A computational analysis of Mahabharata

Debarati Das
UG student,Dept. of CSE
PES
Institute of Technology
Karnataka, India
Debarati.d1994
@gmail.com

Bhaskarjyoti Das
PG Student, Dept. of CSE
Visvesvaraya
Technological University
Karnataka, India
Bhaskarjyoti01
@gmail.com

Kavi Mahesh
Dean of Research and Director
KAnOE-Center for Knowledge Analytics
and Ontological Engineering
PES University, Bangalore
Drkavimahesh
@gmail.com

Abstract

Indian epics have not been analyzed computationally to the extent that Greek epics have. In this paper, we show how interesting insights can be derived from the ancient epic Mahabharata by applying a variety of analytical techniques based on a combination of natural language processing, sentiment/emotion analysis and social network analysis methods. One of our key findings is the pattern of significant changes in the overall sentiment of the epic story across its eighteen chapters and the corresponding characterization of the primary protagonists in terms of their sentiments, emotions, centrality and leadership attributes in the epic saga.

1 Introduction

Large epics such as the Mahabharata have a wealth of information which may not be apparent to human readers who read them for the fascinating stories or spiritual messages they contain. Computational analysis of large texts can unearth interesting patterns and insights in the structure, flow of stories and dynamics of the numerous characters in the intricate stories that make up the epics. Unfortunately, not much attention has been paid to applying natural language processing and other related techniques to carry out computational analyses of Indian epics. In this work, we attempt to carry out detailed analyses of the Mahabharata epic.

Sentiment and social network analyses have been applied mainly to structured texts such as tweets, emails etc. to discover user sentiments or important personalities. Comparatively literary works are less subjected to computational analysis as there are no immediate business incentives. However,similar techniques can be adopted towards appreciating the literary work, to understand underlying social network and to find or validate literary truths. As literary text is built around a social backdrop, it reflects the society the author lives in and reveals a lot about the contemporary social setting.

Unlike SMS and tweets, genre is important in literary text. Amongst the past and recent literary genres, epics and novels have seen most of the work in the Digital Humanity community as the scope is typically large in terms of time, number of events and characters to facilitate computational analysis. The Greek epics Iliad and Odyssey, the English epic Beowulf, novels such as Victor Hugo's Les Miserable and works of William Shakespeare are some of the examples. However, there is no major existing work around Indian epics such as Ramayana and Mahabharata. Hence we have chosen Mahabharata as the target text for a computational analysis effort.

2 Related work

The first important step in computational analysis of a literary text is to identify the protagonists. Next the relatedness of the protagonists can be computed to form the underlying social network. There are essentially two methods to capture social network from a literary text. One option is to capture all social events such as conversations assuming that all characters participating in a social event are socially related. This method does not work well for narrative intensive text. The other method assumes that all characters appearing in a given co-occurrence window have some kind of social relations. This approach ends up considering even insignificant characters but works better for narrative based texts such as epics. Newman and Girvan's work (2004) to detect the communities in Victor Hugo's Les Miserable is the first major effort to find the social network from narratives. Sack (2012) deduced

the plot from network by using concepts of structural balance theory. Elson et al.(2010) proposed dialogue based method to extract social network. Jayannavar et al. (2015) updated Elson's approach by broadening the scope of conversation to social events. Rydberg-Cox (2011) extracted social networks from Greek tragedies. Agarwal et al.(2012) showed that a dynamic network analysis can present more subtle facts. Beveridge and Shan (2016) built the underlying social network for the third book ("A storm of swords") of the TV series "Game of Thrones" with a co-occurrence window of 15 words. Stiller et al. analyzed ten of Shakespeare's plays (2003) also based on the co-occurrence logic. Carron and Kenna (2012) provided a quantitative approach to compare networks. Mac Carron et al.(2014) did a structural analysis of Iliad, English poem Beowulf and Irish epic Tain Bo Cuailnge. P. J. Miranda et al.(2013) has done a structural analysis of underlying social network of Homer's Odyssey. Alberich et al.(2002) have built a social network from Marvel comics.

As Mahabharata is an epic, we must mention Poetics by Aristotle and an excellent commentary provided by Lucas (1968). Aristotle defined literary genres such as poetry, tragedy, comedy and epic. Poetry mimics life. Tragedy is a type of poetry that showcase noble men and their noble qualities as well as values. Epics such as Mahabharata are a type of tragedy and are built around noble men in the form of narratives. A tragedy typically has a plot with a beginning, a middle and an end and other constituents of the text are secondary to the plot. The beginning of the plot typically is a scenario of stability which gets disturbed by some events. The middle is where the disequilibrium comes in along with lot of events and actions by the characters. All the events and actions are towards achieving the end where the problem gets resolved and stability sets in again. Plots have various constituents i.e. suffering, reversal, recognition of new knowledge, surprise. An epic is different from a more recent literary genre like a novel and will have lot of negative sentiment across its breadth but in spite of that conveys a noble theme in the minds of its audience.

One can measure sentence polarity by referring to some standard thesaurus where polarity measures are preassigned by researchers.This approach uses a resource like SentiWordnet (http://sentiwordnet.isti.cnr.it/). Alternatively, in a supervised classification approach labelled data sets from similar domains are utilised. However, this approach works where the training dataset from similar domain is available and this method is not suitable for sentiment analysis for an epic. Emotion analysis finds causes of sentiment. Robert Plutchik(1980) defined the eight basic emotion types. Mohammad and Turney (2010) created the NRC emotion lexicon which is an association of a list of words with these eight basic types of emotion and two types of sentiment. Mohammad (2011) presented an emotion analyzer as a visualization exercise of these emotions in literary text.

Table 1: Key Attributes of Mahabharata Text

Attributes	Value	Remarks
Size in bytes	15,175 K	English translation
Size in bytes	13,947 K	After removing comments
Number of words	28,58,609	Using NLTK
Number of unique words	32,506	Using NLTK
Number of sentences	1,18,087	Using NLTK
Number of chapters	18	"parva"
Number of characters	210	appearing at least 10 times

For our research, we have used the English translation of Mahabharata available at Project Gutenberg site (http://www.gutenberg.org/ebooks/7864). This is a translation by Kisari Mohan Ganguli done between 1883-1896. Mahabharata is larger than Iliad and Odyssey together, compiled many years ago. This has 18 "parva"s or chapters and each "parva" has many sections.

3 The methodology

Mahabharata is not dialogue heavy and is mostly narrative. So, identifying relations between char-

acters is done using co-occurrence algorithm with window size of a sentence.The method we devised for a comprehensive computational analysis of the Mahabharata epic is as follows:

1. Pre-processing

   - Filter out supporting texts such as tables of content, publisher details and chapter summaries.
   - Separate the text into chapters (called "parva") using suitable regular expressions.
   - Separate each "parva" into sections based on the structural elements in the text.

2. Identifying characters

   - Identify all proper nouns using POS tagging
   - Input a list of known characters of the Mahabharata story (widely available on the internet).
   - Input a thesaurus of equivalent names for the characters (also widely known, e.g. Draupadi=Panchali, Arjuna=Phalguni etc.) to merge equivalent names.
   - Filter out a list of known place names in ancient India and its neighbouring regions.
   - Apply a threshold to retain names whose frequency is above a minimum value (resulting in 210 characters for the Mahabharata story).
   - Retain only those characters which are in the top 30 percent of characters mentioned in a given parva (resulting in about 70 characters overall). Same logic is followed for both individual and cumulative analysis of each parva.

   The following steps are carried out separately for each "parva" and also for the entire text.

3. Co-occurrence analysis

   - Compute a co-occurrence matrix for the identified characters using sentence boundaries as windows of co-occurrence.
   - Build a social graph from the co-occurrence matrix.

4. Network analysis

   - Various network metrics are computed for the social graph for each of the 18 "parva"s in both cumulative and standalone way viz. betweenness centrality, closeness centrality, degree centrality, size of maximal cliques, number of detected communities, size of ego networks for main nodes, core periphery analysis, density of the core and overall network etc.
   - Additionally various structural metrics are computed for social graph viz. degree assortativity, percentage size of giant component,average clustering coefficient, average shortest path length etc.

5. Overall sentiment analysis

   - Using syntactic meta data, phrases containing noun, adjective, verb and adverbs are identified.
   - The above text is tokenized using standard NLP techniques.
   - The tokens are POS (parts of speech) tagged and tagged tokens are mapped to synsets in Wordnet in a word sense disambiguation process.
   - The sentiment scores are picked up from SentiWordnet for each synset.
   - Overall sentiment of the parva is derived from these values by summing the constituent sentiment scores.

6. Sentiment analysis for main characters

   - Similarly sentiment analysis of each protagonist is done by extracting the sentences where the protagonist appears. This is done for each parva.

7. Emotion analysis

   - Emotion analysis for the full text and each of the protagonists is done with the help of NRC word-emotion association lexicon. After extracting the relevant part of the corpus,the score is calculated for each POS (part of speech) tagged token for each emotion and finally summed up. The obvious limitation with any lexicon based approach is the limitation imposed by the size of the

221

lexicon itself and this limitation does apply to our analysis as well.

We have used the Python, NLTK(Natural Language Toolkit), various open source libraries (TextBlob, Networkx, Stanford SNAP, Gephi) and data analytics/visualization software Tableau in our work.

4 Analysis of results

4.1 The protagonists

We have tried out 3 different approaches to identify the protagonists.

- **Most frequently mentioned character:** As shown in Figure.1a, this method finds the most frequent characters. However this misses out the protagonists who are unfortunately low on frequency but may be important otherwise.

- **Size of the ego network:** Size of ego network (number of nodes directly connected) calculated from Mahabharata social network produces different results. As shown in Figure.1b, Kripa who is a teacher of the princes, is topping the list. Chieftains like Shalya, Virata, Drupada come towards the top in this list. Kunti(mother of Pandavas), Indra (the king of gods) and Narada (the sage) are also in this list being well connected!

- **Centrality metrics:** The betweenness, eigenvector, closeness and degree centrality are compared. Few observations can be made out of this from Figure.2:

    - Betweenness centrality differentiates the main protagonists whereas other centrality metrics are mostly equivalent.
    - Arjuna, Karna, Krishna, Yudhisthira, Bhisma, Kunti and Drona are the top few in terms of all four centrality. They are the most important protagonists.
    - Some of the personalities with very large ego network are having very low betweenness centrality and not making into the top list (Kripa, Shalya, Drupada, Virata etc.) because their influence is limited to one camp i.e. Kaurava or Pandava. Their importance is mostly local.

- Amongst the princesses and queen mothers, Kunti turns out to be the understated (in the existing literary analysis) power behind the scene (having a large ego network and high centralities). Her low eigenvector centrality leads to false perception that she is not important. Other main lady characters (Gandhari, Madri, Draupadi) are low on betweenness as their influence is limited to one camp.
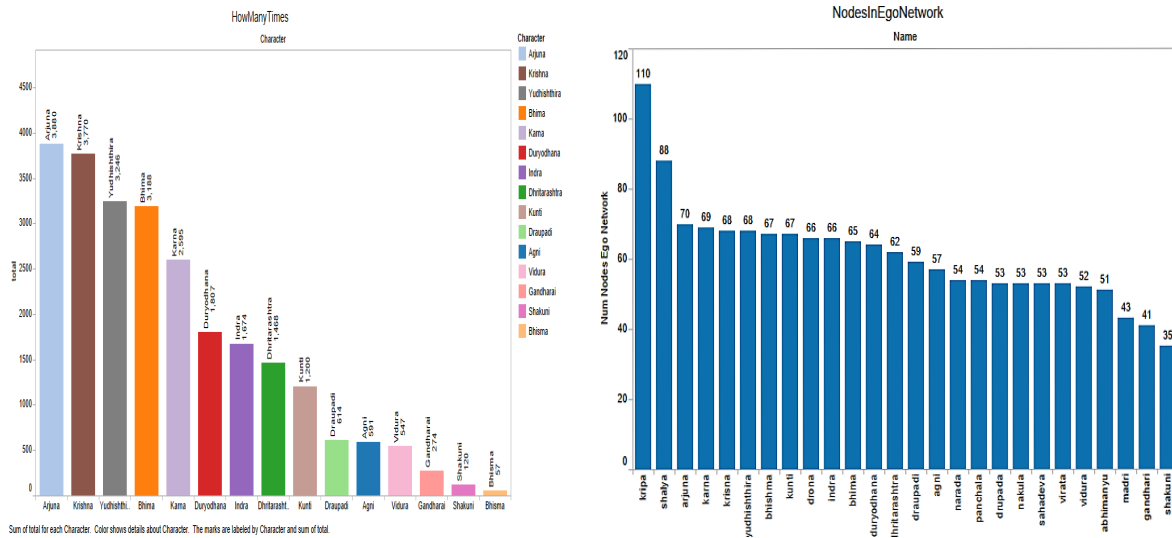
4.2 The words say a lot

Word clouds show a marked difference between the protagonists as shown in Figure.3a to Figure.3d. These are drawn by extracting adjectives from respective corpus.

- Both Arjuna and Bhima are "mighty" and "warrior". But Arjuna has words like "great", "excellent", "capable", "celestial" whereas Bhima has "terrible", "fierce" etc. So Arjuna is the best in his class whereas Bhima is a mighty warrior with "terrible" anger.

- Bhisma has "invincible", "principal", "virtuous" whereas Krishna has "celestial", "beautiful", "illustrious". So, Bhisma sounds more like an invincible warrior famous for his virtue, whereas Krishna is almost godly.

- For Duryodhana, "wicked", "terrible" etc. stand out whereas for Yudhisthira, "virtuous" and "righteous" are key words. Both are leaders of their respective camps but they are poles apart.

4.3 Sentiments across the text

Mahabharata takes the readers through a roller coaster ride of sentiment as shown in Figure.4. "Aadi parva"(1) starts on a positive note but the "Sabha parva" (2) brings lot of negativity with the game of dice. "Vana parva"(3) is again positive as Panadavs in spite of being in exile, make lot of friends and have achievements. "Virat parva"(4) is negative as the Pandavas have to live in disguise doing odd jobs. "Udyog Parva" (5) is again positive with both sides are very hopeful of winning war. After that as elders and leaders get killed in the battle, it is a downward slide of sentiment with Duryodhana's death bringing in positive emotion in "Shalya parva"(9). In "Stri parva" (11), the destruction is complete and sentiment reaches the

(a) Frequency of occurrence

(b) Size of ego network

Figure 1: Finding protagonists by number of mention and ego network

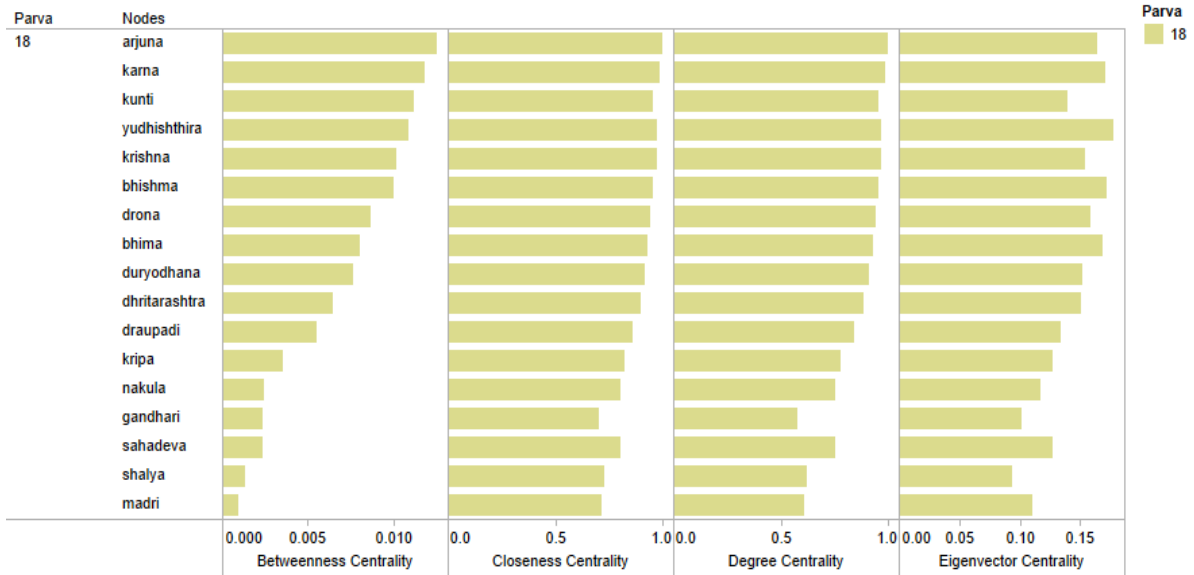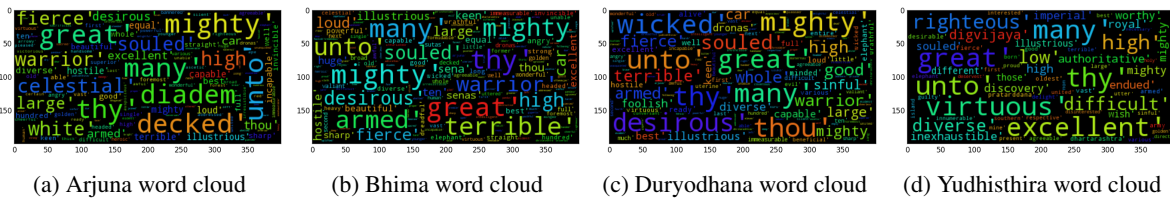Figure 2: Finding protagonists by comparing centrality metrics

(a) Arjuna word cloud

(b) Bhima word cloud

(c) Duryodhana word cloud

(d) Yudhisthira word cloud

Figure 3: Words say a lot
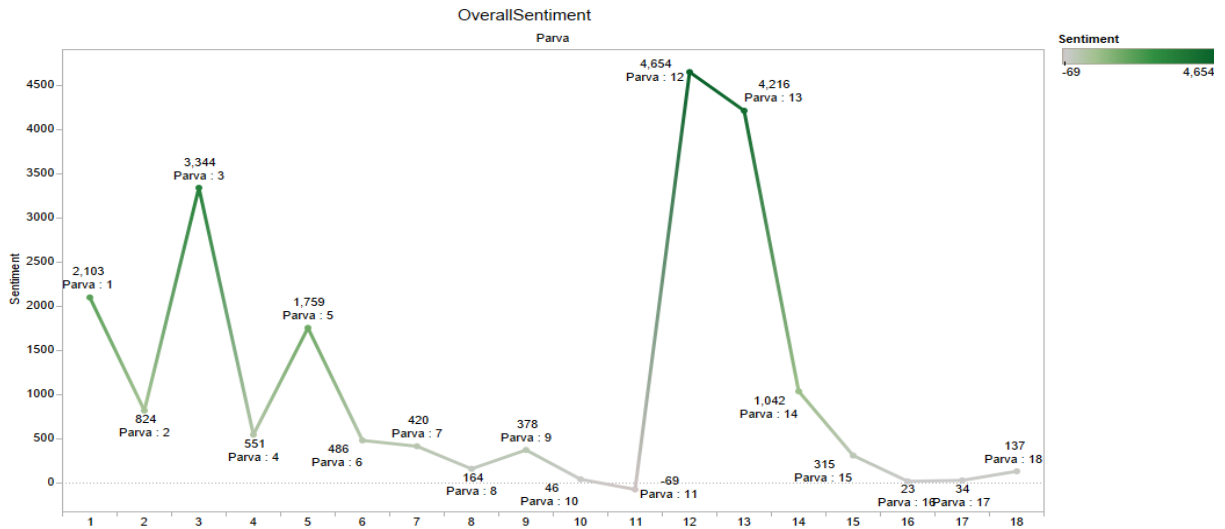
Figure 4: Sentiment across "parvas" of Mahabharata

(a) comparing sentiment: Krishna, Dhritarashtra

(b) comparing sentiment: Drona, Bhisma
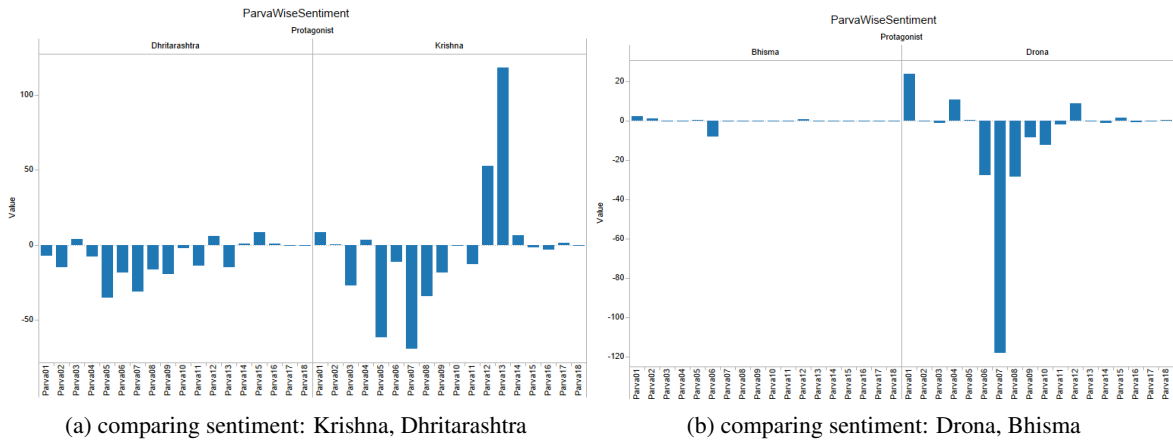
Figure 5: Comparing the sentiments

(a) comparing sentiment: Kunti, Gandhari

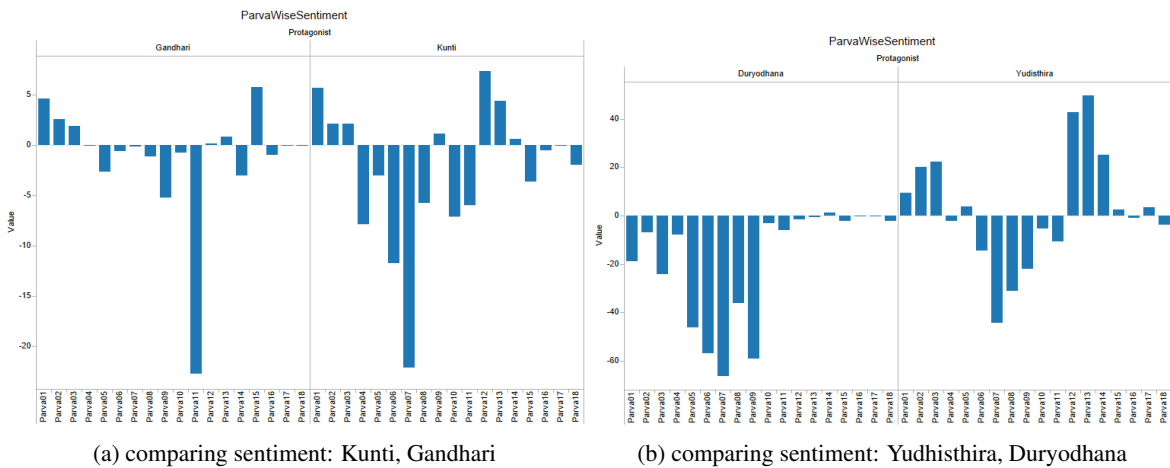(b) comparing sentiment: Yudhisthira, Duryodhana

Figure 6: Comparing the sentiments

lowest level. The "Shanti parva" (12) brings in peak of positive sentiment with coronation of Yudhisthira and many achievements. After that, it is again a downward slide of sentiments with many deaths and even death of Lord Krishna. The sentiment sees an uptick in the last two "parva"'s when Pandavas leave for Himalayas and finally attain divine status. Figure.5a to Figure.6b depict the net sentiment of the main protagonists according to the "parva". It leads to some interesting observations.

- Warriors like Arjuna and Bhima have lot of negativity around them.

- The leaders of the two warring camps Duryodhana and Yudhisthira are clear contrast as Yudhisthira has lot of positive sentiment around him.

- The gods like Indra and Agni have mostly positivity around them as they are mostly neutral on the ground.

- The eldest warrior, Bhisma is mostly neutral whereas Drona is committed to one camp and so is surrounded by negativity. Dhritarashtra, though elder, is mostly surrounded by negative sentiments.

- The two queen mothers Gandhari and Kunti are the sources of positive energy in both camps. Though understated they play pivotal roles. Compared to them, Draupadi is surrounded by negative sentiment.

- Lord Krishna, when he is in the thick of war, has negativity around him but once the battle gets over and larger senses prevail, he brings in sense of karma and lot of positive sentiments.

4.4 The emotions

We have analyzed the emotions both at the global and the protagonist level as shown in Figure.7 to Figure.9. Out of the eight basic emotion types, anger and trust are the key ones as expected in a tragedy that has an epic battle as the mainstay. Anticipation, disgust, fear, sadness come in almost equal proportion. In the scheming world of Mahabharata, there is not much of surprise and joy is kind of overshadowed by the other negative emotions. If we consider the emotions for some of the main protagonists, interesting conclusions can be drawn.

- Amongst the key ladies, Kunti stands out by the richness of positive emotion (trust and joy) and is the bedrock of strength for the Pandavas when they go through all their reversals of fate. Gandhari is relatively low key whereas Draupadi displays all the negative emotions that are key ingredients of a tragedy.

- Amongst the Pandava and Kaurava leaders (Duryodhana and Yudhisthira), Yudhisthira displays trust and joy more than any other emotion. Probably that is why he is perceived as a leader though there are many others with much more bravery and heroics. The contrast between Duryodhana and Yudhisthira is telling.

- Bhima and Duryodhana are very similar in emotions i.e. anger, trust and fear. Arjuna is quite unique and ambidextrous i.e. he displays enough of anger and fear and also large quantity of trust and joy.

- Amongst the elders, Bhisma is a detached persona and he does not show much of emotion. Drona is more attached to one camp and comparatively shows anger more than any other emotion.

- Krishna shows tremendous amount of trust, anticipation and joy in spite of all the tragedies and it is no wonder that he is called an incarnation of god.

4.5 Leadership analysis

We searched for leaders using two criteria viz. high in positive sentiments and high in centrality (degree and/or betweenness) as shown in Figure.10. Our assumption is leaders are not only centrally connected but they also show lot of positivity.

- It becomes very clear why Krishna is supreme as he is the only one who is in the high corner of this target quadrant.

- Closely following Krishna is Yudhisthira. That explains why in spite of not being a great warrior and known addiction for gambling, Yudhisthira is so well respected.

- Going by the same yardstick for leadership, Arjuna, Bhima, Drona, Karna are more of achievers or doers rather than leaders.
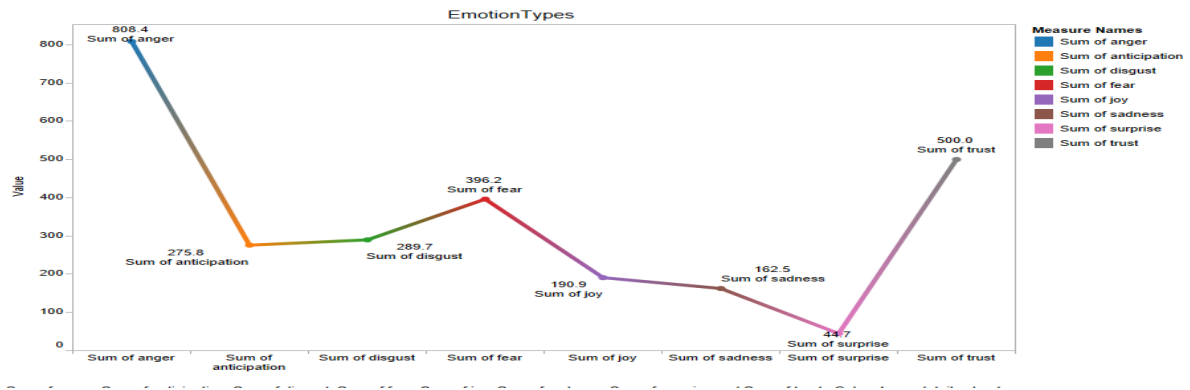
225

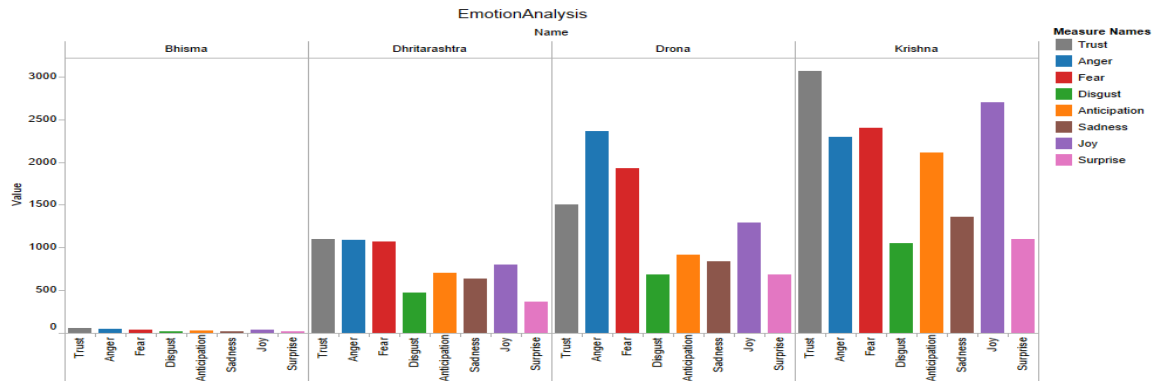Figure 7: Emotions across the text

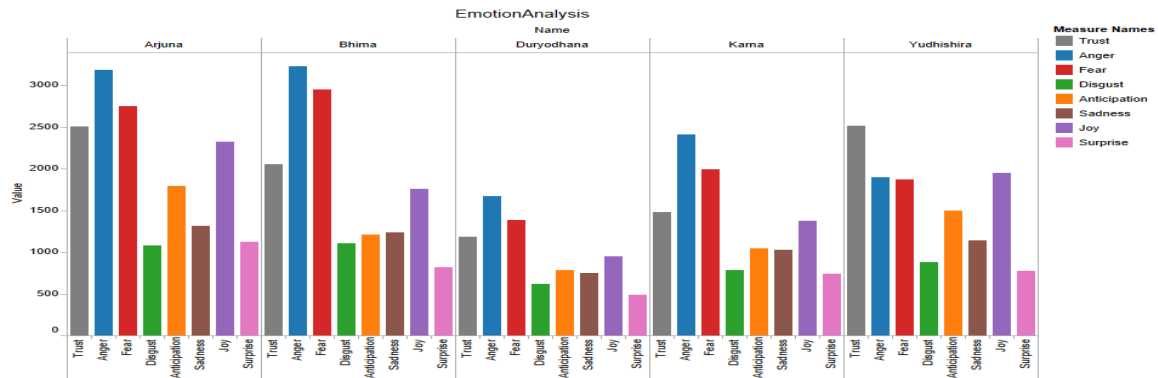Figure 8: Emotion Analysis of Bhisma, Dhritarashtra, Drona and Krishna

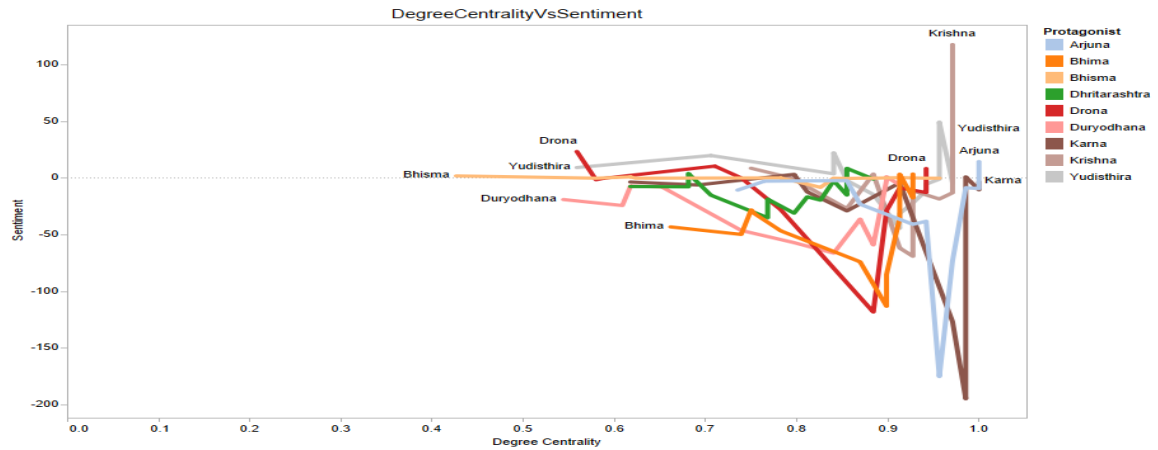Figure 9: Emotion Analysis of Arjuna, Bhima, Duryodhana, Karna, Yudhisthira

Figure 10: Leadership

(a) Considering diameter, degree and edge

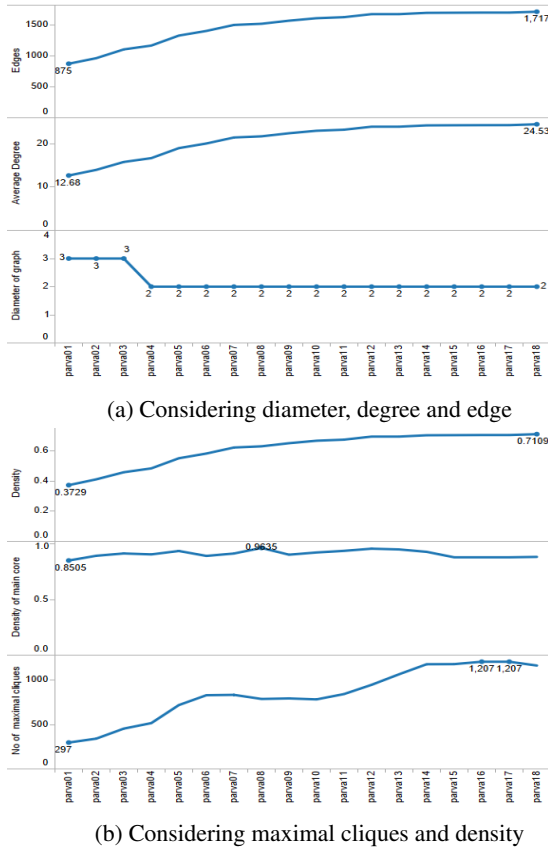(b) Considering maximal cliques and density

Figure 11: Evolution of social network across parvas

- Bhisma is neither great in centrality nor in positivity. He is more of a helpless spectator apart from his hard to find commitment to whatever promise he makes.

- Clearly Kaurava camp lacks in leadership. Duryodhana, the Kaurava leader, shows the lack of it and it is somewhat compensated by the combined effort of the achievers in his camp.

4.6 The social network of Mahabharata

- The "core periphery analysis" of the social network reveals a core of size 52 and consistently high density that remains comparable to the overall density of the network i.e. the plot is built around these members of the core.

- Mahabharata is also the story of three camps as proved by community detection techniques using Louvain algorithm (Blondel et al., 2008). They are the Kauravas, Pandavas and the gods/sages who remained somewhat neutral.

- The story of Mahabharata encompasses many years before the battle, 18 days of battle and around thirty six years after the battle. The evolving social network of Mahabharata across the parvas is analyzed using various structural metrics viz. degree, average degree, number of edges, number of maximal cliques and density of the main core as well as overall density. As shown in Figure.11a and Figure.11b, various structural metrics of the underlying social network tend to stabilize towards the end after becoming destabilised initially following Aristotelian framework of stability-instability-stability.

- Mahabharata network comes out as a small world network(small average shortest path and large clustering coefficient). Transitivity measured is comparable to other random graph of similar size such as Barabassi Albert model. However, modularity is found to be low (mostly 3 communities detected) compared to some real world networks. Also the high positive correlation coefficients for each centrality pair, large giant component and negative degree assortativity indicate large fictional component in Mahabharata.

5 Discussion and conclusion

In this work, we have applied various Natural Language Processing and Social Network Analysis techniques to come up with a computational analysis of the "Mahabharata". We have not only validated what the literary critics have unearthed about the epic but also augmented their findings by discovering subtle facts. Protagonists are identified and analyzed using both statistical and social network parameters such as centrality and ego network. The trajectory of sentiment and various emotions across the length of the text for each protagonist are examined. The findings validate what the literary critics have already found. Additionally this analysis brings out some subtle facts i.e. Kunti is understated in the existing literary analysis but is seen to be playing a pivotal role as displayed by the sentiments, emotions, centrality and large ego network size. We figured out the influence category of various protagonists in terms of local or global influence.

The leadership analysis explains why Yudhisthira is described in such glorious terms in spite of his many weaknesses. We have also looked at

227

leadership quotient of various protagonists by considering their position in the centrality-positivity quadrants and have brought out the leadership contrast between the warring camps in this epic.

The analysis also helps to explain why Mahabharata is an epic. Apart from the sheer number of characters, events, diversity of emotion and sentiment, it is found to conform to the Aristotelian definition of epics having the stability-instability-stability transitions. The analysis of the structural metrics also indicate that Mahabharata is not purely factual and has a large fictional component.

Clearly computational analysis of a literary text does not make the literary analysis redundant. But this provides an additional tool set for the students of literature to validate and augment their findings. The methods used can be easily replicated for other texts.

As a next step, we plan to extend similar analysis to the Indian epic Ramayana and perform similar structural analysis of the underlying social networks.

Acknowledgement

References

Apoorv Agarwal, Augusto Corvalan, Jacob Jensen, and Owen Rambow. 2012. Social network analysis of alice in wonderland. In *Workshop on Computational Linguistics for Literature*, pages 88–96.

Ricardo Alberich, Joe Miro-Julia, and Francesc Rosselló. 2002. Marvel universe looks almost like a real social network. *arXiv preprint cond-mat/0202174*.

Poetics Aristotle. 1968. Introduction, commentary and appendixes by dw lucas. *Oxford*, 125:16.

Andrew Beveridge and Jie Shan. 2016. Network of thrones. *Math Horizons*, 23(4):18–22.

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008.

David K Elson, Nicholas Dames, and Kathleen R McKeown. 2010. Extracting social networks from literary fiction. In *Proceedings of the 48th annual meeting of the association for computational linguistics*, pages 138–147. Association for Computational Linguistics.

Prashant Arun Jayannavar, Apoorv Agarwal, Melody Ju, and Owen Rambow. 2015. Validating literary theories using automatic social network extraction. *on Computational Linguistics for Literature*, page 32.

Pádraig Mac Carron and Ralph Kenna. 2012. Universal properties of mythological networks. *EPL (Europhysics Letters)*, 99(2):28002.

P Mac Carron and R Kenna. 2014. Network analysis of beowulf, the iliad and the táin bó cúailnge. In *Sources of mythology: Ancient and contemporary myths. Proceedings of the Seventh Annual International Conference on Comparative Mythology (15–17 May 2013, Tübingen)*, pages 125–141.

Pedro J Miranda, Murilo S Baptista, and Sandro E de S Pinto. 2013. Analysis of communities in a mythological social network. *arXiv preprint arXiv:1306.2537*.

Saif M Mohammad and Peter D Turney. 2010. Emotions evoked by common words and phrases: Using mechanical turk to create an emotion lexicon. In *Proceedings of the NAACL HLT 2010 workshop on computational approaches to analysis and generation of emotion in text*, pages 26–34. Association for Computational Linguistics.

Saif Mohammad. 2011. From once upon a time to happily ever after: Tracking emotions in novels and fairy tales. In *Proceedings of the 5th ACL-HLT Workshop on Language Technology for Cultural Heritage, Social Sciences, and Humanities*, pages 105–114. Association for Computational Linguistics.

Mark EJ Newman and Michelle Girvan. 2004. Finding and evaluating community structure in networks. *Physical review E*, 69(2):026113.

Robert Plutchik. 1980. A general psychoevolutionary theory of emotion. *Theories of emotion*, 1:3–31.

Jeff Rydberg-Cox. 2011. Social networks and the language of greek tragedy. In *Journal of the Chicago Colloquium on Digital Humanities and Computer Science*, volume 1.

Graham Sack. 2012. Character networks for narrative generation. In *Intelligent Narrative Technologies: Papers from the 2012 AIIDE Workshop, AAAI Technical Report WS-12-14*, pages 38–43.

James Stiller, Daniel Nettle, and Robin IM Dunbar. 2003. The small world of shakespeares plays. *Human Nature*, 14(4):397–408.