Graphing the Mystery: Christie's Narrative Dynamics

Siddheshwari Bankar
Department of Computer Science
Seattle University, Seattle, WA, USA
sbankar@seattleu.edu

ABSTRACT

In the confluence of literary analysis and graph theory, character interactions form the backbone of narrative structures, similar to networks in social and biological domains. This study leverages the classic detective novel "The Murder of Roger Ackroyd" by Agatha Christie to construct a character interaction network, applying a novel metric for quantifying the narrative distance between characters. The metric, based on character co-occurrences within the text, adheres to the foundational principles of distance metrics, ensuring non-negativity. Utilizing this metric, we build a weighted graph to encapsulate the narrative's intricate web of relationships, further subjecting it to community detection via the Louvain community detection algorithm. The result is a visualization of the story's social fabric, which reveals clusters of interaction and character centrality. This approach not only offers a quantitative lens for literary analysis but also enhances our understanding of narrative dynamics through the prism of graph theory.

INTRODUCTION

In this project, we look at how characters in a book are connected, similar to how friends are connected on social media. Imagine each character in a story is a point or dot, and every time two characters appear together, it's like they have a line connecting them. We're using a famous mystery book by Agatha Christie, "The Murder of Roger Ackroyd," as our example.

Network analysis is a fancy term for studying how these nodes (characters) and edges (their connections) form a pattern or network. Just like scientists study networks or patterns in things like the internet or groups of friends, we're doing the same with the story's characters. We're especially interested in something called random graphs, which are like puzzles or models that help us understand how these patterns might look if we just stumbled upon them without knowing the story.

Our main goal is to see which characters hang out together more often, forming their little groups or communities within the book. To do this, we use a special distance metric to measure how close or far apart the characters are in the story. Then, we use a technique, known as the Louvain community detection algorithm, to find these communities.

Think of it like watching who sits with whom in a cafeteria. Some people always sit together, and they form a group or community. By the end of this project, we hope to see who the main groups are in "The Murder of Roger Ackroyd" and how these groups help tell the story.

BACKGROUND

Agatha Christie, an iconic figure in the literary world, is celebrated for her ingenious contributions to the mystery genre. Her novel "The Murder of Roger Ackroyd," first published in 1926, stands as a testament to her mastery of suspense and her innovative use of narrative techniques. Christie's work is characterized by its intricate plots, memorable characters, and unexpected twists, elements that have cemented her status as the "Queen of Crime."

"The Murder of Roger Ackroyd" unfolds in the tranquil English village of King's Abbot, where the peaceful life is shattered by a series of disturbing events. The story, narrated by Dr. James Sheppard, revolves around the mysterious death of Roger Ackroyd, a wealthy widower. Ackroyd's demise follows closely on the heels of his fiancé, Mrs. Ferrars, who confesses to a murder before taking her own life. The plot thickens with blackmail, hidden identities, and intricate alibis, drawing the retired detective Hercule Poirot back into the fray. Christie masterfully leads the reader through a labyrinth of clues and misdirection, culminating in a groundbreaking twist that challenges the conventions of the mystery genre.

METRIC

In this project, we utilized a distance metric to quantify the narrative distance between characters in a literary work, applying it to Agatha Christie's "The Murder of Roger Ackroyd." This metric,

D(x, y), is defined as,

$$D(x, y) = \frac{C(x) + C(y) - 2C(x, y)}{\max(C(x), C(y))}$$

where C(x) and C(y) denote the count of paragraphs in which characters (x) and (y) appear, respectively, and C(x, y) represents the number of paragraphs where both characters appear together. This metric is designed to satisfy the following essential properties of a distance metric:

1. D(x,y) = 0 if and only if x=y
2. D(x,y)>0 if and only if x and y are not the same.
3. D(x,y) = D(y,x)

These conditions are satisfied as below;

- D(x, y) = 0 if and only if (x = y). This property is inherent in our metric since when (x = y), C(x) = C(y) = C(x, y), leading to D(x, y) = 0. Conversely, if D(x, y) = 0, it implies C(x) + C(y) = 2C(x, y), which can only be true if (x) and (y) always appear together, effectively making them indistinguishable within this context.
- D(x, y) = D(y, x) for all (x, y). The symmetry of our metric is straightforward since the formula does not depend on the order of (x) and (y), thereby satisfying this condition by design.

- If x != y the D(x,y) > 0, since x and y are not same there will be some distance between them which is greater than 0 and they will have wvalues for C(x), C(y) and C(x,y), which implies that D(x,y) is positive.

METHODS AND EXPERIMENTS:

In the computational analysis of "The Murder of Roger Ackroyd," the narrative structure is transposed into a network of character interactions to examine the underlying connectivity and community dynamics.

The methodology involves two primary data sets: nodes.csv, containing characters and their occurrence frequencies, and edges.csv, which quantifies the interactions between characters based on their co-occurrences within paragraphs. A Python function is developed to transform this interaction data into a graph format suitable for computational analysis. This conversion relies on a variable threshold, facilitating the construction of multiple graphs that reflect varying intensities of character interactions.

For each threshold-specific graph, network characteristics are computed, including the number of edges, average degree, global clustering coefficient, size of the giant component, average path length, and network diameter. Additionally, the connectivity of each graph is determined to evaluate the consistency of the network's structure as the threshold varies.

The networkx library is instrumental for managing graph-related operations, while the community library provides the means for community detection, employing the Louvain method. This algorithm segments the character network into distinct communities, identifying clusters based on narrative proximity.

The culmination of the process involves the integration of community labels into the original node dataset, creating an augmented dataframe that is then exported as nodes_with_communities.csv. This file serves as the nexus between computational analysis and visualization in Gephi, where the community structures embedded within Agatha Christie's narrative are subject to further exploration and interpretive analysis.

RESULTS

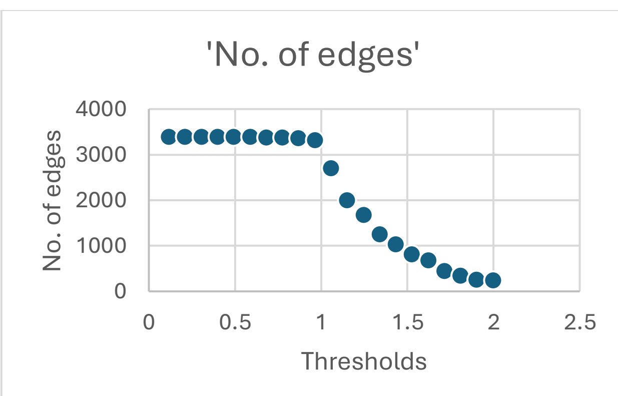
GNC – Graph Not Connected

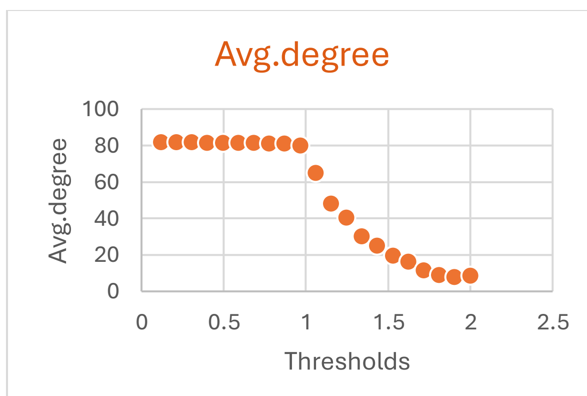
Thresholds	No. of edges	Avg.degree	Global Clustering co-eff.	Size of Giant Component	Avg. Path length	Network Diameter	Connected?
0.12	3385	81.5662	0.9946	83	1.0052	2	Yes
0.214	3384	81.5421	0.9943	83	1.0055	2	Yes
0.308	3382	81.4939	0.9937	83	1.0061	2	Yes
0.402	3380	81.4457	0.9931	83	1.0067	2	Yes

0.496	3379	81.4216	0.9929	83	1.0070	2	Yes
0.59	3377	81.3734	0.9923	83	1.0076	2	Yes
0.684	3374	81.3012	0.9914	83	1.0085	2	Yes
0.778	3365	81.0843	0.9889	83	1.0111	2	Yes
0.872	3354	80.8192	0.9857	83	1.0143	2	Yes
0.966	3310	79.7590	0.9738	83	1.0273	2	Yes
1.06	2695	64.9397	0.8905	83	1.2139	3	Yes
1.154	1996	48.0963	0.8249	83	1.5148	4	Yes
1.248	1673	40.3132	0.8122	83	1.7549	5	Yes
1.342	1250	30.1204	0.7838	83	2.1777	6	Yes
1.436	1028	24.7710	0.7809	83	2.7337	9	Yes
1.53	805	19.3975	0.7217	83	3.6038	12	Yes
1.624	670	16.1445	0.6846	83	4.6926	15	Yes
1.718	442	11.1898	0.7354	54	GNC	GNC	No
1.812	334	8.6753	0.7425	20	GNC	GNC	No
1.906	253	7.5522	0.6991	20	GNC	GNC	No
2.0	239	8.5357	0.7084	20	GNC	GNC	No

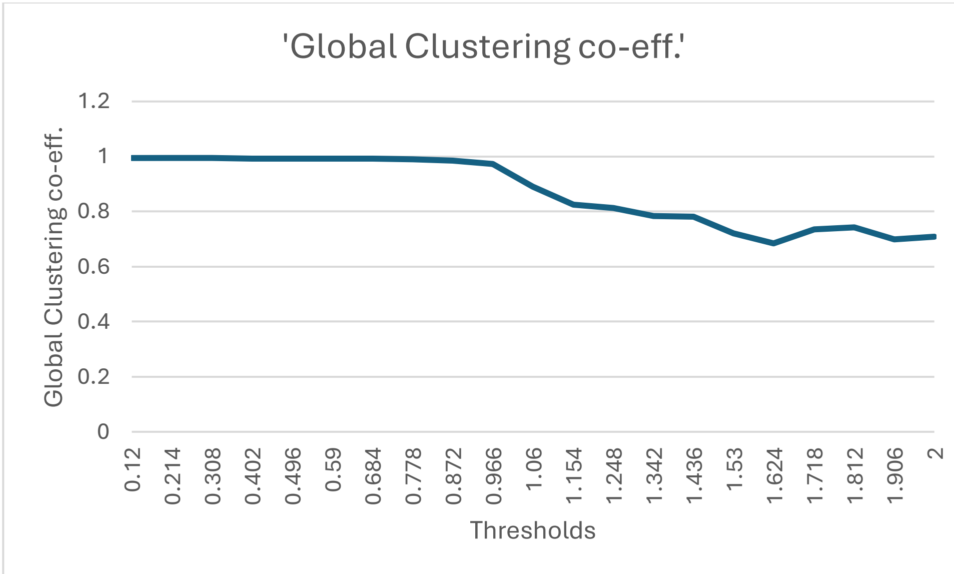
In this study we consider the smallest threshold as T_{min} and the largest threshold as T_{max} with values 0.12 and 2.0 respectively. The step size chosen is 0.094 in order to view changes in the graph smoothly. Having 0.094 step size provides with 21 thresholds until T_{max}. From the table above we have following observations:

As the threshold value increases, the number of edges in the network consistently decreases as shown in the figure below. This trend suggests that fewer character pairs have a narrative distance that falls below the higher thresholds, indicating that the characters are less closely connected in the context of the story. And because the number of edges are decreased, the average degree also decreases.





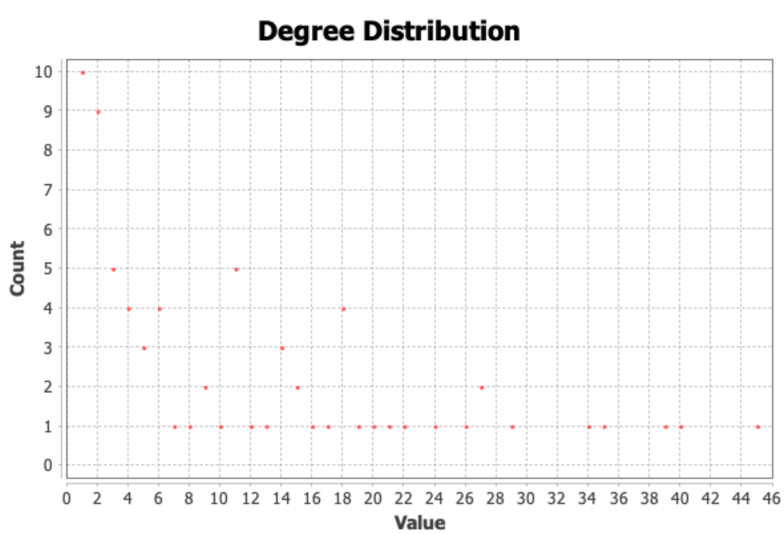
The clustering coefficient generally decreases as the threshold increases. At lower thresholds, there are more tightly-knit clusters of characters. As the threshold increases, the network becomes less dense, and the clustering coefficient decreases, indicating a sparser and less interconnected network.



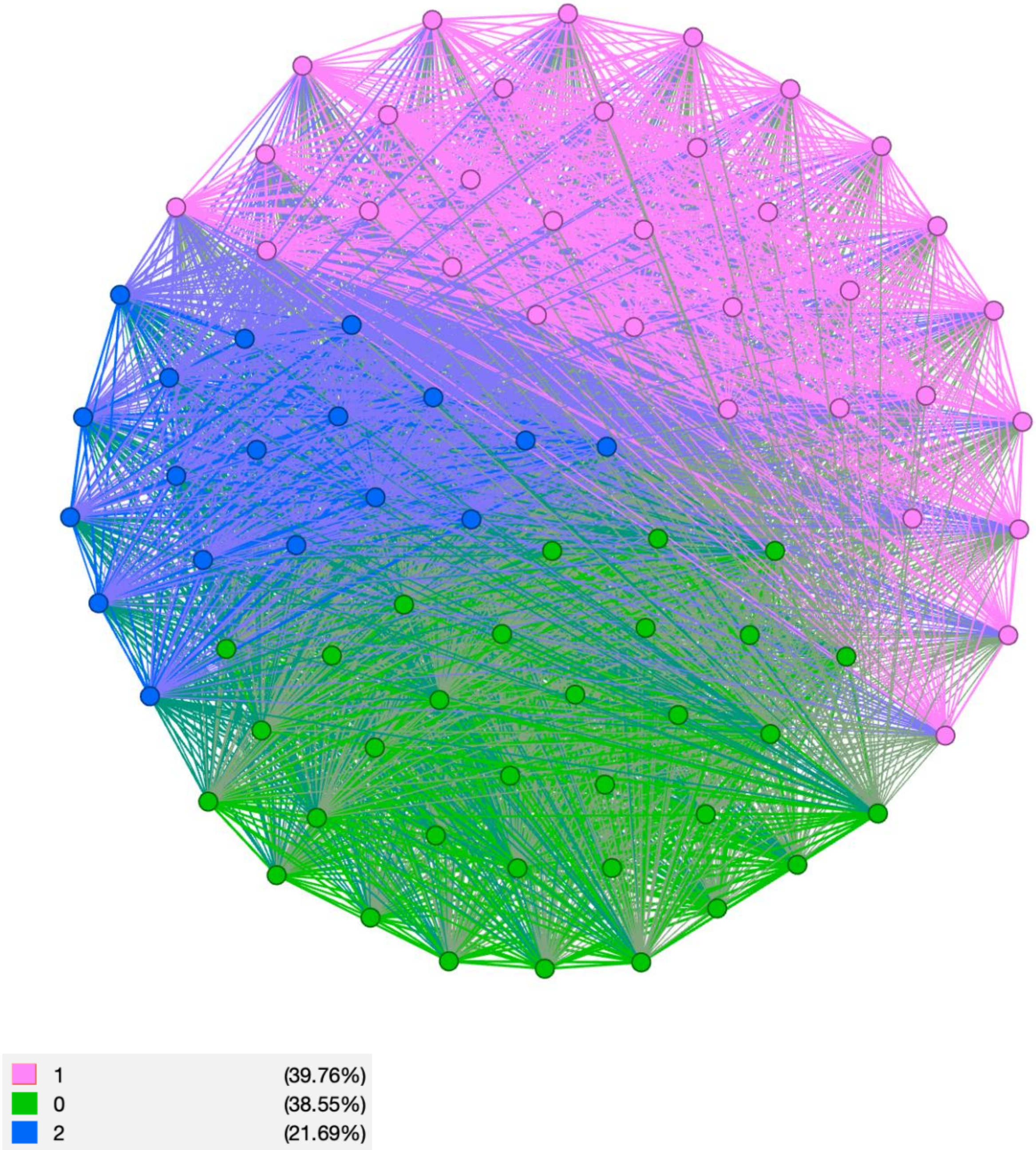
The size of the giant component remains constant at 83 until the threshold reaches 1.718, beyond which the network is no longer connected, and the giant component size drops sharply, indicating the presence of isolated subgraphs or individual nodes.

Both the average path length and the network diameter increase with the threshold. This means that, on average, characters become farther apart in terms of the narrative path as the threshold increases, and the maximum narrative distance between the most distant characters also increases.

If the giant component of a network contains at least 75% of the characters, this suggests a level of connectivity throughout the network. However, determining whether the degree distribution follows a power-law distribution cannot be concluded solely on this basis. In a power-law distribution, we would expect a smoother, long tail off to the right, representing the few nodes with very high degrees.



COMMUNITY DETECTION



Community	Characters
Community 0	Charles Kent, Paton, Ursula, Miss Russell, Major Blunt, Roger Ackroyd, Ralph, Poirot, Dr. Sheppard, James, Flora, Hector Blunt, Fernly, Ralph Paton, Raymond, Inspector Raglan, Sheppard, Miss Ganett, Miss Ackroyd, Hammond, Geoffrey Raymond, Hercule Poirot, Ursula Bourne, Blunt, Ackroyd, M. Poirot, Raglan, Roger, Melrose, Parker, Flora Ackroyd, Caroline, Ferrars

Community 1	Mademoiselle, Mademoiselle Flora, Captain Paton, Davis, Folliott, Russell, Mrs. Ferrars, Miss Flora, Elsie Dale, Abbot, Mah Jong, Major Ellerby, Carter, Annie, Ashley Ferrars, Porrott, Hastings, Miss Flora Ackroyd
Community 2	Hector, Cecil Ackroyd, Mélisande, Clara, Colonel Carter, Miss Sheppard, Rafael Sabatini, Inspector Davis, Cecil, Chow, Diana, Ursula Paton, Miss Caroline, RAGLAN, Monsieur, Hayes, Papa Poirot, Pung, Eh bien, Paddock, Colonel Melrose, ROGER ACKROYD, Ellerby, Bourne, Mademoiselle Caroline, M. Raymond, George Eliot, Elsie, M. le, Mark, ACKROYD, M. l'Inspecteur

For Community 0, we see central characters like Hercule Poirot, Dr. Sheppard, and Roger Ackroyd listed, along with several other characters who are significantly involved in the main storyline of the novel. This community seems plausible as it groups together characters that are central to the plot's development, which likely means they interact more frequently within the narrative. Since the story revolves around the murder of Roger Ackroyd and Hercule Poirot's investigation, it's reasonable to find these characters within the same community.

Community 1 appears to group characters that might share secondary storylines or less frequent interactions with the central figures compared to Community 0. This could indicate a peripheral narrative circle, which the algorithm has appropriately identified as a separate community.

Community 2 contains names that appear to be minor characters or those with less direct involvement in the story's central events. This might suggest that these characters form a narrative background or have interactions that do not heavily influence the main plot, hence forming their distinct community.

The Louvain algorithm uses modularity optimization to maximize the density of connections within communities while minimizing connections between communities. Given this methodology, the algorithm would likely group characters into communities based on the frequency and strength of their interactions.

CONCLUSION

Throughout this analysis, we explored the intriguing community structure within Agatha Christie's "The Murder of Roger Ackroyd" using the Louvain algorithm for community detection. By applying a specific threshold to filter the relationships between characters, we successfully identified distinct communities that reflect the complex web of interactions and relationships central to the narrative. This process not only showcased the novel's intricate character dynamics but also highlighted the potential of network analysis in understanding literary texts.

Observations revealed that characters closely connected within the story were grouped into coherent communities, underscoring the algorithm's effectiveness in capturing narrative relationships. This method offers a novel lens through which to examine the structural elements of storytelling and character development.

For future work, applying this community detection approach to a broader corpus of literary works could yield fascinating insights into narrative structures across genres and authors. Additionally, integrating more sophisticated natural language processing techniques to quantify character interactions could further refine our understanding of community dynamics within literature.

For future work on this project, a comparative analysis across various literary works could be conducted to explore commonalities and differences in character network structures. This could involve expanding the scope to include multiple books by Agatha Christie or contrasting authors from different genres or time periods. Implementing more advanced natural language processing techniques to refine character interaction detection could also provide deeper insights. Additionally, exploring temporal network analysis to see how character interactions evolve over the course of a story could reveal new dimensions of narrative development. Finally, investigating the influence of network structures on readers' engagement and plot perception could bridge the gap between literary theory and computational analysis.

REFERENCES

[1] Gilbert., E. N. (1959). Random graphs. Ann. Math. Stat. 30, 4, 1141–1144.
[2] Yu Xie, Zhiguo Qin, Maoguo Gong, Bin Yu, and Jiye Liang. 2023. Random Deep Graph Matching. IEEE Trans. on Knowl. and Data Eng. 35, 10 (Oct. 2023), 10411–10422.