

# Draft Introduction

David Freed and Samuel Green

May 1, 2016

The wisdom of the crowds is not a new phenomenon. Economists treat the collective opinion of independent rational agents as sacrosanct; efficient market hypotheses are, at their core, expressing a fundamental belief in the wisdom of the crowds. Statisticians can explain the idea as the reduction of systemic risk—the variance of the sum of  $n$  independent and identically distributed random variables is smaller than the variance of any individual r.v. by a factor of  $\frac{1}{n}$ . In politics, prediction markets have quickly become a better predictor of election results than polls or the opinions of the experts<sup>1</sup>.

The last example evidences the importance of being able to properly assess the wisdom of the crowds. While betting on political elections is illegal in the United States, the number of bets on the current presidential race is up fourfold since 2012 in Ireland.<sup>2</sup> Betfair, the largest online betting exchange for U.S. presidential elections, has nearly one million users and records nearly seven million transactions a day, which it claims is “more than all European stock exchanges combined”.<sup>3</sup> In this cutthroat industry, any edge a gambler can get matters.

The betting market on politics pales next to the worldwide sports betting market—a colossal enterprise whose size experts ballpark at around \$1 trillion dollars a year. In a market with roughly the GDP of Indonesia, understanding the underlying statistics (in this case, the relative quality of the teams) is fundamental. Any bettor who can consistently

---

<sup>1</sup><http://www.wnd.com/2016/01/prediction-markets-more-accurate-than-polls>

<sup>2</sup><http://abcnews.go.com/Politics/las-vegas-bets-hillary-clinton-literally/story?id=30897911>

<sup>3</sup><http://corporate.betfair.com/about-us/betfair-facts.aspx>

outperform the crowds can turn a predictable profit; one has to win only 52.4 percent of his bets to break even in Vegas.

In our paper, we tackle the common questions about the wisdom of the crowds from a different angle: instead of asking whether crowds or experts, alone or in aggregate, can predict games before they happen, we look at their ability to understand and predict games that are ongoing. We consider Twitter data from 43 games in the National Collegiate Athletic Association (henceforth, “NCAA”) Men’s Basketball Tournament (henceforth, “March Madness”) to identify metrics that gauge the level of interest and the sentiment of the crowd at any point in time.

The justification for using Twitter as a predictive mechanism is simple: it can update far quicker than standard predictors like the margin of the game. If a star player goes down with injury or picks up a fourth or fifth foul—forcing him out of the game—the immediate effect will not be seen in the margin, but Twitter users will be able to accurately process the effect it will have on the game. Likewise, while Twitter can distinguish margins that are unsustainable (e.g. leads built on fluky plays or low-percentage shots) and those that are not, standard statistical models have a difficult time doing so.

Prior research has taken a cursory look at these questions; previous studies of National Football League (henceforth, “NFL”) games identified that the volume of Tweets before a game is predictive of the final outcome and that Twitter is reactive to big plays in the game. Our work builds upon these analyses, not only by substantiating prior results for a different sport<sup>4</sup> but also by distinguishing between Tweet volume and Tweet sentiment. Prior papers considered only the volume of Tweet in each contest to measure the wisdom of the crowd; we use a sentiment classifier to determine how Twitter feels about both teams and how that evolves over the course of the game. In our Empirical Results section, we demonstrate that Twitter sentiment has predictive power that Twitter volume does not.

---

<sup>4</sup>Given the differences between football and basketball, this is a non-trivial contribution. Since basketball has far fewer breaks than football, one might imagine that Twitter would be significantly slower to reach to big events (fewer timeouts and breaks of play with which to process what happened). We show this not to be true in the Empirical Results section.

Our paper has three main results. First, to establish a baseline for our future analysis, we demonstrate empirically that Twitter is responsive to important game events—Twitter volume and sentiment will increase in response to salient changes in the score. This provides evidence for our initial claim that Twitter is responsive to the events currently going on in the game.

After showing that Twitter can capture what has happened in the past, we argue that Twitter is a useful predictor of what happens in the future. Our second result demonstrates that Twitter sentiment and Twitter volume is a statistically significant predictor of future changes in margin; the aggregate Twitter sentiment in any period<sup>5</sup> is predictive of the change in margin in the next period. Here we find evidence for our prior claim about unobservable data—while margins tend to show mean reversion (i.e. teams that are up by a lot in period  $t$  tend to see their margin shrink in period  $t + 1$ ) as a whole, Twitter sentiment helps to distinguish which leads will continue to increase.

Our final result tests whether Twitter data can be used as an effective predictor of the final result at any point during the game. We use logistic regression models to incorporate the sentiment in a given period. We find that the Twitter sentiment is only statistically significant for about the final quarter of the game, but models that include Twitter sentiment outperform standard models<sup>6</sup> at every single minute of the game. We conclude by showing that incorporating Twitter sentiment is more valuable than simply incorporating raw Twitter volume, as other papers have done.

The remainder of the paper proceeds as follows. In Section 2, we review the brief literature on the subject, demonstrating both the advances and the gaps in prior research. In Section 3, we detail our data collection process, explaining how we matched gametime data (e.g. “7:30 remaining in first half”) to real-time data (e.g. “9:30 PM”) and discussing the construction of both our relevance and sentiment classifiers. In Section 4, we provide an overview of

---

<sup>5</sup>This is roughly measured as the support for one team minus the support for another team.

<sup>6</sup>For the purposes of our analysis, we cite the logistic prediction models rolled out by FiveThirtyEight during this year’s March Madness as the standard model for predicting the end of the game

our empirical results and a thorough discussion of the aforementioned three main results and their significance. In Section 5, we conclude and discuss the important caveats and extensions to our work.