Future Technologies Conference (FTC) 2017
29-30 November 2017| Vancouver, Canada

A Technology Forecasting Framework Enhanced via Twitter Mining

Anthony Breitzman, PhD
Rowan University
Department of Computer Science
201 Mullica Hill Rd, Glassboro, NJ 08028
Email:breitzman@rowan.edu

Patrick Thomas, PhD
1790 Analytics, LLC
130 Haddon Ave., Haddonfield, NJ 08033
Email: pthomas@1790analytics.com

***Abstract*—The amount of information available on new technologies has risen sharply in recent years. In turn, this has increased interest in automated tools to mine this information for useful insights into technology trends, with a particular focus on locating emerging, breakthrough technologies. This paper first outlines an automated framework for technology forecasting developed for the Department of Defense. It then proposes various enhancements to this framework, focusing in particular on utilizing social media data more effectively. Specific topics covered include technology forecasting via Twitter trusted sources and via identification of authoritative Twitter handles. Beyond improving the framework itself, the techniques described in this paper may also be of general interest to researchers using social media data, particularly for technology forecasting.**

Keywords*—*Twitter mining; text mining; emerging technologies

I. Introduction

There is a long history of research into technology forecasting, and a number of methods are well established in this area. For example, the Delphi Method [1] combines opinions from a panel of experts using a systematic, iterative process, based on the premise that forecasts combining multiple expert opinions are likely to be more accurate than forecasts from a single source. Meanwhile, the Bass Model [2] tracks the numbers of ‘innovators’ and ‘imitators’ in a given emerging technology over time, and forecasts diffusion of the technology based on these numbers. Methods developed more recently include strategic planning frameworks such as technology roadmapping [3] and Forecasting Innovation Pathways [4]. Alongside these research efforts, there have also been specific government-sponsored programs directed towards forecasting technological emergence. Notable among these is the European PromTech project, which endeavors to locate emerging technologies via analysis of scientific literature [5]. Meanwhile, in the US, the IARPA FUSE program identified emerging technologies via analysis of scholarly communications; while the IARPA ACE and ForeST programs used crowdsourcing to locate such technologies. Many of the existing methods for technology forecasting rely, in whole or in part, on inputs from subject matter experts. Such methods are often time-consuming, and restricted by the knowledge base of the available experts. Also, expert reviewers risk drowning in the mass of information available in the electronic age, much of which has little relevance to their task. These issues have led to increasing interest in automated approaches to technology forecasting to complement, or help focus, the work of subject matter experts.

With funding from the U.S Department of Defense (DOD) the authors developed an automated Technology Watch/Horizon Scanning (TW/HS) prototype framework. The framework is designed to identify emerging technologies, and locate information relevant to these technologies, in a timely manner. At the same time, the framework acts to reduce the level of noise, which could result from the promotion of non-emerging, mundane technologies, or from the attachment of irrelevant information to otherwise interesting technologies. The framework can operate in one of two modes: Technology Watch (i.e. directed by inputs from analysts, such as keywords and topics) and Horizon Scanning (i.e. undirected by analyst inputs).

The TW/HS framework is based on a hub-and-spoke model, as shown in Fig. 1. The hub in the first generation of the framework consists of patents identified by the Emerging Clusters Model developed by the authors [6]. This model identifies clusters of patents describing emerging technologies, and tracks how these emerging clusters ‘heat’ and ‘cool’ in close to real time. As a result, interesting new technologies can be identified much more quickly than is possible using traditional scientometric techniques.

Under IARPA (FUSE) and DOD funding, the performance of the Emerging Clusters Model was tested over a 25-year period. This test revealed that, in the five years following their selection, patents in emerging clusters are cited as prior art by subsequent patents significantly more frequently than peer patents from the same year and technology (see Fig. 2, [6]). Since citation rates are widely used as proxies for technological impact [7], this result suggests that the Emerging Clusters Model is able to locate high impact technologies in close to real time.

In the first generation of the TW/HS framework, key elements of the emerging clusters in the hub were extracted, and used as seeds to mine other data sources, notably social media, scientific papers, news feeds and financial information. This helped to locate additional information on technologies of interest, while also reducing the time lags associated with patent data (due to the period between when patents are filed, and when their content becomes public). Organizations typically do not discuss new innovations prior to filing patents to protect them, since doing so may affect the patentability

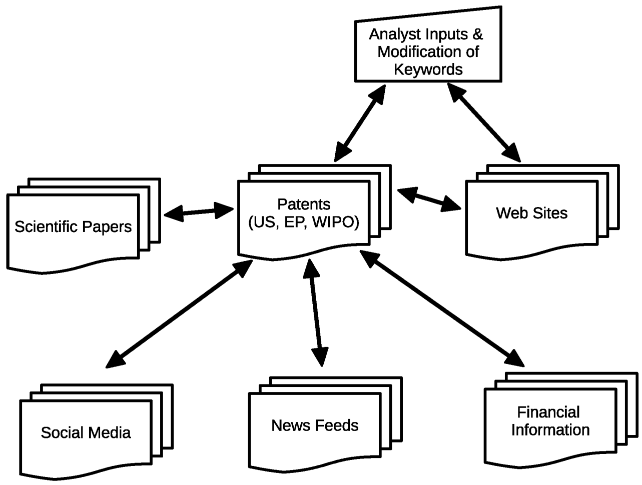

Fig. 1. First generation technology watch/horizon scanning framework.

of these innovations. Once patents have been filed, they may be more forthcoming, and information may start to appear in other forums such as conferences and popular and scientific publications, and may also be highlighted in social media and blogs.

This paper outlines numerous enhancements to the social media spoke of the TW/HS framework. These enhancements differ in terms of their ambition, and the degree to which they alter the operation of the overall framework. At the more conservative end, the enhancements provide more efficient processing and analysis of social media data. More ambitious are enhancements that enable analyst inputs (in the Technology Watch mode of the framework) to be directed by outputs from social media, rather than having to be derived by analysts themselves. Even more ambitiously, some of the enhancements represent stepping-stones towards a framework in which social media data could be used as the hub of the framework, either instead of, or jointly with, patent data.

II. Social Media Enhancements to the TW/HS Framework

Below, we discuss various improvements to the first generation TW/HS framework, with a particular focus on the social media spoke of the framework. The objective of these enhancements is to utilize social media data more effectively, and to make these data more central to the operation of the framework. Fig. 3 shows the proposed updated framework, which can be compared to Fig. 1 showing the first generation framework. Note that, along with the more central role of social media data, a key feature of the updated framework is the concept of curated data (discussed in the enhancements below) replacing the non-curated data used in the first generation of the TW/HS framework.

A. Identifying Emerging Technologies via Twitter Trusted Sources

Social media contains vast quantities of information, most of it irrelevant for technology forecasting. There are a number of reasons for this, not least of which is that sites such as Twitter do not have a specific focus on technology or business. Indeed, Twitter is dominated by celebrities, sportspeople and other public figures, along with people reporting on their day-to-day activities. In order to address the problem of locating relevant information in social media, much recent work has focused on detecting and collating events (e.g. [8]) following a wide variety of both supervised and unsupervised techniques, as summarized by [9]. There have also been numerous efforts to use social media data to identify emerging topics in science and technology [10]. Twitter has been a particular focus of these efforts, since it has become a common communication platform for research scientists to rapidly share scientific information [11] and [12].

The Twitter-based technology forecasting approach described in this paper is based on identifying 'trusted sources' that are dedicated to science and technology. This helps reduce the impact of the overwhelming amount of non-relevant information in Twitter. Also, these trusted sources have an additional advantage, namely, that over 90% of the tweets from them are linked to a web site with a complete story. The tweets thus not only lead to interesting topics, but they serve as a provider of curated articles on these topics. These articles should generally be more authoritative than articles retrieved via a search engine that does not attempt to filter content.

We used the following Twitter handles as trusted sources: @eetimes; @guardianscience; @IEEESpectrum; @newscientist; @ReutersScience; @science; @techreview; @wiredscience; and @wired. Note that we did not include general news sources like the *New York Times*, since these sources may bring in large amounts of non-relevant information. We harvested more than 118,000 tweets from 2012 to March 2017 from these trusted sources. These tweets consist of: 1) every tweet from these sources in 2017 through March; 2) approximately 80% of tweets from these sources in 2016; and 3) about 40-50% of tweets from these sources in 2012-2015. Although not a complete record, this database of tweets is sufficient to provide interesting insights into the possibility of using Twitter data to locate emerging technologies.

Having harvested the tweets, we then carried out a series of pre-processing steps. In the first step, we removed web URLs, emojis, etc., with the resulting output being a line of text. We then replaced contractions, instances of i.e., e.g. and other abbreviations with standard words. We then stemmed, removed stopwords, and built single-word, double-word, and triple-word phrases that do not cross a stopword or punctuation. Take for example the following tweet from IEEE Spectrum: "Brain-Implant Allows Man to Feel Touch on Robotic Hand". This tweet is broken up into the phrases *Brain Implant*, *Man*, *Feel Touch* and *Robotic Hand*, because the stopwords *allows*, *to* and *on* break up each phrase. This particular tweet contains no punctuation, but a period, comma, semicolon or other punctuation mark could also break up a phrase. Note that chunking or chinking could potentially be used to identify noun phrases; however past experience with Twitter mining has led the authors to this simpler method because the non-standard language in tweets often confuses part-of-speech taggers. After these pre-processing steps, we removed duplicates and retweets so that each tweet is only counted once. This brings the initial 118,000+ tweets to 93,675 unique tweets.

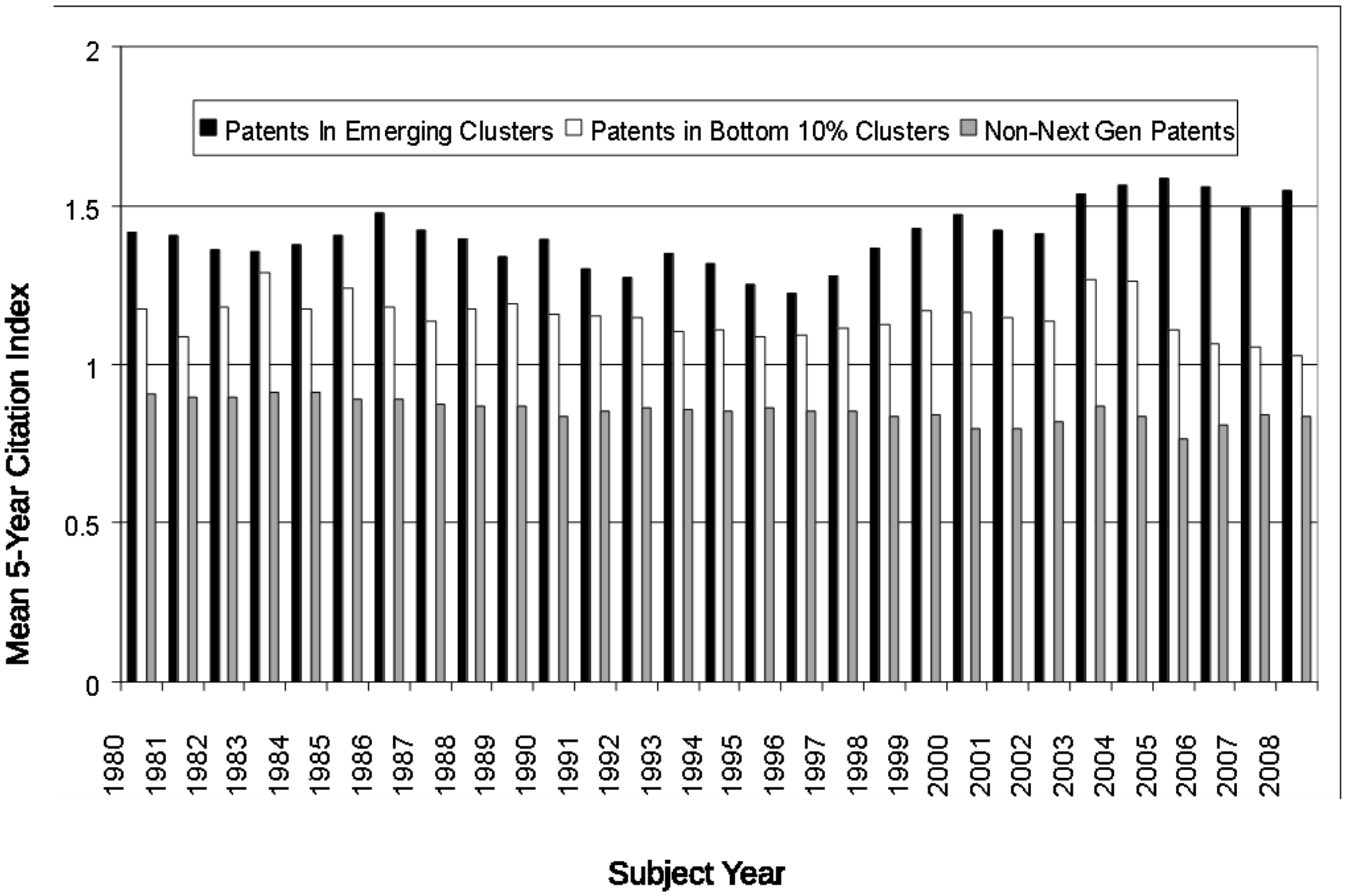

Fig. 2. First generation technology watch/horizon scanning framework.

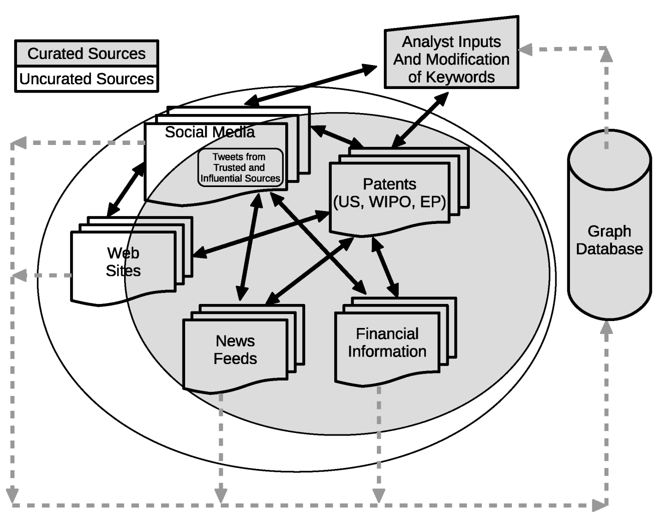

Fig. 3. Proposed updated TW/HS framework.

From these 93,675 unique tweets, we identified all one, two, and three word phrases. We then counted the number of tweets that use each phrase in two time periods: 2012-2015 and 2016-March 2017. Phrases that experience a large increase in mentions in the latter period may be related to technologies that are emerging. Meanwhile, phrases whose usage is decreasing may be related to technologies in decline. A further enhancement, not implemented here, would be to also count the number of different trusted sources in which a phrase appears, to determine the breadth of acceptance of the phrase.

There were 54,077 unique tweets in the first time period, and 39,598 unique tweets in the second period, so we did not use raw counts to compare the two periods. Instead, the unit of analysis we used was the number of mentions per 1,000 tweets. For example, the phrase *deep learning* appeared in 55 tweets in the first time period, and 65 tweets in the second period. This equates to 1.02 mentions per 1000 tweets in the first period and 1.57 mentions per 1000 tweets in the second period.

Table 1 contains a list of the phrases that appear at least once per 1000 tweets in 2016-March 2017, and show at least a 30% increase over the earlier time period. The table has been cleaned up manually to remove phrases that are not useful (with a larger set of stopwords and additional criteria, this clean-up step could be largely automated). A number of the terms in Table 1 are widely-known technologies, such as semiconductors, encryption and electric cars. A number of other technologies are less well-known outside their own fields, and are thus of particular interest. ‘Blockchain’ is a good example. This refers to a distributed database that maintains a continuously growing list of ordered records called blocks. The blocks contain data that cannot be altered retroactively, plus a timestamp and a link to a previous block. These features make blockchains inherently secure, which is why they are the technology behind bitcoin. According to the Harvard Business Review:

"With blockchain, we can imagine a world in which contracts are embedded in digital code and stored in transparent, shared databases, where they are protected from deletion, tampering, and revision. In this world every agreement, every process, every task, and every payment would have a digital record and signature that could be identified, validated, stored, and shared. Intermediaries like lawyers, brokers, and bankers might no longer be necessary. Individuals, organizations, machines, and algorithms would freely transact and interact with one another with little friction. This is the immense potential of blockchain." [13]

TABLE I. PHRASES WITH INCREASED USAGE IN 2016-2017

	Occurrences per 1000 Tweets		
Phrase	**2012-2015**	**2016-2017**	**% Change**
alexa	0.00	1.24	1000+
zika/zika virus	0.02	5.58	1000+
fake news	0.02	1.31	1000+
brexit	0.06	2.58	1000+
trump	0.76	28.18	1000+
google ai	0.06	1.29	1000+
gravitational waves	0.26	2.98	1000+
whatsapp	0.09	1.01	992.50
alibaba	0.17	1.59	855.90
clinton	0.65	5.83	801.30
toshiba	0.18	1.36	637.40
partnership	0.20	1.49	632.40
deepmind	0.15	1.06	616.90
snapchat	0.24	1.57	551.30
blockchain	0.20	1.14	458.60
uber	1.79	7.45	315.30
machine learning	0.57	1.84	221.50
president/presidential	2.09	6.72	221.40
artificial intelligence/ai/#ai	6.86	22.02	220.90
oculus	0.59	1.57	164.50
electric car	0.78	1.94	150.30
encryption	1.26	3.13	149.00
zuckerberg	0.54	1.26	135.40
netflix	1.42	3.28	130.50
auton. cars/ selfdriving cars	3.74	8.28	121.70
elon musk	0.87	1.82	109.20
virtual reality/vr	4.29	8.66	101.90
social media	1.37	2.53	84.50
reef	0.74	1.36	84.30
selfie	0.57	1.01	76.20
semiconductor	0.74	1.29	74.10
coral	0.85	1.46	72.10
iphone	4.88	8.38	71.70
spacex/#spacex/#falcon9	4.25	7.12	67.40
hacking	5.58	8.91	59.60
putin/russia/russian	4.60	7.27	57.90
deep learning	1.02	1.57	53.90
nsa	0.74	1.14	53.60
twitter	5.40	8.28	53.40
silicon valley	2.76	4.07	47.50
america/american	6.27	9.24	47.40
virus	1.65	2.42	47.30
bots	1.24	1.82	46.70
surveillance	1.18	1.67	40.80
cancer	5.47	7.60	38.80
facebook	9.67	13.31	37.60
coal	1.48	2.00	34.80
augmented reality	0.80	1.04	30.20

Overall, Table 1 suggests that artificial intelligence is a hot area right now, with this 60-year old field enjoying a renaissance. The phrase itself appears in Table 1, as do related phrases such as 'deep learning' and 'machine learning'. Also mentioned is DeepMind, which is not a technology, but rather a company that specializes in artificial intelligence and was acquired by Google in 2014.

Also, although the trusted sources are primarily related to science and technology, they are not immune to broader political and economic trends. Table 1 reveals that the phrases 'Brexit' and 'Trump' are both becoming more prominent in these sources, because political actions impact science and technology. For example, there are Brexit-related tweets such as: "Scientists need to wake up to the opportunities of Brexit" and "Excellent science in the UK is at risk if it votes for Brexit". Meanwhile tweets related to Donald Trump include: "How to inoculate people against Donald Trump's fact-bending claims" and "Nobel laureates have spoken out; the battle to defend science against Trump has begun".

Table 2 contains phrases that have been used less frequently in tweets from the most recent time period. This table contains a number of interesting phrases. For example, there are fewer tweets related to Google Glass, as it did not live up to many expectations. There are also fewer tweets surrounding the Higgs Boson discovery, which is understandable given the excitement surrounding this discovery at the time. More broadly, there appears to be a decline in tweets containing phrases related to space (e.g. Saturn Moon, Mars Rover, Spacecraft, Space Station, Space Photos etc.).

Table 3 lists phrases that did not make it into the other tables because they did not increase or decrease significantly in usage between the two time periods. They are worth mentioning because they appear in at least 1% of all the tweets from the trusted sources in at least one time period. As such, they generally represent broad technologies of continuing interest to many researchers. Examples include robot, space, climate and internet.

These results suggest that using trusted sources in Twitter can be a productive approach to locating emerging technologies. The outputs from this process could be used in various ways within a TW/HS framework. They could be provided to analysts as suggestions to direct the framework in Technology Watch mode, rather than the analysts having to derive their own keywords and phrases. For example, highlighting phrases such as 'blockchain' may help analysts select the most fruitful avenues down which to direct the framework. Meanwhile, working in Horizon Scanning mode, the database of tweets from trusted sources is likely to contain more accurate information on a target technology than the regular Twitter stream searched via keywords, which may be full of misleading information or erroneous sources.

B. Identifying Emerging Technologies via Influential Tweeters

The enhancement described in the section above is based largely on trusted sources that represent the social media arms of well-established media outlets, notably leading science and technology publications. Such sources are very useful for general information on a selected technology, but they may be too broad when the target technology is relatively focused. For example, @wired may have some updates on developments in artificial intelligence, but not as many as a source dedicated specifically to this technology. Also, there are many influential

TABLE II. PHRASES WITH DECREASED USAGE IN 2016-2017

	Occurrences per 1000 Tweets		
Phrase	**2012-2015**	**2016-2017**	**% Change**
#mars	1.37	0.00	-100.00
saturn moon	1.07	0.10	-90.60
google glass	1.20	0.13	-89.50
mars rover/curiosity rover	5.75	0.66	-88.60
nanoparticles	1.22	0.18	-85.60
higgs boson	1.02	0.15	-85.20
saturn	2.66	0.56	-79.20
ebola	2.39	0.53	-77.80
social network	1.31	0.35	-73.10
hubble	2.05	0.61	-70.50
reddit	1.04	0.33	-68.30
eclipse	1.78	0.58	-67.30
flash	1.78	0.58	-67.30
flu	1.53	0.51	-67.10
open source	1.72	0.58	-66.30
space photos	1.79	0.61	-66.30
meteor	1.39	0.48	-65.50
kickstarter	1.42	0.51	-64.60
antarctic	1.55	0.56	-64.30
ipad	2.24	0.86	-61.70
drought	1.50	0.61	-59.60
bats	1.48	0.61	-59.10
space station	4.53	1.89	-58.20
3d printer/3d printing	4.64	1.94	-58.20
cloning	1.13	0.48	-57.50
polar	1.13	0.48	-57.50
#robots	1.53	0.66	-57.30
genome	4.09	1.82	-55.60
moore law	1.17	0.53	-54.50
spacecraft	3.37	1.54	-54.30
transistors	1.42	0.66	-53.90
asteroid	4.64	2.15	-53.80
nobel prize	1.83	0.86	-53.10
biotech	1.22	0.58	-52.50
biology	1.37	0.66	-52.10
nanotubes	1.09	0.53	-51.40
supercomputer	1.26	0.63	-49.80
galactic	1.20	0.61	-49.60
collider	1.61	0.83	-48.20
stem cells	1.50	0.78	-47.80
wearable	2.90	1.57	-46.10
watson	1.05	0.58	-44.90
hawking	1.53	0.86	-44.10
fossil	4.59	2.68	-41.70
gmo	1.04	0.63	-39.10
battery	5.64	3.69	-34.70
dark matter	1.59	1.06	-33.40
solar/solar cells	9.76	6.54	-33.10

TABLE III. FREQUENTLY OCCURRING PHRASES

	Occurrences per 1000 Tweets		
Phrase	**2012-2015**	**2016-2017**	**% Change**
robot	30.79	22.45	-27.10
scientists	30.25	18.44	-39.10
space	26.02	15.88	-39.00
google	18.49	20.43	10.40
cars	16.14	20.05	24.20
trump	0.76	28.18	3617.20
brain	14.37	14.52	1.00
artificial intelligence/ai/#ai	6.86	22.02	220.90
data	13.48	14.19	5.20
power	14.74	12.32	-16.40
climate	12.30	13.81	12.30
apple	10.43	15.28	46.40
nasa	16.11	7.83	-51.40
facebook	9.67	13.31	37.60
china	7.90	14.92	89.00
launch	10.47	11.44	9.30
app	10.82	10.81	-0.10
computer	12.28	9.14	-25.60
earth	12.04	8.91	-26.00
mars	14.89	5.53	-62.90
star	12.56	6.89	-45.10
internet	8.67	10.66	22.80
engineers	11.54	6.69	-42.10

sources on Twitter that are not attached to established media outlets, in amongst many unreliable sources. The issue then becomes how to locate technology-specific Twitter sources that can be considered reliable, even if they do not have links to established publications. Such sources may then be tracked for information on the latest developments in a selected technology.

There are a number of standard metrics used in Twitter to measure the influence of different sources, or 'handles'. These include reach (how many users access pages associated with a handle) and engagement (the number of users who have clicked, liked, commented on or shared their posts). It is also possible to track the extent to which posts associated with Twitter handles are re-tweeted by users. Taken together, these various metrics provide a measure of the impact of a Twitter handle.

Impact can be used as an initial screen to help locate influential Twitter handles, but it is insufficient on its own. Even once a target technology is identified, there are still likely to be a large number of irrelevant social media posts. For example, if the target subject is wireless technology, relevant posts discussing developments in this technology may be interspersed with many posts from people complaining about the connectivity of their wireless devices. These may include posts from influential Twitter users with no involvement in wireless technology.

It is possible to distinguish between 'formal' and 'frivolous' tweets based on their language content, for example pronoun usage and presence of slang terms [14]. It is also possible to determine whether the web links in tweets are to scientific outputs, or to advertisements and click-bait sites. From elements such as these, it is then possible to compute a 'gravitas index' for Twitter handles based on their output of 'formal' versus 'frivolous' tweets (this index is the subject of a forthcoming paper by the authors). Hence, when a tweet appears from @xyz, there will be an initial indication whether the tweet is likely to be relevant.

Alongside the impact and gravitas of a Twitter handle, the third element of interest is the extent to which the handle is focused on the target technology. This is measured by the percentage of tweets that use terminologies associated with the technology. These three elements impact, gravitas and focus can then be calculated for each Twitter handle of interest, with high-scoring handles being regarded as good sources of curated information on a selected technology. Handles scoring differently across dimensions may also be useful for different purposes.

As an example, consider natural language processing technology. There is a Twitter handle @NLP_stories created by an NYU student MartinSeongsoon Park. This handle has 42,000+ tweets since 2015, and has 1,600 followers (which is not a particularly large following). Its tweets are almost entirely retweets of stories about natural language processing, deep learning and artificial intelligence. More than 90% of the tweets are linked to content via a URL, and this content includes links to original research, news stories discussing NLP, open source NLP libraries, and links to upcoming NLP conferences. This highly focused, technical content means that @NLP_stories scores very highly for focus and gravitas, although its impact score is relatively low. This suggests it could be a source of highly specialized, technical information on NLP.

A different kind of influential tweeter with coverage of natural language processing is @evankirstel. Evan Kirstel describes himself as a social media influencer, with 113,000 followers and 481,000+ tweets on topics related to all areas of IT, including mobile, robots, internet-of-things, etc. A high percentage of his tweets contain links to business and news sources, which makes @evankirstel influential. However, its gravitas is relatively low, since a number of tweets involve interactions with followers rather than links to content. Also, its focus on NLP is low, since @evankirstel covers many different areas of IT. This suggests that this handle may be a good source for general business and news stories on NLP, but not necessarily technical information about new developments.

C. Improving Linkage Efficiency between Spokes via Graph Databases

In addition to enhancing the social media spoke of the TW/HS framework, it would also be desirable to improve the linkages between information in the various spokes. In the first generation of the framework, data from the various spokes are stored in relational databases. A natural extension of the framework is to instead use a graph database to store all of these data from the various spokes. In basic terms, graph databases have a mathematical graph structure of paths and edges. Every time a relationship is identified between two objects in the database, these objects are linked together via a path, which is then also stored in the database. This is in contrast to a relational database, where data is stored in discrete tables, and complex relationships can only be extracted with queries containing a series of potentially expensive join operations [15].

Graph databases excel in cases where relationships are important. To take a simple example, consider the question: "Who was the actress on *The Sopranos* that played the wife of the actor from that other gangster movie who also appeared in *Field of Dreams* as Shoeless Joe Jackson?". Using a relational database containing information on movies, one would need a series of queries and joins to determine that Ray Liotta played Shoeless Joe Jackson in *Field of Dreams*; and then to search all of his movies and co-performers, plus all actresses on *The Sopranos*, and do a join to get the correct answer (Lorraine Bracco). With a graph database, the process is much simpler. After identifying Ray Liotta as Shoeless Joe Jackson in *Field of Dreams*, a list is made of all actresses that played his wife in any Ray Liotta movie, from which it is determined which have paths back to *The Sopranos*.

Relationships are a key element in technology forecasting, since analysts are often interested in tracking networks of scientists, and determining the subject of their most recent research efforts. An example from the first generation of the TW/HS framework is instructive in this regard. In 2011, one of the highest scoring emerging clusters concerned epothilones, a candidate new chemotherapy treatment. The key inventor in the cluster was Samuel Danishefsky of Memorial Sloan Kettering Cancer Center in New York. Information extracted from the various spokes of the TW/HS framework revealed that Professor Danishefsky's primary research interest was in the total synthesis of compounds i.e. their artificial synthesis from numerous building blocks – and this was the main focus of the patents in the emerging cluster. Epothilones were one of Danishefsky's target compounds, making them cheaper and easier to source than they had been traditionally, since they had to be extracted from slime bacteria.

By storing information in a graph database, rather than a relational database, it would be much more straightforward to link together Danishefskys various patents and scientific outputs, and thus determine his main research focus. It would also be possible to make connections to all of the researchers working with him, plus all venues where he had presented, and research topics on which he and his colleagues are now working. This would result in a much more comprehensive depiction of research activity related to the target technology and associated scientists, generated more quickly and efficiently than is possible using a relational database.

III. Conclusion

This paper outlines various enhancements to an automated TW/HS framework, with a particular focus on the social media spoke of the framework. The objective of these enhancements is to use social media data more effectively, and to make these data more central to the operation of the framework. These enhancements should make the framework more timely and responsive to breakthroughs and milestones in fast-moving, emerging technologies. They should also enable analysts to utilize the framework more effectively, and to retrieve curated information on technologies of interest. Beyond the framework itself, the techniques described in this paper may also be of general interest to researchers using social media data, particularly for technology forecasting.

Future research may involve further exploiting social media data. In particular, one idea would be to build an emerging clusters model using Twitter data, rather than patents. This could be based on similar techniques, identifying articles and news items that are linked to 'hot' tweets that have been retweeted many times. An emerging clusters model based on Twitter data would of course be challenging to build, since Twitter data is less formal than patents, but a resulting model would update particularly quickly, a very useful characteristic in emerging technologies.

Conflict of Interest

The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied of the Department of Defense or the U.S. Government.

Acknowledgment

The research presented here was built upon a framework the authors developed under support by the US Department of Defense (DOD) via contract number HQ0034-12-C-0041 and by the Intelligence Advanced Research Projects Agency (IARPA) via Department of Interior National Business Center contract number D11PC20154. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon.

REFERENCES

[1] H. Linstone and M. Turoff, *The Delphi method: techniques and applications*, Addison-Wesley, Reading, MA, 1975.

[2] F. Bass, "A new product growth model for consumer durables," Management Science, 15, pp. 215–227, 1969.

[3] R. Phaal, C. Farrukh, and D. Probert, "Technology roadmapping - a planning framework for evolution and revolution," Technological Forecasting and Social Change, 71, pp. 5–26, 2004.

[4] D. Robinson, L. Huang, Y. Guo, and A. Porter, "Forecasting Innovation Pathways (FIP) for new and emerging science and technologies," Technological Forecasting and Social Change, 80, pp. 267–285, 2013.

[5] I. Roche, D. Besagni, C. Francois, M. Horlesberger, and E. Schiebel, "Identification and characterisation of technological topics in the field of Molecular Biology," Scientometrics, 82, pp. 663–676, 2010.

[6] A. Breitzman,P. Thomas, "The Emerging Clusters Model: A tool for identifying emerging technologies across multiple patent systems," Research Policy, 44, pp. 195–205, 2015.

[7] B. Sampat, A. Ziedonis, *Patent citations and the economic value of patents, in: Moed H.; Glnzel, W., Schmoch, U. (Eds),* Handbook of Quantitative Science and Technology Research. Kluwer Academic Publishers, Dordrecht, 2004.

[8] L. Aiello, G. Petkos, C. Martin, and D. Corney, "Sensing trending topics in Twitter," IEEE Transactions on Multimedia, 15, 6, pp. 1268–1282, 2013.

[9] F. Atefeh and W. Khreich, "A survey of techniques for event detection in Twitter", Computational Intelligence, 31, 1, pp. 132–164, 2015.

[10] Y. Zhang, G. Zhang, H. Chen, A. Porter, D. Zhu, and J. Lu, "Topic analysis and forecasting for science, technology and innovation," Technological Forecasting and Social Change, 105, pp. 179–191, 2016.

[11] J. Priem and K. Costello, "How and why scholars cite on Twitter," Proceedings of the American Society for Information Science and Technology, 47, 1, pp. 1–4, 2010.

[12] S. Yeo, M. Cacciatore, D. Brossard, D. Scheufele, and M. Xenos, "Twitter as the social media platform of choice for sharing science," 13th International Public Communication of Science and Technology Conference, 2014.

[13] M. Iansiti and K. Lakhani, "The truth about blockchain," Harvard Business Review, Jan-Feb, 2017.

[14] A. Pak and P. Paroubek, "Twitter as a corpus for sentiment analysis and opinion mining," Universit de Paris-Sud, Laboratoire LIMSI-CNRS, 2010.

[15] R. Angles and C. Gutierrez, "Survey of graph database models: ACM Computing Surveys," Association for Computing Machinery, 40(1), 2008.