Ciulla et al. *EPJ Data Science* 2012, **1**:8
http://www.epjdatascience.com/content/1/1/8

EPJ Data Science
a SpringerOpen Journal

EPJ.org

REGULAR ARTICLE Open Access

Beating the news using social media: the case study of American Idol

Fabio Ciulla[1*], Delia Mocanu[1], Andrea Baronchelli[1], Bruno Gonçalves[1], Nicola Perra[1] and Alessandro Vespignani[1,2,3]

*Correspondence: f.ciulla@neu.edu
[1]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA 02115, USA
Full list of author information is available at the end of the article

Abstract

We present a contribution to the debate on the predictability of social events using big data analytics. We focus on the elimination of contestants in the American Idol TV shows as an example of a well defined electoral phenomenon that each week draws millions of votes in the USA. This event can be considered as basic test in a simplified environment to assess the predictive power of Twitter signals. We provide evidence that Twitter activity during the time span defined by the TV show airing and the voting period following it correlates with the contestants ranking and allows the anticipation of the voting outcome. Twitter data from the show and the voting period of the season finale have been analyzed to attempt the winner prediction ahead of the airing of the official result. We also show that the fraction of tweets that contain geolocation information allows us to map the fanbase of each contestant, both within the US and abroad, showing that strong regional polarizations occur. The geolocalized data are crucial for the correct prediction of the final outcome of the show, pointing out the importance of considering information beyond the aggregated Twitter signal. Although American Idol voting is just a minimal and simplified version of complex societal phenomena such as political elections, this work shows that the volume of information available in online systems permits the real time gathering of quantitative indicators that may be able to anticipate the future unfolding of opinion formation events.

1 Introduction

The recent global surge in the use of technologies such as social media, smart phones and GPS devices has changed the way in which we live our lives in a fundamental way. Our use of such technologies is also having a much less visible, but not less significant, consequence: the collection on a massive scale of extremely detailed data on social behavior is providing a unique and unprecedented opportunity to observe and study social phenomena in a completely unobtrusive way. The public availability of such data, although limited, has already ignited a flurry of research into the development of indicators that can act as distributed proxies for what is occurring around the world in real time. In particular, search engine queries or posts on microblogging systems such as Twitter have been used to forecast epidemics spreading [1], stock market behavior [2] and election outcomes [3–6] with varying degrees of success. However, as many authors have pointed out, there are several challenges one must face when dealing with data of this nature: intrinsic biases, uneven sampling across location of interest *etc.* [7–10].

In this paper we intend to assess the usefulness of open source data for the forecasting of societal events. We analyze in depth the microblogging activity surrounding the voting behavior on the contestants in American Idol, one of the most viewed American TV shows as a simple case study of this type of processes. In this program, the audience is asked to choose which contestant goes forward in the competition by voting for their favorites. The well delineated time frame (a period of just a few hours) and frequency (every week) over an extended period (an entire TV season) provides a close to ideal test ground for the study of electoral outcomes as many of the assumptions implicitly used in the analysis of social phenomena are more easily arguable, if not trivially true, in the case of the American Idol competition. In particular, we assume that:

- The demographics of users tweeting about American Idol are representative of the voting pool.
- The self-selection bias, according to which the people discussing about politics on Twitter are likely to be activists scarcely representative of the average voter, seems to become almost a positive discrimination factor in the case of a TV show where the voters are by definition self-selected.
- Voting fans are the most motivated subset of the audience (the population we are trying to probe) that are willing to make an extra effort for no personal reward, and, crucially, they are allowed to vote multiple times.
- Users are not malicious, and engage only in conversations they have a particular interest in.
- The influence incumbency, which strongly affects the outcome of political elections, is not a factor determining the outcome of American Idol.

For the above reasons we can consider TV show competitions as a case study for the use of open source indicators to achieve predictive power, or simply beating the news, about social phenomena. It is thus not surprising that other attempts to use open source indicators in this context have been proposed in the past. Here however we benefit from the constant growth of Twitter that makes it easier to collect significative statistical sample of the population. Furthermore, TV shows are now leveraging on Twitter and other social platform which are becoming in all respects a mainstream part of the show. This amplifies the importance of the indicators one can possibly extract from these media in monitoring the competition. Finally, the increasing use of smartphone and mobile devices produces geolocalized information about Twitter activity that we can mine. We show that including the geographical information is a key ingredient in achieving predictive power. This final consideration clearly points out that the prospective use of Twitter data for predicting social events in other settings shall consider analysis that go well beyond the aggregated number of tweets.

2 Rules and voting system

The first episode of the 11th season of American Idol was aired on January 18, 2012 with a total of 42 contestants. After an initial series of eliminations made by the judges, a final set of 13 participants was selected. All further eliminations were decided by the audience through a simple voting system. During this final phase of the competition, two episodes are aired each week: On Wednesday the participants perform on stage and the public is invited to vote for two hours after the show ends. Voting can take one of three forms: toll-free phone calls, texting and online voting. The rules of the competition only allow for

votes casted by the residents of the US, Puerto Rico and US Virgin Islands. There is no limit to the number of messages or calls each person can make, while the online votes are limited to 50 per computer as identified by its unique IP address. Every week, hundreds of millions of votes are counted and the contestant that gathers the least number is eliminated. The show airs at 8.00 PM local time on each coast. As a result of the time zone difference of three hours between the East and West coast, the total voting window between the first and last possible vote is 10.00 PM-3.00 AM EST. During the season's final performance episode the voting window is extended to four hours after the show airs, resulting in an extended voting window between 9.00 PM-4.00 AM EST.

3 Data

Our fundamental assumption is that the attention received by each contestant in Twitter is a proxy of the general preference of the audience. To validate this assumption, we collected at total of 677,804 tweets containing any of the 51 #tags, @usernames and strings related to the show listed in Table 1.

The main dataset was obtained by extracting matching tweets from the raw Twitter feed used by Truthy [11] for the entire duration of the current season of American Idol, corresponding to 224,189. The feed is a sample of about 10% of the entire number of tweets that provides a, statistically significant, real time view of the topics discussed within the Twitter ecosystem. This allowed us to make a post-event analysis of the last 9 eliminations. This dataset was further complemented by the results of automatically querying the Twitter search API every 10 minutes for tweets containing one or more of the keywords we identified as related to American Idol. The search API data cover the period since May 16, giving us a more detailed view of the last elimination before the season's finale and resulted in 453,615 tweets.

Tweets in our dataset often contain georeferenced location information. They could be GPS coordinates, assigned automatically by smart phones, or self reported one. We consider both. Users with smart phones can use their accounts also in other devices. Geographical coordinates could be then present just in a fraction of their tweets. In order to increase the number of geolocalized tweets we analyzed the whole set of Twitter data col-

Table 1 Text markers in support for each of the contestants

Contestant	Usernames	#Tags
Phillip	@PPhillipsAI11@PhilPhilatic	#phillipphillips #philatic #philatics
Jessica	@JSanchezAI11 @TeamJSanchez @TeamJaaySanchez	#blujays #teambluJay #TeamJessicaSanchez #BBchez #JessicaSanchez
Joshua	@JLedetAI11 @JoshuaLedetNet @TeamJoshua @JoshuaLedetNet @WeLuvLedet @TeamLedet	#jjewels #JoshuaLedet #teamjosh #ledet #teamjoshei #JoshualsLegend
Hollie	@CavanaghAI11	#holliepop #teamhollie #holliepopsfamily #holliecavanagh
Skylar	@SLaineAI11	#skoutlaw #skoutlaws #skollie #skylarlaine
Elise	@ETestoneAI11 @theElitesWorld	#theElites
Colton	@CDixonAI11 @ColtonDSupport	#teamdixon
Deandre	@BrackensickAI11 @deandrefanpage	#DreDreamers
Heejun	@HHanAI11 @MadiHeartHeejun @HeeHangels	#Hangel #Heejunie
Erika	@EVanPeltAI11	#pelters

List of tags and usernames used in the study to detect tweets related to each contestant. The different number of keywords of each contestant does not bias the total number of tweets associated with the contestant. This variety merely reflects the spontaneous emergence within the Twitter community of different ways of referring to a certain contestant.

lected since the beginning of 2012. We mine the data finding the geographical coordinates of all users with at least one geolocalized tweet in the dataset. Using these information we were able to assign geographical references also for tweets that did not contain them, as long they were sent by a user we previously geolocalized.

4 Mapping the fanbase

Studying the set of geolocalized tweets allows us to analyze the spatial patterns in voting behavior. Figure 1 shows a strong geographical polarization in the US towards different candidates. In the weeks preceding the Top 3 show (panels (B) and (C)), for example, Phillip Phillips gathers most of the attention in the Midwest and South, while Jessica Sanchez appears to be popular particularly on the West Coast as well as in the large metropolitan areas across all of the country, and Joshua Ledet is strong in Louisiana. The Top 3 week analysis (panel (A)) shows a disturbance from the previous geographical distribution, perhaps due to the performance of the candidates. As expected, the audience reacts to the events occurring on Wednesday night. On the other hand, and perhaps not surprisingly, the attention basins of each of the three participants always include their origin city (Phillips was born and raised in Georgia, Sanchez is from Chula Vista, California, and Ledet from the Lake Charles metropolitan area in Louisiana) [12].

The geolocalized data also allows for a unique view of the attention devoted to American Idol in the rest of the world. Although one might naively expect interest to be limited to

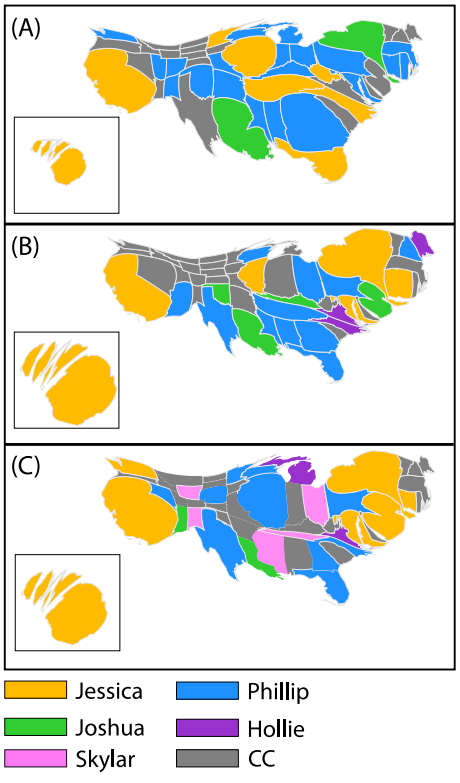

Figure 1 Geographical polarization of the tweets for the Top 3 (A), Top 4 (B) and Top 5 (C) episodes. The area of each state is proportional to the number of geolocalized tweets generated there, while the color represents the contestant with the majority of the vote. The grey represent states we could not assign to a single contestant within the statistical errors (CC).

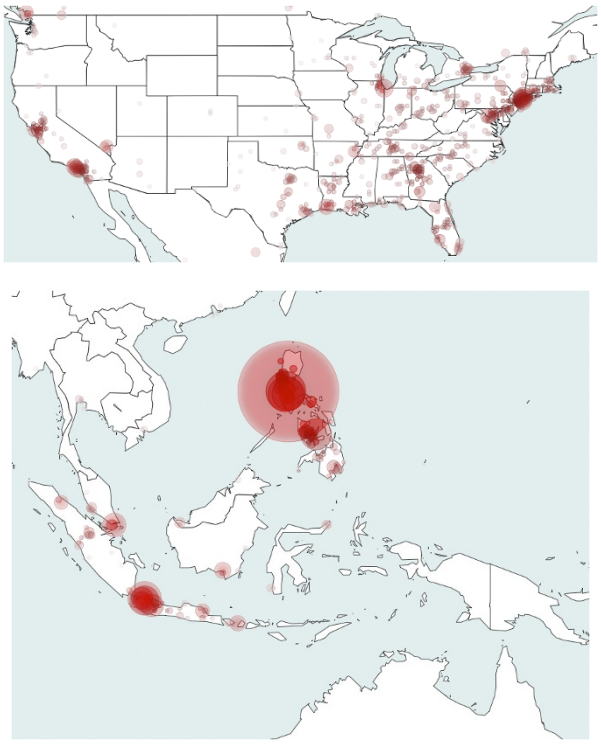

Figure 2 Local and global attention towards American Idol. Top: US data show that the highest Twitter activity is concentrated in the large metropolitan areas, as expected. Bottom: The Philippines are distinctly more active than any other foreign country. It is worth noting also that a remarkable signal is produced in Indonesia, too, which is very active country with respect to Twitter activity in general.

Table 2 Popularity basins

Contestant	USA	World	Philippines
Jessica	45 ± 4	64.2 ± 2.2	92.8 ± 1.9
Joshua	15 ± 3	9.8 ± 1.3	1.4 ± 0.9
Phillip	40 ± 4	26 ± 2.0	5.8 ± 1.7

Data concerns the entire American Idol season up to the morning of May 17 (before the two finalists were announced), and refers to the percentage (%) of popularity within US, the whole World and the Philippines. The geo-localized database for the three candidates contains 3,251 data points. Errors represent the normal confidence interval with a confidence level of 99%.

the US, Figure 2 shows that the show is also popular in several foreign countries and particularly in the Philippines. This can be understood by noting that one of the contestants is of Filipino origin. Jessica Sanchez's mother is originally Filipino, having been born in the Bataan province [13]. Participation in American Idol has made Sanchez so popular in her mother's native country that on May 16 the Philippine President Benigno Aquino III congratulated the singer for her performance and stated, 'Hopefully she really reaches the top.' [14]. Table 2 quantifies this intuition. Jessica Sanchez related Tweets are 45% of the total if only US is considered, while it rises to 64% if the whole World is considered. Officially, Sanchez's popularity abroad should not have any impact on voting, since, as mentioned above, only the US based audience is allowed to take part into the election procedure. However, it is interesting to note that the Filipino-restricted Twitter activity concerning Jessica is strongly peaked in the two voting sessions of American Idol for the East and West timezones, and that numerous websites explicitly address the issue of 'voting tun-

nels': 'How to Vote for Jessica Sanchez from the Philippines and Other Non-US Countries' [15, 16]. Although we have no proof of any irregular voting activity, tweets analysis clearly points out to a possible anomaly that may be a concern.

5 Post-event analysis

Our fundamental, and seemingly naive, assumption is that the number of votes each contestant receives is proportional to the number of tweets that mention her. In other words, the larger the number of tweets referred to a contestant - the Twitter volume - the larger the number of votes she will get. This gives a natural measure to rank each contestant. It is important to note that this is a very simple measure, and that we intentionally choose not to take into account many of the factors that in principle might affect the results, such as the presence of negative or neutral tweets, or attempts to directly affect the counts by spamming the system with automatically generated tweets. In fact, one of the goals of this paper is to test whether or not a minimal set of measures applied to Twitter data can be good indicators of the actual voting outcome. Past attempts have met with ambivalent results and we are interested in testing the limits of this naive approach by building an unsophisticated prediction system assembled in less than one week.

While our dataset spans the entire duration of the current season, we focus only on the top-ten phase of the show, when just 10 contestants remained and test the predictive power of the Twitter proxy against the last 9 eliminations. For 7 of those, the 'bottom-three' contestants, the least three voted contestants (2 in the elimination of May 3rd) were revealed during the iconic part of the show: elimination day. We consider not just the success in predicting the contestant that will be eliminated but also the three that received the least votes.

In order to minimize the noise that might be introduced by discussions after the voting time and especially after the elimination, we considered the number of tweets generated on a specific time window: 8.00 PM-3.00 AM EST each Wednesday. The show airs at 8.00 PM EST. The votes can be submitted until midnight in the West coast which translates to 3.00 AM in the east. In Figure 3 we show the number of tweets related to each of the top three contestants for every minute of the voting window on May 16. Interestingly the number of tweets associated to the eliminated contestant (Joshua) is practically always the

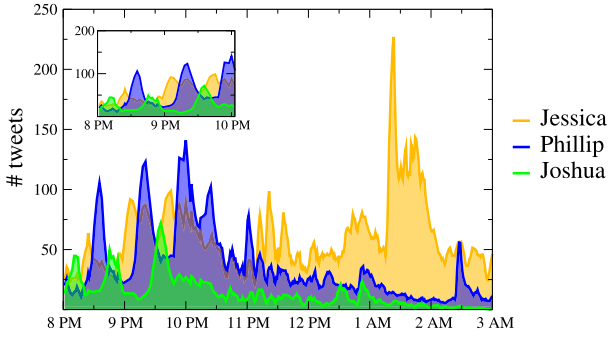

Figure 3 Live popularity of the participants to the Top 3 night. The number of tweets related to each one of the Top 3 contestants is plotted as a function of time, from the start of the show (8 PM EST) and the closing of the voting in the West Coast (3 AM EST), on Wednesday May 16, 2012. The data is plotted with the granularity of the minute. The inset magnifies the two hours of the first airing of the show on the East Coast.

Table 3 Eliminations' recap

Day	Eliminated contestant	Data indicators	Bottom 3
May 17	Joshua	✓	N/A
May 10	Hollie	✓	N/A
May 3	Skylar	✓	✓
April 26	Elise	CC	✓
April 19	Colton	×	✓
April 12	Jessica (saved)	×	(2/3)
April 05	DeAndre	CC	✓
March 29	Heejun	CC	(2/3)
March 22	Erika	CC	(1/3)

We consider the last nine eliminations. In the table we report the date of the elimination, the contestant eliminated, whether the data indicators correctly single out the elimination (✓), it is wrong (×) or provide a too close to call (CC). In the last column we compare the data indications for the bottom three (two) contestants announced during the first seven eliminations. We report when within error bars the signal identifies the bottom three contestants (✓), two out of three (2/3) or one out of three (1/3) contestants.

smallest. The inset provided a detailed view of the live show time period. At this resolution the sequence of peaks of each contestant correlates with time and sequence of their performances that night.

For each of the last 9 weeks, we have integrated the number of tweets related to each user in the *show* + *voting* time window. We then ranked the contestants in decreasing order. The last 3 count as the bottom three and the last contestant is the most likely to be eliminated. We confront our prediction with the real outcomes. In order to account for errors induced by sampling of the real number of tweets we evaluated the 99% confidence intervals assuming a homogeneous and fair sampling and report the results in Table 3. Twitter data serves as a correct indicator for the last three eliminations and identifies correctly most of the bottom three/two contestants.

Twitter signal indications were wrong two times, and we have other four cases in which the confidence intervals in the ranking could not allow to make a prediction ('too close to call'). In Figure 4 it is possible to notice that, as expected, when the number of contestants reduces and the fan base solidifies, the differences between ranks become much clearer and separated.

Our methodology considers the number of tweets as the main indicator of the popularity of each contestant. By construction, n tweets posted by n different users or by the same user counts equally. This might introduce biases in the results, since the simple tweet counting could in principle be affected by bots or spam campaigns. Therefore, in order to test our measure, we analyzed the data ranking each contestant by the number of unique users. The results are unchanged within the statistical errors.

6 And the winner is... Phillip

On May 23 the last episode of the program was aired. Around 10.00 PM EST, the winner of the 11th season was announced: Phillip.

Three days before, on May 20, we submitted to arXiv the first draft of this paper containing the methodology and the post-event analysis of the previous nine eliminations [17]. At that moment, two more episodes of the show were to be aired, the final exhibitions on Tuesday, May 22, and the season finale on the next day.

We made our predictions based on the data collected on May 22, between the beginning of the show on the East at 8.00 PM EST and the end of the voting period in the west, at

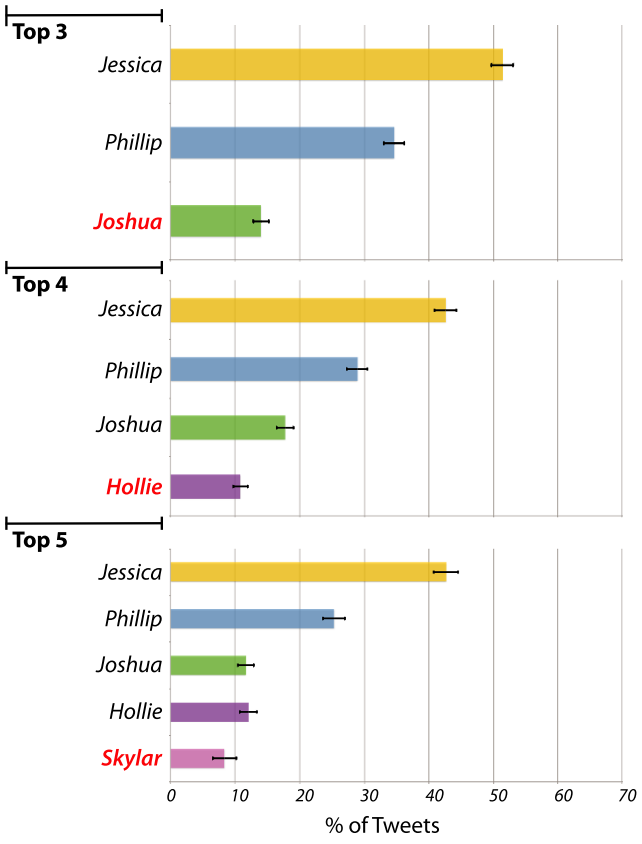

Figure 4 For the last three eliminations we plot the ranking of each contestant measured as the percentage of tweets in the time window 8.00 PM-3.00 AM EST of the last three Wednesdays. We plot the 99% confidence intervals. In red we mark the contestant that was eliminated the next day. It is clear that even considering the errors, the ranking done considering the volume of tweets related to a specific contestant, is sufficient to identify the least preferred.

4.00 AM EST. They were discussed in an updated version of the paper submitted to arXiv hours before the official announcement [18].

In Figure 5 we show our results. They call for a careful inspection. Considering the whole of our dataset, as we have done in the previous analysis, Jessica was clearly the most popular finalist in Twitter during our time window. Henceforth, the criterion we adopted for the previous shows would have implied that Jessica should have been the winner of the show.

However, we realized immediately that the oversimplified analysis could not take into account a crucial factor. As pointed out before, Jessica was the only contestant for which we detected a strong Twitter signal coming from outside of the US (and in particular from the Philippines), with a ramping up trend after the show aired on April the 19th. Given that the voting was restricted to the US only, we had a closer look at the data, and considered the subset of tweets that come with geographical metadata. Although the geolocalized data are a much smaller subset of the total signal (about 10%), this dataset allowed us to provide the contestants' standing restricted to the USA Twitter population. In the US, Phillip had the largest fanbase of the two contestants (see also the cartogram of Figure 6). Then, discarding the possibility of votes coming from abroad, the scenario we found was the opposite: Phillip should have been the winner of the 11th edition of American Idol. This

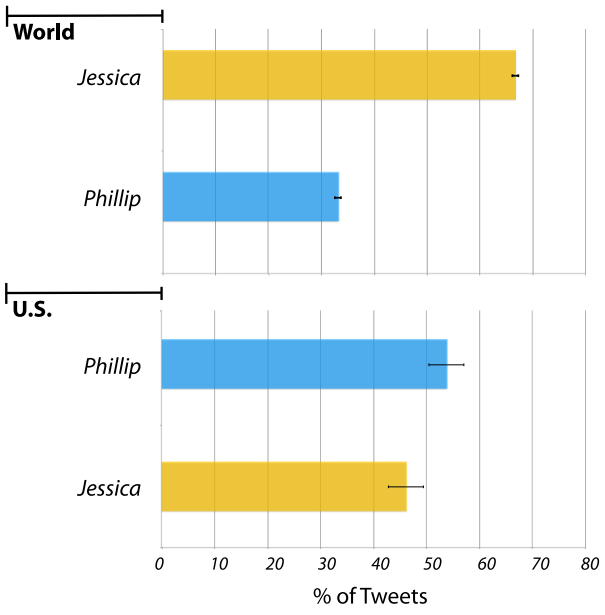

Figure 5 Finale ranking. The ranking of the two contestants of season finale, measured as the percentage of tweets in the time window 8.00 PM-4.00 AM EST, is plotted. The top histogram takes in to account the whole dataset (World), while the bottom one only considers the set of tweets geolocalized in the United States (US). We report the 99% confidence intervals.

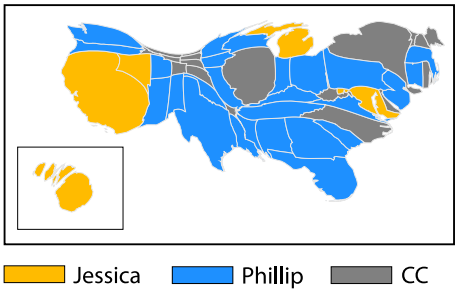

Figure 6 Geographical polarization of the tweets for the Top 2 contestants. The area of each state is proportional to the number of geolocalized tweets generated there, while the color represents the contestant with the majority of the vote. The grey represent states we could not assign to a single contestant within the statistical errors (CC).

turned out to be the case, putting to rest the idea of a large number of users connecting through the Internet from abroad to participate in the voting.

Our analysis shows the importance of geolocated signals. Far from being an additional piece of information, the geographical origin of the tweets turned out to be essential in gaining a clear understanding of the situation we were addressing. This is likely to be a general message, valid for any election-like processes, where the global popularity risks to be a poor indicator, and might induce wrong interpretations/forecasts.

7 Conclusion

We have shown that the open source data available on the web can be used to make educated guesses on the outcome of societal events. Specifically, we have shown that ex-

tremely simple measures quantifying the popularity of the American Idol participants on Twitter strongly correlate with their performances in terms of votes. A post-event analysis shows that the less voted competitors can be identified with reasonable accuracy (Table 3) looking at the Twitter data collected during the airing of the show and in the immediately following hours.

It is worth noting that our analysis aims to be extremely simple in order to establish a valid baseline on what it is possible to deduce by social media. As such, we purposefully do not consider a number of refinements and techniques that could improve the accuracy of our predictions. Distortions due to overactive users can be controlled by evaluating the number of unique users tweeting on each contestant. The text of the tweets could be scrutinized by using sentiment analysis techniques to select and compare only specific positive or negative tweets as a proxy for success/failure. Corrections to the demographic representations of Twitter users could be considered. All these techniques have been or are being developed in the analysis of a wealth of social phenomena and could be tested in a very clear and simple setting such as those of American Idol or similar shows.

Furthermore, we have illustrated that open source data can provide a deeper insight into the composition of the audience, with the eventual possibility of pointing out possible sources of anomalous behaviors. A geographical projection of the data reveals a non-uniform distribution of the basins of fans, and likely of voters, for the different participants. Interestingly, the same inspection highlights also that a strong activity concerning some of the candidates may come from non-US countries, whose audience are officially forbidden to vote.

Finally, our work casts a word of warning on the possible feedback between competitive TV shows and social media. Indeed, while the former rely more and more on the online voting of the audience, and the votes are kept secret and revealed only at the end of the show, all of the data necessary to monitor and even forecast the outcome of these shows is publicly available on the web. Given the large economic interests that lay behind such programs, such as the revenues of betting agencies and the major contracts of the show participants, it is obvious that this situation can lead to a number of undesirable outcomes. For example, the audience could be induced to alter their behavior in function of the situation they observe, and the job of betting agencies could be dramatically simplified. On a more general basis, our results highlight that the aggregate preferences and behaviors of large numbers of people can nowadays be observed in real time, or even forecasted, through open source data freely available in the web. The task of keeping them private, even for a short time, has therefore become extremely hard (if not impossible), and this trend is likely to become more and more evident in the future years.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
All authors designed research, FC and DM performed data analysis. All authors analyzed the results, wrote, reviewed and approved the manuscript.

Author details
[1]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA 02115, USA. [2]Institute for Scientific Interchange Foundation, Turin, 10133, Italy. [3]Institute for Quantitative Social Sciences, Harvard University, Cambridge, MA 02138, USA.

Acknowledgement
The authors would like to thank Duygu Balcan for generating the cartograms used in this manuscript.

Received: 11 June 2012 Accepted: 13 July 2012 Published: 31 July 2012

References

1. Culotta A (2010) Towards detecting influenza epidemics by analyzing Twitter messages. In: Proceedings of the first workshop on social media analytics. ACM, New York, pp 115-122
2. Bollen J, Mao H, Zeng X (2011) Twitter mood predicts the stock market. J. Comput. Sci. 2:1-8
3. Tumasjan A, Sprenger T, Sandner P, Welpe I (2010) Predicting elections with Twitter: what 140 characters reveal about political sentiment. In: Proceedings of the fourth international AAAI conference on weblogs and social media, pp 178-185
4. Livne A, Simmons M, Adar E, Adamic L (2011) The party is over here: structure and content in the 2010 election. In: Fifth international AAAI conference on weblogs and social media
5. Skoric M, Poor N, Achananuparp P, Lim E, Jiang J (2012) Tweets and votes: a study of the 2011 Singapore general election. In: 2012 45th Hawaii international conference on system sciences. IEEE Press, New York, pp 2583-2591
6. Sang E, Bos J (2012) Predicting the 2011 Dutch senate election results with Twitter. In: EACL 2012, p 53
7. Ratkiewicz J, Conover M, Meiss M, Gonçalves B, Flammini A, Menczer F (2011) Detecting and tracking political abuse in social media. In: Fifth international AAAI conference on weblogs and social media, p 297
8. Metaxas P, Mustafaraj E, Gayo-Avello D (2011) How (not) to predict elections. In: Privacy, security, risk and trust (PASSAT), 2011 IEEE third international conference on and 2011 IEEE third international conference on social computing (SocialCom). IEEE Press, New York, pp 165-171
9. Mislove A, Lehmann S, Ahn Y, Onnela J, Rosenquist J (2011) Understanding the demographics of Twitter users. In: Fifth international AAAI conference on weblogs and social media
10. Gayo-Avello D (2012) 'I wanted to predict elections with Twitter and all I got was this Lousy Paper' - A balanced survey on election prediction using Twitter data. arXiv:1204.6441
11. Ratkiewicz J, Conover M, Meiss M, Gonçalves B, Patil S, Flammini A, Menczer F (2011) Truthy: mapping the spread of astroturf in microblog streams. In: Twentieth international World Wide Web conference, vol 249
12. Season 11 contestants. http://www.americanidol.com/contestants/season_11
13. Jessica Sanchez makes it to Idol's top 5; rocker Elise Testone out. http://newsinfo.inquirer.net/183767/jessica-sanchez-makes-it-to-idols-top-5-rocker-elise-testone-out
14. Philippine president Aquino roots for Jessica Sanchez to win 'American Idol' next week. http://ca.news.yahoo.com/philippine-president-aquino-roots-jessica-sanchez-win-american-065124033.html
15. How to vote for Jessica Sanchez from the Philippines and other non-US countries. http://www.starmometer.com/2012/04/19/how-to-vote-for-jessica-sanchez-from-the-philippines-and-other-non-us-countries/
16. Pinoys worldwide can vote for Jessica Sanchez through Facebook? http://www.gmanetwork.com/news/story/258482/pinoyabroad/pinoyachievers/pinoys-worldwide-can-vote-for-jessica-sanchez-through-facebook
17. Ciulla F, Mocanu D, Baronchelli A, Perra N, Gonçalves B, Vespignani A (2012) Beating the news using social media: the case study of American Idol. arXiv:1205.4467v1
18. Ciulla F, Mocanu D, Baronchelli A, Perra N, Gonçalves B, Vespignani A (2012) Beating the news using social media: the case study of American Idol. arXiv:1205.4467v2

doi:10.1140/epjds8
Cite this article as: Ciulla et al.: **Beating the news using social media: the case study of American Idol.** *EPJ Data Science* 2012 **1**:8.