## Digital Methods Summer School

### Bournemouth University Talbot Campus

**1-5th July 2024**

This is a one-week, intensive summer school primarily for postgraduate research students and early career researchers. It is intended for people who do not write code but are utilising digital methods. The Summer School will bring together BU researchers with those working at other institutions. The summer school will be taught by world-leading experts from BU and beyond. The summer school is free and catering will be provided. We will also offer some bursaries of up to £200 for non-BU research students. Places are limited to ensure the overall quality of the learning experience.

The summer school will cover digital media research from data collection through to different analytical approaches and will include sessions on research ethics and industrial collaborations. Students will work on a common project throughout with a goal of generating a joint publication. Previous iterations of this course have led to publications in internationally recognised journals.

Students will learn how to collect website and social media data using both screen-scraping techniques and through APIs. Once collected, students will learn how to clean and then analyse the data using two methods: content analysis and social network analysis. Finally, the course will introduce students to data visualisation techniques.

**Draft Schedule**

### Day 1: Introduction to the Summer School

**Monday 1st July**

**9.15-9.30am Coffee and pastries**

**9.30-10.15am Session 1: Getting to know each other and your PhDs**

**10.15-12: Introducing the Project (Professor Scott Wright and Professor Darren Lilleker)**

In this session we will talk about the project that will being addressed during the week.

**11-11.15 Break**

**12-12.30 Lunch**

**12.30-1.30pm Research Ethics for Digital Media (**[**Professor Richard Berger**](https://staffprofiles.bournemouth.ac.uk/display/rberger) **and Dr Michelle Heward)**

This session will cover the essentials of ethical concerns for digital media and spaces and current debates around APIs (qualitative and quantitative).

**Collecting Website and Social Media Data (Professor Scott Wright)**

***1.30-2.30pm Introduction: the history and politics of digital data***

This session will outline the history of digital data and methods and key events that have shaped our approach from the influence of the ‘Californian Ideology’ through to Cambridge Analytica and the sale of Twitter. It will look at how the politics of data has influenced what researchers study and how this has shaped our understanding of the Internet. It will also talk about attempts by researchers to bypass restrictions and blackboxes, including the ethics and implications of all of this.

The two principal methods for collecting data from websites will then be presented (front end scraping, and back end APIs). This will be underpinned by an explanation of how websites and social media work. The ethical dilemmas and challenges of collecting data will be presented, as well as ongoing changes to APIs; an outline of how scrapers actually work; the limitations of web-scraping; and some of the guidelines rules that must be considered.

*Break 2.30-2.45*

***2.45-4.15 Collecting Digital Media Data* *using APIs***

API Practical: We will learn about methods to collect data from APIs. We are likely to use NodeXL plus one other mechanism. We will learn about the functionality of different APIs and strengths and weaknesses. You will collect sets of tweets and assess the kinds of data that is provided. We will test different search terms and functions and think about the affordances of the APIs.

**Finish 4.15**

**Optional Dinner: Fish and Chips by the Beach (own cost)**

### Day 2:Collecting Website and Social Media Data (Professor Scott Wright)

**Tuesday 2nd July**

**9.15-9.30: coffee and pastries**

**9.30-4.15 Scraping Websites (Professor Scott Wright)**

In these sessions we will learn how to scrape websites from the ‘front end’. That is, you will learn how to scrape webpages and automatically traverse between pages to automate data collection. Using these techniques you can collect visible data. We will learn how to build scrapers in layers and how to traverse using ‘next page’ links and by creating and using a list of URLs. We will learn the strengths and limitations of such techniques and apply them to different websites. We will collect the data for our proposed article (this may continue after the session). What are the implications of any data collection issues for the nature of the articles/Research Questions? (breaks through day)

While scrapers are running we may do some other training in the afternoon, probably Open Refine for data cleaning.

**Collaborating with Industry, NGOs and Government (tbc)**

This session will talk about working with industry, NGOs, and government as part of your research. It will involve an industry speaker and this may be linked to the collective research project. The goals if for this to happen on Tuesday subject to speaker availability.

### Day 3: Automated and Manual Content Analysis

**Wednesday 3rd July**

The third day will focus on different forms of content analysis, and particularly sentiment analysis and manual content analysis. We will discuss the strengths and weaknesses of these approaches.

**9.15-9.30 coffee and pastries**

**9.30-10.45 Sentiment Analysis (Professor Scott Wright)**

The history of sentiment analysis will be presented. We will look at the strengths and weaknesses of sentiment analysis and some examples of sentiment analysis-based studies. We will then undertake a practical demonstration of “senti-strength” on the Twitter data we previously collected. We will look at ways to present and analyse the data.

**10.45-11 Break**

**11-12.15 Computer-Assisted Content Analysis/Corpus Linguistics (tbc)**

Precise details for this session are to be confirmed.

**12.15-1.15 Lunch**

**1.15-4.15pm** **Manual Content Analysis (Professor Dan Jackson and Prof Scott Wright)**

The history and context of content analysis will be discussed. We will look at some classic studies that have used content analysis. We will then go through the methodological principles of content analysis. We will learn how about the importance of intercoder reliability and how to conduct and interpret reliability tests. We will build some codes and apply and test them. For our publication, we will work on a collective codeframe adapted from the literature and do reliability testing on each coder.

### Day 4: Social Network Analysis and visualisation

**9.15-9.30: Coffee and Pastries**

**9.30-10.45: Coding Time**

We will use this to continue coding and address any queries and questions

**10.45-11 Break**

**11-12.45 Social Network Analysis**

In this session we will learn about different approaches to visualising digital media data.

First, we will learn about the theory and principles of social networks. You will learn how to use NodeXL to undertake social network analysis. NodeXL is a powerful tool for collecting, mapping and visualising data. NodeXL is designed to make data visualisation available to people who can't write code, opening the door to key “big data” methods to a broad range of social science and humanities students and academics.

We will discuss the background to, and key principles of, social network analysis. Key debates and studies will be briefly discussed. We will then use NodeXL to visualise/analyse data. We will have already used NodeXL to collect data. Here we will explore the features of NodeXL in more detail, and in particular the mapping functions.

**12.45-1.30 Lunch**

**1.30-4.15pm Data Storytelling Methods (Professor Anna Feigenbaum)**

Details to be confirmed

**Conference Dinner tbc (own cost)**

### Day 5: To Infinity and Beyond: AI and Predictions

**9.15 coffee and pastries**

**9.15-10.45: AI in Research Masterclass (Dr Festus Adedoyin)**

This session will explore some of the key issues and approaches for using AI in research. We will discuss what Generative AI is and how it is shaping digital methods; how to design prompts for generative AI tools; using AI for text and image analysis. We will also from researchers using AI in their own funded grants as well as applications for creative practice. We hope also to talk about legal aspects here.

**10.45-11am break**

**11-12.00 AI in Research Practice (Dr Festus Adedoyin)**

In this session we will do hands on work with different AI tools used in research, building on the first session.

**12.00-12.30 Lunch**

**12.30-3.30 Predictive Modelling (Professor Rick Stafford)**

Predictive models help can help visualise the future, or assess the effect of interventions or policies being made now. While most predictive models are mathematically or computationally complex, we will explore what simple models can tell us – for example, how effective may media communications be in changing behaviour, or what might be some of the unintended consequences of political policies. These models may also aid with digital storytelling or help inform creative practices. They are simple to build and analyse using a new R package – but no skills in programming or R are required.

**3.30-3,45 Break**

**3.45-4.15 Next Steps**

**FINISH**

### Bibliography

*Reading for Data Collection.*

**Hansen, D., Schneiderman, B. and Smith, M. (2010), *Analysing Social Media Networks with NodeXL* Boston, MA: Morgan Kauffmann. We will provide relevant chapters.**

Bradshaw, P. (2012/4), “Scraping multiple pages with ‘next’ links using Outwit Hub”, *Scraping for Journalists*, pp62-69

Bradshaw, P. (2012/4), “Poorly Formatted webpages – solving problems with Outwit”, *Scraping for Journalists*, pp70-83.

*Reading for Twitter and Politicians.*

Again, there is a significant amount of literature. No specific required reading, but preferably two of:

Adi, A., Erickson, K. and Lilleker, D., 2014. Elite Tweets: Analyzing the Twitter Communication Patterns of Labour Party Peers in the House of Lord. *Policy & Internet*, 6 (1), 1-27.

Ausserhofer, Julian, and Maireder, Axel. (2013). National Politics on Twitter. Structures and topics of a networked public sphere. *Information, Communication & Society*, 16 (3), 291-314.

Coleman S, Moss G (2008) “Governing at a distance-politicians in the blogosphere”, *Information Polity.* 13(1-2): 7-20.

Enli, G. and Skogerbo, E. (2013) Personalised Campaigns in Party-Centred Politics: Twitter and Facebook as arenas for political communication*, Information, Communication & Society* 16(5): 757-774.

Graham, T. Between Broadcasting Political Messages and Interacting with Voters: The use of Twitter during the 2010 UK General Election Campaign*, Information, Communication & Society* 16(5): 692-716.

Graham, T., Jackson, D. and Broersma, M. (2018) *The Personal in the Political on Twitter: Towards a Typology of Politicians’ Personalized Tweeting Behaviours.* In Schwanholz, J., Graham, T and Stoll, P-T, (eds.) Managing Democracy in the Digital Age: Internet Regulation, Social Media Use, and Online Civic Engagement. Springer, Cham, Switzerland , pp. 137-157.

Lilleker, D.G. and Koc-Michalska, K., 2013. Online Political Communication Strategies: MEPs, E-Representation, and Self-Representation. *Journal of Information Technology and Politics*, 10 (2), 190-207.

Lilleker, D. and Jackson, N., 2011. Microblogging, constituency service and impression management – UK MPs and the use of Twitter. *Journal of Legislative Studies*, 17, 86-105.

Small, T.A. (2011) What the Hashtag? A Content Analysis of Canadian politicians on Twitter, *Information, Communication & Society* 14(6): 872-895.

Stromer-Galley, J. (2000) Online Interaction and Why Candidates Avoid it, Journal of Communication.

**Content Analysis Reading**

Weare, C. and Lin, W.Y. (2000), “Content Analysis of the World Wide Web: opportunities and challenges”, *Social Science Computer Review*, 18(3): 272-292.

Graham, T. and Wright, S. (2014), “Discursive Equality and Everyday Talk Online: The impact of ‘superparticipants’”, *Journal of Computer-Mediated Communication.* Available at: <http://onlinelibrary.wiley.com/doi/10.1111/jcc4.12016/pdf>.

**Content Analysis Other Reading:**

Berger, A.A. (2011), *Media and Communication Research Methods*, London: Sage. (C11).

Divakaran, A. (2009), *Multimedia Content Analysis: theory and applications* Springer.

Glasgow University Media Group (1980) *Bad News* London: Routledge.

Graham, T. (2008), “Needles in a haystack: A new approach for identifying and assessing political talk in nonpolitical discussion forums”, *Javnost - The Public*, 15(2): 167-36. [Available here.](http://javnost-thepublic.org/article/pdf/2008/2/2/)

Graham, T., Broersma, M and Hazelhoff, K. (forthcoming) “Twitter as an instrument for connected representation”. In: R. Gerodimos, D. Jackson, D. Lilleker & R. Scullion (Eds.), *Agents of (Dis)Empowerment: Media and Civic Engagement* London: Routledge. Available at: <http://www.rug.nl/staff/m.j.broersma/closingthegap_bookversion.pdf>.

Gunter, B. (2000), Overview of media research methodologies: media output. In B. Gunter, Media Research Methods: Measuring Audiences, Reactions and Impact (pp. 55-82). London: Sage.

Hansen, A., Cottle, S. Negrine, R. and Newbold, C. (1998), *Mass Communications Research Methods* Basingstoke: Macmillan. (C5)

Harrison, M. (1985), *TV news whose*bias*? A casebook analysis of strikes, television and*media*studies* Hermitage: Policy Journals.

This book critiques the methodology for media bias – particularly the Glasgow University Media Group.

Krippendorff, K. (2004), *Content Analysis: an introduction to its methodology* London: Sage.

Krippendorff, K. and Bock, M.A. (2008), *The Content Analysis Reader* London: Sage.

Neuendorf, K.A. (2002), *The Content Analysis Guidebook* London: Sage.

Riffe, D., Lacy, S., & Fico, F. (2005), *Analyzing Media Messages: Using Quantitative Content Analysis in Research* (2nded.). Mahwah, NJ: Lawrence Erlbaum.

Weber, R.P. (1990), *Basic content analysis* London: Sage. (ebook)

<http://depts.washington.edu/uwmcnair/chapter11.content.analysis.pdf>

**Sentiment Analysis Reading**

Grimmer, J. and Stewart, B.M. (2013), “Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts, *Political Analysis,* 1-31. Available at: http://web.stanford.edu/~jgrimmer/tad2.pdf.

**Sentiment Analysis Further Reading**

boyd, d., Golder, S., & Lotan, G. (2009). Tweet, tweet, retweet: Conversational aspects of retweeting on Twitter. In Proceedings of the 43rd Annual Hawaii International Conference on Systems Science (HICSS-43). Retrieved November 4, 2010, from http://www.danah.org/papers/ TweetTweetRetweet.pdf

Fink, C.R., Chou, D.S., Kopecky, J.J. and Llorens, A.J. () “Coarse and Fine-Grained Sentiment Analysis of Social Media Text”, *John Hopkins Technical Digest*, 30(1): pp. 22-30. Available at: http://www.jhuapl.edu/techdigest/TD/td3001/Fink.pdf.

Prabowo, R. and Thelwall, M. (UD) Sentiment Analysis: A combined Approach. Available at: http://www.cyberemotions.eu/rudy-sentiment-preprint.pdf.

Samberg, D. (2014), “Kicking Off Super Bowl Week Asking #Whosgonnawin”, Verizon: Available at: [http://www.verizonwireless.com/news/article/2014/01/super-bowl-week- kickoff-empire-state-building.html/](http://www.verizonwireless.com/news/article/2014/01/super-bowl-week-%20kickoff-empire-state-building.html/).

Thelwall, M. and Paltoglou, G. (2011), “Sentiment in Twitter Events”, *Journal of the American Society for Information Science and Technology*, 62(2), 406–418.

Thelwall, M. Sud, P. and Vis, F. (2011), “Commenting on YouTube Videos: From Guatemalan rock to El Big Bang”, *Journal of the American Society for Information Science and Technology*, 63(3): 619-629.

Thelwall, M. (pre-pub), “Big Data and Social Web Research Methods. Available here:

http://www.scit.wlv.ac.uk/~cm1993/papers/IntroductionToWebometricsAndSocialWebA nalysis.pdf. See chapter 6.

Thelwall, M. (2014), “Sentiment Analysis and Time Series Data”. In: K. Weller, A. Bruns, J. Burgess, M. Mahrt and Puschmann, C. (Eds) *Twitter and Society*, Peter Lang.

Vis, F. and Procter, R. (2013), “Reading the Riots on Twitter”, *International Journal of Social Research Methodology*, 16(3): 197-214.

**Social Network Analysis Reading**

Hansen, D., Schneiderman, B. and Smith, M. (2010), *Analysing Social Media Networks with NodeXL* Boston, MA: Morgan Kauffmann.

We will provide relevant chapters.

For another very useful “how to” with NodeXL see Pew and Marc’s team description:

<http://www.pewinternet.org/files/2014/02/How-we-analyzed-Twitter-social-media-networks.pdf>.

**Further Reading**

Fielding, N.G., Lee R.M., and Black, G. (2008) *The Sage Handbook of Online Research Methods* London: Sage. Available as E-book (Chapters 6-9).

Knoke, D, Yang, S. (2008), *Social Network Analysis* London: Sage.

Marin, A. and Wellman, B. (2009), ‘Social Network Analysis: An introduction’, (from Sage Handbook, 2010) <http://softwarepublico.gov.br/5cqualibr/6-publicacoes-e-artigos/view/vetor-ecossistema/sobre-redes-sociais/Analise-de-Redes.pdf>

Marres, N. and Weltrevede, M. (2013), “Scraping the Social: issues in live social research”, *Journal of Cultural Economy*, 6(3): 313-335. Draft available:

<https://wiki.digitalmethods.net/pub/Dmi/PapersPublications/Marres_Weltrevede_Scraping_the_Social_draft.pdf>

Prell, C. (2012), *Social network analysis: history, theory & methodology* London: Sage.

Rogers, R. (2010), ‘Mapping Public Web Space with the Issuecrawler’,. In: C. Brossard and B. Reber (eds.), *Digital Cognitive Technologies: Epistemology and Knowledge Society*. London: Wiley: 115-126.

Scott, J and Carrington, P. (2011), *The SAGE handbook of social network analysis* London: Sage. (e-book)

Scott, J. (2012), *What is Social Network Analysis* Continuum*.* Scraping the Social

Shah, C. and Tayebeh, Y.N. (2011) ‘Politics 2.0 with Facebook – Collecting and Analyzing Public Comments on Facebook for Studying Political Discourses”, *Journal of Information Technology and Politics*, (The Future of Computational Social Science)

The\_Promise\_and\_Peril\_of\_Big\_Data.pdf

Thelwall, M. (2009), ‘Visualization in e-Social Science’. In: N.W. Jankowski (ed) *e-Research: Transformation in Scholarly Practice*, London: Routledge.

Wellman, B. (1983) ‘Network Analysis: some basic principles, *Sociological* Theory, 1: 155-200. Available at: <http://www.gvpt.umd.edu/CITE-IT/Documents/Wellman%201983%20Network%20Analysis.pdf>  (note the age of this piece).

*Predictive modelling reading*

Sokolnicki, J.R., Woodhatch, A.L., Stafford, R., 2022. Assessing Environmentally Effective Post-COVID Green Recovery Plans for Reducing Social and Economic Inequality. Anthropocene Science, 1: 375-383. <https://doi.org/10.1007/s44177-022-00037-x>

Stafford R., Croker A.R., Rivers E.M., Cantarello E., Costelloe B., Ginige T., Sokolnicki J., Kang K., Jones P.J.S., McKinley E. and Shiel C. 2020. Evaluating optimal solutions to environmental breakdown. Environmental Science and Policy. 112: 340-347 https://doi.org/10.1016/j.envsci.2020.07.008