	luke	5,67	jabba	5,33
time	6,33	republic	4,67	force	5,33	main	5,00	artoo	6,00	father	5,67
people	6,33	continuing	7,00	council	8,00	luke	5,33	energy	7,67	master	6,33
naboo	6,67	annie	7,67	time	8,33	artoo	9,67	han	8,67	vader	9,00
master	8,67	time	8,00	windu	8,33	good	10,67	chewie	10,00	threepio	9,33
back	9,33	chancellor	8,33	republic	9,33	chewie	11,00	field	10,33	good	12,33
chancellor	9,33	naboo	11,67	yoda	12,33	shut	11,67	princess	12,00	yoda	14,00

Han loses some ranks, and the reference to the *Princess* (Leia) enter the top. In the last episode, references to one main antagonist, *Jabba* enters the top keywords, and *Threepio* (C-3PO) and *Yoda* enter the top.

Beyond characters, vocabulary related to space vessels appear (*ship*, *main*, *energy*, *field*). The philosophical question of *good* (in opposition to the dark side) is also as an important keyword, in combination with *master* which is core to the structure of Jedi (protagonist) and Sith (antagonist) organizations.

Ranking captions

The ranking of captions, reported in Table 6, suggests that most of the visual similarity between scenes is focused on people's outfit rather than anything else, thanks to the term *wearing* which is almost all of the top captions. Nonetheless, this capture well the visual identity of the movies.

In the prequel series, the appearances of Queen Amidala is remarked from her multiple outfits, and those of her followers. The term *woman* appears a lot in the top captions of SW1 and gradually decreases in the following episodes. SW1 shows a wide range of colors associated with *wearing*: *black*, *red*, *white*, *blue*, and *gray*. The following two episodes mostly bring forward the *black* jacket of Anakin's outfit, and *white* clothes which

Table 6 Top 10 nodes sorted and their influence score of the caption layer G_{CaCa} for each of the 6 SW movies

CAPTIONS G_{CaCa}					
SW1		SW2		SW3	
a,black,jacket, wearing	1,67	a,black,shirt, wearing	1,00	a,black,shirt, wearing	1,00
a,black,wearing, woman	2,00	a,shirt,wearing, white	2,33	a,black,jacket, wearing	2,33
a,red,wearing, woman	3,33	a,black,jacket, wearing	2,67	a,shirt,wearing, white	3,00
a,red,shirt,wearing	4,33	a,black,man, wearing	4,00	a,black,man, wearing	3,67
a,shirt,wearing, white	4,67	a,black,wearing, woman	5,67	a,black,wearing, woman	5,00
a,jacket,red, wearing	6,67	a,wearing,white, woman	6,67	a,man,wearing, white	6,00
a,black,man, wearing	7,33	a,brown,shirt, wearing	7,00	a,red,shirt, wearing	7,33
a,blue,wearing, woman	10,33	a,red,shirt, wearing	7,00	a,blue,shirt, wearing	8,00
a,man,red,wearing	10,67	a,brown,chair, wooden	9,33	a,brown,shirt, wearing	9,67
a,gray,shirt, wearing	12,00	a,blue,shirt, wearing	10,33	a,hat,man, wearing	11,33
SW4		SW5		SW6	
a,shirt,wearing, white	1,00	a,shirt,wearing, white	1,00	a,shirt,wearing, white	1,33
a,man,wearing, white	2,00	a,man,wearing, white	2,00	a,black,jacket, wearing	3,00
a,black,shirt, wearing	3,33	a,black,shirt, wearing	3,33	a,man,wearing, white	3,33
a,black,man, wearing	4,33	a,black,man, wearing	4,33	and,red,sign, white	4,33
a,red,shirt, wearing	5,00	a,red,shirt, wearing	5,00	a,black,shirt, wearing	4,33
a,black,jacket, wearing	5,33	a,black,jacket, wearing	5,33	a,black,man, wearing	5,67
a,wearing,white, woman	7,33	a,wearing,white, woman	7,33	a,hat,man, wearing	7,33
a,helmet,man, wearing	9,67	a,helmet,man, wearing	9,67	a,green,shirt, wearing	7,67
and,bag,black, white	10,00	and,bag,black, white	10,00	a,black,wearing, woman	9,67
a,black,wearing, woman	11,33	a,black,wearing, woman	11,33	a,man,shirt, wearing	9,67

correspond to the numerous clones' armor. We may also notice the introduction of the *brown* outfit that is representative of Jedi knights.

The original series introduces *helmets* or *hat* wearing people, which often matches the outfit of Darth Vader and all the different military people in both the Empire and Rebel armies. Top colors are greatly focused on *black*, which is most represented by Vader, and *white* which is the main color of Luke's outfit. The last episode introduces *green* outfits that are the ones worn by the Rebels in all actions happening in the forests of Endor moon.

Node influence in the multilayer network

We now analyze node influence score from the multilayer networks as reported in Table 7. Interesting nodes in this network highlight and associate different key elements of the story. As illustrated in the global topological analysis of "Topological properties of individual layers" section, the caption layer has order of magnitude differences with all other layers in terms of size, hence strongly influencing the ranking. Our multilayer model allows for investigating this difference by simply checking rankings in the multilayer network $\mathbb{G}' = \mathbb{G} - G_{CaCa}$ with all layers except the caption layer (in Table 8).

Recalling topological properties as displayed in Fig. 6, we may notice that the whole multilayer $\mathbb{G}$ behaves similarly to the caption layer G_{CaCa}, except for diameter which becomes significantly smaller. The multilayer without captions $\mathbb{G}'$ shows a rather low density, but a high clustering coefficient suggesting a of a community structure organization. Most interestingly, it displays a negative assortativity, meaning that high degree nodes tend to connect preferably with low degree nodes. This is probably an effect of the association to location nodes within the graph.

Multilayer network, all layers

The first thing we may notice is that face G_{FF} and character G_{CC} layers are prominent in the results, then comes the caption layer G_{CaCa} and the keyword layer G_{KK}. The fact that captions are not only numerous but cliques generated for each scene reinforces their influence score. However, we have a good amount of redundancy between people over face, script, and keyword detections, confirming these stories are centred around the narration of characters' adventures.

The first movie bring forward all the top characters we may find everywhere, the main protagonists, Qui-Gon, Obi-Wan, with Amidala (through her doppelgangers) and Anakin. The very controversial Jar Jar is often felt as over-represented by the fandom, and we can only confirm this in this ranking. Anakin and Amidala/Padme make the top of the next movie, which revolves over their relationship, and the development of the Jedi training of Anakin, hence the prominent keywords *Master* and *Jedi*. For the last episode of the prequel trilogy, Anakin and Obi-Wan are the top most represented characters (since this episode will lead them to a fight), and their master/Jedi relationship is taking prominence from the keywords. We may notice the introduction of the Jedi master Yoda in the top ranking, a highly central character of the whole series, who is leading the Jedi council in this episode. One main character that was most influential in the face and character layers was Palpatine, but he is absent from the top ranking in the multilayer. This is indicative of his strong connection with a few characters and places in the plot of SW3 for instance with Anakin and mostly on Coruscant. Amidala/Padme is also a central character in SW2 and SW3 but she is stranded on Coruscant for most of the latter

Table 7 Top 10 nodes sorted, with their layer and influence score of the overall multilayer network $\mathbb{G}$ for each of the 6 SW movies

MULTILAYER, ALL LAYERS $\mathbb{G}$								
SW1			SW2			SW3		
QUI-GON	G_{FF}	1,00	ANAKIN	G_{FF}	1,67	ANAKIN	G_{FF}	2,33
A.DOPPELGANGER	G_{FF}	2,33	AMIDALA	G_{FF}	3,00	OBI-WAN	G_{FF}	3,33
ANAKIN	G_{FF}	2,67	master	G_{KK}	3,33	jedi	G_{KK}	5,00
OBI-WAN	G_{FF}	4,00	OBI-WAN	G_{FF}	4,33	anakin	G_{KK}	5,33
QUI-GON	G_{CC}	5,33	a,black,shirt,wearing	G_{CaCa}	6,00	a,black,shirt,wearing	G_{CaCa}	6,67
JAR JAR	G_{FF}	5,67	jedi	G_{KK}	6,33	YODA	G_{FF}	8,00
ANAKIN	G_{CC}	7,00	continuing	G_{KK}	8,00	a,black,jacket,wearing	G_{CaCa}	8,33
a,black,jacket,wearing	G_{CaCa}	9,33	a,black,jacket,wearing	G_{CaCa}	8,33	master	G_{KK}	8,67
a,black,wearing,woman	G_{CaCa}	9,67	PADME	G_{CC}	9,67	a,black,man,wearing	G_{CaCa}	9,33
JAR JAR	G_{CC}	11,67	a,black,wearing,woman	G_{CaCa}	10,33	anakin	G_{KK}	11,67
SW4			SW5			SW6		
a,shirt,wearing,white	G_{CaCa}	1,33	LEIA	G_{FF}	1,00	LUKE	G_{FF}	1,00
LUKE	G_{FF}	1,67	HAN SOLO	G_{FF}	3,00	HAN SOLO	G_{FF}	2,00
a,man,wearing,white	G_{CaCa}	3,00	a,shirt,wearing,white	G_{CaCa}	3,00	LEIA	G_{FF}	3,00
LEIA	G_{FF}	5,33	a,black,shirt,wearing	G_{CaCa}	4,33	HAN SOLO	G_{CC}	5,33
a,black,shirt,wearing	G_{CaCa}	5,67	LUKE	G_{FF}	6,00	luke	G_{KK}	6,33
a,black,man,wearing	G_{CaCa}	7,00	a,black,man,wearing	G_{CaCa}	7,00	C-3PO	G_{CC}	6,67
a,black,jacket,wearing	G_{CaCa}	7,67	a,black,jacket,wearing	G_{CaCa}	7,33	C-3PO	G_{FF}	7,67
a,red,shirt,wearing	G_{CaCa}	8,33	a,wearing,white,woman	G_{CaCa}	9,33	CHEWBACCA	G_{FF}	10,33
LUKE	G_{CC}	10,00	comlink	G_{KK}	9,67	a,shirt,wearing,white	G_{CaCa}	11,67
HAN SOLO	G_{FF}	10,33	a,black,wearing,woman	G_{CaCa}	10,00	LANDO	G_{FF}	12,00

Table 8 Top 10 nodes sorted, with their layer and influence score of the overall multilayer network $\mathbb{G}'$ for each of the 6 SW movies

MULTILAYER, WITHOUT CAPTIONS $\mathbb{G}'$								
SW1			SW2			SW3		
QUI-GON	G_{FF}	1,00	ANAKIN	G_{FF}	1,00	OBI-WAN	G_{FF}	1,00
A.DOPPELGANGER	G_{FF}	2,33	OBI-WAN	G_{FF}	2,67	ANAKIN	G_{FF}	2,00
ANAKIN	G_{FF}	2,67	AMIDALA	G_{FF}	2,67	YODA	G_{FF}	3,00
JAR JAR	G_{FF}	4,33	PADME	G_{CC}	3,67	PALPATINE	G_{FF}	5,33
OBI-WAN	G_{FF}	4,67	ANAKIN	G_{CC}	5,00	B.B.ORGANA	G_{FF}	5,33
QUI-GON	G_{CC}	6,00	OBI-WAN	G_{CC}	6,67	OBI-WAN	G_{CC}	5,67
ANAKIN	G_{CC}	7,00	M.WINDU	G_{FF}	7,33	ANAKIN	G_{CC}	6,67
PANAKA	G_{FF}	8,00	YODA	G_{FF}	9,00	DVQSD	G_{LL}	9,67
SHMI	G_{FF}	9,00	jedi	G_{KK}	9,00	M.WINDU	G_{FF}	11,00
PADME	G_{FF}	10,33	master	G_{KK}	10,33	PALPATINE	G_{CC}	11,00
SW4			SW5			SW6		
LUKE	G_{FF}	1,00	LEIA	G_{FF}	1,00	HAN SOLO	G_{FF}	1,33
LEIA	G_{FF}	2,33	LUKE	G_{FF}	2,00	LUKE	G_{FF}	2,67
LUKE	G_{CC}	2,67	HAN SOLO	G_{FF}	3,67	LEIA	G_{FF}	3,00
H.SOLO	G_{FF}	4,00	HAN SOLO	G_{CC}	5,33	C-3PO	G_{FF}	3,67
C-3PO	G_{FF}	5,00	MHMFC	G_{LL}	6,33	CHEWBACCA	G_{FF}	5,67
C-3PO	G_{CC}	6,67	HRBCC	G_{LL}	6,67	HAN SOLO	G_{CC}	6,00
O.WAN	G_{FF}	8,00	CHEWBACCA	G_{FF}	7,00	C-3PO	G_{CC}	7,67
CHEWBACCA	G_{FF}	8,67	LUKE	G_{CC}	8,67	LANDO	G_{FF}	8,33
H.SOLO	G_{CC}	10,00	C-3PO	G_{FF}	10,33	LUKE	G_{CC}	9,33
MFC	G_{LL}	10,67	YODA	G_{FF}	10,67	HAN SOLO	G_{KK}	10,00

film, whereas her and Anakin where travelling a lot in the former. There is no specific conclusion from the captions' perspective, other than black outfits are dominating this series.

The two first episodes of the original series see much more captions being brought forward. Beyond the black and white outfits we discussed in the previous section, we may notice the introduction of *red shirts* which are none other than the uniform of the Rebels. Luke, Leia, and Han Solo are the most represented characters, following the cast distribution. We may also notice in SW5 the mention to *comlink* because the characters and separated in different sites throughout the movie, and communicates a lot through this device. The last episode unifies subplot in which secondary characters also play more important roles (such as delivering Solo, or cutting the power from Endor) and we see this in the introduction of other charismatic characters: C-3PO, Chewbacca, and Lando.

Although the location layer nodes are not represented, the influence of the layer through links to characters may be observed. Prominent character nodes (whichever the layer) that are brought forward often correspond to those traveling a lot between locations. For example, although Amidala is central in SW3, she enters the top in SW2 where she travels a lot, and the other around is true for Yoda who travels a lot in SW3.

Multilayer network, without the caption layer

The ranking of nodes in $\mathbb{G}'$ (Table 8) is very close to those of the full multilayer $\mathbb{G}$ (Table 7), with the exception of all captions being taken out of the top. We can however observe a few locations making their place into the top ranking, but less keywords.

From the first episode in the prequel series, the main changes are the following. The ranking of Jar Jar has increased a bit, but we can mostly notice the inclusion of *Shmi*, who is Anakin's mother, a central character in the whole segment concerning Tatooine. Queen Amidala, under her name *Padme*, is also entering the ranking. *Panaka* is the guard who accompanies Amidala/Padme all along to protect her, and take a long participation in most action scenes. In SW2, *Obi-Wan* gains a few ranks, probably for his numerous travels (checking on the clone army). The leaders of the Jedi council, *Mace Windu* and *Yoda* enter the ranking too, and for the next movie. The keywords *Jedi* and *master* are still maintained, underlining the other thema of this movie which revolves around the Jedi training of Anakin. The last movie of the prequel does not show the persistence of these keywords in the top ranking, but sees major introductions of first *Palpatine* who corrupted Anakin, and of *Bail Organa*, a senator organizing the resistance against Palpatine, who will harbour one child of Anakin after his turning to the dark side. A location appears in this movie rankings, which is Darth Vader's Quarter Star Destroyer, in a scene at the ending that exists only in the script, and was finally deleted.

The original series also sees a lot new nodes replacing captions, above all, *Chewbacca* and *C-3PO*, companions of the main characters, entering all top rankings. In SW4, *Obi-Wan* also enters the ranking, since he guides the young Luke all along this adventure. Most importantly, the *Millenium Falcon Cockpit* (*MFC*) the vessel which caries all characters through their adventure is the main location which enters this ranking. The *comlink* keyword disappears of SW5 but *Yoda* appears in this ranking, since Luke makes the trip to receive training from him during this episode. Two locations enter in the ranking, *Main Hangar - Millenium Falcon - Cockpit* (*MHMFC*) and *Hoth - Rebel Base - Command Center* (*HRBCC*) where most characters regroup during the first part of the movie, before

being separated then. In the last episode, nothing changes much except that *Han Solo* takes the leadership of the ranking.

Community detection

Our preliminary results on global topological properties in "Topological properties of individual layers" section suggest the existence of communities especially given the clustering coefficient of the different layers (Orman et al. 2013a). To study clustering in the individual layers and the overall network, we use the modularity-based (Girvan and Newman 2002) community detection algorithm often referred to as the Louvain method (Blondel et al. 2008), which has been generalized to multilayer networks too (Domenico et al. 2014). Figure 7 reports the number of communities with the modularity per layer for each movie of the saga. Not surprisingly, the captions have the highest number of communities and highest modularity due to their definition which are cliques on each scene. It is however more surprising to see a high modularity for locations. Keywords best clusterize during SW5. Character and faces layers are social networks, displaying some potential for clustering. Captions also have a very high modularity, due to their nature as a collection of cliques. Despite receiving a strong influence from the caption layer with comparable number of communities, the multilayer graph $\mathbb{G}$ shows overall modularity close to the keywords and faces. Without the caption layer, the multilayer graph $\mathbb{G}'$ seems very close to the community structures induced by faces, association to locations through cut order of the movie probably reinforces the importance of face co-occurrences.

The third episode of the prequel trilogy (Lucas 2005) is an interesting point in the series, where we can observe the main character of the whole saga, Anakin, turning into the dark version of himself that will be known as Darth Vader. We will observe how the different communities we measure may reflect this division. Communities of this episode are illustrated in Figs. 8 and 9 and with Gephi (Heymann 2014) for SW3 only, all other episodes are also illustrated in the Additional file 1.

Starting with the character layer, we may notice three major communities. One community (pink) is centred around *Padme* and *Obi-Wan* and would correspond to the Jedi council that is represented by *Yoda*, *Mace Windu*, *Ki-Adi*, together with the clone army

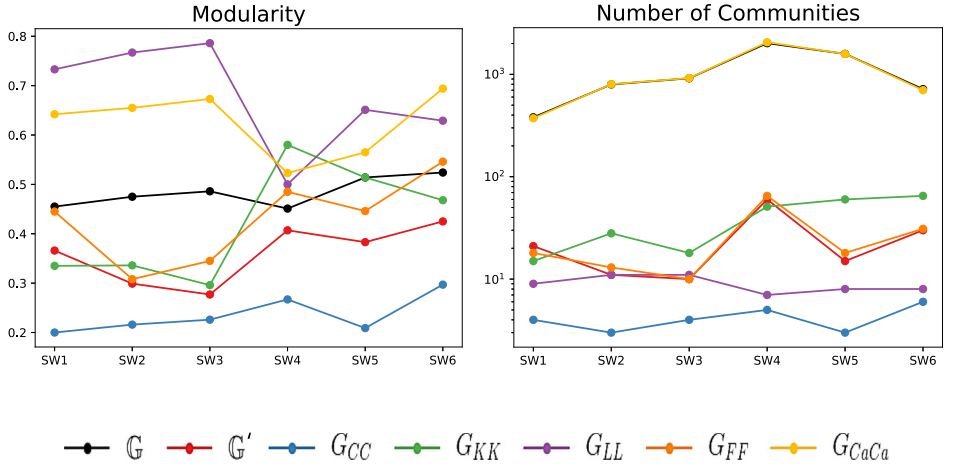
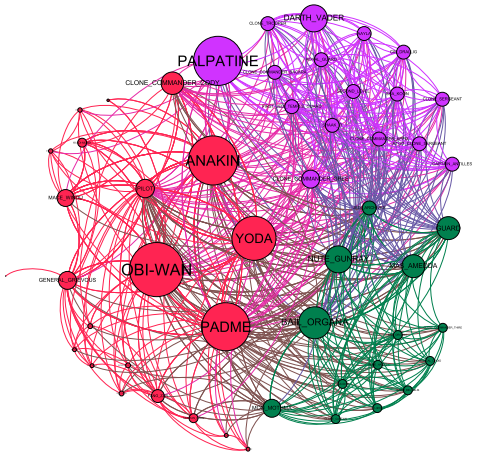


Fig. 7 Number of communities and modularity per layer for each movie of the saga

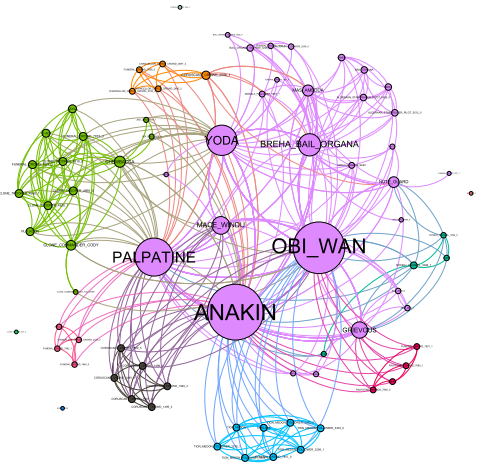


(a) G_{CC}

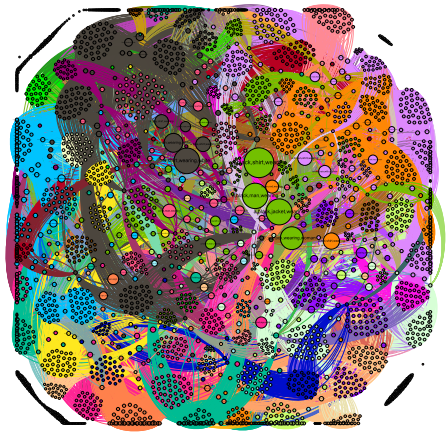
(b) G_{KK}

(c) G_{LL}

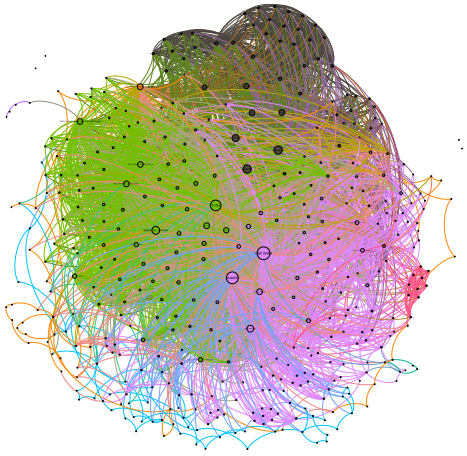
Fig. 8 The networks are better seen zoomed on the digital version of this document. Visualization of communities in different layers of Episode III - Revenge of the Sith (2005) (Lucas 2005). The size of each node corresponds to its degree. **a** The character layer G_{CC}. **b** The keyword layer G_{KK}. **c** The location layer G_{LL}



(a) G_{FF}



(b) G_{CaCa}



(c) $\mathbb{G}'$

Fig. 9 The networks are better seen zoomed on the digital version of this document. Visualization of communities in different layers of Episode III - Revenge of the Sith (2005) (Lucas 2005). The size of each node corresponds to its degree. **a** The face layer G_{FF}. **b** The caption layer G_{CaCa}. **c** The multilayer without captions $\mathbb{G}'$, with the node label encoding: CHARACTER(C), FACE(F), keyword, and LOCATION-

they are leading, represented by *Clone Commander Cody*. Their antagonist, *General Grevious* is also put in this community, because one major plot of this episode is the fight of the Jedi against Grevious. A second community (green) is centred on politics and revolves around the senate on Coruscant, with *Bail Organa*, and *Mas Amedda*. A last community (purple) regroups the Sith side, with the major characters *Palpatine* and *Darth Vader*.

On the contrary, the face layer does not make the distinction between Anakin and Vader. It shows 7 communities, with the main one (purple) formed from the main actors who are constantly interacting during the movie (*Obi-Wan*, *Anakin*, *Palpatine*, *Yoda*, *Mace Windu*, *etc.*). Other communities are formed around secondary characters or crowds such as the clone army together with *Cody*. We also find as smaller tight communities such the Jedi council as a community, Coruscant politicians, crowds and followers. These minor characters are often presented together in one same scene creating such cliques.

Although the location layer gets a total of 10 communities, a few stand out. The locations are often connected by geographical proximity, as a sequence of scenes will follow a particular character or action that evolve in a small, continuous environment. On a larger scale, this is the temporal proximity that emerges. Sequences of events taking place at the same time but in different places connect the related locations. In particular, one community (purple) relates to the end of the film. At this point, the action is concentrated on the duel between Anakin and Obi-Wan on *Mustafar* and the one between Yoda and The Emperor at the *Senate* and *Palpatine's office*. The *Mustafar main control center* is one key location of the fight but is also cut while Jedis are shown being executed by clones all across the galaxy, and Anakin is killing the last separatist leaders. This community also includes the *Alderaan starcruiser*, the protagonists last stand at the end of the movie. Another community (green) consists of locations used to showcase the battle at the beginning of the movie in *space* while cutting to the inside of *Obi-Wan's starfighter cockpit* as well as *Anakin's starfighter cockpit*. In the film, once they localize the *Trade Federation cruiser* where *Palpatine* is held hostage, they head inside. We can see this transition occur via the *hangar* of the ship. The next community exposes the inside of the cruiser, such as the *bridge* and the *elevator* that lead the protagonists to the *Senator's room* and eventually *General quarters*. At the end of their confrontation, General Grievous escapes through the *pod bay*, returning the action to *space*. The sequence ends with Anakin navigating a damaged ship through the *skies above Coruscant*. From this point on, the characters go on different adventure which is why the other communities are not as geographically focused. Yoda is on *Kashyyyk*, Obi-Wan goes to *Utapau* and Anakin remains on *Coruscant*.

The keyword layer presents 10 communities corresponding to different topics. The largest community (light green) may be related to Anakin's emotional journey with words such as *anakin*, *kill*, *padme*, *obiwan*, *love*, *destroy*, *save*, *lost*, *etc.* A second community (pink) groups around the political intrigue with *jedi*, *chancellor*, *senate*, *dooku*, *etc.* Confirming our observations on the character layer, another community (purple) is on the organization of the Jedi council *master*, *kenobi*, *windu*, *etc.*, and of course another one is focused on the dark side with *force*, *power*, *sith*, *apprentice*, *darth*, *etc.*

Captions are clustered by scene in a large number of communities. Each scene has a number of captions which describe what happened in this scene. Observing communities does not offer much more interpretation beyond the colors clothing community. Since it

impacts a lot the multilayer structure, we are more interesting in observing communities in the multilayer network $\mathbb{G}'$ that excludes this layer. There is a total of 12 communities. Four major communities regroup from 52 to 140 nodes, with very little overlaps between layers. In a first community (green), we have 52 main and secondary characters (all from G_{CC}) interacting together during the movie. In a second community (light green), we have 77 locations mostly from the end of the movie, with a handful of keywords related to the last dual (*fight*, *late*, *inside*, *chamber*, *burning*, *koon*), and two extra characters. In a community (purple) of 86 nodes that combines all layers and regroups vocabulary attached to the force from both Sith and Jedi sides (e.g. *master*, *force*, *afraid*, *feel*, *great*, *lord*, *powerful*, *order*, *dark*, *control*, *strong*, *anger*, *etc.*) and the locations where Anakin is turned *Opera* and *Lobby to Chancellor's Office*. A last community of 140 nodes also regroup most layers, with just a little bit of characters, a lot of faces of people in situation with battles and crowds, with people from crowds, such as *Obi-Wan*, *Grevious*, *Cody*, *etc.* The locations are very varied, and the vocabulary attached tends to be more technical of battles, including *droids, clones, contact, move, platform, hold, attack, break, hangar, squad, commander, troops, escape, fire, mission, surface, front, engage, missiles, fighter, etc.* All in all, we can see a difference between the last two major communities that underline the two worlds, centred on Anakin, and that clash at the end of the movie. One is closer to the world of Padme/Amidala, with the senate politics and Organa, the other is closer to the Palpatine side, fights and adventure. The main reason might be the very little interactions between Anakin and Organa on one side, and between Padme and Palpatine on the other side.

Conclusion

In this paper, we introduce a multilayer model with movie elements *characters*, *locations*, *keywords*, *faces* and *captions* are in interaction. Unlike single layer networks which usually focus only on *characters* or *scenes*, this model is much more informative. It completes the single character network analysis with a new topological analysis made of more semantic elements that brings us a global broad picture of the movie story. We also propose an automatic method to extract the multilayer network elements from the script, subtitles, and movie content. In order to enrich the previous model, additional multimedia elements are included, such as face recognition, dense captioning and subtitle information. We have publicly released all our multilayer network datasets and made them available at http://github.com/renoust/multilayermovies.

On a model side, we have not fully discussed another contribution of Kivelä's model (Kivelä et al. 2014) which are *aspects*. Aspects could be understood as another discrete dimension of the multilayer network model, and this completely captures the notion of time depicted by the different episodes of the saga. In addition, one could consider furthermore the media modality from which we extract information to be another *aspect* dimension, this is actually, what we are doing when separating the faces network from the character network. In future work, we will focus on questioning the coupling across these aspects.

So far we have not proposed any fusion of nodes through layers, such as face and characters, but we considered them separately, especially since some characters correspond to different personas (Anakin/Vader, Padme/Amidala/Doppelgangers). This alignment will show its usefulness in further studies. The locations are typically hierarchical in the way

they are depicted (e.g. *planet - location - room*) and would deserve further treatment. This will be necessary to propose one full analysis at the level of the 6 movies taken at once.

We have deployed the model on the popular 6-movies of the Star Wars saga. Results of a brief analysis of the extracted networks confirmed the effectiveness of the model. So far, we have considered the succession of scenes to be the time granularity. We may however extend this notion and attempt to recover time as represented *in* the movie world. This will require more complex processing of the events in the movie, and would help untangle complex movies like *Memento* or *Pulp Fiction* which have complex timelines, or like *the Lord of the Rings* which has many parallel plots. It could be used as a support to study the location of characters along the plot and to enable a better transition between places: imagine a plot divided into multiple parts with parallel actions, we wish to recover this parallel nature (currently the location network may only form looping chains by definition). Note that much more information can be gained by a deeper topological analysis, for example, deriving a co-occurrence network of characters in the same location, a directed network of conversations, or mention of characters, *etc.* As for the time granularity, we wish to get done to the level of shots and even seconds, to help deploy dynamic analysis. Our future work will also include a larger set of multilayer dedicated metrics, such as node entanglement (Renoust et al. 2014), and centrality measures designed for modular networks (Gupta et al. 2016a; Ghalmane et al. 2019b). Furthermore, in the future, we plan to deploy our tool on larger collections, such as tv-series, or even a larger collection of movies so we may obtain a higher view at collection level of artistic styles (Sigaki et al. 2018).

Apart from movie representation for network analysis purposes, we believe that the model opens a numerous of new research directions. Indeed, it can also be used to characterize movie genres, or directors, and even correlate with acting careers from public databases such as IMDB. Furthermore, we can imagine automatically generate the movie trailer by searching important scenes where all movie characters are present. We also are working on including another layer to this multilayer network through emotions, which could help characterize characters and movie genres. Other layers from different media are left so far to explore, such as the actual sound component, the DVD chapter decomposition, and even language comparison if we consider different languages of the subtitle tracks. Fusing all sources of information like the proposed model does should come handy in supporting machine learning tasks, such as face recognizers, and movie classification (Gorinski and Lapata 2018; Viard and Fournier-S'niehotta 2018).

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1007/s41109-019-0226-0.

Additional file 1: Supplementary Materials.

Acknowledgements
Not applicable.

Authors' contributions
YM is the main author of this paper, he has implemented most of the experiments and wrote the original draft . LV is responsible for the implementation regarding the face detection and tracking. OR has led the use case analysis. BR, HC, MEH designed the model, the framework and the experiments. BR participated to the experiments implementation, and the writing of the original draft. HC and MEH did the review and editing of the first draft. They also proposed additional units of analysis. All the authors have read and approved the final manuscript.

Authors' information
Not applicable.

Funding
Not applicable.

Availability of data and materials
Not applicable.

Competing interests
The authors declare that they have no competing interests.

Author details
[1]LRIT URAC 29, Mohammed V University, Rabat, Morocco. [2]Institute for Datability Science, Osaka University, Osaka, Japan. [3]LIB, University of Burgundy, Dijon, France.

Received: 23 August 2019 Accepted: 18 October 2019
Published online: 23 December 2019

References

Auber D, Archambault D, Bourqui R, Delest M, Dubois J, Lambert A, Mary P, Mathiaut M, Mélançon G, Pinaud B, Renoust B, Vallet J (2017) Tulip 5:1–28

Al Omran FNA, Treude C (2017) Choosing an nlp library for analyzing software documentation: a systematic literature review and a series of experiments. In: Proceedings of the 14th International Conference on Mining Software Repositories. IEEE Press. pp 187–197. https://doi.org/10.1109/msr.2017.42

Bao J, Zheng Y, Wilkie D, Mokbel M (2015) Recommendations in location-based social networks: a survey. GeoInformatica 19(3):525–565

Bioglio L, Pensa RG (2017) Is this movie a milestone? identification of the most influential movies in the history of cinema. In: International Workshop on Complex Networks and their Applications. Springer. pp 921–934. https://doi.org/10.1007/978-3-319-72150-7_74

Blei DM, Ng AY, Jordan MI (2003) Latent dirichlet allocation. J Mach Learn Res 3(Jan):993–1022

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech Theory Exp 2008(10):10008

Cao Q, Shen L, Xie W, Parkhi OM, Zisserman A (2018) Vggface2: A dataset for recognising faces across pose and age. Automatic Face & Gesture Recognition (FG 2018). In: 2018 13th IEEE International Conference on. IEEE. pp 67–74. https://doi.org/10.1109/fg.2018.00020

Castellano B (2012) PySceneDetect. http://github.com/Breakthrough/PySceneDetect. Last Accessed 20 June 2019

Cavnar WB, Trenkle JM (1994) N-gram-based text categorization. In: Proceedings of SDAIR-94, 3rd annual symposium on document analysis and information retrieval Vol. 161175

Chen R-G, Chen C-C, Chen C-M (2019) Unsupervised cluster analyses of character networks in fiction: Community structure and centrality. Knowl Based Syst 163:800–810

Chen B-W, Wang J-C, Wang J-F (2009) A novel video summarization based on mining the story-structure and semantic relations among concept entities. IEEE Trans Multimed 11(2):295–312

Cherifi H, Palla G, Szymanski BK, Lu X (2019) On community structure in complex networks: challenges and opportunities. arXiv preprint. arXiv:1908.04901

Demirkesen C, Cherifi H (2008) A comparison of multiclass svm methods for real world natural scenes. In: International Conference on Advanced Concepts for Intelligent Vision Systems. Springer. pp 752–763. https://doi.org/10.1007/978-3-540-88458-3_68

Domenico M, Porter M, Arenas A (2014) Centrality in interconnected multilayer networks. In: *CoRR* Vol. 10

Domenico MD, Sol-Ribalta A, Omodei E, Gmez S, Arenas A (2013) Centrality in interconnected multilayer networks. In: CoRR

EAC Jr., Marinho VQ, Amancio DR (2019) Semantic flow in language networks. CoRR abs/1905.07595. 1905.07595

Ester M, Kriegel H-P, Sander J, Xu X (1996) Density-based spatial clustering of applications with noise. Int Conf Knowl Discov Data Min 240

Eude T, Cherifi H, Grisel R (1994) Statistical distribution of dct coefficients and their application to an adaptive compression algorithm. In: Proceedings of TENCON'94-1994 IEEE Region 10's 9th Annual International Conference on:'Frontiers of Computer Technology'. IEEE. pp 427–430. https://doi.org/10.1109/tencon.1994.369265

Flint LN (1917) Newspaper writing in high schools: Containing an outline for the use of teachers. Pub. from the Department of journalism Press in the University of Kansas

Ghalmane Z, El Hassouni M, Cherifi C, Cherifi H (2019) Centrality in modular networks. EPJ Data Sci 8(1):15

Ghalmane Z, El Hassouni M, Cherifi C, Cherifi H (2019) Centrality in complex networks with overlapping community structure. Sci Rep 9(10133). https://doi.org/10.1038/s41598-019-46507-y

Ghalmane, Z, Cherifi C, Cherifi H, El Hassouni M (2019) Centrality in complex networks with overlapping community structure. Sci Rep 9(1):15

Girvan M, Newman ME (2002) Community structure in social and biological networks. Proc Natl Acad Sci 99(12):7821–7826

Gisbrecht A, Schulz A, Hammer B (2015) Parametric nonlinear dimensionality reduction using kernel t-sne. Neurocomputing 147:71–82

Gorinski PJ, Lapata M (2018) What's this movie about? a joint neural network architecture for movie content analysis. In: Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). pp 1770–1781. https://doi.org/10.18653/v1/n18-1160

Guo Y, Liu Y, Oerlemans A, Lao S, Wu S, Lew MS (2016) Deep learning for visual understanding: A review. Neurocomputing 187:27–48
Gupta N, Singh A, Cherifi H (2016) Centrality measures for networks with community structure. Phys A Stat Mech Appl 452:46–59
Gupta N, Singh A, Cherifi H (2016) Centrality measures for networks with community structure. Phys A Stat Mech Appl 452:46–59
He J, Xie Y, Luan X, Zhang L, Zhang X (2018) Srn: The movie character relationship analysis via social network. In: International Conference on Multimedia Modeling. Springer. pp 289–301. https://doi.org/10.1007/978-3-319-73600-6_25
He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp 770–778. https://doi.org/10.1109/cvpr.2016.90
Heymann S (2014) Gephi. Encycl Soc Netw Anal Min:612–625. https://doi.org/10.1007/978-1-4614-7163-9_299-1
Jhala A (2008) Exploiting structure and conventions of movie scripts for information retrieval and text mining. In: Joint International Conference on Interactive Digital Storytelling. Springer. pp 210–213. https://doi.org/10.1007/978-3-540-89454-4_27
Jiang H, Learned-Miller E (2017) Face detection with the faster r-cnn. Automatic Face & Gesture Recognition (FG 2017). In: 2017 12th IEEE International Conference on. IEEE. pp 650–657. https://doi.org/10.1109/fg.2017.82
Johnson J, Karpathy A, Fei-Fei L (2016) Densecap: Fully convolutional localization networks for dense captioning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp 4565–4574. https://doi.org/10.1109/cvpr.2016.494
Jung B, Kwak T, Song J, Lee Y (2004) Narrative abstraction model for story-oriented video. In: Proceedings of the 12th annual ACM international conference on Multimedia. ACM. pp 828–835. https://doi.org/10.1145/1027527.1027720
Kadushin C (2012) Understanding social networks: Theories, concepts, and findings. https://doi.org/10.5860/choice.49-6586
Kipling R (1909) "The Elephant's Child, Just So Stories". *Illustrated by R. Kipling. London: Tauchintz. (1902)*
Kivelä M, Arenas A, Barthelemy M, Gleeson JP, Moreno Y, Porter MA (2014) Multilayer networks. J Complex Netw 2(3):203–271
Knuth DE (1993) The stanford graphbase: a platform for combinatorial computing. AcM Press, New York
Krishna R, Zhu Y, Groth O, Johnson J, Hata K, Kravitz J, Chen S, Kalantidis Y, Li L-J, Shamma DA, Bernstein MS, Fei-Fei L (2017) Visual genome: Connecting language and vision using crowdsourced dense image annotations. Int J Comput Vision 123(1):32–73
Kurzhals K, John M, Heimerl F, Kuznecov P, Weiskopf D (2016) Visual movie analytics. IEEE Trans Multimed 18(11):2149–2160
Latapy M, Viard T, Magnien C (2018) Stream graphs and link streams for the modeling of interactions over time. Soc Netw Anal Min 8(1):61
Li J, Zhang K, et al. (2007) Keyword extraction based on tf/idf for chinese news document. Wuhan Univ J Nat Sci 12(5):917–921
Lucas G (1977) Star Wars: Episode IV - A New Hope. Twentieth Century Fox Film Corporation. https://doi.org/10.1007/978-1-349-92604-6_73
Lucas G (1980) Star Wars: Episode V - The Empire Strikes Back. Twentieth Century Fox Film Corporation
Lucas G (1983) Star Wars: Episode VI - Return of the Jedi. Twentieth Century Fox Film Corporation
Lucas G (1999) Star Wars: Episode I - The Phantom Menace. Twentieth Century Fox Film Corporation
Lucas G (2002) Star Wars: Episode II - Attack of the Clones. Twentieth Century Fox Film Corporation
Lucas, G (2005) Star Wars: Episode III - Revenge of the Sith. Twentieth Century Fox Film Corporation
Lv J, Wu B, Zhou L, Wang H (2018) Storyrolenet: Social network construction of role relationship in video. IEEE Access 6:25958–25969
Markovič R, Gosak M, Perc M, Marhl M, Grubelnik V (2018) Applying network theory to fables: complexity in slovene belles-lettres for different age groups. J Complex Netw 7(1):114–127
Mish B (2016) Game of Nodes: A Social Network Analysis of Game of Thrones. https://gameofnodes.wordpress.com. Accessed 2016
Mourchid Y, Renoust B, Cherifi H, El Hassouni M (2018) Multilayer network model of movie script. Springer. pp 782–796. https://doi.org/10.1007/978-3-030-05411-3_62
Nadeau D, Sekine S (2007) A survey of named entity recognition and classification. Lingvisticae Investigationes 30(1):3–26
Newman ME (2006) Modularity and community structure in networks. Proc Natl Acad Sci 103(23):8577–8582
Orman K, Labatut V, Cherifi H (2013) An empirical study of the relation between community structure and transitivity. Complex Netw:99–110. https://doi.org/10.1007/978-3-642-30287-9_11
Orman K, Labatut V, Cherifi H (2013) An empirical study of the relation between community structure and transitivity. In: Complex Networks. Springer. pp 99–110. https://doi.org/10.1007/978-3-642-30287-9_11
Park S-B, Oh K-J, Jo G-S (2012) Social network analysis in a movie using character-net. Multimed Tools Appl 59(2):601–627
Pastrana-Vidal RR, Gicquel JC, Blin JL, Cherifi H (2006) Predicting subjective video quality from separated spatial and temporal assessment. Hum Vision Electron Imaging XI 6057:60570. International Society for Optics and Photonics
Ren H, Renoust B, Viaud M-L, Melançon G, Satoh S (2018) Generating "visual clouds" from multiplex networks for tv news archive query visualization. In: 2018 International Conference on Content-Based Multimedia Indexing (CBMI). IEEE. pp 1–6. https://doi.org/10.1109/cbmi.2018.8516482
Renoust B, Kobayashi T, Ngo TD, Le D-D, Satoh S (2016) When face-tracking meets social networks: a story of politics in news videos. Appl Netw Sci 1(1):4
Renoust B, Le D-D, Satoh S (2016) Visual analytics of political networks from face-tracking of news video. IEEE Trans Multimed 18(11):2184–2195
Renoust B, Melançon G, Viaud M-L (2014) Entanglement in multiplex networks: understanding group cohesion in homophily networks. Soc Netw Anal Community Detect Evol:89–117. https://doi.org/10.1007/978-3-319-12188-8_5

Rital S, Cherifi H, Miguet S (2005) Weighted adaptive neighborhood hypergraph partitioning for image segmentation. In: International Conference on Pattern Recognition and Image Analysis. Springer. pp 522–531. https://doi.org/10.1007/11552499_58
Salton G, Wong A, Yang C-S (1975) A vector space model for automatic indexing. Commun ACM 18(11):613–620
Sekara V, Stopczynski A, Lehmann S (2016) Fundamental structures of dynamic social networks. Proc Natl Acad Sci 113(36):9977–9982
Sigaki HY, Perc M, Ribeiro HV (2018) History of art paintings through the lens of entropy and complexity. Proc Natl Acad Sci 115(37):8585–8594
Simply Scripts. www.simplyscripts.com. Last Accessed 20 June 2019
Tan MS, Ujum EA, Ratnavelu K (2014) A character network study of two sci-fi tv series. AIP Vol. 1588. pp 246–251. https://doi.org/10.1063/1.4866954
The Internet Movie Script Database (IMSDb). https://doi.org/10.1108/err.1999.3.5.56.52. www.imsdb.com. Last ccessed 20 June 2019
Tran QD, Jung JE (2015) Cocharnet: Extracting social networks using character co-occurrence in movies. J UCS 21(6):796–815
Viard T, Fournier-S'niehotta R (2018) Movie rating prediction using content-based and link stream features. CoRR abs/1805.02893. 1805.02893
Waumans MC, Nicodème T, Bersini H (2015) Topology analysis of social networks extracted from literature. PloS ONE 10(6):0126470
Weng C-Y, Chu W-T, Wu J-L (2009) Rolenet: Movie analysis from the perspective of social networks. IEEE Trans Multimed 11(2):256–271
Yang L, Tang K, Yang J, Li L-J (2017) Dense captioning with joint inference and visual context. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Vol. 2. https://doi.org/10.1109/cvpr.2017.214
Yang S, Luo P, Loy CC, Tang X (2016) Wider face: A face detection benchmark. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr.2016.596
Yeung M, Yeo B-L, Liu B (1996) Extracting story units from long programs for video browsing and navigation. Multimedia Computing and Systems, 1996. In: Proceedings of the Third IEEE International Conference on. IEEE. pp 296–305. https://doi.org/10.1016/b978-155860651-7/50117-0
Yuepeng L, Cui J, Junchuan J (2015) A keyword extraction algorithm based on word2vec. e-Sci Technol Appl 4:54–59

Publisher's Note