Temporal Network Analysis as a Method to Explore Australian World War I Diaries

Ashley Dennis-Henderson, Matthew Roughan, Jonathan Tuke

School of Computer and Mathematical Sciences, The University of Adelaide

19 July 2023

THE UNIVERSITY
of ADELAIDE

Talk Outline

- Motivation

- Data

- Date Extraction

- Methods and Results

Motivation

Ph.D. Aim

To analyse Australian World War I diaries using mathematical and computational techniques in order to understand what the soldiers wrote about and how they felt over the course of the war.
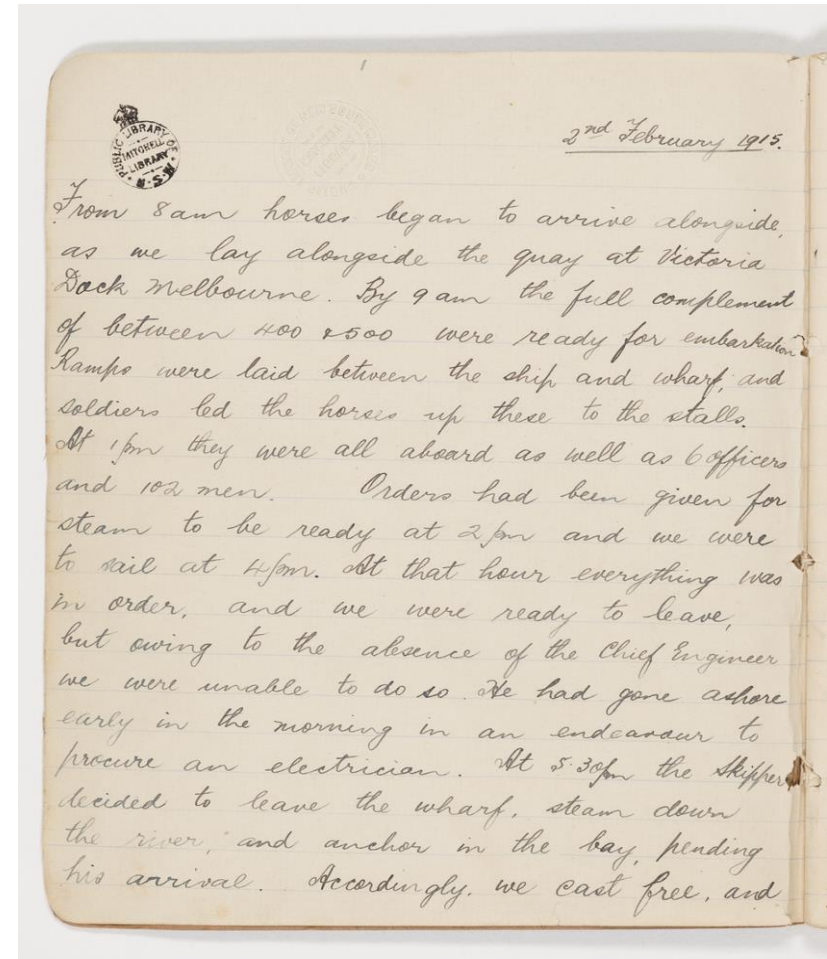
Research Questions

1. Can network analysis be used to find days of interest which should be further explored by close reading?

2. Can network analysis be used to find diaries which are representative of this collection?

Data

State Library of New South Wales Collection

- European War Collecting Project

- Contains over 1,600 transcribed documents from Australians in WW1, including 550 diaries

- Currently 161 cleaned diaries are ready to be analysed

Figure: Gilroy, Norman. Norman Thomas Gilroy war diary, Volume 1: 2 February-29 April 1915. Mitchell Library, State Library of New South Wales.

Date Extraction

Date Extraction Aim

Date	Entry
02/02/1915	from 8am horses began to arrive alongside, ...
08/02/1915	was so ill during this period that i was unable ...
28/03/1915	many large french and british transports ...
05/04/1915	at 5.30a we were towed out from the wharf ...
25/07/1915	arose about 6.30 and went down to church ...
27/07/1915	at 5am we were shifted to a berth further down ...

Date Extraction Challenges

"Tuesday 9th February 1915. At 6am woke to find we were "

"Wed Jan 30th Left line tonight being relieved by the 54th BN."

"Wednesday 24th Feby. Very disappointed when we received a"

"Tuesday May 8.5.17. Stand to 5.30 A.M. cool morning fed up"

"5th. Last night I had a party of 20 men repairing the blown in"

"17: Pte. Sellars fell down hold 50 ft., dying four hours later,"

"Wed. Kit inspection getting fixed up before we arrive."

"8th Field Ambulance. Breakfast had been previously arranged"

Date Extraction Approach

1. Use regular expression to extract date-like objects.

2. Use an optimisation program to fill in missing data, fix mistaken dates, and (highlight) non-dates so they can be removed.

Date Extraction Approach

However, this method does not work well on:

- Out-of-order diaries

- Diaries where a large proportion of dates only contain the day of the week

Removing these diaries leaves us with 139 cleaned diaries to work with.

Methods and Results

Building a Temporal Network

On each day:

- Nodes are individual diaries (who have an entry on that day)

- An edge exists between two nodes if those diaries have similar entries on that day

- The edge weight is the similarity score between those entries

Building a Temporal Network

To determine the similarity between entries:

1. Entries are converted to vector representation using BERT

2. The cosine similarity is calculated between entries on the same day

3. This produces a score from -1 (complete opposite) to 1 (complete similarity), with 0 being no similarity

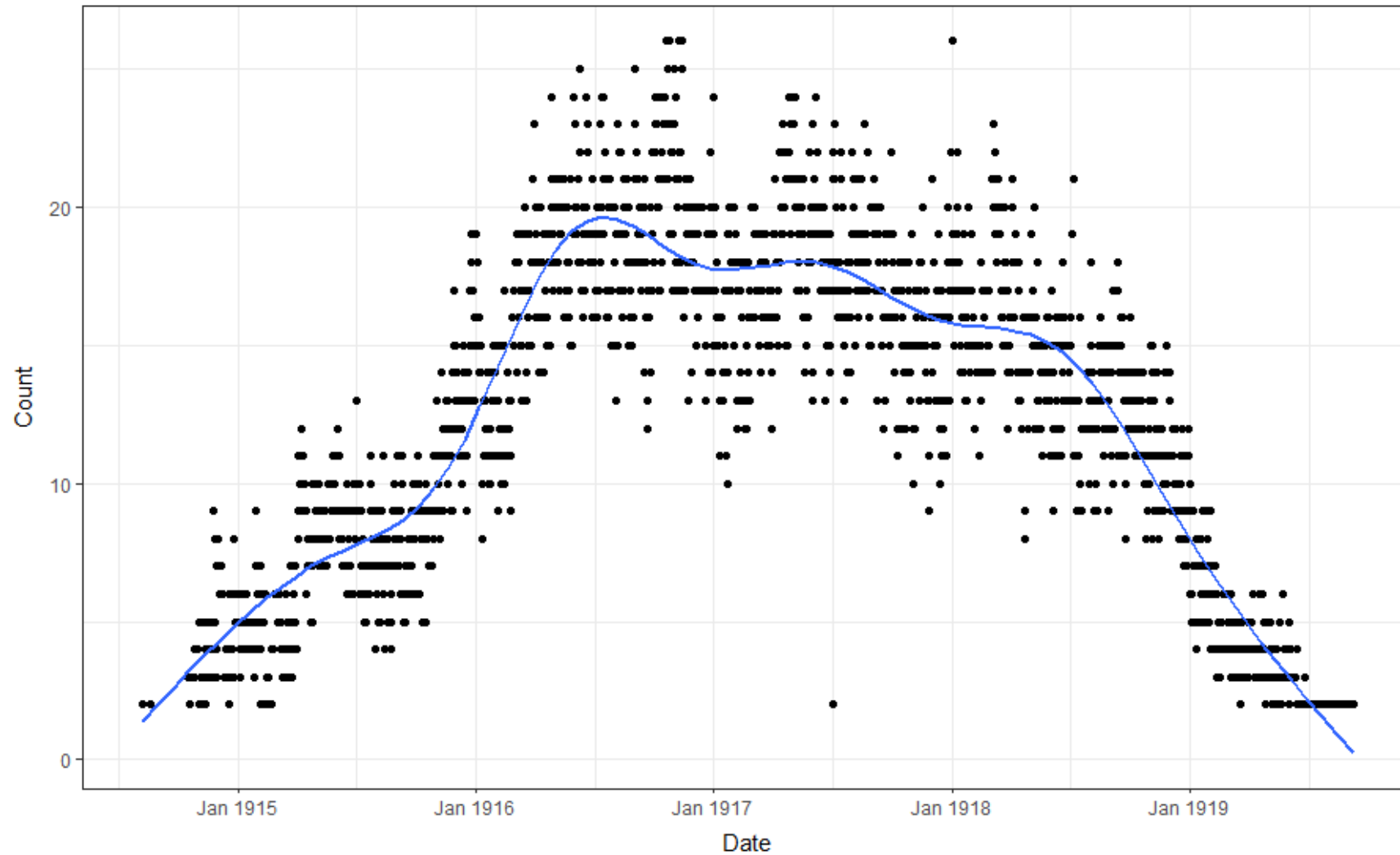4. Edges are created between nodes where the similarity score is above the threshold (median similarity $\approx$ 0.47)

Question 1

Can network analysis be used to find days of interest which should be further explored by close reading?

Question 1: Results

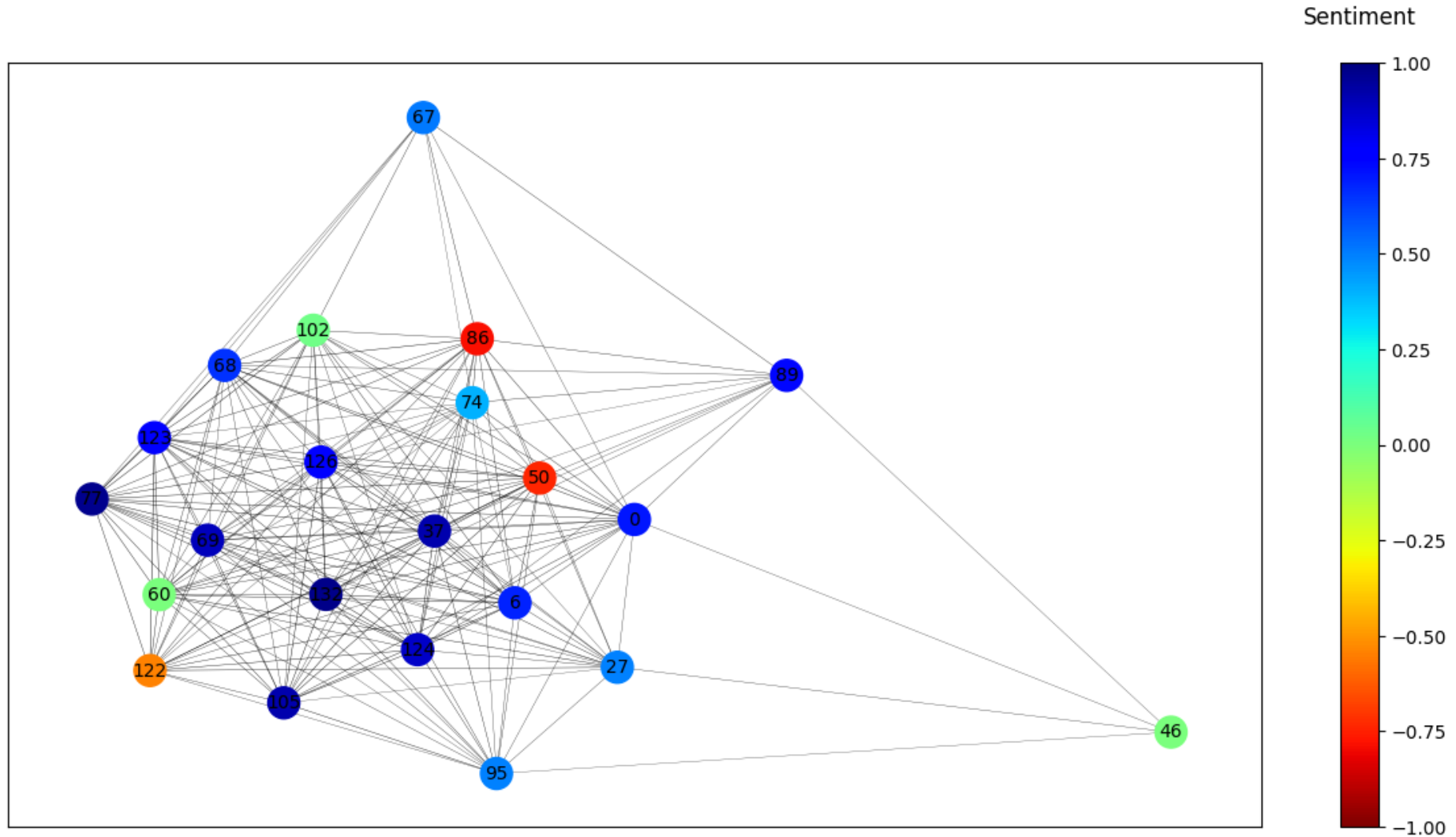| Date	# Nodes	# Edges	Density	Topics
16 Oct 1916	24	225	0.82	Marching/travelling. Conscription vote.
14 Nov 1916	26	218	0.67	Marching/travelling. Hospitals. Fighting. Work details.
12 Nov 1916	25	201	0.67	Marching/travelling. Parades/inspections. Camps. Fighting.
09 Nov 1916	26	198	0.61	Marching/travelling. Work details. Enemy aircraft. Fighting. Poor weather. Sailing.
25 Dec 1916	22	188	0.81	Christmas. Parcels/food. Poor weather.

Question 1: Results

Question 1: Results

Date	# Nodes	# Edges	Density	Event
25 Dec 1915	19	170	0.99	Christmas. Food/parcels. Egypt. News of the evacuation of Gallipoli.
11 Nov 1918	14	89	0.98	Armistice. Celebrations.
15 Feb 1916	10	42	0.93	Egypt/Suez canal.
02 Feb 1915	6	14	0.93	Egypt. Horses/camels/mules. Sailing.
06 Sep 1915	6	14	0.93	Fighting on Gallipoli.

Question 1: Results (25th Dec 1916)

Question 1: Future Considerations

- How does changing the similarity threshold change our results?

- Are there other metrics which are more suited to this task?

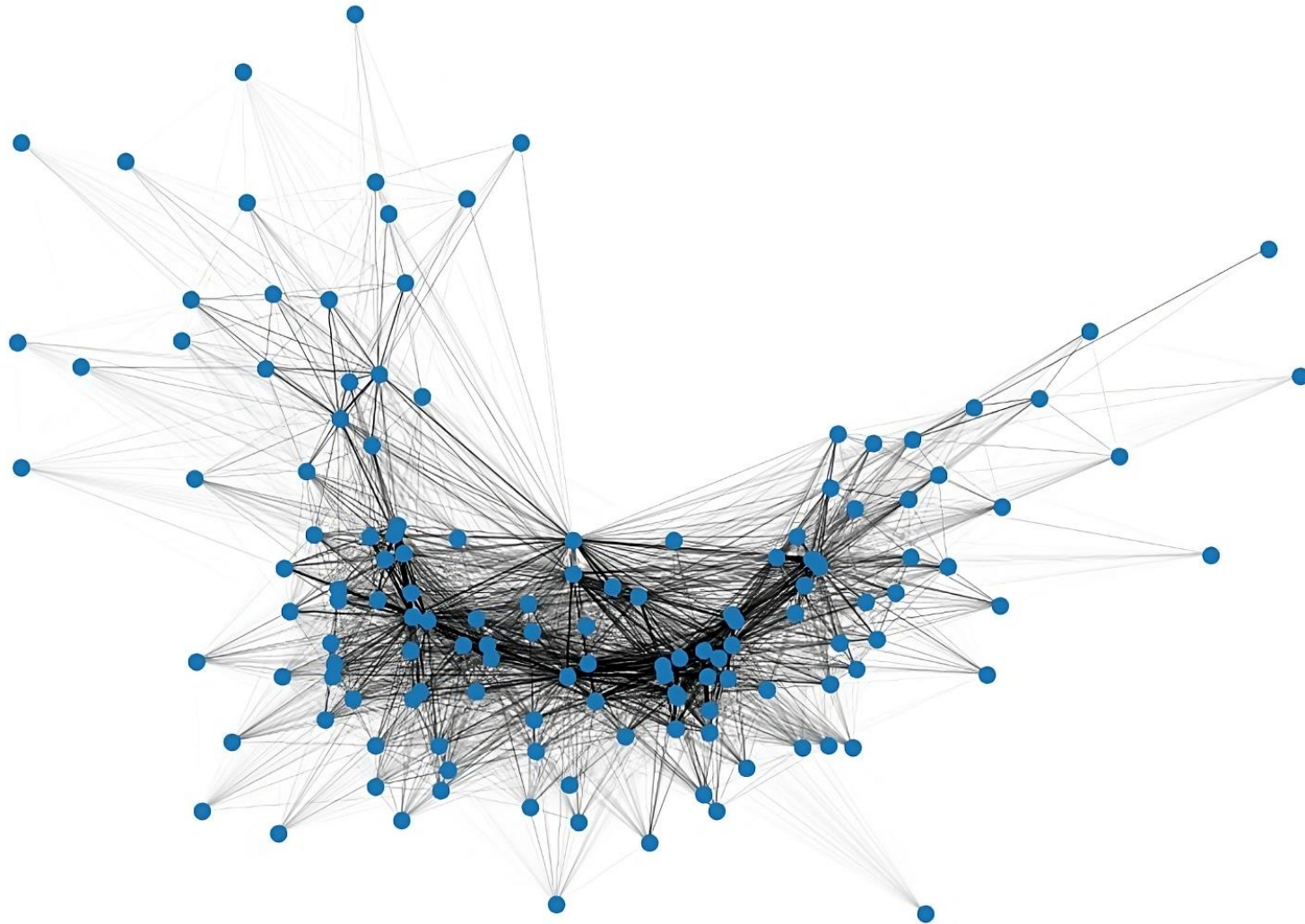- Does having an even number of entries across time change the results?

Question 2

Can network analysis be used to find diaries which are representative of this collection?

Question 2: Method

The temporal network is converted into a single static network where:

- Each node is a diary

- An edge exists between two nodes if those diaries have similar entries on at least one day

- Edge weights are the number of days in which those diaries have similar entries

Question 2: Network Visualisation

Number of Nodes: 139

Number of Edges: 3,551

Density: 0.37

Transitivity: 0.66

Number of Connected Components: 1

Question 2: Results

Diary ID	Degree	Weighted Degree
69	135 (1)	3,755 (3)
95	133 (2)	3,566 (4)
38	124 (3)	3,205 (10)
37	96 (8)	4,639 (1)
105	88 (13)	2,751 (13)
60	86 (14)	3,259 (9)

Question 2: Results

Diary ID	Number of Days Covered	Number of Entries	Topics
69	1,852	697	New Guinea, sailing, troop transport
95	1,626	607	Sailing, Egypt, Gallipoli, Western Front
38	1,326	449	Sailing, Egypt, Palestine
37	354	495	Sailing, Egypt
105	618	333	Sailing, Egypt, Western Front
60	566	294	Sailing, Egypt, Western Front

Questions?

Acknowledgements

I would like to thank the State Library of New South Wales for providing the transcriptions of these diaries.

I would also like to thank the School of Computer and Mathematical Sciences at the University of Adelaide and the Weir family (through the Hugh Martin Weir Prize) for supporting this trip.

References

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), Minneapolis, 2019, pp. 4171–4186.

J. Han, M. Kamber, and J. Pei, "Getting to Know Your Data," in Data Mining, 3rd ed., J. Han, M. Kamber, and J. Pei, Eds. Boston: Morgan Kaufmann, 2012, pp. 39–82.

C. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," Proceedings of the International AAAI Conference on Web and Social Media, vol. 8, no. 1, pp. 216–225, 2014.