Mapping Biases in Baseball Tales: The Black Sox Scandal

Sarah Hackney¹
¹Pratt Institute

Abstract
This research investigates how the use of network visualizations can be used to highlight potential bias in narratives and citation networks that prioritize the use of a single ‘definitive’ source, using the Black Sox Scandal as an example. Using the network visualization tool Gephi, a picture of player involvement in the scandal is depicted, as described in the first-person account of Chick Gandil. Reweighting the network using eigenvector centrality, it is shown how Gandil uses his own narrative to position himself outside the central nexus of blame for the events. The next step of this project is to use these same tools to evaluate the network of research that has taken place around the Black Sox Scandal, to determine whether the culpabilities and accepted series of events is weighted or distorted by Eight Men Out, the definitive popular work on the subject.

Keywords: Baseball; network analysis; research networks; information bias; visualization

do: 10.9776/16553

Copyright: Copyright is held by the author.
Contact: shackney@pratt.edu

1 Background
This research looks at a first-person narrative describing events surrounding the 1919 World Series, in which the Chicago White Sox allegedly deliberately lost to the Cincinnati Reds, in exchange for financial compensation from various underworld entities. This is commonly known as the Black Sox Scandal. For further reading on the Black Sox Scandal, refer to the “Resources” section at the end of this paper.

In 1957, Arnold “Chick” Gandil wrote an essay for Sports Illustrated, giving his version of the events. Gandil was a player on the 1919 White Sox team, and is considered by scholars to be the central figure in planning and executing the fix. Although this account was penned almost 40 years after the events it describes, it remains one of the few first-person narratives of the scandal. In it he outlines how the Black Sox players were approached by gamblers about fixing the World Series, were promised money they were never given, and eventually decided to “play clean” and lost on their own merit, rather than by choice. This description of events is in direct conflict with the accepted series of events, which describe Gandil as the one who approached mobsters and gamblers.

Having been fingered by his teammates as the primary instigator of the fix, Gandil uses his own narrative to shift the blame for the scandal away from himself. By creating a visual representation of the events from Gandil’s perspective, we are able to better understand how he constructed his own position within the scandal, and the biases that his story reflects.

Baseball lends itself to network analysis. It can be divided into entities on the player, team, division or league level, and each of those entities is constantly interacting with other entities in describable ways. Additionally, baseball is an exceptionally well-documented sport, in both historically and in terms of gameplay metrics and measurements. In 2011, Beckman and Chi published a paper in the Baseball Research Journal, the flagship journal for SABR, the Society for American Baseball Research, outlining their preliminary findings that baseball players who are better connected— that is, have played on teams with a wider variety of teammates, perform better offensively. (Beckman & Chi 2011)

Additionally, there is a deep vein of statistical scholarship attempting to quantify the gameplay during the 1919 World Series in order to determine which players were “in on the fix.” Much of the scholarship on this subject focuses on “Shoeless” Joe Jackson, the White Sox’s left fielder, one of the great players of his generation. The Society for American Baseball Research has a committee dedicated to continued historical and statistical research on the subject of the Black Sox Scandal, but very little work has looked at Gandil’s own account of the events.

2 Methods
Using the first-person account published in Sports Illustrated in 1957, and retrieved using the Internet Archive’s Wayback Machine, a list of all the people and places that Gandil names in his account of the scandal was made. This included the players on the White Sox, players on the Cincinnati Reds, baseball
ownership and management, members of the press, gamblers and other underworld entities, and places such as hotels, racetracks and baseball parks.

This was followed by the creation of a table of relationships between the entities as described by Gandil. At this point, the focus of the project was narrowed to involve just the interactions directly related to securing the fix, betting and exchange of money. All of the on-field activities during the playing of the 1919 World Series were discarded, including gameplay for both the White Sox and Cincinnati Reds. Bennett (1991) explores the gameplay element of the series in depth, particularly as it relates to “Shoeless” Joe Jackson. The elements of baseball play and the interactions of the players on the field are not within the scope of this project, and by focusing more closely on the interpersonal relationships between players off the field, with their management and other outside forces, the resulting visualization was able to hone in on how blame is distributed by Gandil in the network his story creates.

This narrowed the data down to 24 nodes, and 114 edges. The edges represent relationships ranging from “spoke to” all the way to “banned from professional baseball for life.” As a result, these edges are weighted on a 4-point scale, with 1 representing amiable relationships, 0 neutral or passing interactions, -1 unfriendly relationships, and -2 reserved for the lifetime bans issued by Kenesaw Mountain Landis to eight Black Sox players.

A visualization was then created. Because Gandil’s account is, at its essence, a description of a series of social interactions, network analysis, as visualized by Gephi, was chosen as the ideal means of representing the series of events that Gandil describes. Creating visualizations of complex social data makes identifying and isolating significant persons and interactions a more intuitive process for researchers and audiences alike (Card, Mackinlay, and Shneiderman 1999).

3 Findings

By inputting the relationships, as represented by nodes and edges, into the network visualization software Gephi, a picture of what Gandil describes is created. As seen in Figure 1 (below), the persons involved with the scandal are divided into three groups.

The top group, with light blue nodes, contains the players and management of the Chicago White Sox who were not involved with the World Series fix. These are relatively separate from the group of White Sox players with grey nodes, who were involved with the scandal—the Black Sox. This is a representation of the divide in the White Sox clubhouse between those who got along with owner Charles Comiskey, and those who did not, well-documented by Gandil (1957) as well as by Asinof (1963) and Linder (2007). Additionally, the gamblers, with green nodes, originating from Arnold Rothstein, the money, and possibly the mastermind, behind the fix, connect to the Black Sox players, but not to the ‘clean’ players or staff. In addition to these groups, there are a few other outside entities who act upon the Black Sox players, with red nodes: Jake Lingle, a reporter, The Cook County Grand Jury, who acquitted the players who went to trial, and Kenesaw Mountain Landis, who banned them from baseball for life. This network creates a shape for the story as Gandil tells it, showing a group of men who were already separated from the good graces of their sport, being attacked on all sides by offers of dirty money, rumors and other indignities.
When eigenvector centrality is run on the network, and the size of the nodes adjusted to reflect the results, another story emerges. Eigenvector centrality measures the relative "importance" of a node in a network, measured by how well a particular node is connected to other important nodes. In figure 1 (above), Gandil’s eigenvector centrality is the lowest (his ‘dot’ is smallest) among all the Black Sox players. Additionally, Eddie Cicotte’s is largest. We can interpret this as the slant that Gandil brings to his telling of the story. He says at the end of his account, “I held a deep resentment against Cicotte [...] I felt I would never forgive the guy.” (Gandil 1957) By continuing to position Cicotte as the central figure in the scandal, Gandil shifts focus away from himself and onto Cicotte, which is measured in Figure 1 by the eigenvector centrality of each man.

4 Discussion & Next Steps
This visualization serves to highlight how the bias of Chick Gandil shapes the network he describes in his account. In all cases, the information we use to conduct research and evaluate options on an everyday basis are biased. There is no such thing as “neutral information.” The use of eigenvector centrality in this case creates a quantifiable way of measuring that bias, by showing the relative closeness of players in the events as Gandil describes them.

The next steps of this project is to evaluate the realm of research taking place surrounding the Black Sox Scandal in a similar manner, by examining the network of authors, researchers, and citations.
on this subject. This information will be gleaned from the archives of the Baseball Research Journal, the academic publication of SABR, the Society for American Baseball Research. Articles and resources related to the Black Sox Scandal will be drawn from the archives through keyword search of the text of the Baseball Research Journal, and via the Index of SABR Publications (1972-2005). By analyzing these publications, this project hopes to not only evaluate the state of connectedness in research concerning the Black Sox Scandal, but also to consider the narrative biases in Black Sox research potentially created by the central popular narrative of the events, that of Eliot Asinof’s *Eight Men Out*.

5 References

6 Resources
Baseball Research Journal Archives http://sabr.org/content/baseball-research-journal-archives
Gephi, https://gephi.github.io/
Index of SABR Publications (1972-2005) viewable at https://drive.google.com/file/d/0BzJhDMhwdOaWeENWNmxyVDU2R3M/view?usp=sharing