Social Media Audience Metrics as a New Form of TV Audience Measurement

Darryl Woodford, Ben Goldsmith, and Axel Bruns

The Uncertain Business of Audience Measurement

Understanding and acting on the behavior of media audiences is a multi-billion dollar business. Broadcasters and other media providers, advertisers, advertising agencies, media planners and audience research companies have significant financial stakes in the collection and analysis of audience data. In addition, policy makers, academics, and audience members themselves have interests in the technologies and methodologies used to measure audiences, as well as in the data itself and the uses to which it is put. But the audience rating convention – the necessary consensus among stakeholders about who and what is counted, how the counting is done, how the data is interpreted and how it is valued – is under pressure as never before. Digitization, media convergence and audience fragmentation have dramatically disrupted the business of audience measurement. New metrics and analytical systems have been developed to answer some of the questions raised by technological change, but they are also posing challenges to stakeholders about their capacity to deal with the explosion of raw and customized data on audience behavior. The volume of information that is available for aggregation and analysis has grown enormously, but with that growth has come a host of uncertainties about audience measurement, and in particular about the broadcast ratings system.

Uncertainty has driven an extraordinary research effort, a flight to accountability, in which a proliferating number of information and research companies have tried to make sense of the accumulating data about media use, often with conflicting results. This was one of the reasons behind what Alan Wurtzel, President of Research at American broadcaster NBC, has called the “crisis in measurement” (Wurtzel, 2009), although the wealth of data and the efforts being made to analyze it may mean that this period could be looked back on as a golden age if the industry’s ideal scenario – the collection, cross-tabulation and fusing of massive amounts of data and large datasets – can be realized. This would potentially produce the advertising industry’s holy grail: single source, or consumer-centric holistic measurement (WFA, 2008), although serious questions would also arise, not least about privacy and audience members’ and consumers’ awareness of the data that is being collected (Andrejevic, 2007, 2013).

In some respects, the current state of uncertainty is nothing new. Historically, the introduction or expansion of commercial broadcasting services, changes to the structure, economics, technologies, or the policy field of broadcasting, and evolving patterns of audience behavior have all spurred the development of new technologies, methodologies and rationales for quantifying television audiences. For various reasons primarily to do with establishing the parameters for the buying and selling of airtime in predominantly commercial or mixed public service/commercial broadcasting markets in countries around the world, consensus has tended to form around the need for an authoritative, simple measure of exposure – who is watching television, which channel or service are they watching, and for how long. There has long been great (and recently, increasing) interest in measuring audience members’ engagement with programming and advertising – how much attention they are paying, what their opinion is about what they are watching, and what impact the program or commercial has on them – but exposure has remained the standard for measuring broadcast ratings and the core of the ratings convention ever since Archibald Crossley’s first survey of American radio listeners in 1929 (Balnaves, O’Regan, & Goldsmith, 2011). Despite the contemporary crisis, which is multifaceted, ratings data are still and will continue to be in demand because there will always be the need for common currencies for buying and selling advertising and program content. There will undoubtedly be changes in the practicalities of audience measurement, particularly given the challenges presented by the likely spread of broadband-enabled set-top boxes, which have been described by the Council for Research Excellence as the “wild west” of research (Council for Research Excellence 2010, p.4).

The availability of multiple channels through subscription or free to air television, coupled with ever-increasing online video options, amplifies viewer/consumer choice and consequently distributes the available audience much more widely than earlier broadcasting systems. Napoli (2003, p. 140) argues that this fragmentation increases the disparity between the predicted and the measured audience and reduces the reliability of data collected in traditional sample-based methods. It certainly increases what has been called the “research ask,” and complicates the carefully calibrated equations that produce the ratings. Although mass audiences can still be assured for certain major events, often live international sports championships, audiences in general have dispersed. Content providers, advertisers and research organizations have had to track not only timeshifting and catch-up TV, but also migration across platforms, and even beyond the home. Audience fragmentation has precipitated proliferations of data, methods, metrics and technologies that in turn have allowed samples of a few hundred, in panels or diaries, to multiply into surveys of millions of subscribers and produce competing currencies. Opportunities for advertisers to reach consumers through media and other touchpoints have proliferated, while advertisers’ and content providers’ desire for solid numbers and discontent with the prevailing currency and methods have opened spaces for research and analysis.

Public service broadcasters have typically been more interested than their commercial counterparts in qualitative research that provides detailed information about audience enjoyment and engagement with programming. This, for example, was the focus of audience research conducted by and for the BBC from the mid-1930s (Silvey, 1951). For commercial broadcasters – as, eventually, for public service broadcasters too – ratings have served a range of purposes, from measuring the popularity of particular programs, providing guidance in program planning and scheduling, informing service delivery, keeping abreast of change in audience tastes and practices, and establishing the value of time sold for advertising. Ratings can also act as a proxy for the broadcaster's share price and an indicator of (and influence upon) its overall financial health (Balnaves et al., 2011). For advertising agencies, media planners and advertisers themselves, ratings help determine how much will be spent on advertising on a particular channel or network, as well as where and when advertisements will be placed. But ratings are not only used within broadcasting. They are also of interest to the mainstream media and the public at large for what they appear to reveal about the success or otherwise of programs and broadcasters, as well as to academics, media critics and public authorities who "use, quote, debate and question" ratings (Bourdon & Méadel, 2014, p. 1). In Canada, for example, the media regulator uses ratings as one measure to judge the success of CanCon (Canadian Content) drama policies, and as the basis on which funding for future production is allocated (Savage & Sévigny, 2014). Criticism of the ratings has come from many quarters and taken many forms, from theoretical and technical questioning of the methodologies and technologies deployed over time to concerns about the business practices of data suppliers, and the tendency of those who use the ratings to "endow the audience with a reality and thereness it does not possess" (Balnaves et al. 2011, p. 229). And yet, despite the disruption wrought by digitization, a variety of parties continue to maintain a variety of interests in the collection of robust, reliable and commonly agreed upon metrics about audiences, as well as in agreeing on what counts as an audience.

As Alan Wurtzel observed of the situation in the US in 2009,

> A couple of years ago, Nielsen delivered a single TV-rating data stream. Today, Nielsen routinely delivers more than two dozen streams (yes, we counted them) and countless more are available for any client willing to pull the data. Moreover, set-top boxes (STB), moving closer and closer to second-by-second data, will produce a staggering amount of new information. And, with internet and mobile metrics as well, it's not the amount of data that is the problem; it's the quality and utility. (2009, p. 263)

This is the key challenge for ratings providers in the future: providing quality and useful data. But given that so much is in flux, including common understandings of "quality" and "utility," it appears for the moment as though multiplication of research vehicles and partnerships will inevitably continue as ratings companies jostle over currencies and simultaneously provide bespoke services to individual clients.

In addition to Nielsen's multiple streams and the wealth of other services available, broadcasters, content providers and advertisers must also contend with the power of bottom-up systems of recommendation and rating that have emerged with the Internet. From Facebook's "Like" option which allows readers to signify in a single click their approval or appreciation of something posted by a friend (importantly, there is no option to "Dislike") through sharing and retweeting on Twitter to supporting (or Digg-ing) something posted to Digg.com, the opportunities for audiences to register opinions or rate all sorts of things on the Internet are many and varied. To varying degrees, research companies, advertisers and content providers are realising the importance of social media in gauging audience opinions about the quality of content. The characteristic online behavior of countless people now routinely involves what futurist Mark Pesce (2006) calls "the three Fs:" finding, filtering and forwarding information found online to contacts (or followers in Twitter-speak, friends on Facebook). In contrast to more restricted media such as free-to-air broadcast television, audiences can now find desired audiovisual content, or close approximations, online. The actions of tagging, rating and recommending function as forms of feedback, often for the principal benefit of the audience's own network. But ever more sophisticated and insistent forms of monitoring behavior and turning it into useful data are capturing this information, adding it to databases for dissection and fusion. In terms of quality, audience members who follow or forward content on multiple media are exactly the audiences that producers of media content are trying to cultivate, in part because of the ratings they may provide in the future. It is the most committed, the most voracious of the online explorers or pioneers, the keenest edge of the community that evolves around content, who can shape the media choices of those around them, who will be most highly valued by producers, if not always by advertisers. All of these developments point to the likelihood that measures of popularity in social media will become more extensive in the future. In the remainder of this chapter, we discuss the ways in which particular forms of social media analysis can produce useful and actionable data about engagement with television that augment and extend the ratings' core focus on exposure.

Toward Social Media-Derived Audience Metrics

Traditional television ratings schemes provide a standardized and broadly reliable, but ultimately limited and one-sided, measure of audience interest; historically, they provide information on what audience research could readily and regularly quantify, but fail to offer any fine-grained, in-depth evaluation of audience activities even from a quantitative perspective, much less from a qualitative one. In the emerging multi-channel, multi-platform, multi-screen environment, they become manifestly insufficient.

Audiences for televisual content now access their shows through a range of channels: in addition to conventional reception of the live television broadcast, they may also utilize streaming cable or broadband catch-up services, time-shifted pause and rewind functionality, or (official or unauthorized) video downloads. Such services may be offered by a wide range of providers and platforms, including the original domestic broadcasters, their counterparts in other geographic regions (where new shows may screen ahead of the domestic broadcast date, and become accessible to users outside the region through the use of geo-masking VPN services), by video streaming platforms such as YouTube (where content may have been uploaded by production companies, one or several regional broadcasters, or fans), and by download services from the Apple iStore to Bittorrent filesharing sites.

Audience engagement with such content remains identifiable and quantifiable in most of these cases: on-demand platforms from official catch-up services to unauthorized filesharing sites each generate their own usage metrics, even if they are not always shared publicly. To date, however, such metrics have yet to be aggregated and standardized in any reliable form; a number of scholarly and industry research projects have attempted to do so for individual platforms, but several such studies, especially by industry-affiliated market research organizations, are also flawed by an underlying agenda to promote fledgling on-demand services or prove the impact of content piracy.

Further, significant challenges exist in ascribing meaning to these metrics. Mere figures describing the number of requests for specific on-demand video streams or downloads may be highly misleading if they turn out to be inflated by multiple requests from the same user due to poor server performance or broadband throughput; even unique user figures may be misleading if there is a significant influx of audiences from outside the intended region of availability through the use of VPNs and other mechanisms. Recent research suggests, for example, that on-demand movie and television streaming service Netflix has already gained a 27% share of the overall on-demand market in Australia, even though Netflix does not officially operate in that country (Ryall, 2014). Australian Netflix users' activities are therefore likely to inflate the metrics of the U.S. platform to which they have managed to connect.

Figures for on-demand requests and downloads also fail to accurately capture the quality of engagement with the televisual content thus accessed: was a downloaded video actually watched? Did viewers watch the entirety of the program? Here, in spite of their own limitations, even conventional television ratings provide a more comprehensive picture of audience engagement, because they are at least able to track audience sizes at regular intervals during a broadcast, and thus to offer a glimpse or audience attrition or accretion rates. In their use of demographically representative panels of television households, such conventional ratings also continue to provide more detailed data on the popularity of specific programming with particular audience segments; this is likely to be absent from the metrics for alternative channels, where demographic data are often rudimentary at best.

Such information is especially crucial for broadcasters in the public service media sector, where an application of conventional ratings to programming which is often deliberately designed to address specific niche interests and audiences can significantly misjudge the ability of such programming to meet its intended aims. Here, evaluating the forms and quality of audience engagement is often more important than simply measuring the total size of the audience. But for commercial television channels, too, such information provides important clues which feed back into the design and production of new programming; there is therefore a significant need to move beyond the limitations of merely quantitative audience measurements.

Media, communication, and cultural studies scholarship has a long history of recognising the active audience of mass media programming (Fiske, 1992), but has traditionally found it difficult to measure the extent and impact of audience activities or provide comprehensive qualitative evidence beyond individual small-scale case studies. That is, scholarship in this field has established the necessary conceptual tools for evaluating and understanding diverse forms of audience engagement, but has so far lacked access to a substantial base of evidential data on audience activity to which such tools may be usefully applied in order to determine and categorize the forms of audience engagement with media content which are prevalent in the contemporary media ecology, or to evaluate their meaning and relevance in the context of the specific public service and/or commercial aims pursued by media organizations.

This situation has shifted markedly in recent years, especially due to the emergence of second-screen engagement through social media as an audience practice that accompanies the viewing of televisual content. Such engagement has turned the active audience of television into a *measurably* active audience that generates a rich trail of publicly available evidence for its activities, and this trail can be gathered through the Application Programming Interfaces (APIs) of mainstream social media platforms, or internally from the access logs of the engagement platforms operated by broadcasters themselves. With the computational turn (Berry, 2011) in humanities research, such data may now be used to test and verify the conceptual models for audience engagement that have been developed by media, communication, and cultural studies disciplines, in order both to quantify the level of such activity for individual broadcasters and their programming, and to benchmark the quality of this engagement against the aims and ambitions set by the content producers.

This focus on using social media activities as an indicator of audience engagement is not without its own limitations, however. In the first place, social media audience metrics require active television audiences also to be active *on social media*, and may thus privilege particular audience demographics that are especially likely to be using platforms such as Facebook and Twitter to discuss their television viewing. Further, social media-based engagement with televisual content is likely to be greatest when individual users are able to engage with other viewers of the same programming in close to real time; such metrics continue to privilege live or close to live viewing (through conventional broadcast or streaming services) rather than significantly time-shifted access. For major television events, a considerable social media audience around a shared televisual text is likely to persist at least for several hours, perhaps even days, before and especially after the live broadcast, so that the measurement of social media audience activities need not necessarily require exactly simultaneous engagement with the same text. This is demonstrated for example by the global social media response to television events such as new episodes of popular series from *Doctor Who* to *Game of Thrones*, which are typically screened in different timeslots but in close temporal proximity to each other in different broadcast regions around the world.

If such limitations inherent in the data derived from television-related social media activities can be successfully negotiated, then a range of new opportunities for quantifying as well as qualifying audience engagement with televisual content emerge. First, a number of comparatively simple audience metrics may be established, including the volume of postings that relate to specific programming, and the number of unique users generating such audience responses. Here, the substantially improved precision of public social media data compared to conventional ratings data makes it possible to identify almost to the second which moments in a particular broadcast generated the greatest audience response, and thus how such activity ebbed and flowed with the progress of the show; a measurement of unique active users over the course of the broadcast also offers first insights into the influx or exodus of viewers. Various contextual factors must also be considered in such analysis, however – different program types and formats may lend themselves more or less well to continued social media activity, for example: audiences may be glued to the screen during drama programming, and post social media updates only during commercial breaks, while during political talk shows they may be more prepared to respond to the panelists' statements on a continuous basis.

Additionally, publicly available background data derived from the social media platforms themselves may also be brought to bear on the analysis: for example, in addition to measuring the total number of users participating in a social media conversation about a given show, it would also be possible to determine the number of social media friends or followers for each user's account, and thus to evaluate the extent to which the broadcast has been able to attract highly networked (which may be read as "influential") participants. Similarly, if background data exist not just about the size of such friendship networks, but also about their structure (as Bruns, Burgess, & Highfield, 2014, have developed it for the Australian Twittersphere, for example), it becomes possible both to pinpoint the location of individual users within that network, and to determine the total footprint of a particular programme within the overall social media platform.

Such indicators begin not just to quantify total engagement, but also to provide a post-demographic alternative to the audience segmentation models of conventional ratings: as social media networks are often structured not primarily according to geographic or sociodemographic factors, but by similarities in interests, this approach to analysing social media-based audience activities offers insights into whether a specific broadcast was able to achieve deep engagement with those sections of the overall network which are particularly concerned with the broadcast's topics, and/or whether it managed to generate broad engagement irrespective of users' day-to-day interests and preferences (cf. fig. 1). Depending on broadcaster and program type, either or both of these objectives may be desirable: a political talk show on a niche public broadcast channel may seek deep engagement with a narrowly defined group of so-called "political junkies" (Coleman, 2003), for example, while a broad-based entertainment show on a major commercial station would aim for responses from as broad a public as possible. Again, it should be noted that such analyses assume that engagement by the social media audience either provides a reasonable approximation of engagement by the wider television audience beyond specific social media platforms, or that it is possible to correct for the demographic and post-demographic skews in the measurement of audience interests and activities that such a focus on social media-based engagement activities produces.

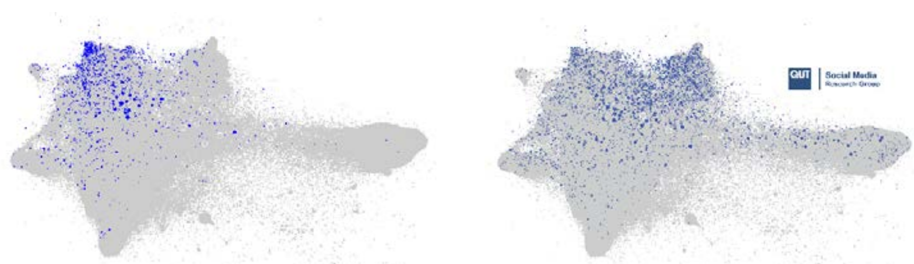

Fig. 1: Social media footprints of different TV programming in Australia – Twitter-based engagement with political talkshow Q&A *(left) and the 2014 Grand Final of the Australian Football League (right). Against the backdrop of a follower network map for the 140,000 most connected Twitter accounts in Australia (in gray), actively tweeting accounts for either broadcast are shown in blue.* Q&A *tweeters are recruited predominantly from a network cluster focusing on politics (top left); Grand Final tweeters from a cluster focusing on sports (top center), but with much wider take-up across the Australian Twittersphere.*

Finally, the immediate availability of audience members' social media responses to specific televisual programming also enables a qualitative analysis of their reactions beyond mere engagement metrics. It becomes possible, for example, to extract from the content of audience posts the key themes and topics of their responses, which may highlight the names of popular (or at least controversial) public figures, organizations, and actors, and to chart their relative centrality to the programming over the course of individual episodes or entire seasons. This can also feed back into programming choices, from featuring popular journalists and presenters in current affairs programming to enhancing storylines for favorite characters in drama series. Such approaches may also seek to explore the use of sentiment analysis, in order not only to determine the volume of mentions for specific themes or persons, but also to identify the tone and context in which they are mentioned (Is a reality TV contestant controversial or popular? Is the coverage of a topic appreciated or criticized?); it should be noted in this context, however, that the effectiveness of current sentiment analysis techniques in processing the very short texts of social media posts remains disputed (Liu, 2012; Thelwall, 2014).

Context-Sensitive Approaches to Measuring Audience Engagement

To a great extent, and perhaps as a reflection of the persistence of ratings thinking in social media audience measurement, existing approaches by commercial research enterprises to the analysis of social media data around television are largely based on relatively simplistic volumetric measurements. Nielsen, for example, uses the SocialGuide platform to rank shows according to what the company terms the "Unique Audience" of a show; that is, the estimated number of Twitter users who could have seen a tweet about a show (Nielsen Social, 2014). But this measurement fails to account for the different contexts in which shows air: for example, in the US it compares shows screening on the less subscribed USA Network to those broadcast by the mainstream national network ABC, and places moderately popular FOX afternoon sporting events on an equal footing with primetime pay-per-view wrestling broadcasts. UK operator SecondSync, which has now been purchased by Twitter, Inc., similarly ranks social media activity by two volume-based metrics: total tweets, and tweets per minute; in both cases, it also compares shows on different types of network without accounting for their underlying differences (SecondSync, 2014).

Such approaches to social media audience metrics are clearly and significantly limited in their ability to measure engagement effectively. For instance, a simple ranking of shows by the total number of tweets they have received ignores the number of tweets posted per user, and thus fails to differentiate between, on the one hand, broad but shallow engagement by a large number of moderately committed viewers and, on the other, deep but narrow engagement by a dedicated niche audience of fans. These generic metrics also implicitly assume that the mode of engagement with a show is the same for viewers of all formats; that is to say, they assume that audiences engage in the same way with a reality TV show as they do with a drama, for instance. But this is disproved by SecondSync's own data, which show that the peak of audience activity for drama broadcasts often occurs after the conclusion of an episode, whereas for reality TV viewers are more likely to tweet during a show (Dekker, 2014). Although a ranking of shows by their tweets-per-minute average may allow for such genre-specific variations in audience engagement, it does not incorporate any evidence of sustained engagement with a show; a show that flat-lines except for a moment of major social media controversy would rank highly by this metric, compared to a broadcast which receives solid and steady engagement throughout.

Metrics that seek to quantify sustained audience engagement, and do so with regard to the specific characteristics of that engagement, would then already be a significant improvement over currently available measurements. When seeking to understand the social media footprint of television shows, it is important that contextual factors that affect social media users' engagement with television content are accounted for. In particular, it would be desirable to normalize available measures of the volume and dynamics of content posted through social media, and thus of social media engagement with a show, by accounting for underlying systemic factors such as the geographic reach of a broadcast network, the weekday and month of a broadcast, the broadcasting genre, or the show's time slot. In this way, viewer engagement with a high-budget primetime drama on a major television network could be benchmarked more meaningfully against the social media activities around a reality TV show airing on cable television. Rather than simply comparing raw volume figures, which will always favor major channels and primetime broadcasts, comparisons could thus be based on measurements of a show's social media performance relative to the long-term average for engagement broadcasts on the same channel, in the same time slot, and/or of the same genre.

Given that this critically depends on accounting more comprehensively for the broadcast context of a given show, it is logical to consider other fields in which contextualizing statistics is significant. Noteworthy new impulses for the further development of social media engagement analytics come from the field of sports metrics, where data analysts have long faced a very similar challenge to that which underlies audience measurement: separating the signal from the noise (Silver, 2012). Sporting analytics has addressed this challenge by seeking to account for the fact that traditional measurements of team performance (wins and losses) and players (individual statistics) can be influenced by a wide range of factors beyond the skill level and performance of a player on the field, including the skill of other players on a team's roster, the standard of the opposition, and the playing conditions of a specific match.

Despite recent developments in ice hockey, basketball and American football (Moskowitz & Wertheim, 2012), as well as soccer (Anderson & Sally, 2013), baseball analytics remains the most developed of these fields, through the work of researchers such as James (1982), Silver (2003-2009), and Tango, Lichtman, & Dolphin (2007). The field of baseball analytics that has emerged from their efforts is called Sabermetrics (named after the Society for American Baseball Research, SABR); we therefore refer to our adaptation of these methods to the study of television audience engagement on social media as *Telemetrics*.

For the purpose of interpreting and improving contemporary audience engagement metrics, the most useful sporting analytics we may draw on are those that seek to separate a player's actual performance from the contextual factors outside the player's control that may have affected it. In baseball, pitchers have historically been evaluated through a statistic called ERA, or Earned Run Average, which is calculated by dividing the number of Earned Runs[i] conceded by the number of innings pitched. However, this metric has been shown to be inferior to contemporary, context-based metrics. One measure of the validity of a statistic that evaluates performance is the extent to which it is predictive of future performance. However, research has shown that ERA (Swartz, 2012), as a measure of pitching ability, is not as predictive of the pitcher's future performance as those metrics which account for context. A number of competing statistics have been developed which account for particular elements of the pitcher's context, such as the quality of the fielders, the random distribution of errors, and the performance of the opposition batters whom the pitcher faced on a given day. These alternative metrics include measures such as xERA (expected ERA), FIP (Fielder Independent Pitching) and xFIP (expected Fielder Independent Pitching). The statistical measure that is most commonly used in contemporary Sabermetrics is SIERA (Skill-Interactive ERA), which measures pitching ability by taking into account only those metrics which are solely under a pitcher's control.

In measuring television audience engagement through social media, it is vital to control for the systemic boost in social media activity caused by a broadcast's time slot, network, and other factors. To do so, we can draw on a range of sporting metrics that account for such factors by weighting the standard measures accordingly. PERA, or Peripheral ERA (Baseball Prospectus Team of Experts, 2004), is one example of this: it recognizes the inherent "park factors" of each stadium where baseball is played. Essentially, this is calculated by benchmarking each playing statistic for the home and away teams in a given stadium against their overall performance away from that stadium: for example, a stadium with a Home Runs park factor of 112 sees 12% more home runs than the average stadium. Each pitcher's PERA can then be calculated by adjusting the counts of hits, walks, strikeouts and home runs that underlie the standard ERA measure by the park factors of the stadium where the game was played, thus eliminating any such location-specific contextual factors.

Translating this methodology to the measurement of social media activities relating to broadcast content, we have developed a similarly context-independent metric to quantify Twitter-based audience engagement, the Weighted Tweet Index (Woodford & Prowd, 2014). Using this approach, we have been able to identify a number of the contextual factors that influence social media activity levels, including the multiplier effects resulting from the specific television network, the genre, the time of day and year, and the location of a specific episode within the seasonal cycle of a show. The Weighted Tweet Index builds on large longitudinal datasets for a wide range of US television series during the 2012-13 broadcast seasons, including Twitter activity metrics published by Nielsen SocialGuide and data collected directly from the Twitter API. Drawing on data for 9,082 individual episodes over 21 months (April 2012 - January 2014), we calculated a range of contextual broadcast factors analogous to the park factors described for PERA, allowing us to understand the influence of these factors on the volume of social media audience engagement. These factors are normalized to an index value of 1: thus, a factor of 1.28 represents overall engagement 28% above average.

Unsurprisingly, the largest influence on social media engagement observed in our dataset was the broadcast channel itself: we identified a significant difference in the baseline social media activity levels for shows aired on major networks (e.g. CBS), and those shown on cable channels such as MTV. Our data contained shows on 161 US television channels, with major networks such as ABC (1.15), CBS (1.09) and NBC (0.80) differing substantially from cable channels such as BBC America (0.09), Nickelodeon (0.12) and VH1 (0.47). A second key factor that influences engagement with shows on social media is the time at which an episode airs. This affects the size of television audiences more generally: networks have defined seasons for new shows, pause shows over Christmas and New Year when audiences traditionally fall, and rarely air prime shows on Fridays. Quantifying the differences between these times is key to both evaluating historic engagement values and predicting future activities; in our analysis (Table 1), the factors for monthly engagement varied from 0.587 (April) to 1.286 (January), and daily variation ranged between 0.24 (Friday) and 1.51 (Tuesday).

Table 1: Selected underlying broadcast factors affecting social media audience engagement in the US

Month	*Index Factor*
January	1.29
February	1.1
March	1.2
April	0.59
May	0.62
June	0.87
July	0.86
August	1.03
September	0.97
October	1.27
November	1.1
December	1.26

Day	*Index Factor*
Monday	1.45
Tuesday	1.51
Wednesday	1.18
Thursday	1.16
Friday	0.24
Saturday	0.58
Sunday	0.88

Network	*Index Factor*
ABC	1.15
CBS	1.09
ESPN	1.05
FOX	0.96
NBC	0.8
TNT	0.69
MTV	0.56
VH1	0.48
ABCF	0.36
MTV2	0.3

These long-term factors are exceptionally valuable for any attempts to move beyond a simplistic ranking of shows based on their raw social media activity metrics: for the first time, they enable a benchmarking of the social media-based audience engagement with television content that is able to compare prime time and daytime broadcasts, mainstream and cable content, drama and reality TV genres without merely coming to the obvious conclusion that mainstream content generates more tweets, likes, and comments. The Weighted Tweet Index provides a valuable starting-point for advancing beyond the basic metrics generated by commercial analysts such as Nielsen SocialGuide and SecondSync, and constitutes a key tool for the evaluation of shows on a like-for-like basis, and for predictions of how a successful cable show might fare if aired on a mainstream network. Its weighted metrics allow networks and producers to benchmark their shows against others, not just on raw numbers, but by controlling for the other factors which influence audience engagement. However, it is important to note the limitations of this approach. Key among these is that such weightings can never account for the content of a specific episode. For example, in the 2013 season of *Big Brother* (US) we saw a large spike in social media activity that was attributable to a controversy over racism; such acute events are impossible to account for through purely quantitative approaches. Necessarily, the existing weightings can also be further refined, just as the sporting analytics frameworks we have drawn from were developed over a number of years.

A particular focus of related sporting analytics has been the prediction of future performance, on both the team and the player level, for a variety of purposes. Team executives need to make decisions on roster composition, contract values, and other issues; sporting media and fan sites are tracking the performance of teams and seek to contribute insightful commentary; participants in fantasy sports and gambling markets may have significant financial investment in players' performances – these all subscribe to data sites that offer performance predictions based on cutting-edge data analytics approaches. One example of this is Baseball Prospectus's PECOTA (Player Empirical Comparison and Optimization Test Algorithm), which uses advanced Sabermetric statistics to predict players' performances several seasons into the future. These player-level statistics can then be used with the Pythagorean expectation formula, developed by Bill James (1980), to estimate the games a team *should* have won, in order to calculate expected wins and losses for teams over the course of a season.

By determining the contextual broadcast factors that influence social media engagement and applying them to the long-term social media engagement averages for a show once the scheduling of upcoming episodes is known, it is similarly possible to generate predictive measures of the expected social media volume for these episodes. Predictive measures can serve a number of purposes: for the viewer, they enable the selection of shows that are likely to have an active social media audience to engage with; for broadcasters, television producers, and social media strategists, they provide a benchmark to measure whether a show has been as successful on social media as it should have been; and for advertisers, they offer a tool for more targeted promotions, both through traditional commercials and directly through social media-based advertising that reaches a specific social media demographic. Although current social media audience measurement systems remain imperfect and are as yet unable to meet all of the demands of all of the various stakeholders and interested parties – producers, broadcasters, advertisers, advertising sales agents, media buyers and planners, audience research agencies, academics and audiences themselves – they can nonetheless already illuminate new forms of audience behavior and provide insights into particular audiences' levels of engagement with screen content.

Our new approach draws on developments in sports metrics to develop a method for comparing both the performance of particular television content and measuring audience engagement through computational analysis of social media data. Our findings to date indicate that, for the moment at least, social media derived television metrics are no cure-all for the current shortcomings of traditional television audience metrics. Ratings systems for commercial television will continue to be used for as long as the various stakeholders are able to extract value from them. New measurement systems such as Telemetrics that are based on social media analysis are unlikely to replace the ratings; rather, such systems will co-exist with and complement each other as the media industries' long quest to understand their audiences continues.

References

Anderson, C., & Sally, D. (2013). *The numbers game: Why everything you know about soccer is wrong*. London, UK: Penguin Press.

Andrejevic, M. (2007). *iSpy: Surveillance and power in the interactive era*. Lawrence, KS: University Press of Kansas.

Andrejevic, M. (2013). *Infoglut: How too much information is changing the way we think and know*. New York, NY: Routledge.

Balnaves, M., O'Regan, T., & Goldsmith, B. (2011). *Rating the audience: The business of media.* London, UK: Bloomsbury Academic.

Baseball Prospectus Team of Experts. (2004). *Baseball prospectus*. New York, NY: Workman Publishing Company.

Berry, D. (2011). The computational turn: Thinking about the digital humanities. *Culture Machine, 12*. Retrieved February 25, 2015 from http://www.culturemachine.net/index.php/cm/article/view/440/470

Bourdon, J., & Méadel, C. (2014). Introduction. In J. Bourdon & C. Méadel (Eds.), *Television audiences across the world: Deconstructing the ratings machine* (1-30). Basingstoke, UK: Palgrave Macmillan.

Bruns, A., Burgess, J., & Highfield, T. (2014). A ‘big data’ approach to mapping the Australian Twittersphere. In K. Bode & P. Arthur (Eds.), *Advancing digital humanities*. Basingstoke: Palgrave Macmillan.

Coleman, S. (2003). A tale of two houses: The House of Commons, the *Big Brother* house and the people at home. *Parliamentary Affairs*, *56*(4), 733-758.

Council for Research Excellence. (2010). *The state of set-top box viewing data as of December 2009*. Retrieved February 24, 2015 from http://www.researchexcellence.com/files/pdf/2015-02/id90_set_top_box_study_ 2_24_10.pdf

Dekker, K. (2014). 2014 – The top tweeted shows so far. *SecondSync Tumblr*. Retrieved January 8, 2015, from http://secondsync.tumblr.com/post/80172270097/2014-the-top-tweeted-shows-so-far

Fiske, J. (1992). Audiencing: A cultural studies approach to watching television. *Poetics*, *21*(4), 345-359.

James, B. (1980). *The Bill James abstract*. Self-published.

James, B. (1982). *The Bill James baseball abstract*. New York, NY: Ballantine Books.

Liu, B. (2012). *Sentiment analysis and opinion mining*. San Rafael, CA: Morgan & Claypool.

Moskowitz, T., & Wertheim, L. J. (2012). *Scorecasting: The hidden influences behind how sports are played and games are won.* New York, NY: Three Rivers Press.

Napoli, P. (2003). *Audience economics: Media institutions and the audience marketplace*. New York, NY: Columbia University Press.

Nielsen Social. (2014). Nielsen SocialGuide Intelligence. Retrieved January 8, 2015, from http://www.nielsensocial.com/product/social-guide-intelligence/

Pesce, M. (2006). You biquity. Keynote address from the Webdirections South Conference, Sydney, AU.

Ryall, J. (2014, 16 July). How Netflix is quietly thriving in Australia. *Sydney Morning Herald*. Retrieved January 8, 2015, from http://www.smh.com.au/entertainment/tv-and-radio/how-netflix-is-quietly-thriving-in-australia-20140716-ztirm.html

Savage, P., & Sévigny, A. (2014). Canada’s audience massage: Audience research and TV policy development, 1980-2010. In J. Bourdon & C. Méadel (Eds.), *Television audiences across the world: Deconstructing the ratings machine* (69-87). Basingstoke, UK: Palgrave Macmillan.

SecondSync. (2014). Twitter leaderboard. Retrieved January 8, 2015, from https://secondsync.com/leaderboard.html

Silver, N. (2003-2009). Nate Silver author archives. *Baseball Prospectus*. Retrieved January 8, 2015, from http://www.baseballprospectus.com/author/nate_silver

Silver, N. (2012). *The signal and the noise: Why so many predictions fail – but some don't*. London, UK: Penguin Press.

Silvey, R. (1951). Methods of viewer research employed by the British Broadcasting Corporation. *Public Opinion Quarterly*, *15*(1), 89-94.

Swartz, M. (2012). Are pitching projections better than ERA estimators? *Fangraphs.* Retrieved January 8, 2015, from http://www.fangraphs.com/blogs/are-pitching-projections-better-than-era-estimators/

Tango, T. M., Lichtman, M. G., & Dolphin, A. E. (2007). *The book: Playing the percentages in baseball*. Lincoln, NE: Potomac Books.

Thelwall, M. (2014). Sentiment analysis and time series with Twitter. In K. Weller, A. Bruns, J. Burgess, M. Mahrt, & C. Puschmann (Eds.), *Twitter and Society* (83-96). New York, NY: Peter Lang.

Woodford, D., & Prowd, K. (2014). Everyone's watching it: The role of hype in television engagement through social media. Paper presented at the Social Media and the Transformation of Public Space conference, Amsterdam, NL.

World Federation of Advertisers (WFA). (2008, June). Blueprint for consumer-centric holistic measurement. Brussels, BE: World Federation of Advertisers. Retrieved January 8, 2015, from http://www.wfanet.org/media/pdf/Blueprint_English_June_2008.pdf

Wurtzel, A. (2009). Now. Or never. An urgent call to action for consensus on new media metrics. *Journal of Advertising Research*, *49*(3), 263-265.

---

i Earned runs differ from total runs in that they exclude any runs given up after a fielding error prevented the third out of an inning.