

VISUAL SUMMARIZATION OF STORIES

Singapore Science and Engineering Fair 2019

Jiang Zhiheng, Tan Yi Kai, Huang Junwei

SS014

Introduction

Visual Summarization of Stories has the aim of creating network summaries of the characters from three different detective stories, A Study in Scarlet by Arthur Conan Doyle, The Mysterious Affair at Styles by Agatha Christie and The Circular Staircase by Mary Roberts Rhinehart, to observe patterns in these stories.

Definitions:

Agents - Characters who carry out actions and influence the storyline due to active role in story

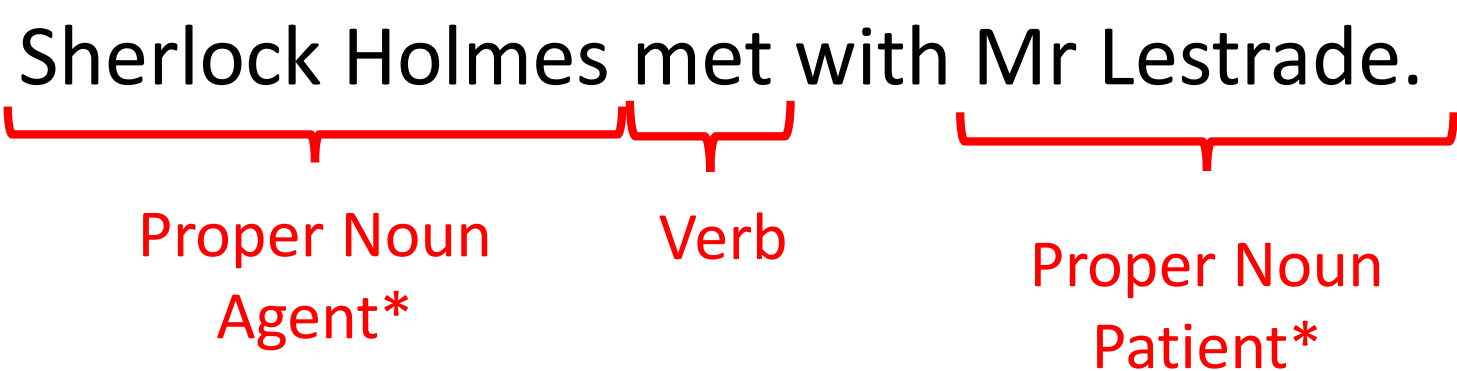
Patients - Characters who are acted upon by agents and have a more passive role in the story

Methodology

Data Processing:

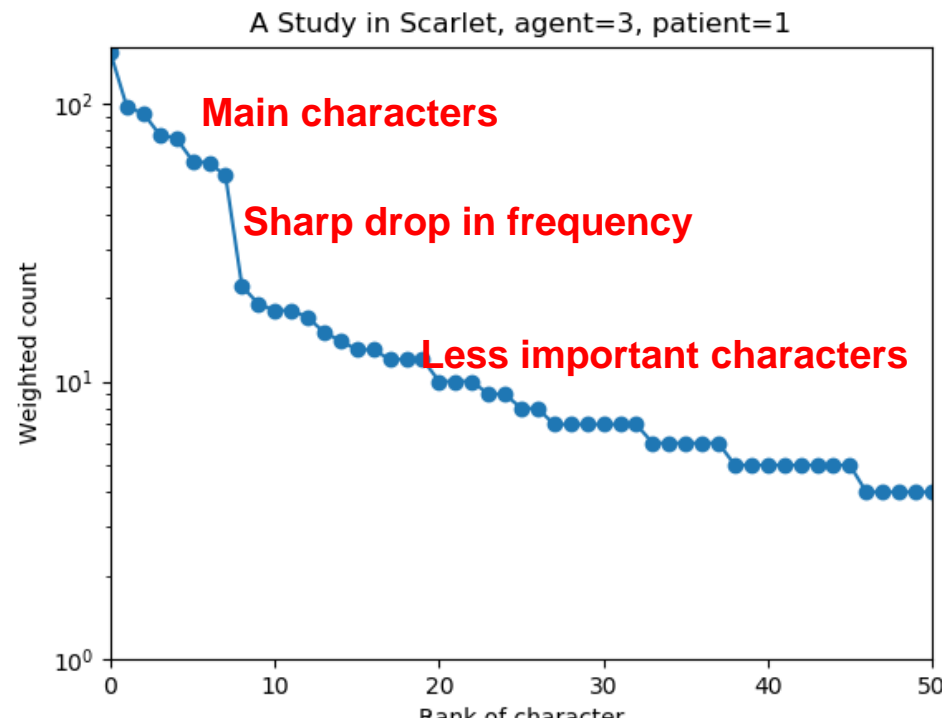
All texts are cleansed from Project Gutenberg, where books have expired copyrights and can be downloaded legally. See “Data” for more details.

Identification of Key Entities:

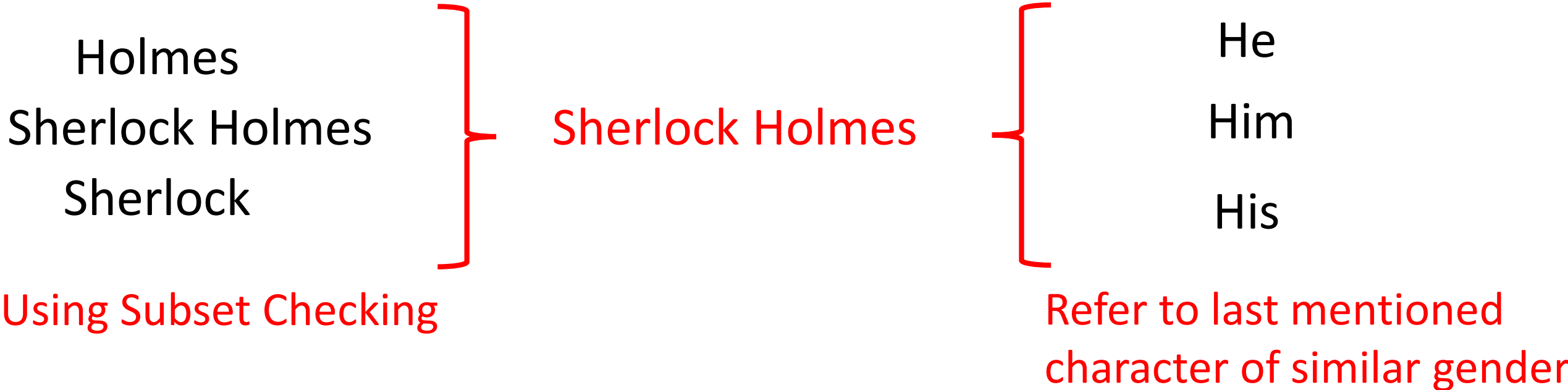


ABOVE: Agents are identified by Proper Noun + Verb combination, if character is not an agent, he is identified as a patient, using Python NLTK** word tagging.

RIGHT: We plotted a log-y frequency graph based on sorted list of characters to identify the main characters before the sharp drop in frequency.



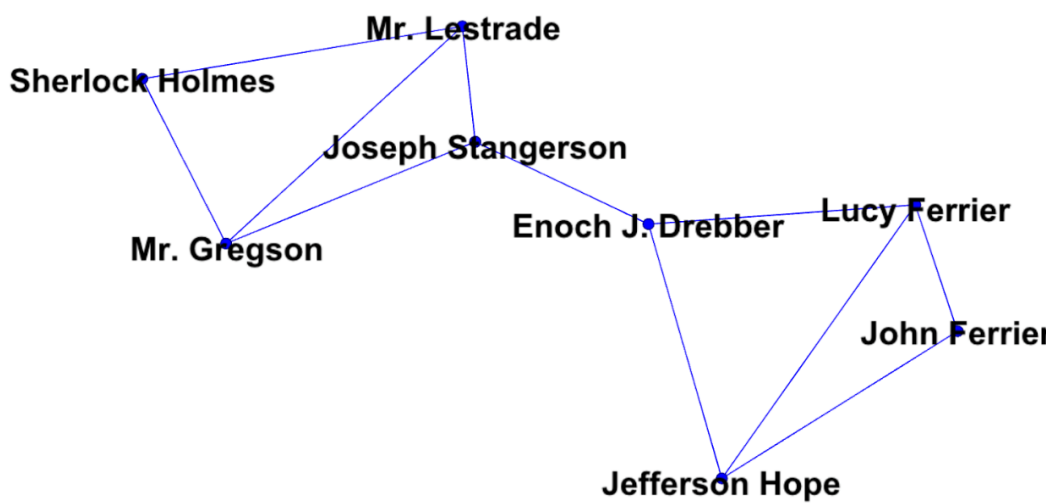
Disambiguation:



ABOVE: A character can be referred to in many ways, disambiguation ensures that the same character is not counted as separate characters.

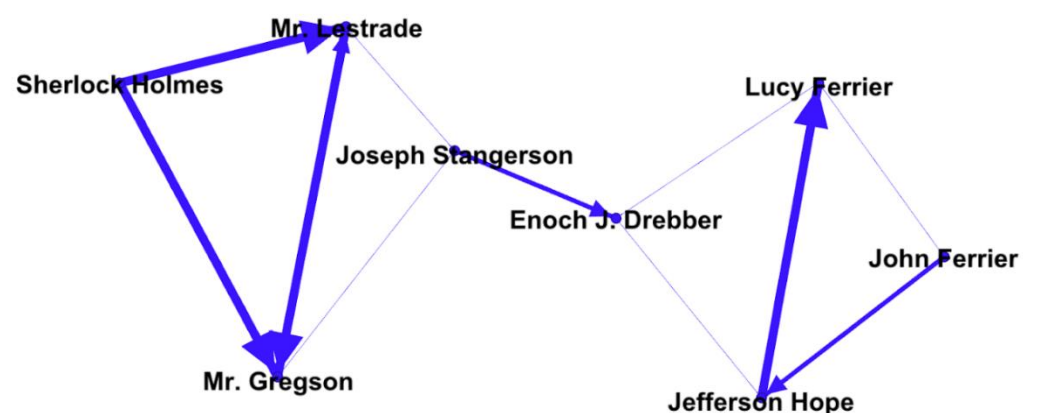
Construction of network:

Unweighted Undirected Network



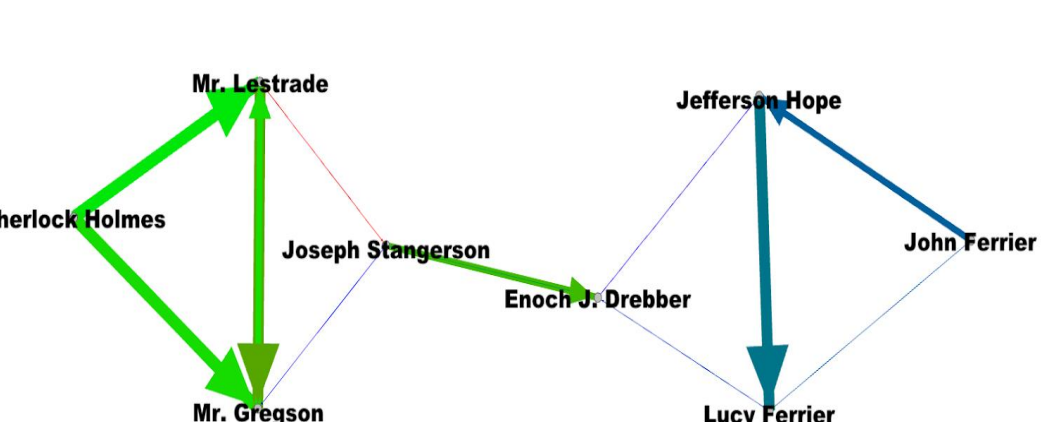
We plot out the main characters as nodes. If there are interactions between two characters in a sentence, we draw an edge.

Weighted Directed Network



We add weights to the edges, symbolizing the frequency of interactions. If the character is an agent, it is represented by an outgoing arrow.

Coloured Weighted Directed Network



We use VADER Sentiment Analysis to find average sentiment of interactions between any two characters, and colour the edges, red being most negative, blue being most positive and green being neutral. Graph was created using Gephi.

*Refer to the definition of agents and patients in “Introduction” section

**Natural Language Tool Kit

Data

Title of Book	A Study in Scarlet	A Mysterious Affair at Styles	The Circular Staircase
Author	Arthur Conan Doyle	Agatha Christie	Mary Roberts Rinehart
First published	1887	1920	1908
First publisher	Ward Lock and Co.	John Lane	Bobbs-Merrill
Source	https://www.gutenberg.org/ebooks/244	https://www.gutenberg.org/ebooks/863	https://www.gutenberg.org/ebooks/434
No. of Words	43,414	53,507	56,378
Format			.txt

Results and Discussion

Key:

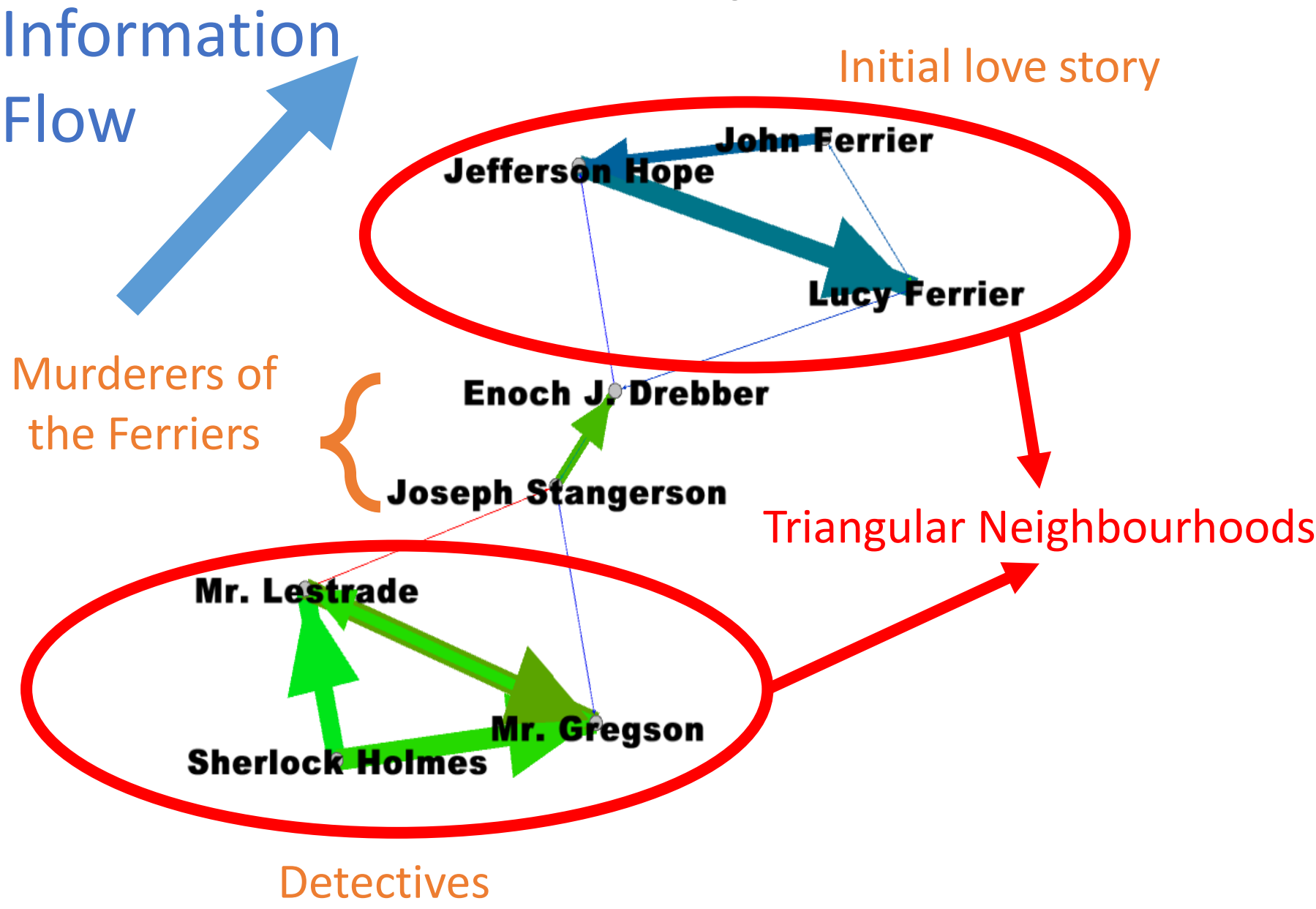
Outgoing edge: Character is exercising agency (carrying out an action) on another character

Ingoing edge: Character is being acted upon and is therefore defined as a patient

Colour of edge: Sentiment of each interaction from red (most negative) to blue (most positive)

Thickness of each edge: The thicker the edge, the greater the frequency of interaction

Network Representation of A Study in Scarlet



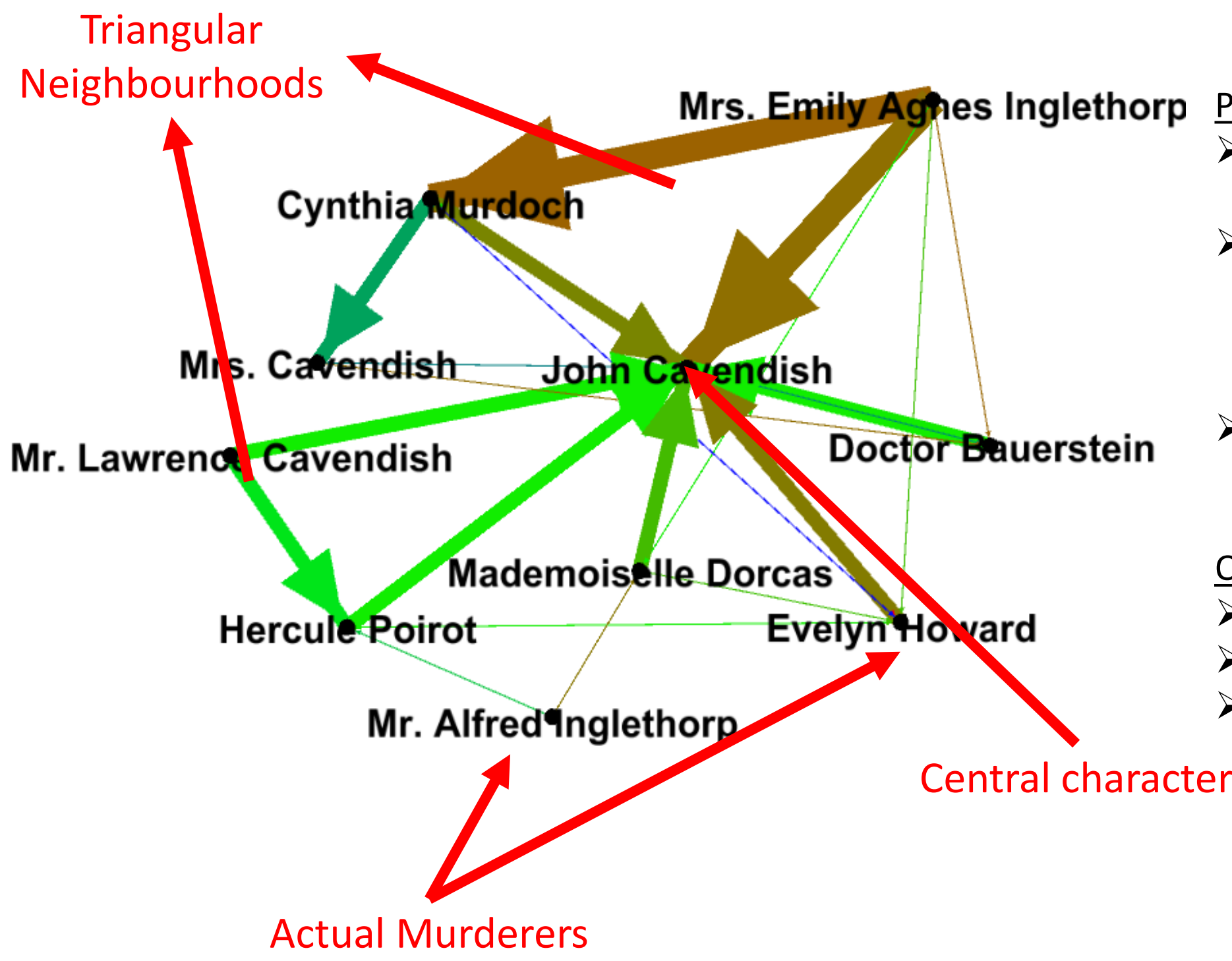
Plot outline and Discussion

- Hope was in love with Lucy Ferrier, daughter of John Ferrier as shown by positive(blue) relationships
- Stangerson, with his accomplice Drebber, shown by outgoing green arrow showing agency over Drebber, caused the deaths of the Ferriers
- Hope plotted a murder to kill Stangerson and Drebber for revenge

Observations

- Two main network neighbourhoods
- Information flow from detective neighbourhood to the Hope-Ferrier neighbourhood via the link between Stangerson and Drebber
- The detectives find out about Hope via the murdered Stangerson and Drebber

Network Representation of The Mysterious Affair at Styles



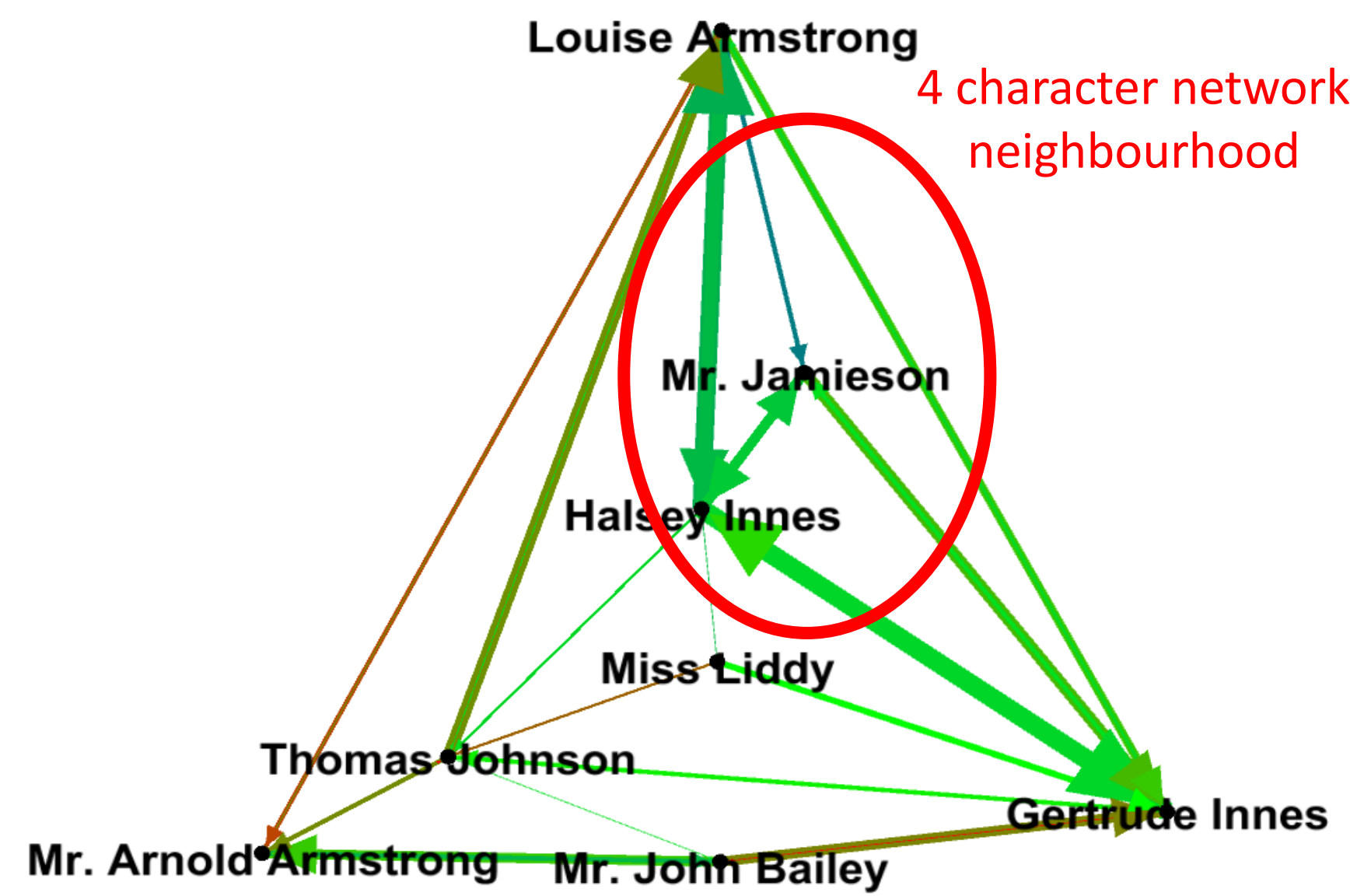
Plot outline and Discussion

- Mrs Cavendish (later Emily Inglethorp) inherits late husband's wealth and was poisoned
- Evelyn Howard and Alfred Inglethorp work together to murder Mrs Cavendish, but it did not appear they worked together until the end of the story, explaining no link between characters at all
- John Cavendish was suspected by many therefore he appears to be the central character

Observations

- Two main network neighbourhoods
- Characters are connected to a central character
- Actual murderer (Alfred Inglethorp) seems to have few links with the crime while accomplice Evelyn Howard exercises agency to frame John Cavendish, as shown by outgoing arrow from Howard to John Cavendish

Planar Network Representation of The Circular Staircase



Plot outline and Discussion

- Plot outline was very confusing
- Mixture of romance and murder story as it had romantic subplots, could not identify similar patterns to the other detective stories
- A planar graph was generated from original graph as shown on the right to remove less significant edges, as original graph had too many interactions, making it too confusing to interpret

Observations

- Four-character network neighbourhood between Louise Armstrong, Mr Jamieson, Halsey Innes and Gertude Innes
- Murderer was not present as it was diluted by a love story
- Halsey Innes and Gertude Innes shared strong bidirectional links as they were siblings, may have exercised joint agency, where characters act together
- Rachel Innes (narrator) is the protagonist, which does not appear in this graph as it would dilute all other interactions

Conclusion

The three detective stories we have chosen are in increasing order of difficulty. For the first two stories, we have successfully analyzed the plot of the story and identified the main network neighbourhoods. We could clearly see how the characters interacted with each other via the network, and we also observed the degree of agency of each character to give us an accurate visual summary of the story, after comparing our summary to the actual plot of the story. However, for the third story, the interactions were confusing and difficult for our method to understand, especially when the murder is mixed together with romantic subplots, so we could not identify similar patterns as other stories.

All diagrams and figures were self-created unless explicitly stated otherwise.