Network analysis of Twitter data to predict donation activity in a charity campaign

Introduction

Many charities have long regarded the most valuable donor age group to be the over 50s but recent research suggests that younger demographic groups now donate a considerably higher proportion to charity campaigns (May, 2018). Furthermore, traditional methods of giving physical money ‘on the high street’ are in decline, replaced increasingly with web and mobile transactions through digital campaign engagement (Green, 2019). In an increasingly digitised and datafied society (Kennedy, 2016), this study will investigate whether charities can leverage social media data to predict donation activity during a charity campaign.

In recognition that the income of smaller charities in the UK is in decline whilst the income of large ‘super charities’ continues to increase (Hornung, 2019), the ultimate ambition of the study is to create insights that can be used by charities of any size to maximise the value of their engagement with potential donors via social media. However, to generate a large social media dataset, the study will focus on Sport Relief 2020, a biennial campaign by Comic Relief – one of the UK’s largest charities. The Twitter social network will be used for the study due to the availability of data and demographic suitability, the majority of Twitter users being younger than 35 years old with above average income (Tien, 2018).

Literature review

In the UK, the Charities Bill 2005 defines what is a charity and a ‘charitable purpose’, a progression of over 400 years since the earliest legal definition of a charity as described in the Charitable Uses Act of 1601 (House of Commons, 2005). Unsurprisingly therefore, the role of charity in society and of people’s relationship to and engagement with donating has long been a matter of social interest. A large corpus of academic research reflects the vast prior investigations into people’s reasons and impulses for donating to charity, demonstrating a complex decision potentially affected by gender, social demographics, emotional engagement, peer pressure, political motivation, spare time and financial status (Hsu et al., 2005; Kottasz, 2004; Lee and Chang, 2007; Radley and Kennedy, 1995). Further research has considered how donating varies depending whether the donation context is one of ongoing giving, disaster response or a promoted charity campaign (Manesi et al., 2019; Martin, 2013; Yörük, 2012).

With audience engagement moving from traditional media to digital channels, charities now increasingly use social media to promote campaigns and interact with potential donors. As Baym observes, “audiences have always been more engaged than measures of size could capture, but on the internet that engagement is visible [and] traceable” (2013, p. 4), thereby offering new opportunities to understand donation impulses. Social media not only offers a new way to engage with existing audiences but also for engagement with new audiences, leading Wallace et al to state that “charities and non-profit organisations recognise the value of online social media platforms for influencing consumer responses, particularly among younger consumers” (2017, p. 90). Although understanding donation impulses remains a complicated subject, using social media data for research has become an established and broadly utilised method, offering a variety of recognised digital research practices to study such scenarios (Lupton, 2015; Marres, 2017; Rogers, 2013).

Much of the research into donating behaviour on social media has used a qualitative ‘content analysis’ approach, often analysing the text of users’ social content when engaging with charities (or charity campaigns). Due to the nuances and complexities of expression on social media, methods typically range between two alternative approaches: a labour-intensive manual content analysis method, inherently limiting the amount of content that can be studied but suitable for complex content; or a programmatic method such as ‘sentiment analysis’, which can process much more data but is usually better suited to easily classifiable, subjective, opinion based content (Rogers, 2013). Furthermore, the public nature of social media also introduces added complexities, such as the increased visibility of peers, causing Bigsby et al to recognise that “a large body of research in psychology and sociology has explored how social networks impact individuals' decisions” (2019, p. 5). Comparable research methods have had varied results in different contexts – ranging from a study of charity engagement concluding that Twitter content did not reflect donation activity (Korolov et al., 2015) to a study of football player loyalty declaring that “Twitter data consistently added value to predictive models” (Bigsby et al., 2019, p. 10).

Recognising the complex nature of using big data or social media data in research, eminent digital scholars boyd and Crawford assert that “taken out of context, Big Data loses its meaning” (2012, p. 670). However, in response, Seaver demonstrates that context is often central to big data applications, many of which are highly successful, and therefore when dealing with big data it is how context is retained and/or constructed that matters (2015, p. 4). The issue of context is of interest to the charity campaign scenario because the context is relatively well defined – simplistically, an engaged audience with an understood objective – yet seemingly this well-defined context does not simplify the research task of using social media data to create new knowledge through content analysis.

A current gap in charity engagement research using social media data is to investigate whether patterns within the network of engaged users can provide insight into donation activity. Network analysis using social media data has been used in prediction modelling in other areas such as stock market predictions (Qian Li et al., 2016) or football player loyalty (Bigsby et al., 2019), but not in the study of donation activity. By siting the study in the context of a time-constrained charity campaign and using a specific known-text tweet as a marker of a successful donation, this study aims to analyse the social media network of recognised donors as opposed to the social content generated during their engagement with the charity campaign.

Research objectives and methodology

The objective of this research is to define a model that predicts the probability of someone donating as part of a charitable campaign based only on summary analytics and network analysis of their Twitter activity. The study will use the 2020 Sport Relief charity campaign, a high visibility campaign conducted by Comic Relief in the UK, culminating in an ‘all evening’ live TV broadcast on 18/03/2020. As a campaign with huge social media engagement and a fixed timeframe represented by the TV show, it represents an ideal case study for this study.

As a form of ‘big data’, social media data is suitable to inform predictive models due to its attributes such as high volume; real-time availability; variety; granularity; and relationality (Kitchin, 2014). However, these attributes also mean that traditional manual research approaches are unsuitable for analysing big data for social research and programmatic tools and methods are considered more suitable (Marres, 2017). Typically, rather than using previously collected and stored data, programmatic approaches use a real-time connection to retrieve social data using an Application Programming Interface (API) that provides predefined functions for interacting with data for a specific social network (Rogers, 2013). Researchers are therefore able to collect and make sense of huge datasets whilst also analysing the granular insights within, though scholarly critique has observed that the outcomes of such research approaches are intrinsically affected by the choices of the technical dependency owners (Baym, 2013; Rogers, 2013; Stieglitz et al., 2014; Vis, 2013). This study will adopt a data mining and analysis approach widely observed as the knowledge discovery process (Sharma et al., 2012), through which data is collected, processed and analysed to generate meaningful outcomes and knowledge.

Data collection

Twitter provides a tiered range of APIs for interacting with social data on the Twitter network, ranging from a free standard tier with restrictions up to commercial tiers with unconstrained data access. This study will use Twitter's standard Search API, and will therefore be restricted to imposed limitations such as limiting the number of requests allowed, returning only the majority of tweets and only from the previous 7 days of data (Thelwall, 2015; Twitter Developer documentation, n.d.). Tools such as Mozdeh and NodeXL have been developed to assist non-technical people to interact with technical interfaces such as the Twitter API, but for greater efficiency, flexibility and ethical considerations this study will use a bespoke node.js application to collect and process only the specific data required.

The data collection process can be outlined as follows:

1. From 04/03/20 (14 days before Sport Relief), retrieve any tweets that contain the text phrases "#sportrelief", "Sport Relief" or "@sportrelief".
2. Store the Twitter User ID for each user who posted a retrieved tweet, and store which of the text phrases their tweet included.
3. Asynchronously, retrieve and store Twitter User IDs for the Twitter users followed by that user.
4. From 17/03/20 (24 hours before Sport Relief) until 21/03/20 (3 days after Sport Relief) retrieve any fixed-text 'post-donation declaration' tweets containing the exact text shared after officially donating to Sport Relief.
5. Store the Twitter User ID who shared each 'post-donation declaration' tweet retrieved and set/update the user as a recognised donor.
6. Programmatically generate and store a numerical Anonymous ID for all users e.g. 1000000001, 1000000002, etc.
7. Capture the Anonymous ID and Twitter username of identified official influencer users.
8. Delete all Twitter User IDs.
9. Generate summary statistics for data exploration prior to network analysis.
10. Generate GraphML data for subsequent network analysis.

Data analysis

After data collection and processing, summary statistics of the dataset (Diagram 1) will be produced to highlight potential areas for investigation during data analysis.

Diagram 1: Example of summary statistics (not using actual data)

anon_id	donated	match_hashtag	match_text	match_username	matching_tweet_latest	matching_tweet_during_event
1000000001	FALSE	2	0	0	17/03/2020	FALSE
1000000002	TRUE	11	1	2	18/03/2020	TRUE
1000000003	TRUE	3	2	0	16/03/2020	FALSE
1000000004	TRUE	5	0	2	18/03/2020	FALSE
1000000005	FALSE	1	0	0	11/03/2020	FALSE

The data will be analysed using a network analysis method, conducted using Gephi, a tool that enables the exploration and investigation of network data through the visualisation and analysis of the nodes (Twitter users), edges (connection between users) and related metadata (did they donate; how many times did they use the hashtag) (Jacomy et al., 2014). The main hypothesis of the study is that a donation prediction pattern will be observable based on the attributes of a donor's network. Through visualising the data with Gephi using the ForceAtlas2 graph, exploration and analysis of the network graph will be investigated to generate insights such as:

- are there observable patterns when separating donors and non-donors?
- are there distinct outliers (non-conforming users)?
- are there observed clusters (groups) of users (see influencers below)?
- is the distance (number of connections) between a user and an influencer significant?
- are the in-degree/out-degree edge directions (from/to another user) significant?

Within a network analysis approach using social media data, it is typical to identify key 'influencers' – users with whom a connection may have a significant impact on the impact of the network (Rainie and Wellman, 2012). Previously identified 'official' users (such as @comicrelief or @sportrelief) are expected influencers and will be highlighted manually in the graph using their known Anonymous ID but newly observed anonymous influencers will also be recognised during the analysis and their significance considered on the prediction model. Influencer significance is usually regarded as proportional to the number of connections, though academic research has yet to investigate recent advances in social media strategy suggesting that influence is more greatly affected by the strength of relationship to the connected user and therefore 'nano' influencers may be more significant than 'macro' influencers (Ismail, 2018).

Ethics

Despite Facebook's founder Mark Zuckerberg declaring in 2010 that privacy is "no longer a social norm" (Johnson and Vegas, 2010) and that most content posted to Twitter is publicly visible, using social media data for social research raises many ethical concerns in comparison to traditional social research approaches (Marres, 2017). Furthermore, due to the personal, political, moral and emotional associations with charitable giving (Paulin et al., 2014), this study places huge emphasis on the potential impacts of using social media data.

During the design of this study, the ethical impacts were initially assessed by using the Association of Internet Research ethics guidelines, a framework for key considerations when conducting research using social media data (AoIR, 2012). Further demonstrating the importance and frequency of the need to adopt good ethical practice in digital social research, Zook collaborated with other leading digital scholars to describe “ten simple rules for responsible big data research” (2017) and these have also been used to guide the ethical considerations of this study.

As this study will use a programmatically retrieved dataset, it will not be practical (or even possible) to ask users’ consent to participate and therefore it will be essential to ensure appropriate data anonymization and obfuscation practices are applied (Marres, 2017). The bespoke data collection approach used for this study ensures that ethical considerations are central to the study by deliberately and significantly minimising the amount of data retrieved and stored. Common ethical concerns of social data research such as ‘reverse engineering’ quoted content or metadata to identify a user (Zimmer, 2018) will be vastly reduced by limiting the use of tweet content to only identify engagement with the campaign rather than to conduct content analysis. Regarding usernames, identified official public accounts (e.g. @sportrelief) are considered as valid for later identification, but all other usernames will be deleted after data collection is complete and replaced with a programmatic anonymised User ID as this will be adequate for subsequent network analysis. Where stored, tweet dates will be stored as date only (without time) which will avoid identification due to the high volume of similar tweets on dates during the period of data collection.

From a data storage perspective, the ethical considerations above ensure that the final dataset to be used for analysis will be suitably anonymised. The data retrieved and stored during data collection will be stored and processed in a database but once the final datasets are produced the database will be deleted. The outcome will be that it will not be possible to de-anonymise or in any way update the final dataset.

From a more holistic ethical perspective, one concern is that if a suitably convincing model is demonstrated that predicts donation probability by network analysis, the approach could be (ab)used in future with the intent of adversely manipulating the act of donating or the amount donated.

Challenges and concerns

Researching social media data is fraught with complexity. In deconstructing remarks made by the Twitter CEO, Vis observes conflicting positions whereby Twitter presents itself as a visible representation of *the* world, whilst it might more accurately be regarded as *a* world - constructed by the business rules and algorithms developed and controlled by Twitter (2013). Consequently, whilst social researchers are keen to recognise the value and opportunity presented by social media research, it is essential to perpetuate the position that data made available through official APIs only reflect the “specific world” of that social network and are further dictated by the specific rules and features of the API or the software through which they are accessed. Furthermore, a current concern is that Twitter (and other social networks) are increasingly limiting and constantly changing open access to social media data, even when explicitly known to be for academic purposes. Although commercial services are available that remove some or all of the limitations of free APIs, digital sociologists argue that, even if funding were unlimited, as social media is so prevalent in society it should be openly available to research (Lupton, 2015).

Considering the topic of this study, attempting to research donation behaviour around a specific charity campaign is more complicated and potentially less informed without the involvement of the charity. For example, using only the social media data accessible via the Twitter API and using the ‘post-donation tweet’ as a signifier of a donation cannot recognise users who donate but choose not to tweet about it, probably resulting in a significant amount of false negatives in the data. Whilst these ‘false negative’ users should conform to the same patterns as the donors in the analysis, it will be impossible to validate and as such they will reduce the clarity of the patterns of known donors. By using other data available to the charity, this issue could probably be addressed and extra data such as amount donated would enable more granular segmentation and analysis.

More generally regarding the research approach, it is not clear at this proposal stage whether the aim to create insights that could be used by smaller charities can be achieved whilst studying a very large charity campaign – a necessity due to the need for large amounts of social media data. At the very least, further research would be required on a varied type and scale of charity campaigns, as well as considering other charity campaign scenarios such as disaster response campaigns. Additionally, creating accurate and effective predictive models in any context with the variance introduced by human involvement is difficult if even possible, and even if the model works for one charity or one campaign it may not translate to other scenarios.

Regarding the research methods used, the summary analytics would be more insightful by conducting full logistical regression analysis prior to the network analysis. The value of key signifiers or metadata within the initial social media data might have been overlooked, such as capturing the number of tweet replies, tweet comments or retweets (Bigsby et al., 2019). Similarly, considering the recognised context specificity, conducting content analysis – potentially just on a specific subset of the social media data - could provide further insights to support the predictive model when combined with the existing network analysis.

From a computing perspective, network analysis using Gephi (or similar alternatives) is a highly resource intensive process and the social media data used in this study could be relatively large. It is unclear without conducting preliminary tests whether processing the data is feasible with the defined data collection process. Of added concern, the ambition to put ethics at the centre of the research design means that retrospectively amending the data collection is more difficult once the original Twitter usernames have been deleted. The data collection process would ideally therefore be tested and iterated before conducting the main study.

Conclusion

This research proposal positions the study in the context of social media engagement with a charity campaign, with the objective of producing a predictive model of donation activity using network analysis of Twitter social media data. The importance, context and approach of the study are positioned within existing literature and demonstrate a gap that the study aims to fill. The data collection and analysis methods to be used in the study are accurately defined, and are informed strongly by the ethical considerations central to the research approach. Recognising that the objective and approach of the study combines the complex aspects of predictive modelling, the use of social media data and a complicated subject, some of the challenges and concerns of the study have been highlighted and reflected upon.

References

AoIR (2012) Ethics [Online]. Available at https://aoir.org/ethics/ (Accessed 20 May 2019).

Baym, N. K. (2013) ‘Data not seen: The uses and shortcomings of social media metrics’, First Monday, vol. 18, no. 10 [Online]. DOI: 10.5210/fm.v18i10.4873 (Accessed 20 May 2019).

Bigsby, K. G., Ohlmann, J. W. and Zhao, K. (2019) ‘The turf is always greener: Predicting decommitments in college football recruiting using Twitter data’, Decision Support Systems, vol. 116, pp. 1–12 [Online]. DOI: 10.1016/j.dss.2018.10.003 (Accessed 19 May 2019).

boyd, danah and Crawford, K. (2012) 'Critical Questions for Big Data', Information, Communication & Society, vol. 15, no. 5, pp. 662–679 [Online]. DOI: 10.1080/1369118X.2012.678878 (Accessed 20 May 2019).

Green, C. (2019) Online giving in the UK increases 5.5% in 2018 [Online]. Available at https://www.charitydigitalnews.co.uk/2019/02/25/online-giving-in-the-uk-increases-5-5-in-2018/ (Accessed 21 May 2019).

Hornung, L. (2019) Small charities: Key findings from our data | NCVO Blogs [Online]. Available at https://blogs.ncvo.org.uk/2019/01/21/small-charities-key-findings-from-our-data/ (Accessed 21 May 2019).

House of Commons (2005) House of Commons - Explanatory Note [Online]. Available at https://publications.parliament.uk/pa/cm200506/cmbills/083/en/06083x--.htm (Accessed 21 May 2019).

Hsu, J. L., Guan-Yu, L. and Tien, C.-P. (2005) 'Social Concerns and Willingness to Support Charities', Social Behavior and Personality; Palmerston North, vol. 33, no. 2, pp. 189–199 [Online]. DOI: 10.2224/sbp.2005.33.2.189 (Accessed 21 May 2019).

Ismail, K. (2018) Social Media Influencers: Mega, Macro, Micro or Nano [Online]. Available at https://www.cmswire.com/digital-marketing/social-media-influencers-mega-macro-micro-or-nano/ (Accessed 21 May 2019).

Jacomy, M., Venturini, T., Heymann, S. and Bastian, M. (2014) 'ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software.', PLoS ONE, vol. 9, no. 6, p. e98679 [Online]. DOI: 10.1371/journal.pone.0098679 (Accessed 21 May 2019).

Johnson, B. and Vegas, L. (2010) 'Privacy no longer a social norm, says Facebook founder', The Guardian, 11th January [Online]. Available at https://www.theguardian.com/technology/2010/jan/11/facebook-privacy (Accessed 20 May 2019).

Kennedy, H. (2016) Post, Mine, Repeat: Social Media Data Mining Becomes Ordinary, London, UK, Palgrave Macmillan Limited [Online]. Available at http://ebookcentral.proquest.com/lib/sheffield/detail.action?docID=4720259 (Accessed 21 May 2019).

Kitchin, R. (2014) The data revolution: big data, open data, data infrastructures and their consequences, Los Angeles, SAGE.

Korolov, R., Peabody, J., Lavoie, A., Das, S., Magdon-Ismail, M. and Wallace, W. (2015) 'Actions are louder than words in social media', 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pp. 292–297 [Online]. DOI: 10.1145/2808797.2809376.

Kottasz, R. (2004) ‘Differences in the Donor Behavior Characteristics of Young Affluent Males and Females: Empirical Evidence From Britain’, Voluntas: International Journal of Voluntary and Nonprofit Organizations, vol. 15, no. 2, pp. 181–203 [Online]. Available at http://www.jstor.org/stable/27927905 (Accessed 21 May 2019).

Lee, Y.-K. and Chang, C.-T. (2007) ‘Who Gives What to Charity? Characteristics Affecting Donation Behavior’, Social Behavior and Personality; Palmerston North, vol. 35, no. 9, pp. 1173–1180 [Online]. DOI: 10.2224/sbp.2007.35.9.1173 (Accessed 21 May 2019).

Lupton, D. (2015) Digital sociology, London, Routledge.

Manesi, Z., Van Lange, P. A. M., Van Doesum, N. J. and Pollet, T. V. (2019) ‘What are the most powerful predictors of charitable giving to victims of typhoon Haiyan: Prosocial traits, socio-demographic variables, or eye cues?’, Personality and Individual Differences, vol. 146, pp. 217–225 [Online]. DOI: 10.1016/j.paid.2018.03.024 (Accessed 19 May 2019).

Marres, N. (2017) Digital Sociology: The Reinvention of Social Research, Oxford, Polity Press.

Martin, J. A. (2013) ‘Disasters and Donations: The Conditional Effects of News Attention on Charitable Giving’, International Journal of Public Opinion Research, vol. 25, no. 4, pp. 547–560 [Online]. DOI: 10.1093/ijpor/eds044 (Accessed 19 May 2019).

May, M. (2018) Younger people now give more than over 55s, report shows [Online]. Available at https://fundraising.co.uk/2018/05/03/younger-people-now-give-more-than-over-55s-report-shows/ (Accessed 20 May 2019).

Paulin, M., Ferguson, R. J., Schattke, K. and Jost, N. (2014) ‘Millennials, Social Media, Prosocial Emotions, and Charitable Causes: The Paradox of Gender Differences’, Journal of Nonprofit & Public Sector Marketing, vol. 26, no. 4, pp. 335–353 [Online]. DOI: 10.1080/10495142.2014.965069 (Accessed 19 May 2019).

Qian Li, Bing Zhou and Qingzhong Liu (2016) ‘Can twitter posts predict stock behavior?: A study of stock market with twitter social emotion’, 2016 IEEE International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), pp. 359–364 [Online]. DOI: 10.1109/ICCCBDA.2016.7529584.

Radley, A. and Kennedy, M. (1995) ‘Charitable Giving by Individuals: A Study of Attitudes and Practice’, Human Relations, vol. 48, no. 6, pp. 685–709 [Online]. DOI: 10.1177/001872679504800605 (Accessed 21 May 2019).

Rainie, H. and Wellman, B. (2012) Networked: the new social operating system, Cambridge, Mass., MIT Press.

Rogers, R. (2013) Digital Methods, Cambridge, Massachusetts, MIT Press [Online]. Available at https://ieeexplore.ieee.org/servlet/opac?bknumber=6517069 (Accessed 5 January 2019).

Seaver, N. (2015) ‘The nice thing about context is that everyone has it’, Media, Culture & Society, vol. 37, no. 7, pp. 1101–1109 [Online]. DOI: 10.1177/0163443715594102 (Accessed 21 May 2019).

Sharma, S., Osei-Bryson, K.-M. and Kasper, G. M. (2012) ‘Evaluation of an integrated Knowledge Discovery and Data Mining process model’, Expert Systems with Applications, vol. 39, no. 13, pp. 11335–11348 [Online]. DOI: 10.1016/j.eswa.2012.02.044 (Accessed 21 May 2019).

Stieglitz, S., Dang-Xuan, L., Bruns, A. and Neuberger, C. (2014) ‘Social Media Analytics’, Business & Information Systems Engineering, vol. 6, no. 2, pp. 89–96 [Online]. DOI: 10.1007/s12599-014-0315-7 (Accessed 20 May 2019).

Thelwall, M. (2015) ‘Evaluating the Comprehensiveness of Twitter Search API Results: A Four Step Method’, Cybermetrics: International Journal of Scientometrics, Informetrics and Bibliometrics, no. 18–19, p. 1 [Online]. Available at https://dialnet.unirioja.es/servlet/articulo?codigo=6550233 (Accessed 20 May 2019).

Tien, S. (2018) Top Twitter Demographics That Matter to Social Media Marketers in 2018 [Online]. Available at https://blog.hootsuite.com/twitter-demographics/ (Accessed 21 May 2019).

Twitter Developer documentation (n.d.) Standard search API [Online]. Available at https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.html (Accessed 20 May 2019).

Vis, F. (2013) ‘A critical reflection on Big Data: Considering APIs, researchers and tools as data makers’, First Monday, vol. 18, no. 10 [Online]. DOI: 10.5210/fm.v18i10.4878 (Accessed 20 May 2019).

Wallace, E., Buil, I. and de Chernatony, L. (2017) ‘When does “liking” a charity lead to donation behaviour?: Exploring conspicuous donation behaviour on social media platforms’, European Journal of Marketing, vol. 51, no. 11/12, pp. 2002–2029 [Online]. DOI: 10.1108/EJM-03-2017-0210 (Accessed 19 May 2019).

Yörük, B. K. (2012) ‘The Effect of Media on Charitable Giving and Volunteering: Evidence from the “Give Five” Campaign’, Journal of Policy Analysis and Management, vol. 31, no. 4, pp. 813–836 [Online]. Available at http://www.jstor.org/stable/41653797 (Accessed 19 May 2019).

Zimmer, M. (2018) ‘Addressing Conceptual Gaps in Big Data Research Ethics: An Application of Contextual Integrity’, Social Media + Society, vol. 4, no. 2 [Online]. DOI: 10.1177/2056305118768300 (Accessed 20 May 2019).

Zook, M., Barocas, S., boyd, danah, Crawford, K., Keller, E., Gangadharan, S. P., Goodman, A., Hollander, R., Koenig, B. A., Metcalf, J., Narayanan, A., Nelson, A. and Pasquale, F. (2017) ‘Ten simple rules for responsible big data research’, PLoS Computational Biology, vol. 13, no. 3 [Online]. DOI: 10.1371/journal.pcbi.1005399 (Accessed 20 May 2019).