Extraction and Analysis of Fictional Character Networks

Vincent Labatut
Laboratoire Informatique d'Avignon (LIA)
339 chemin des Meinajariès, BP 1228
84911 Avignon Cedex 9, France
vincent.labatut@univ-avignon.fr

A *character network* is a graph extracted from a narrative, in which vertices represent characters and edges correspond to interactions between them. As shown by the expanding related literature, a number of narrative-related tasks can be conducted by considering character networks as a relational modeling framework of the narrative. I distinguish three main types of approaches.

First, in the context of *Narrative Analysis*, character networks are generally used to describe the plot structure at various levels, detect relevant patterns and narrative events, identify character roles, assess the validity of literary theories, and produce graphical representations. They are used not only to describe individual plots, but also to perform comparative analysis.

Second, another category of works also adopts a descriptive and comparative approach, but relying on a *Complex Systems* paradigm. These authors consider that character networks are a type of *Complex Network*, and as such they apply the standard tools developed to analyze them, and/or propose new ones. The network itself is the object of the study. Many articles of this type compare the topological properties of character networks with those of other kinds of complex networks, e.g. real-world social networks, random models, or other fictions.

Third and finally, a number of works originate from the *Artificial Intelligence* domain. They focus more on automating the network extraction process, which requires solving various text, speech, image, and/or video processing problems, depending on the media used. These works also consider character networks as models of the plot, and take advantage of this to automate higher-level tasks, such as genre classification, storyline detection, scene segmentation, summarization, and recommendation systems.

Character networks have many uses, but their reliable extraction from works of fiction is a difficult task, especially in an automatic way. This is first because narratives are complex objects conveying many forms of information, both explicit and implicit. Moreover, works of fiction possess specific properties making them different from other narratives, which adds an extra layer of difficulty by possibly invalidating generic extraction methods. Such differences can be stylistic. For instance in text narratives, literary prose is considered as more complex than journalistic prose. This significantly affects the performance of generic methods on a variety of NLP tasks, including text summarization, named entity recognition, and co-reference resolution. Similarly in video narratives, generic tools are often designed and trained to handle only very controlled conditions. For instance, face recognition software work for certain lights, colors, and angles, which are too limited when considering the complexity of film grammar.

This talk aims at presenting the methods used in the literature to extract, analyze, and harness character networks from works of fiction. We will first describe the extraction process in a generic way, and explain how its constituting steps are implemented in practice, depending on the medium of the narrative, the goal of the network analysis, and other factors. We will then review the descriptive tools used to characterize these networks, with a focus on the way they are interpreted in this context. We will illustrate the relevance of character networks by also providing a review of applications derived from their analysis. Finally, we will identify the limitations of the existing approaches, and the most promising perspectives.