Network Analysis of the Matrix Trilogy

Nataliia Romanenko[1], Cole Johnson[1], Sawyer Marchand[1], Dr. Megan Heenehan[2], Dr. Garrett M. Dancik[3]

Program in [1]Computer Science; [2]Department of Mathematical Sciences and [3]Department of Computer Science

Eastern Connecticut State University, Willimantic, Connecticut, USA

Introduction

Network science is an interdisciplinary and growing field that studies complex networks by combining graph theory and data science to gain insight into the connections and relationships around us. Here we present both a network analysis and a sentiment analysis of the Matrix trilogy: The Matrix (1999), The Matrix Reloaded (2003), and The Matrix Revolutions (2003).

Background

Network analysis of character interactions in a movie is used to answer questions about the social structure of the movie and to identify different communities of characters and relationships. This analysis helps to better understand the social structure and social dynamics across movies.

Sentiment analysis is used to identify and quantify affective states and subjective information through text analysis. We used sentiment analysis to better understand the emotional valence of main characters across movies.

Methods

The programming language R was used for data collection, data cleaning, statistical and sentiment analysis. The software package Gephi was used for network analysis and visualization.

A network for each movie was created based on character interactions extracted from the movie scripts. Each node in the network is a character in the movie, and interactions between characters are represented by edges between nodes.

We identified important characters by calculating different network measures such as weighted degree centrality (number of interactions) and betweenness centrality (the amount of influence a node has over the flow of communication).

Modularity was used to identify communities (groups of nodes that are more interconnected).

Results

Sentiment Analysis

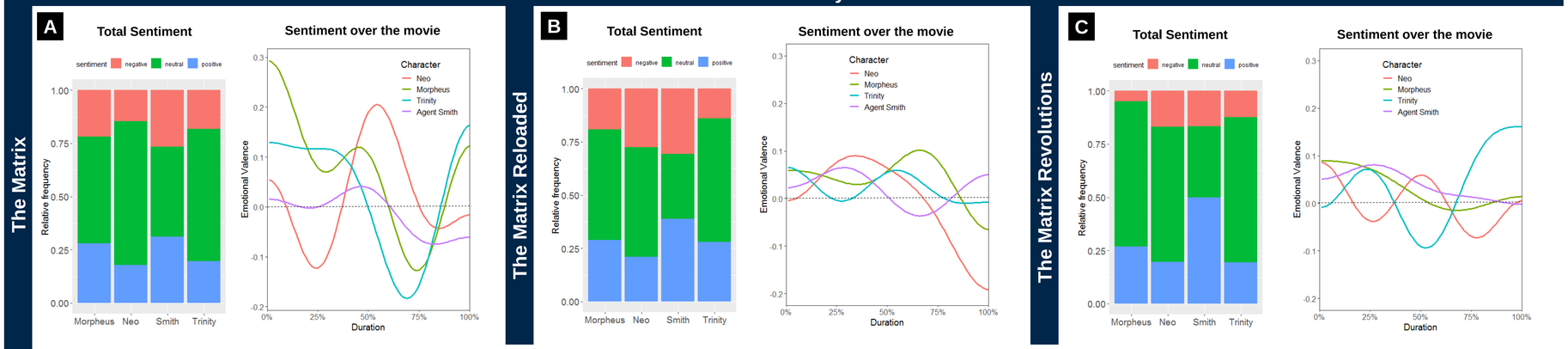

Figure 1. Sentiment analysis of the main characters (Neo, Trinity, Morpheus, Agent Smith) in (A) The Matrix, **(B)** The Matrix Reloaded, and **(C)** The Matrix Revolutions.

Movie Networks

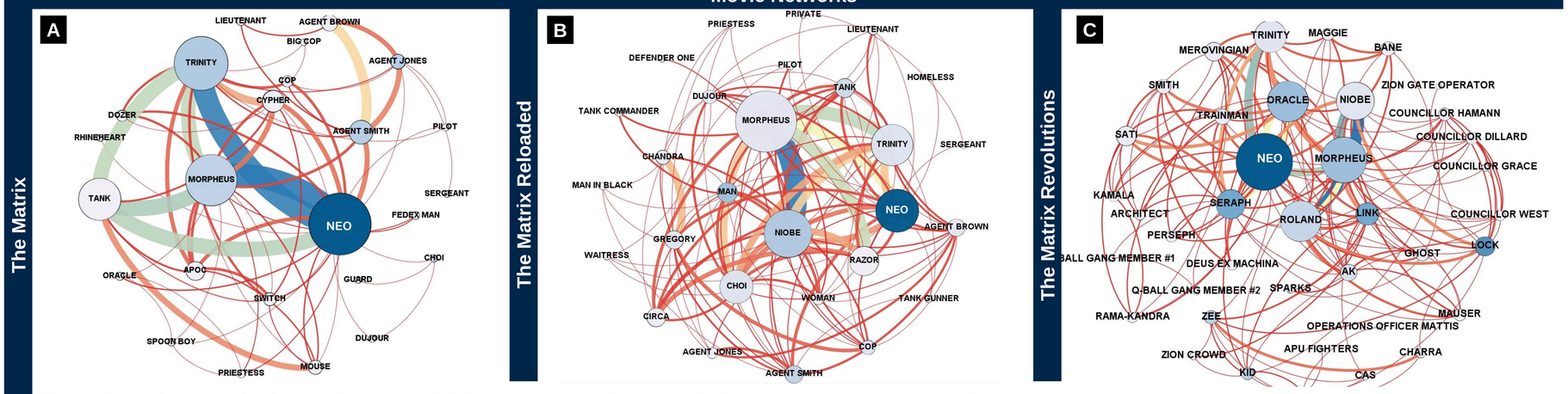

Figure 2. Movie Networks for (A) The Matrix, **(B)** The Matrix Reloaded, and **(C)** The Matrix Revolutions. Node size corresponds to weighted degree, node color reflects betweenness centrality (the darker the color the higher the betweenness centrality). Edge color and thickness reflect the number of interactions between characters.

Communities

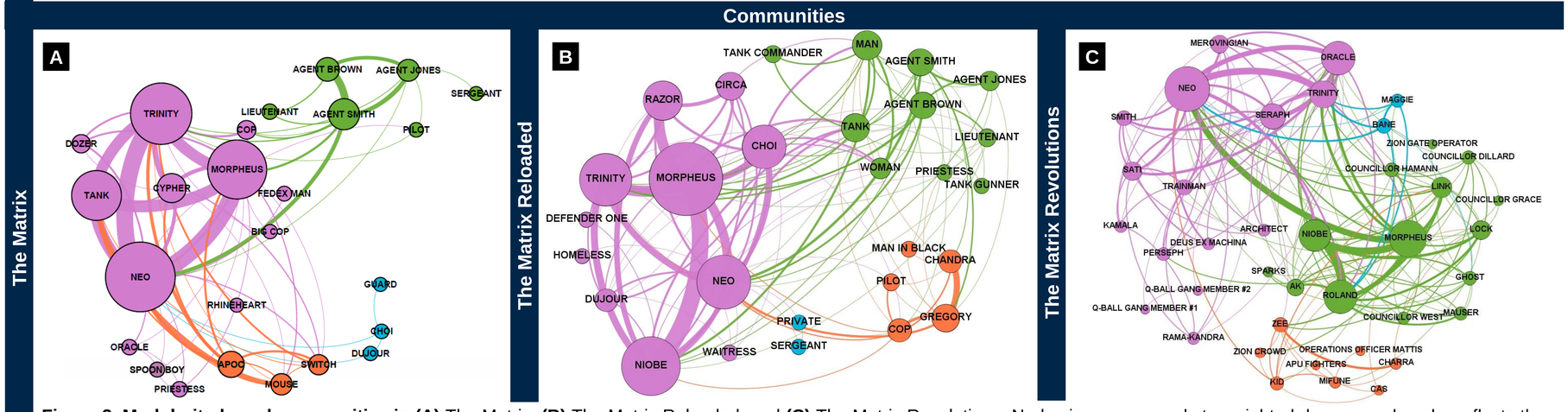

Figure 3. Modularity-based communities in (A) The Matrix, **(B)** The Matrix Reloaded, and **(C)** The Matrix Revolutions. Node size corresponds to weighted degree, node color reflects the community to which the character belongs. Edge color is based on community and edge thickness reflects the number of interactions between characters.

References: Barabási, Albert-László and Pósfai, Márton. Network science. Cambridge: Cambridge University Press, 2016.

The scripts for analysis were downloaded from the following sources: http://www.scifiscripts.com/scripts/matrix_97_draft.txt (The Matrix), http://nldslab.soe.ucsc.edu/charactercreator/film_corpus/film_2012xxxx/imsdb.com/Matrix-Reloaded_-The.html (The Matrix Reloaded), and https://www.matrixfans.net/movies/the-matrix-revolutions/transcript/ (The Matrix Revolutions).