Social Network Analysis for One Piece

Qingwen Zeng[a,1], **Qiwen Huang**[a,2], **and Adrian Lopez Pirvu**[a,3]

This manuscript was compiled on January 11, 2025

One Piece is one of the largest and most influential anime series today. With a serialization span of over 27 years, it has produced more than 1,000 episodes, and over 1,100 chapters, and features a vast array of characters. However, gaining a deep understanding of the character relationship network within such a vast story arc and across the overall storyline is a challenge. To address this, this paper employs web scraping techniques to extract detailed information from the One Piece Wiki Fandom pages. Using this data, we construct social network graphs that reflect the relationships both within individual story arcs and across the overall storyline. We analyze the internal network using centrality and betweenness centrality metrics and apply network degree distribution, Louvain community detection, and sentiment analysis methods to explore the external relationships within the broader narrative. Ultimately, we identify key characters within individual arcs, uncover character relationship communities in the external storyline, and examine the connections between community networks and the narrative, along with the overall sentiment dynamics of these communities. Our analysis offers new insights into the complex web of relationships that drive the series' overarching narrative.

Social network analysis | Community detection | Sentiment analysis

One Piece's expansive narrative and rich character development make it one of the most popular and enduring series in modern anime. Spanning 27 years, with over 1,000 episodes and 1,100 chapters, the series features hundreds of characters and tightly connected storylines. However, the complexity generated by this large accumulation of text over such a long period of time poses significant challenges for analyzing character relationships, narrative structure, and overarching themes.

Social network analysis provides a powerful framework for understanding the crisscrossing complex relationships in complex systems. By treating characters as nodes and their interactions as edges, it is possible to quantitatively explore relationships, influences, and structures in large-scale narratives. Recent studies have successfully applied similar methods to analyze social structures in other media forms such as providing insights into central characters, community formation, and sentiment trends. (1)(2)(3)

Based on these methods, this study aims to systematically study social networks in One Piece using data from the One Piece Wiki page. We focus on two perspectives: the internal network, which uses centrality and betweenness centrality to tease out character relationships in a single storyline by studying character data in a single arc, and the external network, which uses degree distribution modeling, community detection, and sentiment analysis to capture the overall relationship of the entire narrative by combining all character data at different story stages. We aim to discover key characters, narrative structure, and emotional tone that define the storytelling of One Piece to provide a deeper analysis of work with a simple and interesting storyline but diverse and complex character portrayals.

Our results reveal important patterns in the evolution of the social network, including the increase in character interactions over time, the emergence of tightly connected communities, the changes in the interactions of key characters across storylines, and the changes in the fitted model of the network. In addition, community segmentation and TF-IDF analysis reveal the connections between different node groups and the storyline of the anime, and sentiment analysis highlights the overall neutral-optimistic tone of the series. These findings not only enhance our understanding of One Piece but also contribute to the broader discussion of using computational methods to analyze complex narratives.

Results

Data Collection and Processing.

Data Collection. This paper uses the List of Canon Characters page from the One Piece Wiki to gather information about

Significance Statement

This study applies social network analysis and sentiment analysis to unravel the intricacies of the One Piece storyline, focusing on key character dynamics, narrative structures, and emotional tones across arcs. By leveraging metrics like centrality, community detection, and sentiment dynamics, we identify pivotal characters and reveal how the narrative evolves over time—from a centralized network to a more evenly connected structure. This research not only provides a novel analytical framework for understanding fictional worlds but also highlights the power of computational methods in exploring cultural phenomena, offering insights relevant to fields like narrative studies, digital humanities, and network science.

Author affiliations: [a]Technical University of Denmark Lyngby, DK-2800 Kongens Lyngby

1 Q.Z. organized the exploration of the network degree distribution and sentiment analysis of the characters and communities. 2 Q.H. collected the data through web APIs. Build the inner network for each arc and analyze the inner network for each arc. 3 A.L. collected the data through web APIs to build an external network to analyze the overall growth of the external network and partition the external network to explore the cross-relationships between communities and arcs. Three people completed the writing of the paper together and contributed equally throughout the whole project.

The authors declare no conflict of interest.

[1] To whom correspondence should be addressed. Email: s232892@dtu.dk

all the characters and chapter content in One Piece. The One Piece Wiki Fandom is a fan-driven encyclopedia dedicated to the One Piece series, offering detailed information about the series. The List of Canon Characters page lists all characters considered canon in One Piece. By following the character name links on this page, we can be directed to specific character pages. These character pages provide basic information such as the character's name, aliases, backstory, abilities, appearances in chapters or episodes, and relationships with other characters. Using web scraping tools, we constructed an initial social network of all One Piece characters. Each character from the List of Canon Characters page was treated as a node, and links to other characters mentioned on each character's page were used to create edges between them. As a result, we built an initial social network with 1,486 nodes and 28,974 edges.

Data Processing. One of the challenges in understanding the One Piece universe is the sheer number of characters in the story. Many of these characters have significant influence throughout the series, while some contribute little to the narrative or the broader network of relationships. To better understand the network, we designed a filtering method based on four metrics: Out-References, Word Count, In-References, and Intersection References. Specifically, Out-References represents the number of other characters mentioned on a character's wiki page; Word Count is the number of words in the character's wiki entry; In-References refers to the number of characters that mention this character; and Intersection References represents the number of characters that this character mentions, who also mention the original character.

These metrics are combined using a weighted sum, with certain attributes assigned higher weight due to their relevance. For example, In-References are considered more indicative of a character's importance because being mentioned by other characters highlights their influence in the story. Ultimately, each character's score is the sum of their rankings in each metric, ordered in ascending order. The character with the highest score is considered the most important.

We calculated the metrics for the preliminary social network extracted earlier and the individual scores for each character in the network. We experimented with thresholds ranging from 800 to 1500 and found that setting the threshold to 1200 allowed for the maximum filtering of minor characters while retaining major ones. Ultimately, we identified a network of 639 main character nodes and 2,760 edges representing the social relationships of the main characters.

Inner Network.

Construction of Inner Networks. This study constructs and analyzes social networks based on the story arcs of One Piece. The first step is to extract the storyline content of each arc. Using the homepage content of all characters on the One Piece Fandom website, we extract each character's "History" section, which is organized based on arc narratives. Each arc is listed as a subheading, detailing the character's background relationships and events involving other characters within that arc. Leveraging this structure, we employed web scraping to collect all the arc-related content from character homepages.

For each arc, when the name of another character appears in the content, it indicates an interaction or relationship between the primary character of the homepage and the mentioned character within the arc. Based on this, we construct nodes and undirected edges between these characters in the arc's social network. In total, 32 arc-specific social networks were constructed. These networks are plotted in chronological order, with the number of nodes and edges in each arc's network shown on the x-axis, as illustrated in Fig.1C. From the Figure, we observe that as the timeline progresses, the number of characters and edges generally increases. This indicates that the storyline expands over time, involving more characters and interactions. Early arcs, such as Romance Dawn Arc and Orange Town Arc, have fewer characters and edges, reflecting the limited scope of the narrative in its initial stages. In contrast, later arcs, such as Wano Country Arc and Dressrosa Arc, feature a peak in the number of characters and interactions, reflecting their complexity and scale. Interestingly, while the overall scale increases, fluctuations in network size are evident between adjacent arcs. For example, the Post-Enies Lobby Arc shows a temporary decrease in the number of characters and relationships, reflecting transitional phases in the story where fewer characters are involved.

Inner Network Analysis. Focusing on the social networks of all arcs, we analyze five key arcs that are central to the storyline: Arabasta Arc, Enies Lobby Arc, Marineford Arc, Dressrosa Arc, and Wano Country Arc. These arcs represent significant milestones in the narrative. The Arabasta Arc occurs in the first third of the story and marks the protagonist's first major victory against a critical enemy. The Enies Lobby Arc focuses on the protagonists' efforts to rescue a comrade, showcasing emotional intensity and teamwork. The Marineford Arc serves as the central climax, transitioning the story to a new sea for further adventures. Within the new sea, the Dressrosa Arc and the Wano Country Arc are two of the most expansive chapters, culminating in the growth of the protagonist's team.

To analyze the networks, we use degree centrality (4) (5) and betweenness centrality (6). Degree centrality measures the number of connections a node has with other nodes, reflecting the importance and influence of characters within the chapter. Betweenness centrality evaluates a node's role as a bridge in the shortest paths between other nodes, indicating its importance in facilitating interactions and information flow. We calculated the top 10 characters with the highest degree and betweenness centrality in the networks of the five selected arcs, as shown in Fig.1.

From Fig.1A, the centrality analysis identifies the most critical characters in each chapter. For example, in the Wano Country Arc, Kaidou exhibits the highest degree centrality, emphasizing his dominant role as the arc's primary antagonist. Interestingly, the protagonist, Luffy, does not have the highest centrality in any single arc, reflecting that his interactions with other characters are often fewer than those of newly introduced key characters in specific arcs.

From Fig.1B, betweenness centrality analysis reveals characters crucial for information flow within the storyline. While characters with the highest degree centrality often have high betweenness centrality, exceptions exist. For instance, in the Arabasta Arc, although Crocodile, the main antagonist, has the highest degree centrality, Vivi has the highest betweenness centrality. This highlights Vivi's role as the central figure driving the conflict between the protagonist

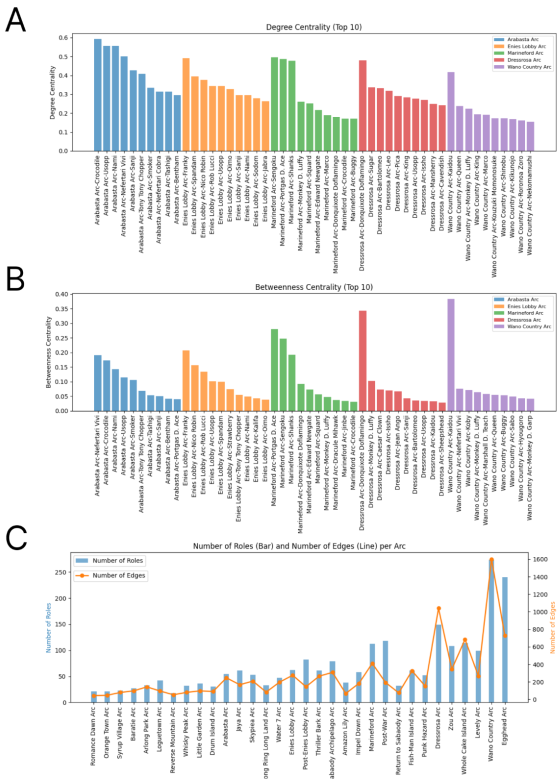

Fig. 1. Analysis of Centrality and Scale in Character Networks Across Different Story Arcs. (A) Shows the top 10 characters ranked by degree centrality in different story arcs, reflecting the extent of their direct connections with other characters in the network. (B) Shows the top 10 characters ranked by betweenness centrality in different story arcs, indicating the importance of these characters as mediators or bridges in the network. (C) Compares the number of characters (bars) and the number of relationships (edges, lines) in different story arcs, highlighting the complexity of the arcs and the richness of character interactions.

and the antagonist, playing a pivotal role in information and relationship dynamics within the arc.

External Network.

Degree Distribution of the External Network. The construction of the external network is based on all the characters that appear in the consecutive arcs, instead of analyzing the individual character composition of each arc, we divide all arcs into 5 equal parts and keep accumulating the new characters. By extracting the degree distribution of the external network at each period, we can further explore its growth process, meanwhile, With the help of the angle distribution, we try to fit the Scale-Free Distribution (7) to the external network in different periods to determine the nature of the network and understand the changing trends of the points and angles in the network by observing the changes in the parameter α(in the book (8) it is called λ). For the degree distribution changes in Fig.2A, the number of nodes in the external network shows a trend of continuous increase, and the distribution of angles gradually shows a stronger diversity. In particular, the number of nodes with low angles (0-10) and medium angles (10-20) has increased significantly, while the number of high-angle nodes (greater than 30) has been relatively sparse, although there are more and more high-degree nodes appearing. The degree distribution presents an

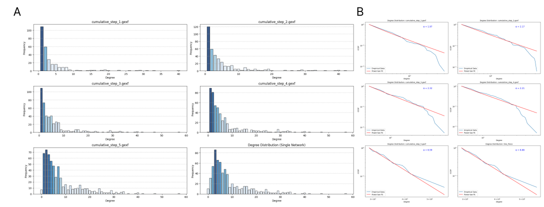

Fig. 2. Degree Distribution Analysis (A) The degree distribution of the external network shows the degree distribution of 5 different external network periods. (B) The model fit of the external network shows the estimation α of the fitted Scale-free Distribution and their fitting effect

obvious long-range distribution, namely the degree of most of the nodes is lower, but there are a few high-level nodes(hubs). This kind of distribution characteristics match the network characteristics of the Scale-free Distribution. To Confirm the external network model and determine the impact of these hubs on the entire external network, we need to refer to the strength of the parameters α in Fig.2B. The CCDF curve agrees well with the power-law fitting line within a certain range, indicating that the node degree distribution of these networks conforms to the Scale-free Distribution. As the external network grows, the alpha value continues to increase. Generally speaking, when $\alpha \leq 2$, a few super hub nodes dominate the entire network. $2 < \alpha < 3$ is a typical scale-free network. When $\alpha \geq 3$, the network distribution tends to be uniform, the number of hub nodes decreases, and it is closer to a random network. Therefore the external network gradually changes from a scale-free network with highly centralized angles to a random network with uniform angle distribution, which gradually weakens the influence of the hub.

Network Partition. The study of network partitioning in the context of One Piece is extremely appealing. To carry out this analysis, we use the Louvain algorithm with a resolution of 1. This algorithm is based on the maximization of modularity, identifying subgroups of nodes that are more densely connected among themselves than with the rest of the network.

We observe that 24 communities are identified. However, 14 of them correspond to fewer than 9 nodes, while 10 of them account for 601 of the 639 total nodes. Upon manually inspecting the results of the less populated communities, we identify characters with a limited number of relationships, positioned on the periphery of the central network. Therefore we center the following analysis on just the populated networks.

Community and Arc. To explore the relationship between communities and the sagas in One Piece, a table was constructed in Fig.3A, which shows the distribution of characters between the different communities. This table illustrates how characters from a specific saga tend to cluster in certain communities, indicating a strong correlation between the narrative divisions and the structure of the network.

To perform a deeper analysis of the dominant themes within each community, the TF-IDF (Term Frequency-Inverse Document Frequency) technique was used. This method identifies the most representative keywords for each community. The results obtained are presented in Fig.3B

The TF-IDF analysis was repeated for characters grouped by the first saga in which they appeared. The results obtained are presented in Fig.3C. Upon comparing the word clouds created from both community-based grouping and saga-based grouping, and jointly presenting the word clouds of the community and the saga with the highest number of shared characters, it becomes clear that there is a significant overlap of words.

Community 1 and Dressrosa Saga, Community 1 and Dressrosa Saga agree on the word with the highest TF-IDF value: Doflamingo. This term corresponds to the main villain of the arc, also known as Donquixote. Other significant terms include Dwarf, Kyros, Law, Bartolomeo, and Rebecca, which represent the main allies of the Straw Hat Pirates during this saga. Similarly, Pica, Diamante, Carrot, Trebol, and the entire Donquixote family align with the TF-IDF values within the community and the saga. An intriguing difference emerges with Momonosuke, who shows a high TF-IDF value in the Dressrosa Saga but not within the community. This observation is logical, as Momonosuke makes his debut in Dressrosa but gains greater importance in subsequent arcs. Thus, while he does not stand out as a central character in Dressrosa, his role becomes crucial later in the story.

The Community 3 and East Blue Saga, both highlight Pirate as the word with the highest TF-IDF value. This term captures the essence of the saga and community, reflecting the introduction of the central theme of One Piece: the quest to become the Pirate King. Additional key terms include Ace, Whitebeard, and Marine, representing major factions and characters. Furthermore, Buggy and Shank are notable, as they are prominent figures in the East Blue Saga, where Buggy appears as a significant antagonist and Shank as a powerful and influential pirate. Crew underscores the importance of Straw Hat Pirates in both contexts. In the East Blue Saga, terms like Zoro, Usopp, and Nami are particularly frequent, as these characters are introduced in this arc, marking the beginning of the Straw Hat crew's formation. Meanwhile, Community 3 emphasizes a broader range of terms, such as Admiral, Koby, and Sabo, which reflect the expanding world of the Marines and the Revolutionary Army. The prominence of Whitebeard and Ace also highlights the ongoing importance of the Whitebeard Pirates and their connection to Luffy's journey. The East Blue Saga strongly emphasizes the early members of the Straw Hat crew (Zoro, Usopp, Nami), while Community 3 shifts focus toward powerful pirates and the Marines, signaling a broader and more mature phase in One Piece. The higher TF-IDF values for Marine, Whitebeard, and Ace in Community 3 illustrate the evolution of the story from small-scale adventures to large-scale conflicts and power struggles.

In addition, the analysis shows that in Community 7, 72 characters make their first appearance in the Yonko Saga. However, this saga also includes 58 characters from Community 8. A key overlap is Sanji (a member of the Straw Hat Pirates) and Big Mom, whose nickname represents one of the Yonkou, a major and powerful figure in One Piece. Interestingly, there are notable differences between the saga and the community. Terms such as cake, brûlée, and pudding—Big Mom's favorite foods—appear more prominently in the community than in the saga. This suggests that the community more accurately reflects the characters

A

	1	2	3	7	8	11	13	15
EAST Blue Saga	1	1	30	1	4	11	15	2
Alabasta Saga	0	1	4	2	2	0	3	35
Sky Island Saga	2	6	12	1	1	28	0	1
Water 7 Saga	3	25	9	0	0	0	1	6
Thriller Bark Saga	0	1	14	1	1	0	2	1
Summit War Saga	4	5	23	3	8	0	16	10
Fish-Man Island Saga	0	0	0	5	1	0	28	0
Dressrosa Saga	59	2	1	5	12	0	0	0
Yonko Saga	0	16	10	72	58	0	1	3

B

C

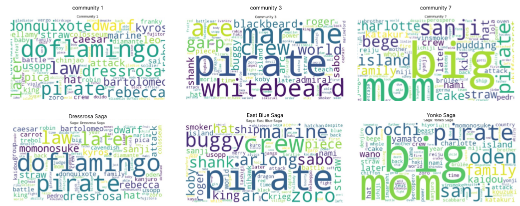

Fig. 3. (A) Community and Saga Association Matrix. (B) Community Sentiment TF-IDF Analysis Results. (C) Saga's TF-IDF Analysis Results.

associated with Big Mom. Conversely, the saga incorporates another major storyline that forms its own distinct community. This is exemplified by terms like Kaido, Wano, and Oden, which hold greater significance in the latter part of the saga, constituting a separate community independent of Big Mom's storyline.

Sentiment Analysis. Sentiment analysis of the main characters and Arc is implemented based on the external network and its community partition. We mainly conducted sentiment analysis on the top 9 characters and the main character of the comic Monkey D. Luffy to explore whether they belong to the righteous side or the evil side. We also analyzed the top 10 communities. Based on their association with the arc, we can infer whether the tone of different arcs is happy or sad. LabMT (9) is a limited sentiment dictionary, which makes it difficult to deal with large-scale text in specific comic scenes. We chose the TextBlob (10) library for sentiment analysis. TextBlob uses the Naive Bayes model for sentiment analysis. Combining the dictionary and context can more flexibly adapt to the text to derive more accurate sentiment scores. The result is shown in Fig.4.

The polarity is a value between -1 to 1, positive values represent positive sentiment, and negative values represent negative sentiment, the value represents the intensity of the Sentiment. According to the Fig.4. Both individuals and the community show a very slight positive tendency, that is, most of the important characters and arcs in One Piece often present neutral or slightly positive content.

Discussion

This study constructs social networks based on One Piece story arcs, using data from character homepages on the One Piece Fandom website. We extracted the "History" sections of each character, organized by arc, to build undirected networks with nodes as characters and edges representing relationships. A total of 32 arc-specific networks were created, showing an increase in characters and interactions as the series progresses. Early arcs, like Romance Dawn and Orange Town, have fewer interactions, while later arcs like Wano Country and Dressrosa show more complexity.

In the inner network, we analyzed five key arcs: Arabasta, Enies Lobby, Marineford, Dressrosa, and Wano Country, using degree centrality and betweenness centrality to identify

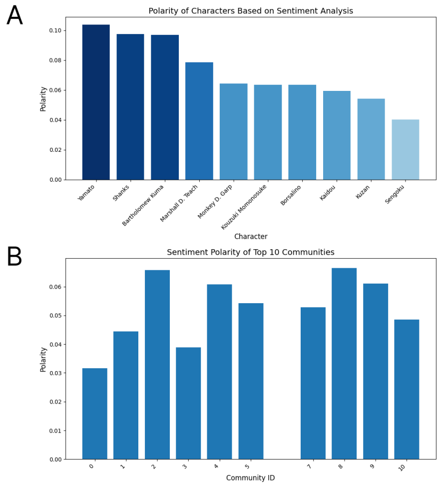

Fig. 4. Sentiment Analysis (A) The sentiment analysis of characters includes 9 characters with the top degree and the main character Monkey D. Luffy. (B) The sentiment analysis of communities includes the sentiment analysis of the top 10 communities

important characters. Degree centrality highlights characters with many connections, while betweenness centrality identifies those central to interactions. Kaidou, for example, has the highest degree of centrality in the Wano Country Arc, while Luffy's centrality is lower compared to new key characters in each arc. In the Arabasta Arc, Vivi has the highest betweenness centrality, playing a crucial role in the narrative flow despite fewer connections. This analysis offers insights into the evolving character relationships and key figures driving the plot of One Piece.

Compared with the inner network's more detailed exploration of character relationships, the external network focuses on the overall characteristics and changes of the network.

As the external network grows, its degree distribution from a very strong Scale-free Distribution gradually approaches a random network, which is surprising, but considering our cautiousness in connecting nodes between networks (removing a large number of characters in the anime that only have a few connections to others, only considering connections between two characters that reference each other, etc.), this phenomenon can be reasonably explained, that is, if only the interactions that are important to the characters are considered (two-way references), the characters in One Piece will basically be limited to his teammates and a small number of important enemies. Although new characters appear very quickly, most of them and important characters are only tools to set off the main characters rather than establishing long-term cooperative relationships with important characters. This network structure also brings benefits to the subsequent community analysis. We don't need to worry about the important character nodes showing too strong dominance in the network (connecting most of the nodes in the network) and causing the segmentation algorithm to be unable to correctly measure the community where the important characters are located.

Some further insights emerge through network partition, the resulting community has a wonderful reaction between the different arcs, the underlying logic of this phenomenon is that there is a strong correlation between the characters and arcs in One Piece. Apart from the protagonist and his party, other important characters basically only appear in their own arcs. When we partition communities, we divide the groups belonging to these important characters, and they dominate their respective arcs, resulting in arcs and communities sharing many keywords.

The sentiment analysis has not brought much meaningful information to us since most sentiments of characters and communities are slightly positive around the same level. This may be caused by the character creation of the author of One Piece and the writing style of the One Piece Wiki website. Most characters in One Piece avoid overly stereotyped creations such as heroes and villains. Each character has his own character arc. In addition, the complexity of the characters' personalities leads to the fact that most characters' sentiments are neutral. In addition, One Piece itself is a relatively inspirational and joyful work, so most of the content will have a slightly optimistic color.

Materials and Methods

DataSource. The data source is the wiki page of One Piece: One Piece Wiki - Fandom

Github. The explainer notebook is in Github

Scale-free Distribution. The formula of the Scale-free Distribution is a Power-Law Distribution as $P(k) \sim k^{-\alpha}$.

1. M Dihyat, K Malik, M Khan, B Imran, Detecting ideal instagram influencer using social network analysis. *arXiv preprint arXiv:2107.05731* (2021).
2. A Bristow, Fragments and links: Organizational actor-world of the harry potter phenomenon. *Cult. Organ.* **13**, 313–325 (2007).
3. E Coiera, Social networks, social media, and social diseases. *Bmj* **346** (2013).
4. H Ahn, M Park, H Kim, KP Kim, Implementing closeness centrality measurements on workflow-supported enterprise social networks in *2015 17th International Conference on Advanced Communication Technology (ICACT).* pp. 763–767 (2015).
5. X Zhao, S Guo, Y Wang, The node influence analysis in social networks based on structural holes and degree centrality in *2017 IEEE International Conference on Computational Science and Engineering (CSE) and IEEE International Conference on Embedded and Ubiquitous Computing (EUC).* Vol. 1, pp. 708–711 (2017).
6. P Suppa, E Zimeo, A clustered approach for fast computation of betweenness centrality in social networks in *2015 IEEE International Congress on Big Data.* pp. 47–54 (2015).
7. AL Barabási, R Albert, Emergence of scaling in random networks. *science* **286**, 509–512 (1999).
8. AL Barabási, *Network Science.* (Cambridge University Press), (2016).
9. PS Dodds, KD Harris, IM Kloumann, CA Bliss, CM Danforth, Temporal patterns of happiness and information in a global social network: Hedonometrics and twitter. *PloS one* **6**, e26752 (2011).
10. S Loria, Textblob: Simplified text processing (https://textblob.readthedocs.io/) (2018) Accessed: 2024-12-11.