

COMPREHENSIVE TEXT ANALYSIS REPORT: THE ATHLETIC



Andrew D'Armond

HULT INTERNATIONAL BUSINESS SCHOOL | TEXT ANALYTICS

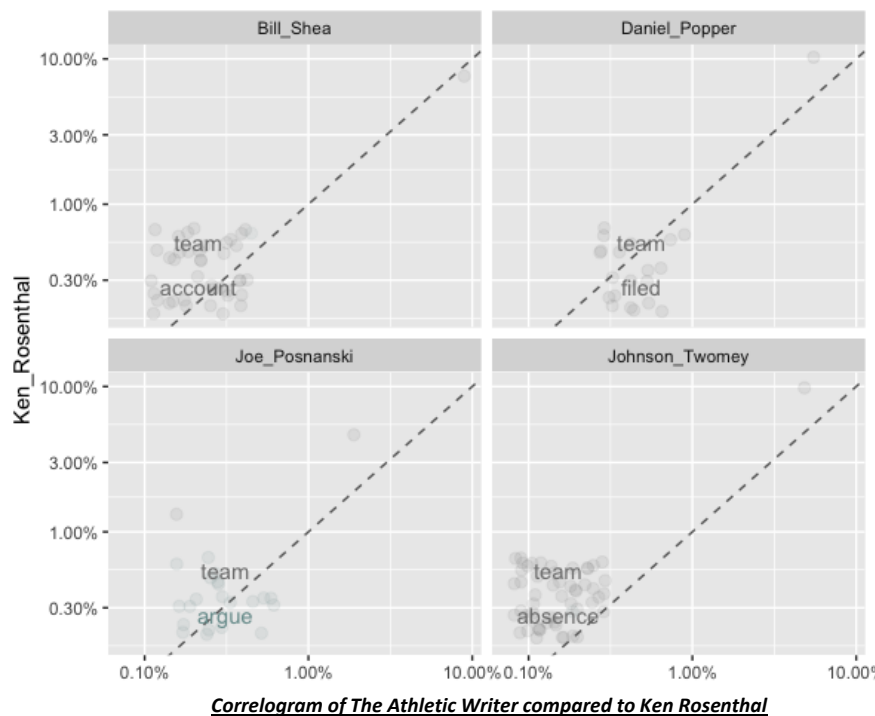
Introduction

The Athletic is a newly founded sports publication that is setting waves in the sports media world and for a simply annual subscription of around \$25, it's a bargain and a half. Founded in January of 2016, The Athletic features league wide beat writers that are highly regarded in David Aldridge, Jay Glazer & Ken Rosenthal. This is a report on 5 articles written by some of their respected writers and all are under the umbrella of "Sports Business" theme articles. Most sports columnists write on subjects that are more specific to teams and leagues if they are lucky to reach national scale. The idea to research this theme is because of the possibility of wide variety. This report will feature a variety of text analytics practices to mine through the articles in a combined and singular fashion in order to understand trends and provide insights into this new sports website. Both individually and collective will the authors be understood through the power of text analytics. It's imperative to understand that this is not an overarching review of the entirety of the sports publication but does provide a snapshot into the style of writing. The goal is understanding the publication thoroughly and provide conclusions of possible opportunities to further leverage their brand in styled sports writing.

Ken Rosenthal the Standard

The base of this analysis centers around a multitude of sports business-related issues, sport varies from 4 different sports in baseball, basketball, football, and soccer. All of the articles scored a 99% correlation score to the random chosen Y variable of Ken Rosenthal article on the Joc Pederson arbitration meeting. For those who may not know, Ken Rosenthal is one of the most revered baseball writers of today and surely will continue. He covers the World Series every year for Fox and constantly creates earthquakes with his hot stove news. The conclusion of Rosenthal's article through text analytics was that Pederson's arbitration is directly linked to Mookie Betts. This is due to the recent trade for the superstar outfielder and surely will cause future headaches worth having for the Dodger's front office. Taking a step back, the first idea of this analysis was to show the variety of the site under a wide topic in sports business. However, these 5 separate authors all have a correlation of 99% to each other which is incredible. The insight is the outstanding management that the Athletic takes in the organization of its publication. When comparing the other articles to Rosenthal's piece in a correlogram the most common word that appeared for Rosenthal is team. This is due to the relative nature of

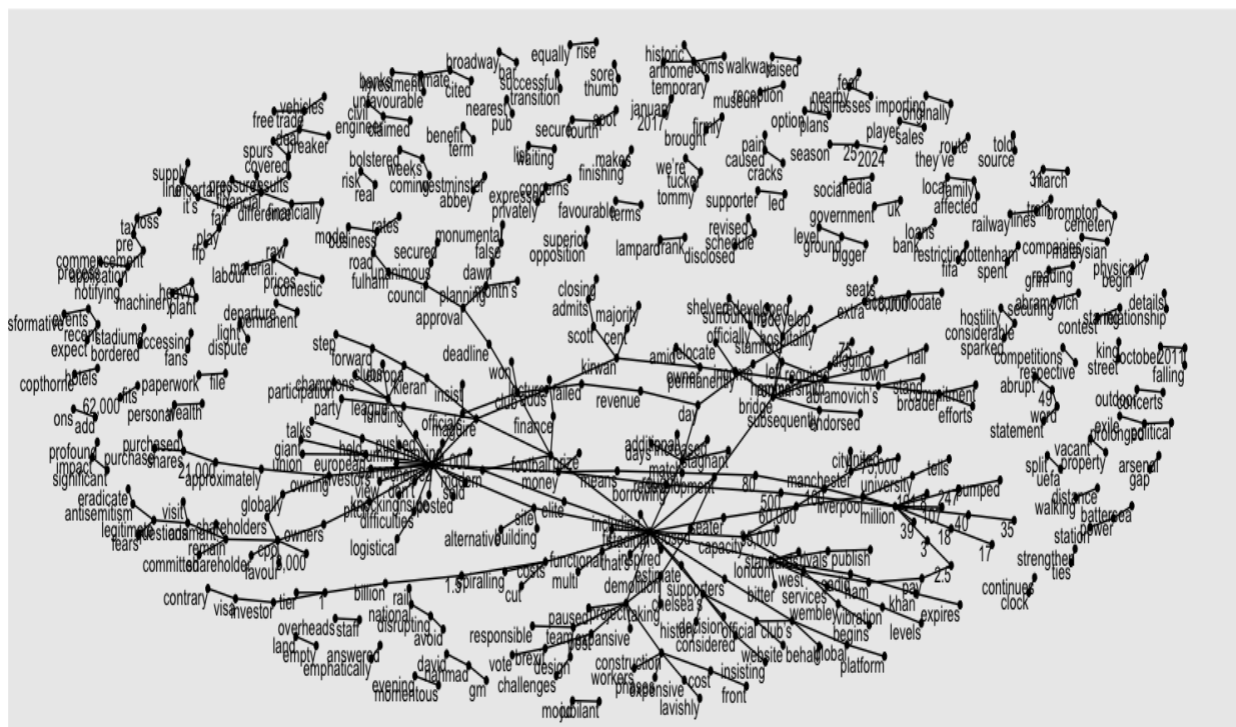
comparing team vs player talks and the combination of referencing to other situations in baseball namely the Red Sox. When studying the quadrograms you can see that the use of awkward arbitration meeting and the captivating mega deal for the Dodger's in the piece. This surprisingly led to a very unexciting sentiment AFINN score of 0.45. The scale is from 5 to -5 when looking at pieces of text such as this so you would think that article would be slightly higher but shows the quality of penmanship to keep the reader levelheaded. The last caveat of the article is the fact the bigram network is very scarce compared to the others yet to be discussed. This may be a lack of overarching content for the article, but still Rosenthal sets the standard.



Sports Analytical Writers

Drilling deeper, the other articles are no small fry as the total years of sports writing reach over 50 years between the next 5 authors over 4 articles. Joe Posnanski writes about a long-time author in the sports world that he would entertain his nightstand with. The books by Roger Kahn would be centered around the business of baseball as well. Daniel Popper writes about the tragedy of the NFL Game Pass usage during the Super Bowl. Johnson & Twomey write a scintillating piece on the construction of Chelsea FC Stamford Bridge. Finally, Bill Shea writes about Michael Redd, long time NBA player, and his relationship with Kobe Bryant in Venture Capital. All except Popper wrote pieces that scored above 0 in an AFINN sentiment test,

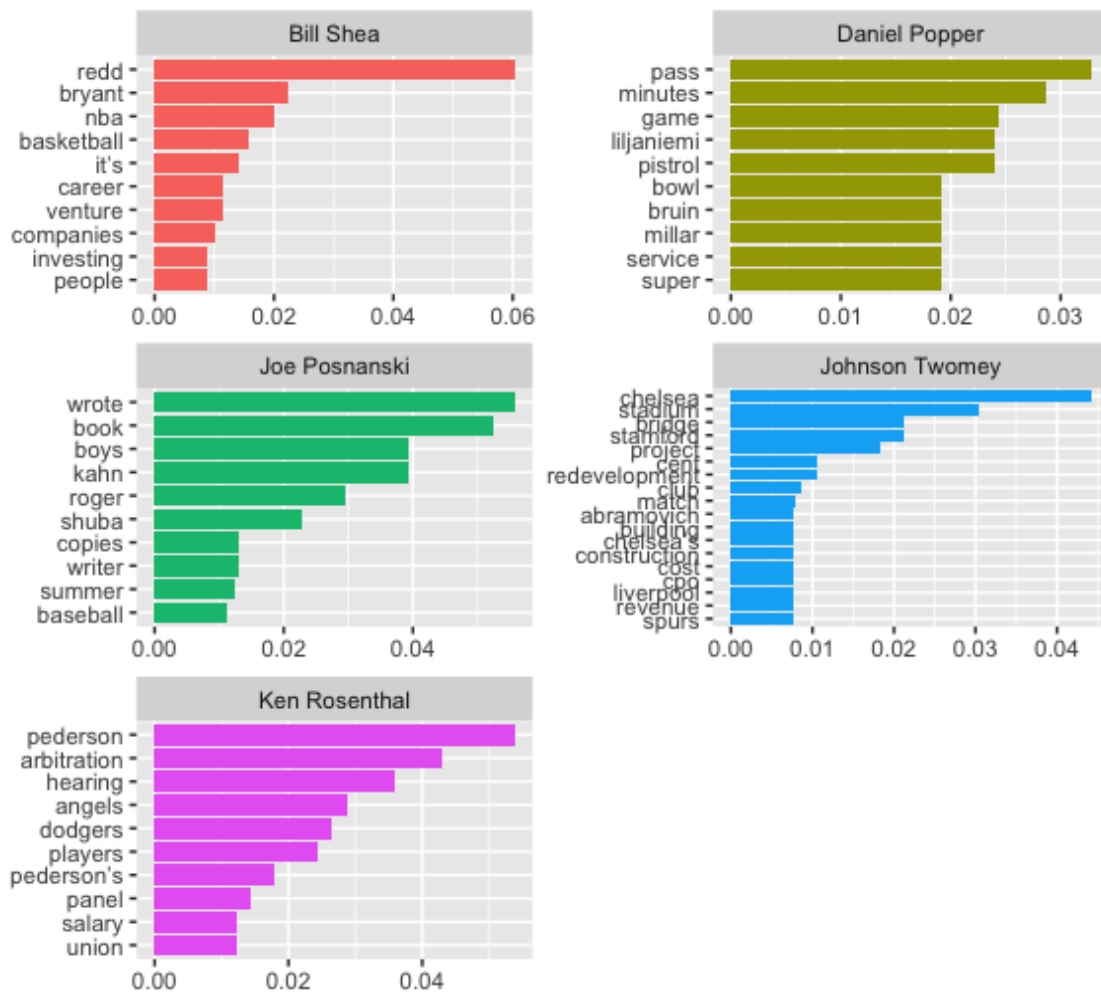
which Popper scored -0.17, none scored above a 1. Considering the theme of the articles as sports business this makes the literature interesting around with a hint of delight. Looking at all the articles in bigrams and quadrograms a few interesting trends appear. In Shea's article on VC in the NBA on over 50 occasions are deals of \$1 Billion mentioned. Surely showcases the heights professional sports athletes have rightfully reached today. Poppers article showcases that 181 countries used the service for NFL Game Pass to use for the Super Bowl along with a company called Bruin Sports being one of the major accomplices of the disaster in the stream's failure. Posnanski has stats abound in his piece as he writes about Pete Rose's 4200 hits along his total plate appearances of over 47,000. Finally, Johnson & Twomey share the insights about Chelsea receiving unanimous council aid along with the expected costs figures of 500 million pounds for the new bridge. Below is a bigram network chart of the article and is centered around a few obvious key words, but is highly connected, providing deep insights such as 21,000 shares were bought of Chelsea to facilitate the liquidity in assets to begin construction. The stadium is expected to open in the season of 2024-2025 or its capacity is 61,000. The insights are endless.



The redevelopment of Stamford Bridge. Written by: Johnson & Twomey

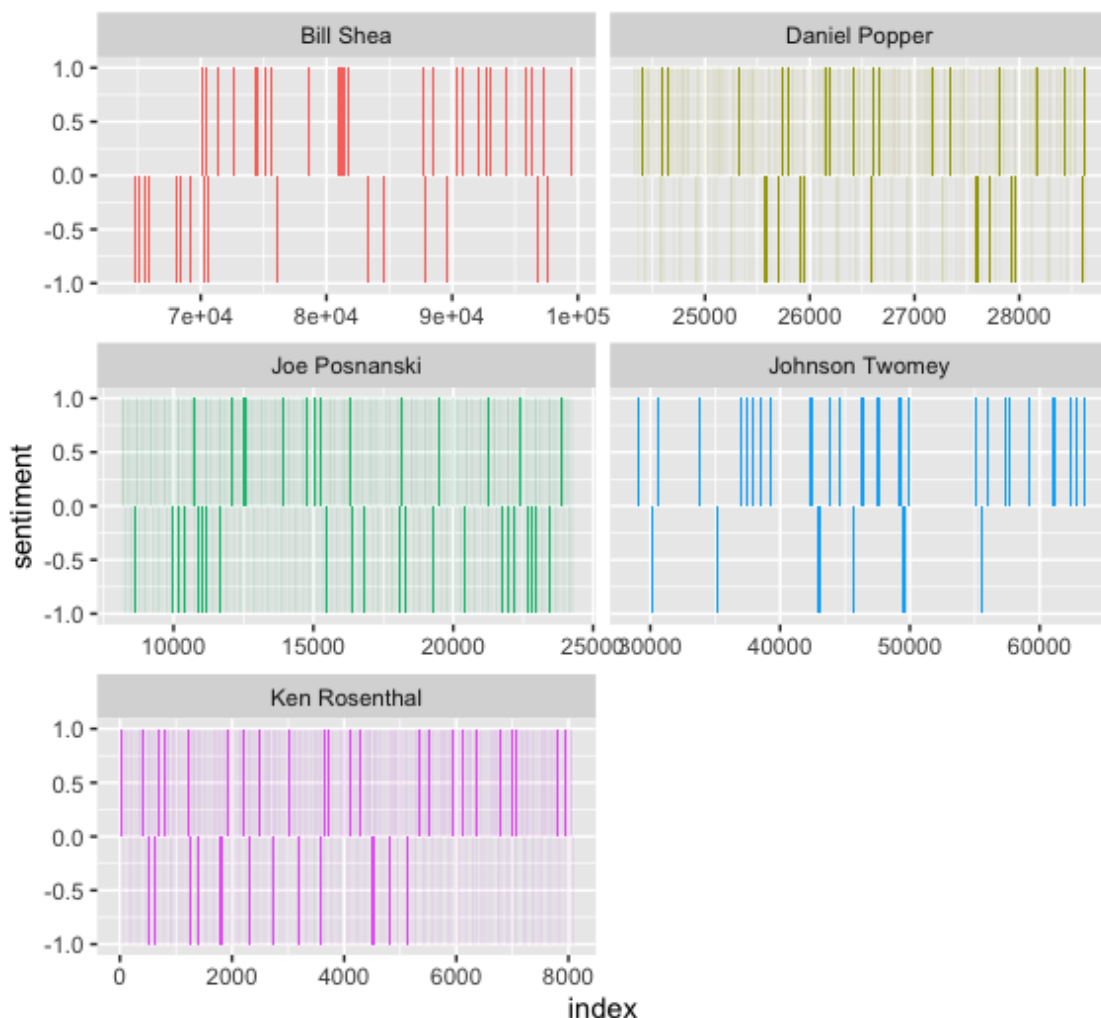
The Athletic

To further leverage the power of words we must include a comprehensive overview of the publication itself. This will begin with the understanding of the main practices we used for individual writers. This will be able to give us the insights we desire for actually using the manipulated text properly. First, one of the most powerful frameworks that one can use in this field is the tf-idf framework. In short it is a numerical statistic that is intended to reflect how important a word is to document, which is obviously very powerful when mining. From our other findings the writing is very expansive, and the extensive use of repeated words is unlikely, but that doesn't mean that patterns won't be found. As you can see below the framework has given us a few key insights. First being the Bill Shea writing gave us the most influential words of all the 5 articles with "Redd" as in Michael Redd. This makes logical sense when considering his choice of writing but power that it holds in comparison to its nearest competitors from other



TF-IDF Analysis for all writers

articles mainly from Johnson & Twomey in the word “Chelsea” or “Pederson” from Ken Rosenthal. Both of these words hover around .05 of influence, which is very high when considering the framework if you have used it in any case. Next, sentiment analysis is a terrific way to understand the tone of a writer or in any form of text. The below graph illustrates the sentiment from each author and shows the trajectory the author takes throughout their respective articles. As you can see the most sentimental article using NRC sentiment is Johnson/Twomey. Bill Shea and Daniel Popper were more negative in their respective introductions, but more importantly takeaway is that all of the authors try to leave on an upbeat not in their articles especially Rosenthal and Johnson/Twomey.



Sentiment across each writer's article trajectory

Conclusion

Sports provide most people an escape from everyday life, whether it is problems with work, school, family, etc. Yet, many people fail to highlight the impact that has come from the business end of sports. Sports drives local and national charity groups, creates personal connection that is rivaled by music, and drives reform on national issues. The business of sports is booming, and it is because of elegant literature that captivates communities, ingrowths lunch time, water cooler talks, and ingratiate world society today. The Athletic continues to receive new seed funding just within the last year and why? Simple, sports change lives. Many publications have washed away due to the overwhelming bias of their penmanship and drastic misguided views on where right and wrong stands today. Sports writers whether you like them or not have a binary code that guides them: Win or Lose. Whether an individual, team, or country win or lose the ink that dries from the parchment has ever lasting influence just like the bulbs that emit the final score. It's this that shape their speech and can forever make them more influential to consumers as they have a code that determines the philosophy of their literature. I read the athletic because I am sports fan and I will never repress the emotion that comes from my body, so it lifts me to write this report, which I hope you enjoyed. The Athletic publication can be leveraged by looking at the R code on my GitHub, reviewing the literature, and instigating policies such as maintaining low sentiment values for pieces in this category or setting standards of network images such as the one above. All of this is possible through text analytics and used wisely and can create insights such as the ones listed in this report to drive clicks, shares, and smiles.

References:

- Popper, D. (2020, February 7). NFL Game Pass crashes for international fans during final... Retrieved February 8, 2020, from <https://theathletic.com/1587426/2020/02/07/nfl-game-pass-crashes-for-international-fans-during-final-three-minutes-of-super-bowl/>
- Posnanski, J. (2020, February 8). In remembrance of 'The Boys of Summer' author Roger Kahn,... Retrieved February 8, 2020, from <https://theathletic.com/1592608/2020/02/07/in-remembrance-of-boys-of-summer-author-roger-kahn-who-found-beauty-in-the-stories-of-underdogs/>
- Rosenthal, K. (2020, February 8). Rosenthal: Delayed trade and awkward arbitration hearing... Retrieved February 8, 2020, from <https://theathletic.com/1592997/2020/02/08/rosenthal-delayed-trade-and-awkward-arbitration-hearing-prompt-more-friction-between-players-and-the-league/>
- Shea, B. (2020, February 6). Ex-Bucks star-turned-venture capitalist Michael Redd... Retrieved February 8, 2020, from <https://theathletic.com/1589082/2020/02/06/ex-bucks-star-turned-venture-capitalist-michael-redd-discusses-friendship-with-kobe-bryant-investing-career/>
- Twomey, L., & Johnson, S. (2020, February 7). Inside Chelsea: The clock continues to tick on the... Retrieved February 8, 2020, from <https://theathletic.com/1585623/2020/02/06/inside-chelsea-stamford-bridge-redevelopment/>

Please Visit my GitHub for all code to verify at your discretion.

<https://github.com/andrewdarmond/Andrew-DArmond>

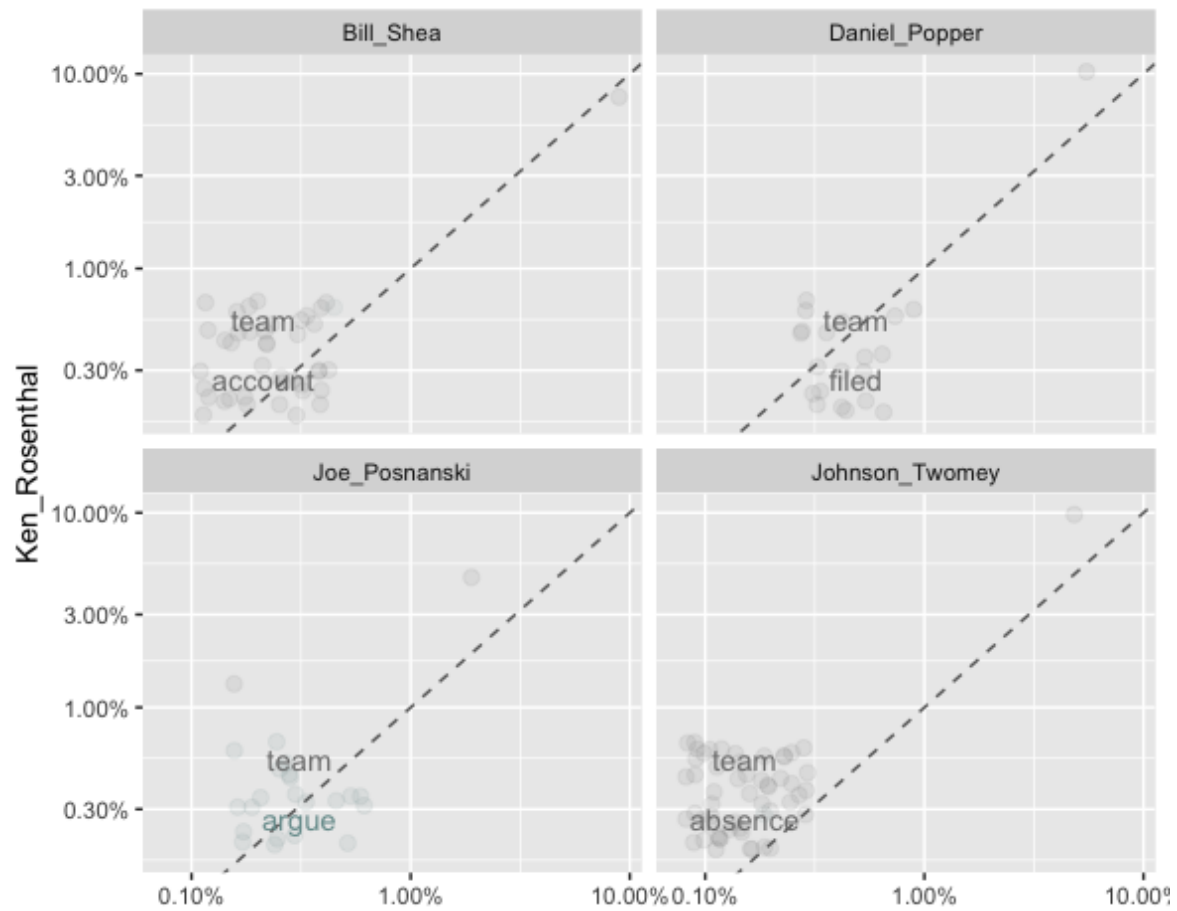


Figure 1 Correlogram

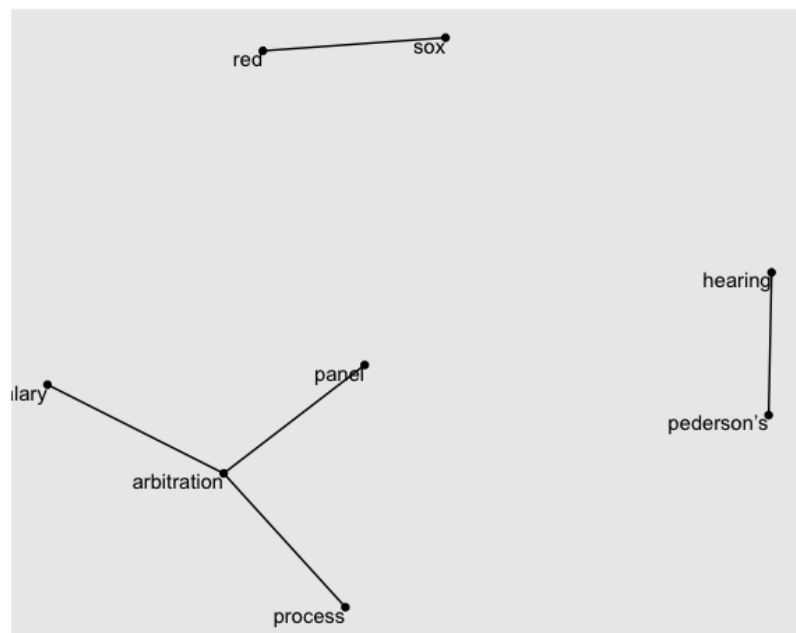


Figure 2 Rosenthal BIGRAM NETWORK

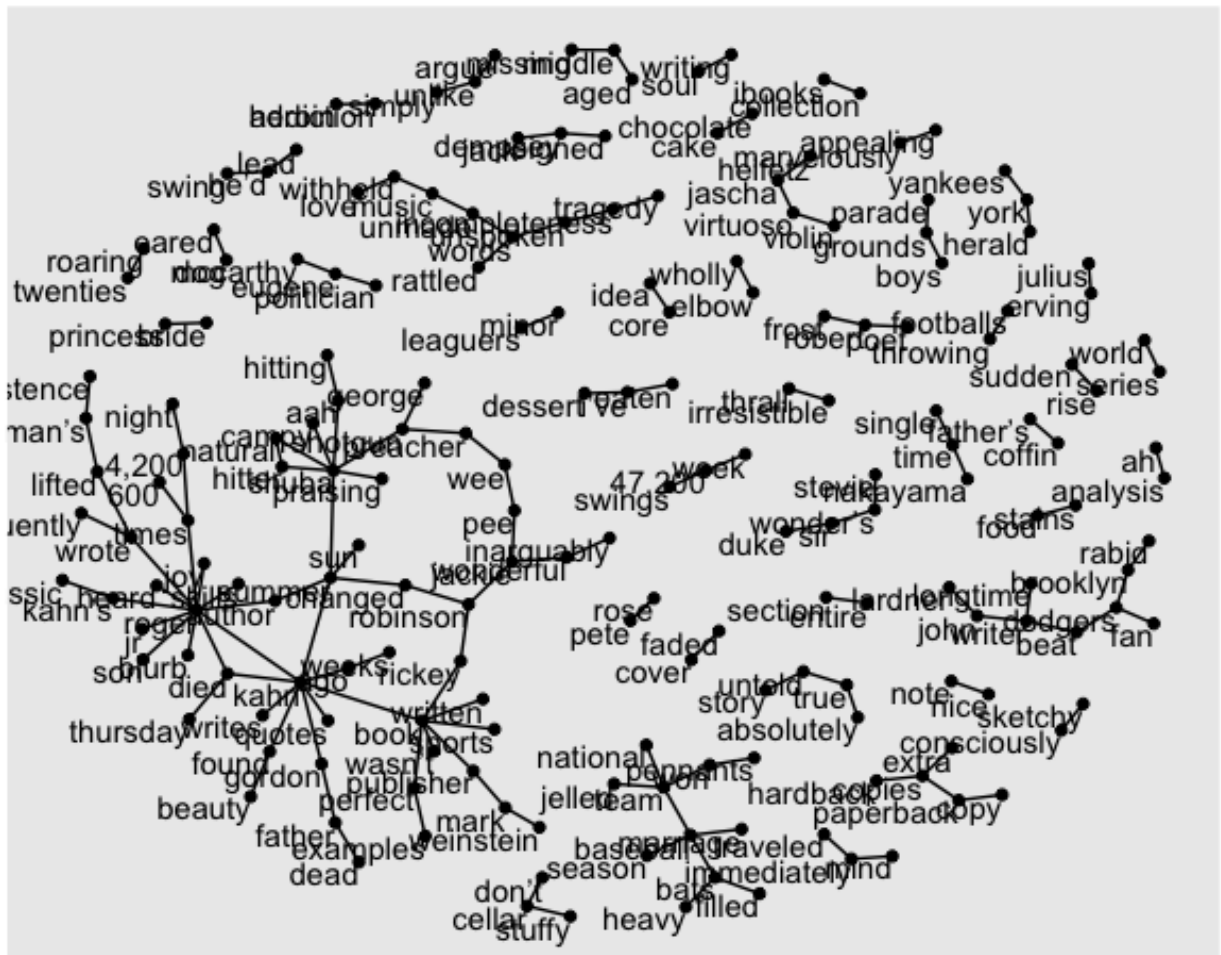


Figure 3 Posnanski BIGRAM NETWORK

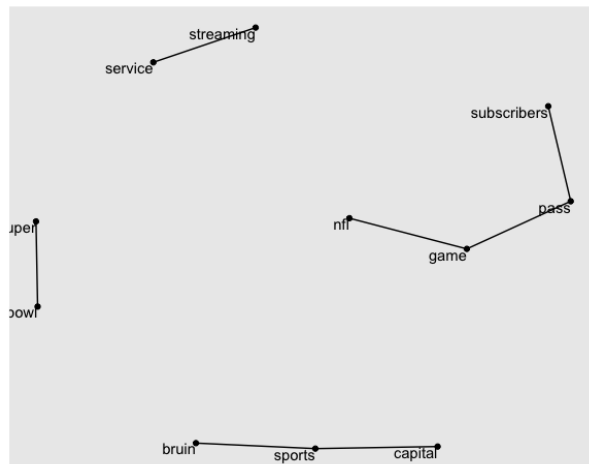


Figure 4 Popper BIGRAM NETWORK

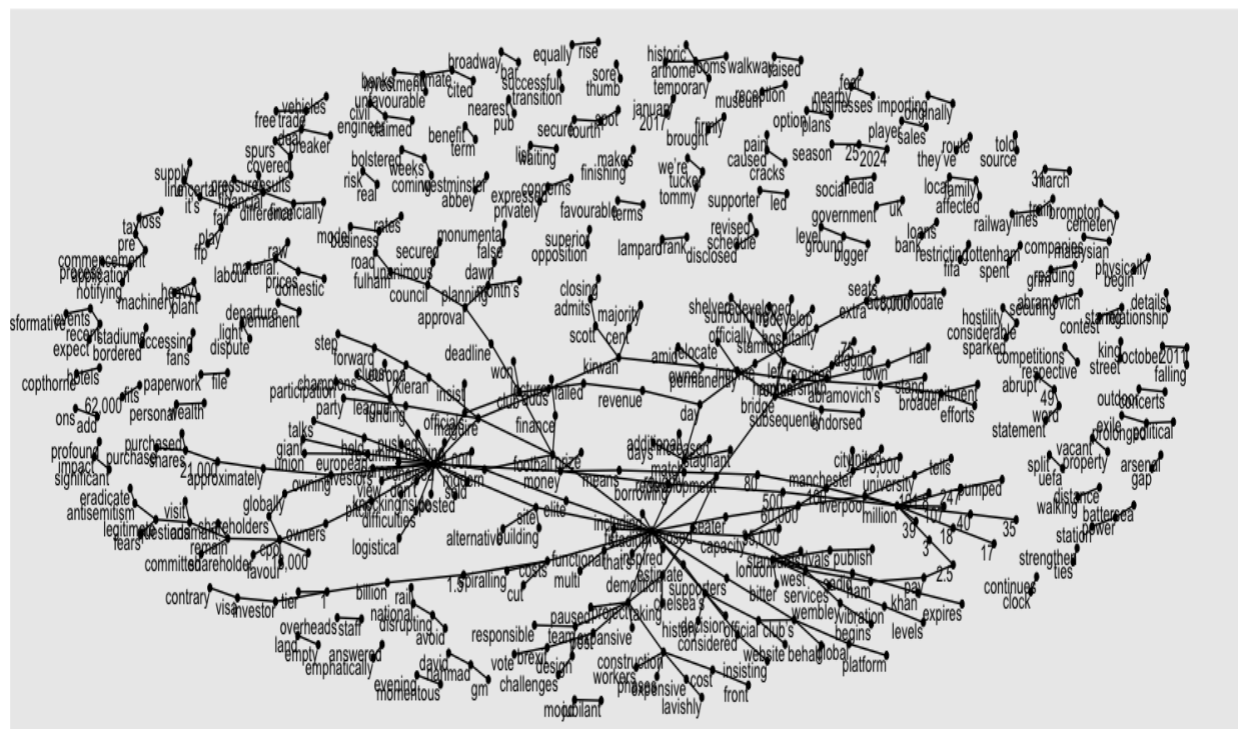


Figure 5 Johnson/Twomey BIGRAM NETWORK

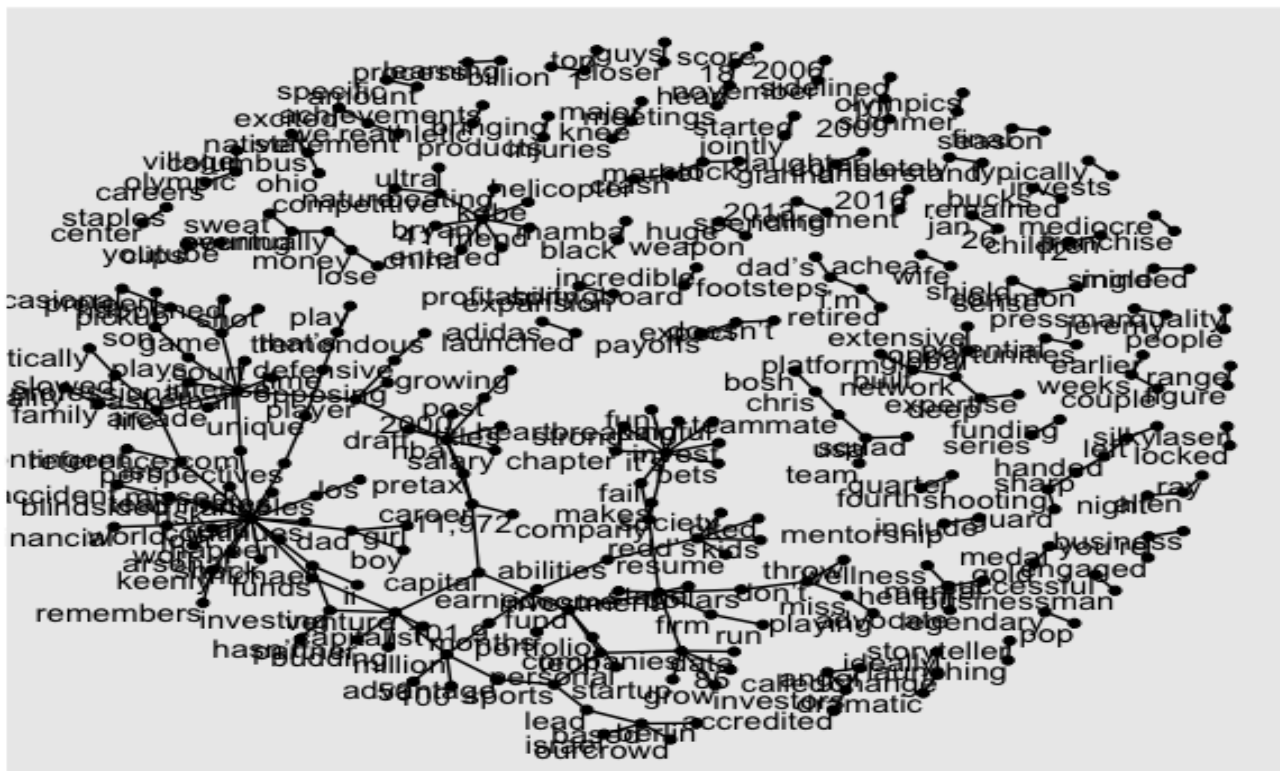
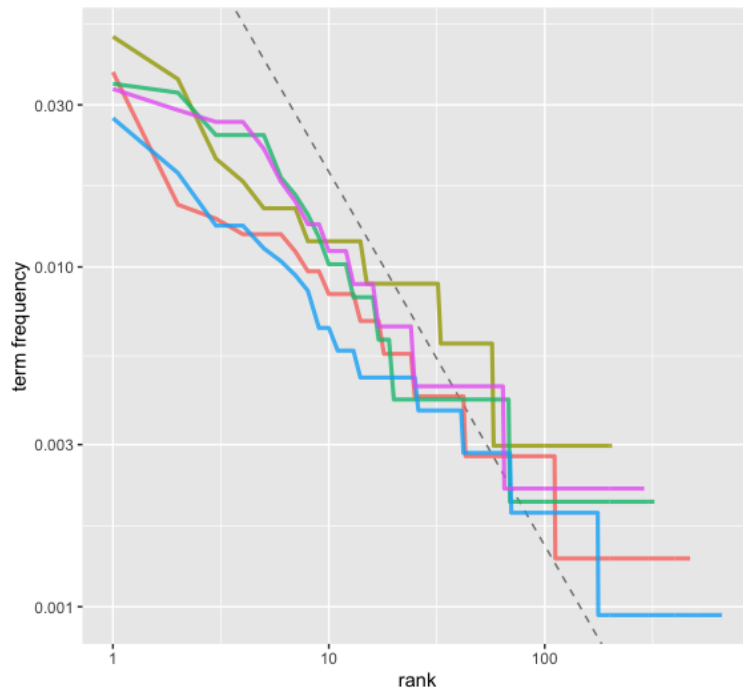
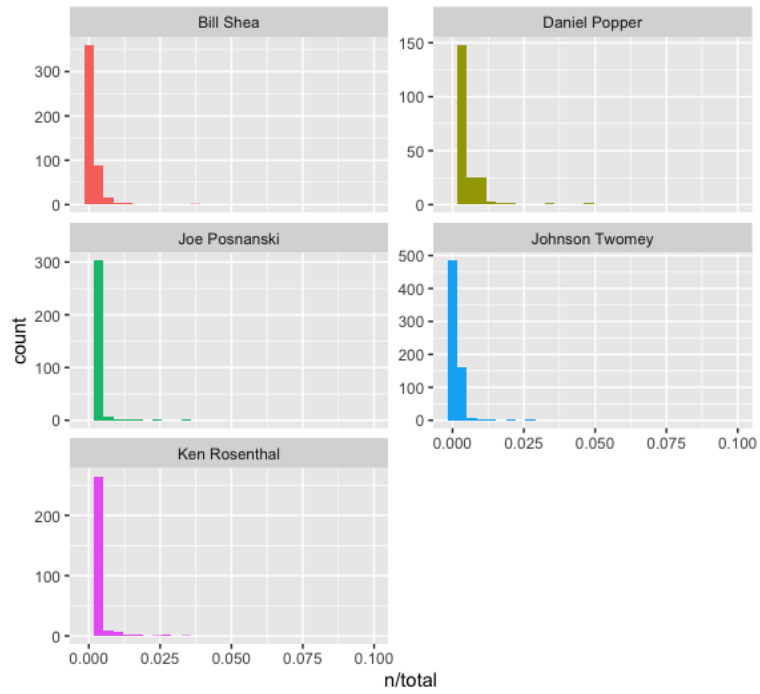
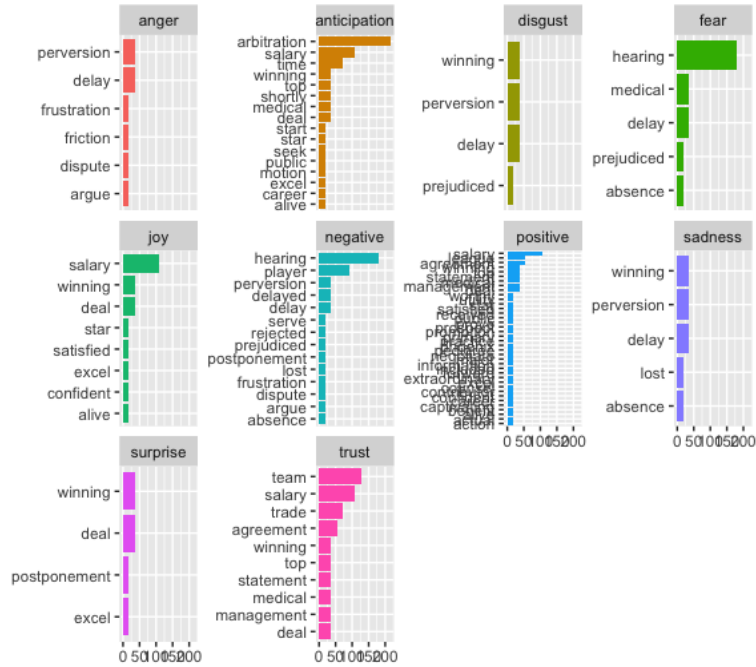
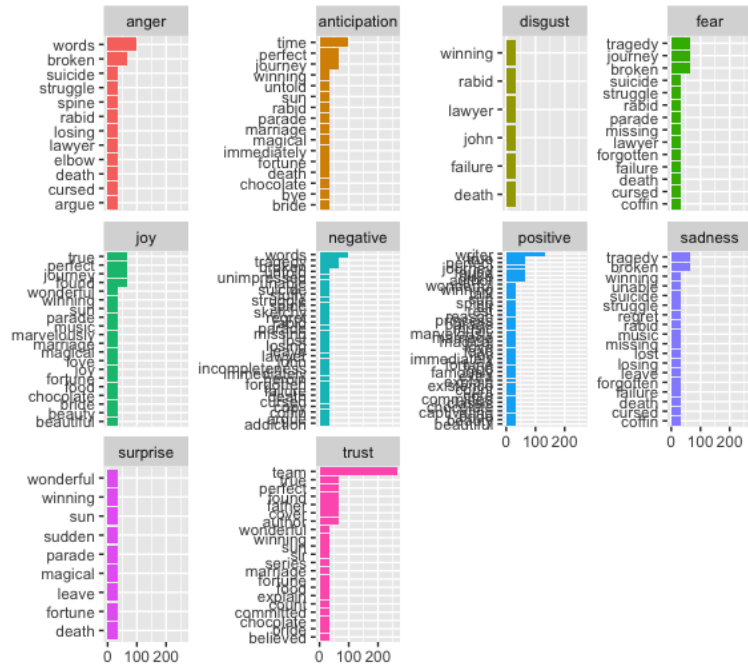


Figure 6 Shea BIGRAM NETWORK

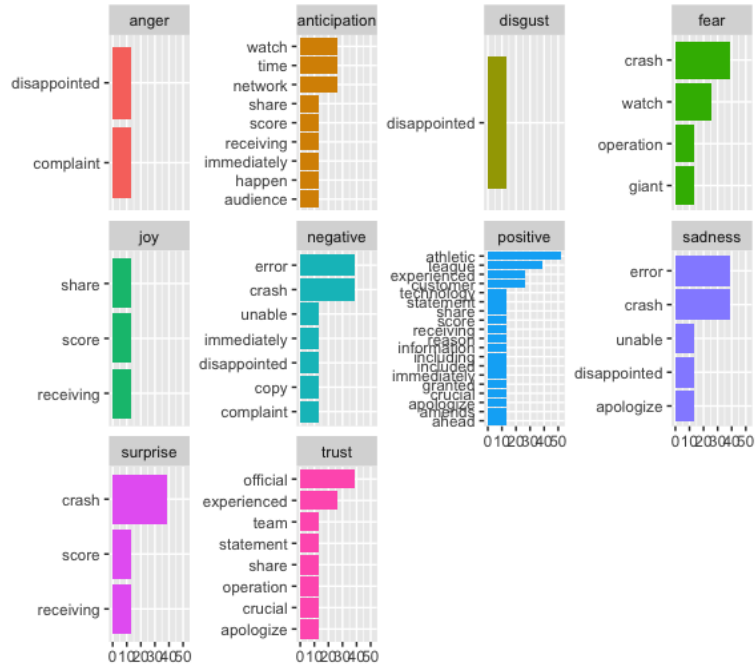




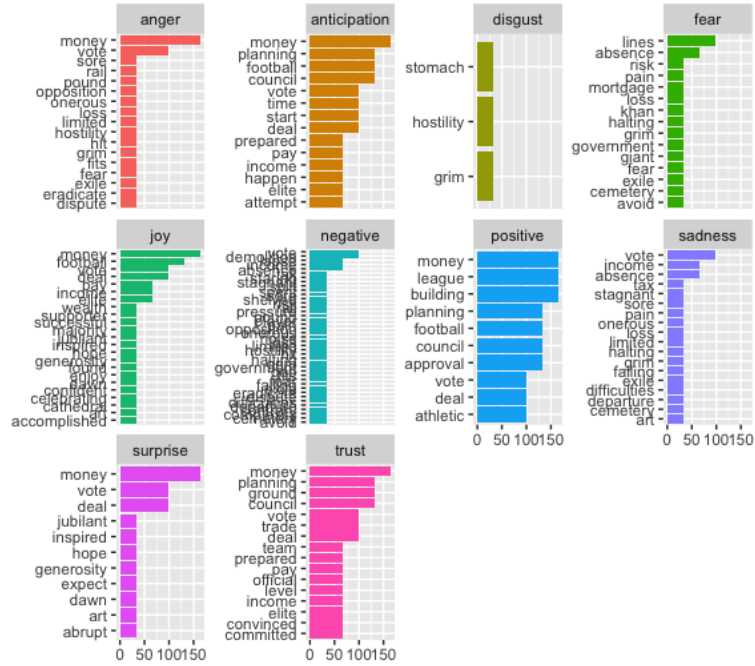
Contribution to sentiment



Contribution to sentiment



Contribution to sentiment



Contribution to sentiment

