MECO4114 Research Methods

Francesco Bailo

The University of Sydney

27 Mar 2024

Observing behaviour

This section is largely based on Chapter 2 of *Bit by bit* by Matthew Salganik (2018).

Whatever the subject of your research, there are mainly three ways to collect data:

1. Running experiments

Whatever the subject of your research, there are mainly three ways to collect data:

1. Running experiments
2. Asking questions

Whatever the subject of your research, there are mainly three ways to collect data:

1. Running experiments
2. Asking questions
3. Observing behaviour ←

Whatever the subject of your research, there are mainly three ways to collect data:

1. Running experiments
2. Asking questions
3. Observing behaviour ←

- Observational data are collected without interfering with either
    - the subject of the investigation or
    - the environment of the subject of the investigation.

Observation of something or somebody is the primordial way of investigating what we are interested in (and usually the beginning of an investigation).

Using instruments and sensors to observe and record what we observe is not new.

What is new is the number of instruments and sensors that monitor and record human behaviour.

Figure 1: *'Galileo Galilei showing the Doge of Venice how to use the telescope', Bertini, 1858*

The combination of *flow* and the *stock* of data produced by these instruments and sensors is often called **Big Data**.

Big data are *big* on three dimensions:

- **V**olume
- **V**ariety
- **V**elocity

- So... can you think of any example of big data or source of big data?

- Big data are not only data generated by the online activity of users and are not only created by companies.

- Big data are not only data generated by the online activity of users and are not only created by companies.

1. Big data are generated online and offline every time a sensor records a human behaviour.

- Big data are not only data generated by the online activity of users and are not only created by companies.

1. Big data are generated online and offline every time a sensor records a human behaviour.
2. Big data are created by companies and governments.

‘Big data are created and collected by **companies** and **governments** for purposes other than research. Using this data for research therefore requires repurposing.’ (Salganik, 2018, p. 14)

$\neq$

Australian Bureau of Statistics Census
census.abs.gov.au
OUR MOMENT TO MAKE A DIFFERENCE

According to (Salganik, 2018), Big Data shares ten characteristics.

1. Big Data are **big**: Rare events, heterogeneity, small differences. *But* how data were created?

According to (Salganik, 2018), Big Data shares ten characteristics.

1. Big Data are **big**: Rare events, heterogeneity, small differences. *But* how data were created?

2. Big Data are **always-on**: Unexpected events and real-time estimates. *But* the systems that collected the data are constantly changing (see *drifting* later)!

According to (Salganik, 2018), Big Data shares ten characteristics.

1. Big Data are **big**: Rare events, heterogeneity, small differences. *But* how data were created?

2. Big Data are **always-on**: Unexpected events and real-time estimates. *But* the systems that collected the data are constantly changing (see *drifting* later)!

3. Big Data are **nonreactive**: Measurement is less likely to change behaviour. *But* a social desirability bias persists.

According to (Salganik, 2018), Big Data shares ten characteristics.

1. Big Data are **big**: Rare events, heterogeneity, small differences. *But* how data were created?
2. Big Data are **always-on**: Unexpected events and real-time estimates. *But* the systems that collected the data are constantly changing (see *drifting* later)!
3. Big Data are **nonreactive**: Measurement is less likely to change behaviour. *But* a social desirability bias persists.
4. Big Data are **incomplete**: No demographic information, no information on behaviour on other platforms, and no data to operationalise theoretical constructs (e.g. 'intelligence').

According to (Salganik, 2018), Big Data shares ten characteristics.

1. Big Data are **big**: Rare events, heterogeneity, small differences. *But* how data were created?
2. Big Data are **always-on**: Unexpected events and real-time estimates. *But* the systems that collected the data are constantly changing (see *drifting* later)!
3. Big Data are **nonreactive**: Measurement is less likely to change behaviour. *But* a social desirability bias persists.
4. Big Data are **incomplete**: No demographic information, no information on behaviour on other platforms, and no data to operationalise theoretical constructs (e.g. 'intelligence').
5. Big Data are **inaccessible**: Access is controlled and conditional.

6. Big Data are **non-representative**: Data do not come from a probabilistic random population sample.

6. Big Data are **non-representative**: Data do not come from a probabilistic random population sample.
7. Big Data are **drifting**: Population drift, behavioural drift, system drift. Systems keep changing all the time!

6. Big Data are **non-representative**: Data do not come from a probabilistic random population sample.
7. Big Data are **drifting**: Population drift, behavioural drift, system drift. Systems keep changing all the time!
8. Big Data are **algorithmically confounded**: Engineering choices impact user behaviours. Also, performativity issues.

6. Big Data are **non-representative**: Data do not come from a probabilistic random population sample.
7. Big Data are **drifting**: Population drift, behavioural drift, system drift. Systems keep changing all the time!
8. Big Data are **algorithmically confounded**: Engineering choices impact user behaviours. Also, performativity issues.
9. Big Data are **dirty**: Dirty data can be created unintentionally or intentionally (e.g. bots).

6. Big Data are **non-representative**: Data do not come from a probabilistic random population sample.
7. Big Data are **drifting**: Population drift, behavioural drift, system drift. Systems keep changing all the time!
8. Big Data are **algorithmically confounded**: Engineering choices impact user behaviours. Also, performativity issues.
9. Big Data are **dirty**: Dirty data can be created unintentionally or intentionally (e.g. bots).
10. Big Data are **sensitive**: The potential sensitivity of the data is always difficult to assess.

Network analysis

Relations, not attributes. Networks, not groups.

[S]ocial network analysts argue that causation is not located in the individual, but in the social structure. While people with similar attributes may behave similarly, explaining these similarities by pointing to common attributes misses the reality that *individuals with common attributes often occupy similar positions in the social structure*. That is, *people with similar attributes frequently have similar social network positions*. Their similar outcomes are caused by the **constraints**, **opportunities** and **perceptions** created by these similar network positions. (Marin & Wellman, 2011, p. 13)

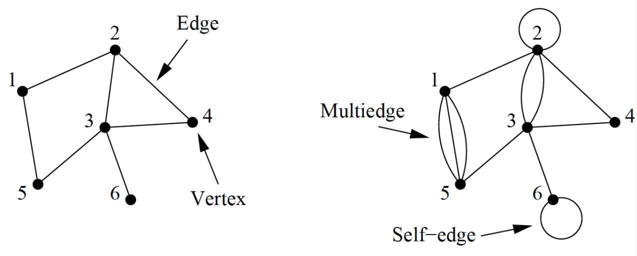


Figure 2: Traditional visualisation of two small networks...

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 1 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}$$

Figure 3: ... and the adjacency matrix of the left-hand network (Newman, 2010, p. 111).

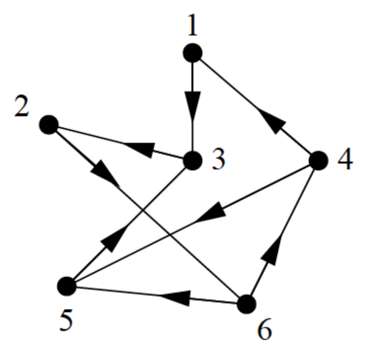


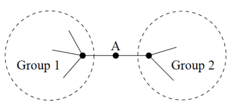
Figure 4: *A directed network...*

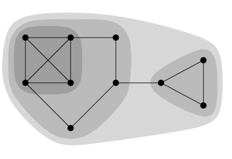
$$\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 0 \end{pmatrix}$$

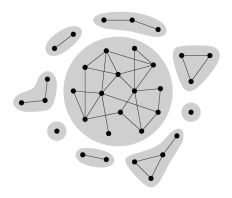
Figure 5: *... and its adjacency matrix (not symmetric!) (Newman, 2010, p. 112).*

Network measures

Degree of a vertex number of connections



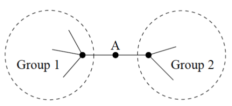


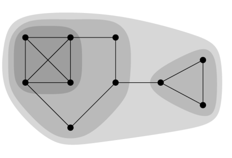


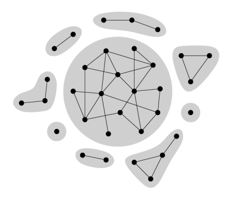
Network measures

Degree of a vertex number of connections

Authority of a vertex number of important connections





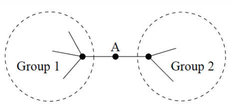


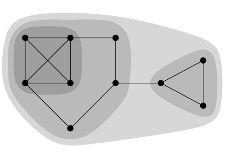
Network measures

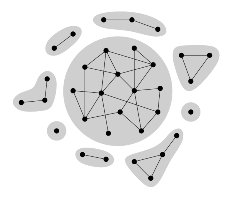
Degree of a vertex number of connections

Authority of a vertex number of important connections

Closeness of a vertex mean distance to other vertices







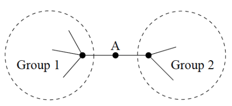
Network measures

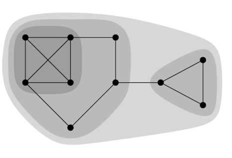
Degree of a vertex number of connections

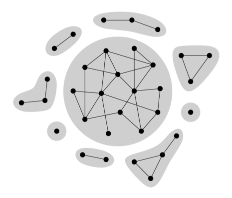
Authority of a vertex number of important connections

Closeness of a vertex mean distance to other vertices

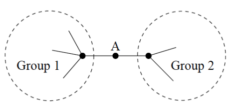
Betweenness of a vertex extent to which a vertex lies on paths between other vertices







Network measures



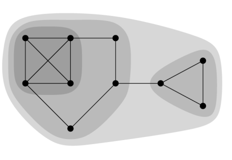
Degree of a vertex number of connections

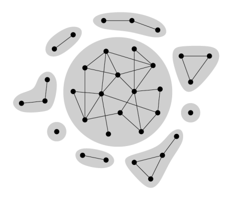
Authority of a vertex number of important connections

Closeness of a vertex mean distance to other vertices

Betweenness of a vertex extent to which a vertex lies on paths between other vertices

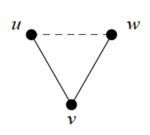
Group of vertices

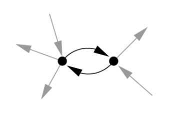


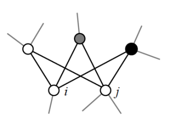


Network measures

Transitivity of edges Alice *friend of* Bob *friend of* Cat *friend of* Alice



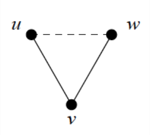


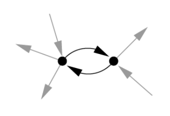


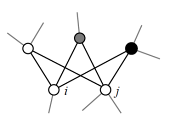
Network measures

Transitivity of edges Alice *friend of* Bob *friend of* Cat *friend of* Alice

Reciprocity of edges Alice *friend of* Bob *friend of* Alice





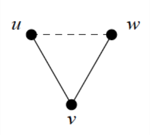


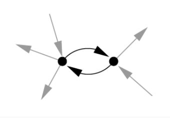
Network measures

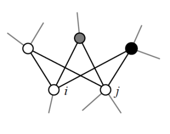
Transitivity of edges Alice *friend of* Bob *friend of* Cat *friend of* Alice

Reciprocity of edges Alice *friend of* Bob *friend of* Alice

Similarity of vertices extent to which the *neighbourhood* of vertices is similar







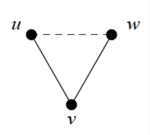
Network measures

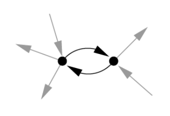
Transitivity of edges Alice *friend of* Bob *friend of* Cat *friend of* Alice

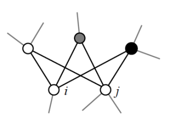
Reciprocity of edges Alice *friend of* Bob *friend of* Alice

Similarity of vertices extent to which the *neighbourhood* of vertices is similar

Homophily of vertices tendency to associate with similar vertices







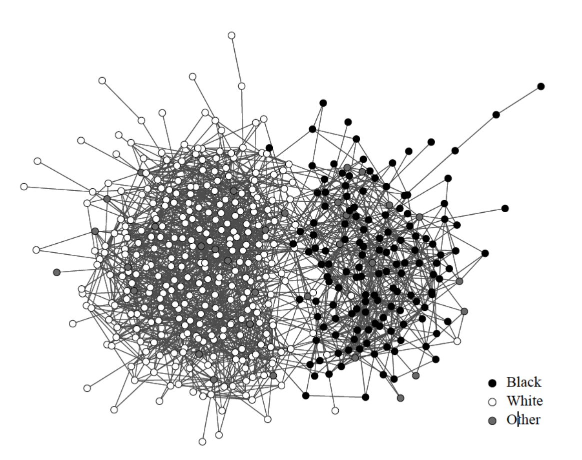


Figure 6: *Friendship network at a US high school (Newman, 2010, p. 221).*

Community detection

The goal of a community detection algorithm is simply to separate nodes into groups that have only a few edges *between* them and many edges *within*.

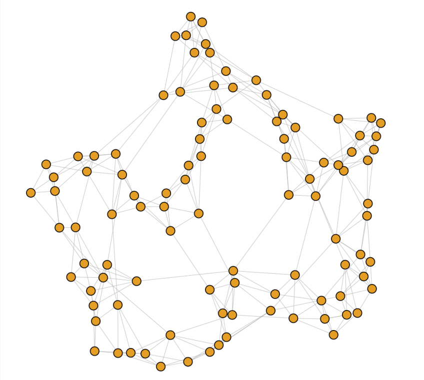
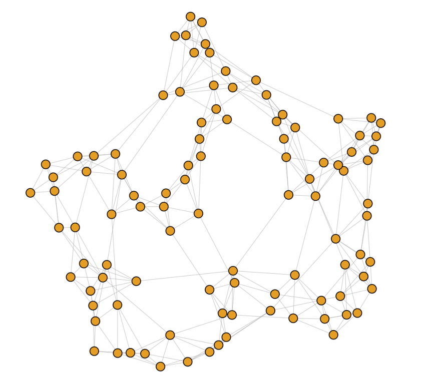


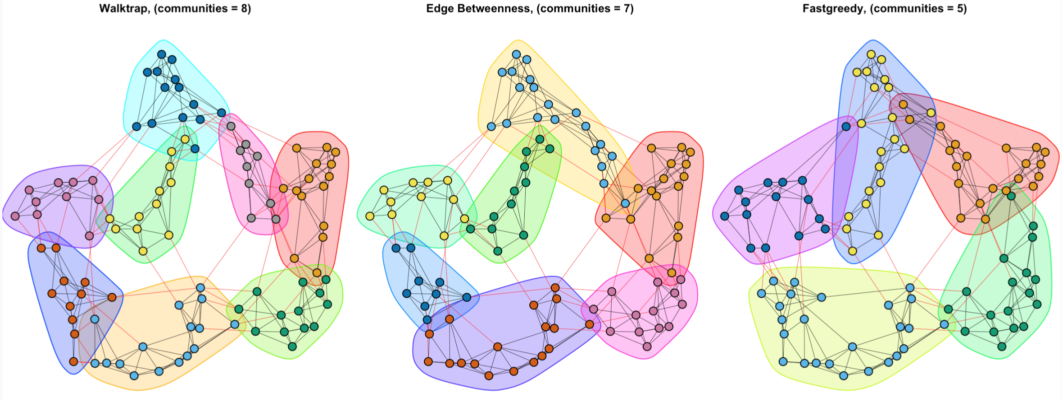
Figure 7: *A randomly generated network with 100 vertices and 300 edges*

Community detection

How many communities do you see in this network?



Community detection



INFORMATION, COMMUNICATION & SOCIETY, 2017
VOL. 20, NO. 11, 1660–1679
http://dx.doi.org/10.1080/1369118X.2016.1252410

Routledge
Taylor & Francis Group

Check for updates

Hybrid social and news media protest events: from #MarchinMarch to #BusttheBudget in Australia*

Francesco Bailo and Ariadne Vromen

Department of Government and International Relations, University of Sydney, Sydney, Australia

ABSTRACT
Public protest events are now both social media and news media events. They are deeply entangled, with news media actors – such as journalists or news organisations – directly participating in the protest by tweeting about the event using the protest hashtag; and social media actors sharing news items published online by professional news agencies. Protesters have always deployed tactics to engage the media and use news media agencies' resources to amplify their reach, with the dual aim of mobilising new supporters and adding their voice to public, mediatised debate. When protest moves between a physical space and a

ARTICLE HISTORY
Received 1 December 2015
Accepted 11 October 2016

KEYWORDS
Social movements; social media; news; social networking

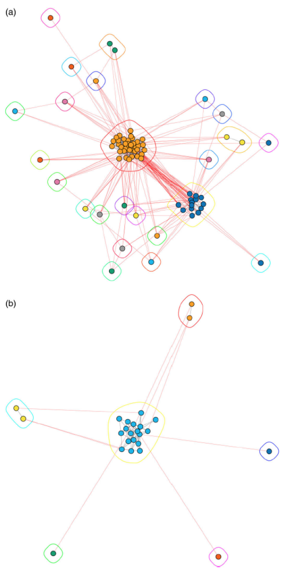
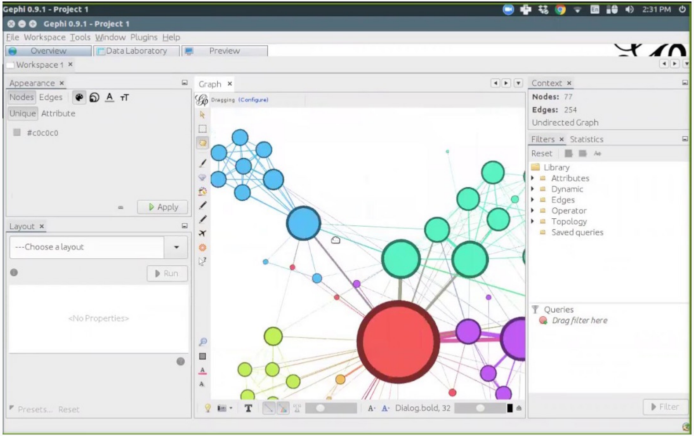
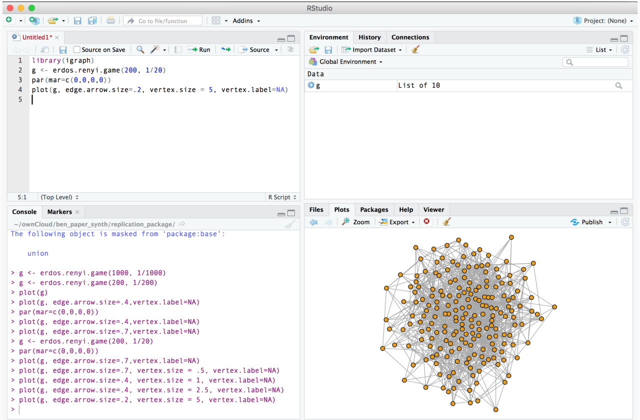


Figure 4. Mutual relations among friends and communities of a news media account (a) and regular users (b).

Easy, small n: Gephi (gephi.org)



Hard, big n igraph package (igraph.org) in R (www.r-project.org) or Python (www.python.org)



Getting started bibliography:

Easy Scott, *What is social network analysis?*, 2012

Important Marin and Wellman, 'Social network analysis: An introduction', 2011

Hard Newman, *Networks*, 2010

Tutorials for beginners by Katherine Ognyanova (Rutgers University):

- Network visualisation with Gephi (kateto.net/sunbelt2016)
- Network visualization with R (kateto.net/network-visualization)
- Network Analysis and Visualization with R and igraph (kateto.net/networks-r-igraph)

Text analysis

Quantitative text analysis is necessary when the manual coding of documents is not feasible or acceptable.

When you face a large **corpus** of **documents**, you might want some methods to automatically:

1. Find patterns within the documents,
2. Compare (and maybe group) documents.

Finding patterns

A textual pattern is as simple as `dog`.

- Finding patterns doesn't involve any statistical analysis.
- But you might need to use regular expressions (a.k.a. 'regex') if your pattern is complex.

Finding patterns

Let's say that you want to find in your corpus all the instances of `dog` and `cat`.

You want to find 'I have two dogs and a cat' or 'Cats are felines'

(link to interactive example)

Finding patterns

Let's say that you want to find in your corpus all the instances of `dog` and `cat`.

You want to find 'I have two dogs and a cat' or 'Cats are felines'

But you don't want to find 'the categorization of syntactic categories'

(link to interactive example)

Finding patterns

Let's say that you want to find in your corpus all the instances of `dog` and `cat`.

You want to find 'I have two dogs and a cat' or 'Cats are felines'

But you don't want to find 'the categorization of syntactic categories'

You need a regular expression like: `\b(cats?|dogs?)\b`

(link to interactive example)

Finding patterns

A few simple regex topics:

Quantifier `?`

Exercise: Go to `regexr.com/3os9b` (not with Explorer) and enter a regular expression to match 'France' but also 'French'.

Finding patterns

A few simple regex topics:

Quantifier `?`

- `abc?` matches a string that has ‘ab’ followed by zero or one ‘c’

Exercise: Go to `regexr.com/3os9b` (not with Explorer) and enter a regular expression to match ‘France’ but also ‘French’.

Finding patterns

A few simple regex topics:

Quantifier `?`

- `abc?` matches a string that has 'ab' followed by zero or one 'c'

OR operator `|`

Exercise: Go to `regexr.com/3os9b` (not with Explorer) and enter a regular expression to match 'France' but also 'French'.

Finding patterns

A few simple regex topics:

Quantifier `?`

- `abc?` matches a string that has 'ab' followed by zero or one 'c'

OR operator `|`

- `a(b|c)` matches a string that has 'a' followed by 'b' or 'c'

Exercise: Go to `regexr.com/3os9b` (not with Explorer) and enter a regular expression to match 'France' but also 'French'.

Finding patterns

A few simple regex topics:

Quantifier `?`

- `abc?` matches a string that has 'ab' followed by zero or one 'c'

OR operator `|`

- `a(b|c)` matches a string that has 'a' followed by 'b' or 'c'

Boundaries `\b`

Exercise: Go to `regexr.com/3os9b` (not with Explorer) and enter a regular expression to match 'France' but also 'French'.

Finding patterns

A few simple regex topics:

Quantifier `?`

- `abc?` matches a string that has 'ab' followed by zero or one 'c'

OR operator `|`

- `a(b|c)` matches a string that has 'a' followed by 'b' or 'c'

Boundaries `\b`

- `\babc\b` matches only a whole word

Exercise: Go to `regexr.com/3os9b` (not with Explorer) and enter a regular expression to match 'France' but also 'French'.

Comparing documents

Comparing documents involves statistical analysis and matrix algebra (while finding patterns doesn't). It usually relies on Natural-language processing (NLP), the branch of computer science that studies the human language and its interactions with the machines.

In its most primordial application, NLP treats documents as **bag-of-words** (BoW):

- The *position* of terms within the document is disregarded,
- What counts is the *frequency* of the terms.

Comparing documents

Let's see how we process documents in a common NLP application.

- We remove from the documents all the stop-words;

Comparing documents

Let's see how we process documents in a common NLP application.

- We remove from the documents all the stop-words;
- doc1 = "drugs hospitals doctors"
  doc2 = "smog pollution environment"
  doc3 = "doctors hospitals healthcare"
  doc4 = "pollution environment water"

Comparing documents

Let's see how we process documents in a common NLP application.

- We remove from the documents all the stop-words;
- doc1 = "drugs hospitals doctors"
  doc2 = "smog pollution environment"
  doc3 = "doctors hospitals healthcare"
  doc4 = "pollution environment water"
- We count the frequency of each term in each document, and we produce a term-document matrix

Comparing documents

	doc1	doc2	doc3	doc4
doctor	1	0	1	0
drug	1	0	0	0
environ	0	1	0	1
healthcar	0	0	1	0
hospit	1	0	1	0
pollut	0	1	0	1
smog	0	1	0	0
water	0	0	0	1

Table 1: *Term-document matrix. Terms were stemmed.*

Bag of Words (BoW)

- **Concept:** Text representation as a bag of its words, ignoring the order.
- **Representation:** Fixed-length vectors, counting word occurrences or indicating presence/absence.
- **Advantages:** Simple, good for specific tasks like spam detection.
- **Limitations:** Ignores context and semantics, leading to sparse, high-dimensional vectors.

Embeddings (used by the Large Language Models)

- **Concept:** Dense, low-dimensional vectors representing words, capturing semantic meanings.
- **Representation:** Continuous vectors that reflect context and relationships between words.
- **Advantages:** Captures semantics, reduces dimensionality and is versatile for various NLP tasks.
- **Limitations:** Requires more computational resources, less intuitive.

- Nvivo (www.qsrinternational.com/nvivo)
- Regular Expression (regexr.com)
- R (www.r-project.org) or Python (www.python.org)
- ChatGPT and other Large Language Models...

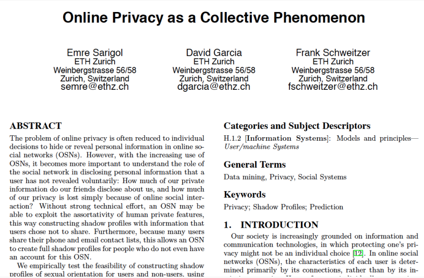
Introductory Jockers, *Text analysis with R for students of literature*, 2014

Introductory Bird et al., *Natural language processing with Python*, 2009

Hard Manning et al., *Introduction to information retrieval*, 2008

Ethics

Issues with relational data



Online Privacy as a Collective Phenomenon

Emre Sarigol
ETH Zurich
Weinbergstrasse 56/58
Zurich, Switzerland
semre@ethz.ch

David Garcia
ETH Zurich
Weinbergstrasse 56/58
Zurich, Switzerland
dgarcia@ethz.ch

Frank Schweitzer
ETH Zurich
Weinbergstrasse 56/58
Zurich, Switzerland
fschweitzer@ethz.ch

ABSTRACT

The problem of online privacy is often reduced to individual decisions to hide or reveal personal information in online social networks (OSNs). However, with the increasing use of OSNs, it becomes more important to understand the role of the social network in disclosing personal information that a user has not revealed voluntarily: How much of our private information do our friends disclose about us, and how much of our privacy is lost simply because of online social interaction? Without strong technical effort, an OSN may be able to exploit the assortativity of human private features, this way constructing shadow profiles with information that users chose not to share. Furthermore, because many users share their phone and email contact lists, this allows an OSN to create full shadow profiles for people who do not even have an account for this OSN.

We empirically test the feasibility of constructing shadow profiles of sexual orientation for users and non-users, using

Categories and Subject Descriptors

H.1.2 [Information Systems]: Models and principles—*User/machine Systems*

General Terms

Data mining, Privacy, Social Systems

Keywords

Privacy; Shadow Profiles; Prediction

1. INTRODUCTION

Our society is increasingly grounded on information and communication technologies, in which protecting one's privacy might not be an individual choice [12]. In online social networks (OSNs), the characteristics of each user is determined primarily by its connections, rather than by its in-

Figure 8: *Sarigol et al., 'Online privacy as a collective phenomenon', 2014*

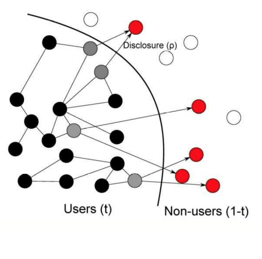


Figure 4: Schema of the full shadow profile construction problem.

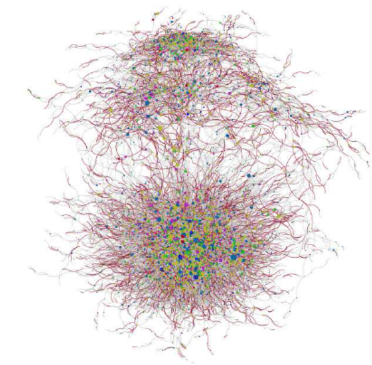


Figure 1: The network for a subset of Friendster users. The red edges represent assortativity, where the endpoint nodes are in the same sexual orientation class. The node colors correspond to the sexual orientation class.

Ethics in the digital age: Open issues

- Public and Private space. What about online fora (e.g. Facebook public pages?)
- Informed consent.
- Right to privacy. But who owns the data?

My research on social media

- Kong, Q., Booth, E., Bailo, F., Johns, A., & Rizoiu, M.-A. (2022).**Slipping to the extreme: A mixed method to explain how extreme opinions infiltrate online discussions.** *Proceedings of the International AAAI Conference on Web and Social Media*, *16*(1), 524–535. https://doi.org/10.1609/icwsm.v16i1.19312
- Bailo, F., Johns, A., & Rizoiu, M.-A. (2024).**Riding information crises: The performance of far-right Twitter users in Australia during the 2019–2020 bushfires and the COVID-19 pandemic.** *Information, Communication & Society*, 27(2), 278–296. https://doi.org/10.1080/1369118X.2023.2205479
- Johns, A., Bailo, F., Booth, E., & Rizoiu, M.-A. (2024).**Labelling, shadow bans and community resistance: Did Meta's strategy to suppress rather than remove COVID misinformation and conspiracy theory on Facebook slow the spread?** *Media International Australia*. https://doi.org/10.1177/1329878X241236984

Slipping to the extreme

Kong, Q., Booth, E., Bailo, F., Johns, A., & Rizoiu, M.-A. (2022).**Slipping to the extreme: A mixed method to explain how extreme opinions infiltrate online discussions.** *Proceedings of the International AAAI Conference on Web and Social Media*, *16*(1), 524–535. https://doi.org/10.1609/icwsm.v16i1.19312

- This study addresses the infiltration of extreme opinions in online discussions, leveraging machine learning algorithms alongside qualitative research methods.
- Aims to bridge the gap between depth of qualitative insights and breadth of quantitative analysis in understanding problematic online speech.

Methodology

- Initial qualitative study constructs an ontology of problematic speech, identifying key themes and opinions in social media posts.
- Large-scale data collection from Facebook, Twitter, and YouTube, followed by iterative dataset augmentation using a human-in-the-loop approach.
- Machine learning models classify and augment the dataset, expanding the initial qualitative study's findings.

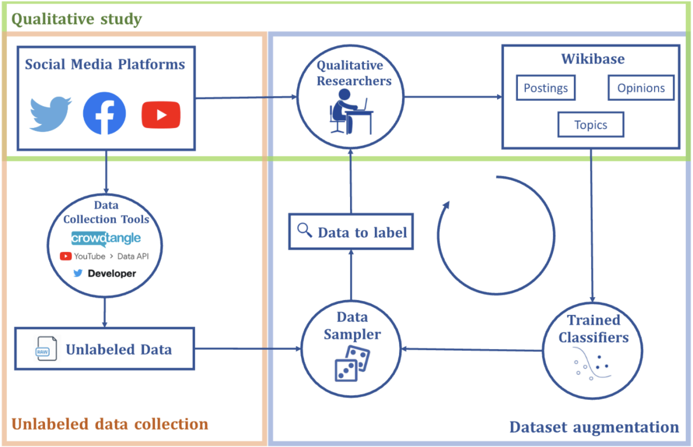


Figure 1: The pipeline of machine learning accelerated qualitative research where the human-in-the-loop machine learning algorithms are employed for dataset augmentation.

Findings

- The mixed-method approach successfully identifies and expands the dataset, revealing detailed case studies of problematic speech dynamics in specific online communities.
- Analysis of opinion emergence and co-occurrence suggests pathways through which extreme opinions enter mainstream online discourse.

Labelling, shadow bans and community resistance

Johns, A., Bailo, F., Booth, E., & Rizoiu, M.-A. (2024).**Labelling, shadow bans and community resistance: Did Meta's strategy to suppress rather than remove COVID misinformation and conspiracy theory on Facebook slow the spread?** *Media International Australia*. https://doi.org/10.1177/1329878X241236984

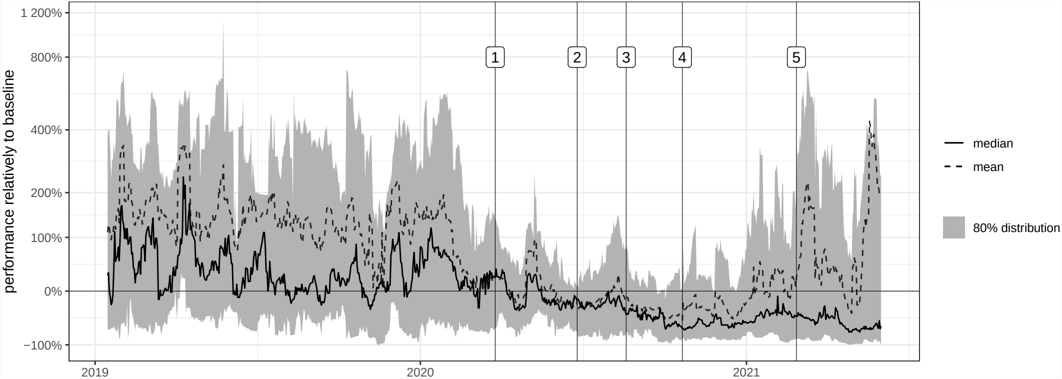
- Focus on the effectiveness of Meta's content moderation on Facebook during COVID-19.
- Analysis of 18 Australian right-wing/anti-vaccination pages between January 2019 and July 2021.
- Integration of engagement metrics, time series analysis, and content analysis.

Sampling

- Utilised CrowdTangle to collect data from 21 identified Australian Facebook public accounts.
- Final analysis included 18 accounts after excluding those with less than 1% of relevant posts.
- A total of 34,202 postings were analysed.

Data Analysis

- Performance analysis via CrowdTangle's 'overperforming score' and average number of shares-per-post.
- Content and thematic analysis on comments from two overperforming public pages.
- Latent Dirichlet Allocation (LDA) for topic modelling and exploration.


performance relatively to baseline
1 200%
800%
400%
200%
100%
0%
−100%
2019
2020
2021
1
2
3
4
5
median
mean
80% distribution

Policy Review

- Examination of Meta's content moderation and recommendation policy announcements.
- Focus on policies introduced from January 2020 to July 2021 relevant to Facebook.
- Analysis juxtaposed against key policy announcements and page performance.

Key Findings

- Meta's content moderation systems showed partial effectiveness.
- Identified trends in content labelling and 'shadow banning' resistance by communities.
- Highlighted the importance and challenges of transparent and consistent moderation policies.

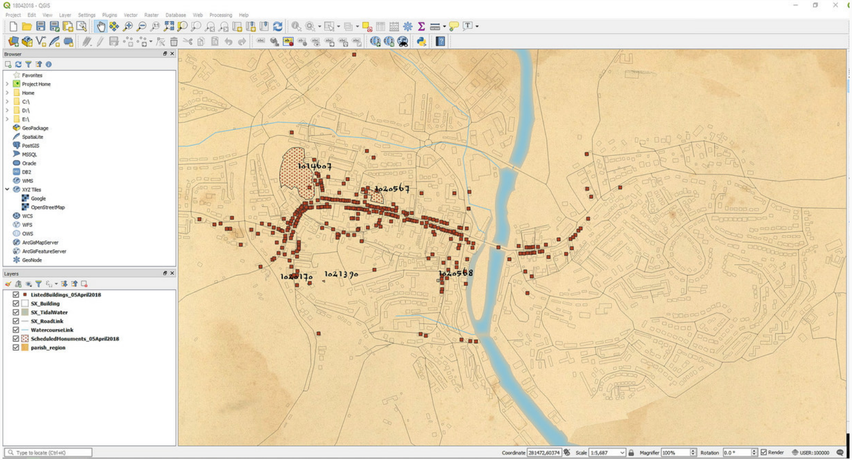
Bonus: Spatial analysis

Spatial analysis

Redrawing of John Snow's map of cases of cholera during the London outbreak of 1854 (Tufte, 2001, p. 24)

Tool for spatial analysis

- QGIS (qgis.org)



References

Bailo, F., Johns, A., & Rizoiu, M.-A. (2024).**Riding information crises: The performance of far-right Twitter users in Australia during the 2019–2020 bushfires and the COVID-19 pandemic.** *Information, Communication & Society*, 27(2), 278–296. https://doi.org/10.1080/1369118X.2023.2205479

Bailo, F., & Vromen, A. (2017).**Hybrid social and news media protest events: From #MarchinMarch to #BusttheBudget in Australia.** *Information, Communication & Society*, *20*(11), 1660–1679. https://doi.org/10.1080/1369118X.2016.1252410

Bertini, G. (1858). **Galileo Galilei showing the Doge of Venice how to use the telescope.** Retrieved May 1, 2018, from https://commons.wikimedia.org/wiki/File:Bertini_fresco_of_Galileo_Galilei_and_Doge_of_Venice.jpg#/media/File:Bertini_fresco_of_Galileo_Galilei_and_Doge_of_Venice.jpg

Bird, S., Klein, E., & Loper, E. (2009, July). ***Natural language processing with Python.*** O'Reilly Media.

Jockers, M. L. (2014). ***Text analysis with R for students of literature.*** Springer.

Johns, A., Bailo, F., Booth, E., & Rizoiu, M.-A. (2024).**Labelling, shadow bans and community resistance: Did Meta's strategy to suppress rather than remove COVID misinformation and conspiracy theory on Facebook slow the spread?** *Media International Australia.* https://doi.org/10.1177/1329878X241236984

- Kong, Q., Booth, E., Bailo, F., Johns, A., & Rizoiu, M.-A. (2022).**Slipping to the extreme: A mixed method to explain how extreme opinions infiltrate online discussions.** *Proceedings of the International AAAI Conference on Web and Social Media*, *16*(1), 524–535. https://doi.org/10.1609/icwsm.v16i1.19312
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). ***Introduction to information retrieval.*** Cambridge University Press.
- Marin, A., & Wellman, B. (2011). **Social network analysis: An introduction.** In J. Scott & P. J. Carrington (Eds.), *The Sage handbook of social network analysis* (pp. 11–25). SAGE Publications.
- Newman, M. E. J. (2010). ***Networks: An introduction.*** Oxford University Press.
- Salganik, M. J. (2018). ***Bit by bit: Social research in the digital age* [OCLC: 1012406622].** Princeton University Press.
- Sarigol, E., Garcia, D., & Schweitzer, F. (2014).**Online privacy as a collective phenomenon.** *arXiv preprint arXiv:1409.6197*. Retrieved October 6, 2014, from http://arxiv.org/abs/1409.6197

Scott, J. (2012). ***What is social network analysis?*** Bloomsbury Publishing.

Tufte, E. R. (2001). ***The visual display of quantitative information.*** Graphics Press.