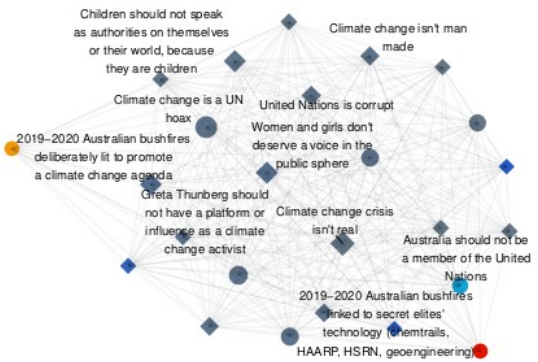
Behavioral Data Science

UTS Data Science

UTS UNIVERSITY OF TECHNOLOGY SYDNEY



Mapping Online Problematic Content: Mixing Qualitative Approaches with State-of-the-art Machine Learning

Marian-Andrei Rizoiu | Behavioural Data Science
Marian-Andrei.Rizoiu@uts.edu.au
https://www.behavioral-ds.science

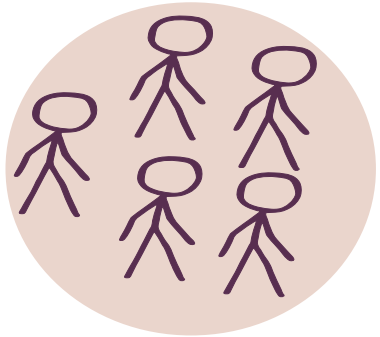
The research group

Behavioral Data Science

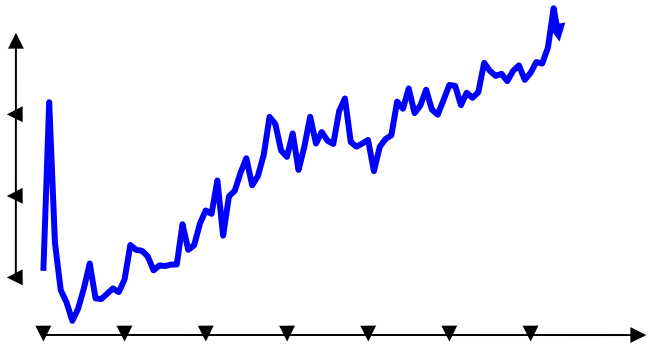
1 PostDoc, 6 PhD, 3 Masters, 1 assistant prof.

The Behavioral Data Science

1.

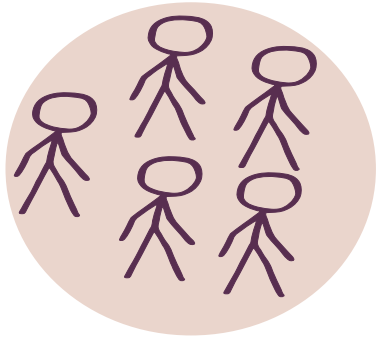


information diffusion
epidemics spreading
behavioral modeling

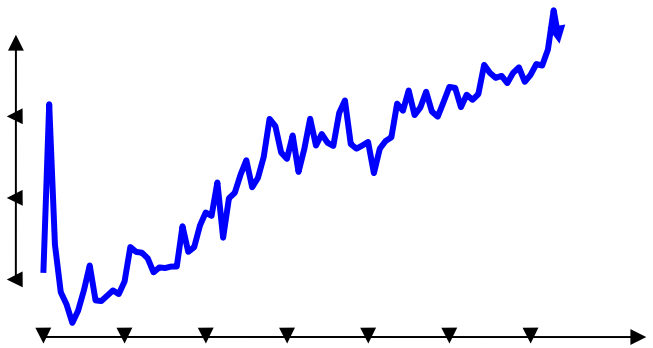


The Behavioral Data Science

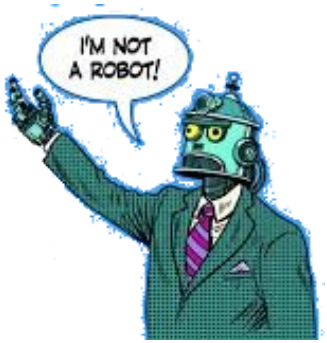
1.



information diffusion
epidemics spreading
behavioral modeling



2.

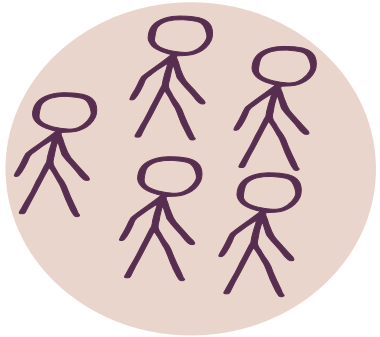


[Rizoiu et al ICWSM'18]

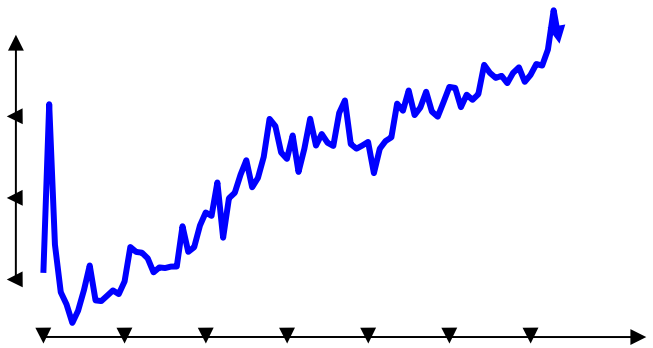
[Kim et al Journ.Comp.SocSci'19]

The Behavioral Data Science

1.

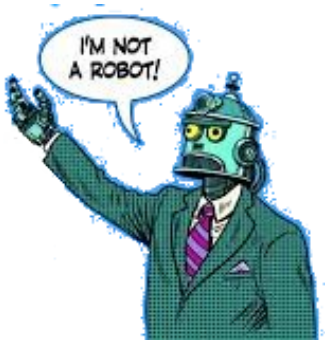


information diffusion
epidemics spreading
behavioral modeling



3.

2.



[Rizoiu et al ICWSM'18]

[Kim et al Journ.Comp.SocSci'19]

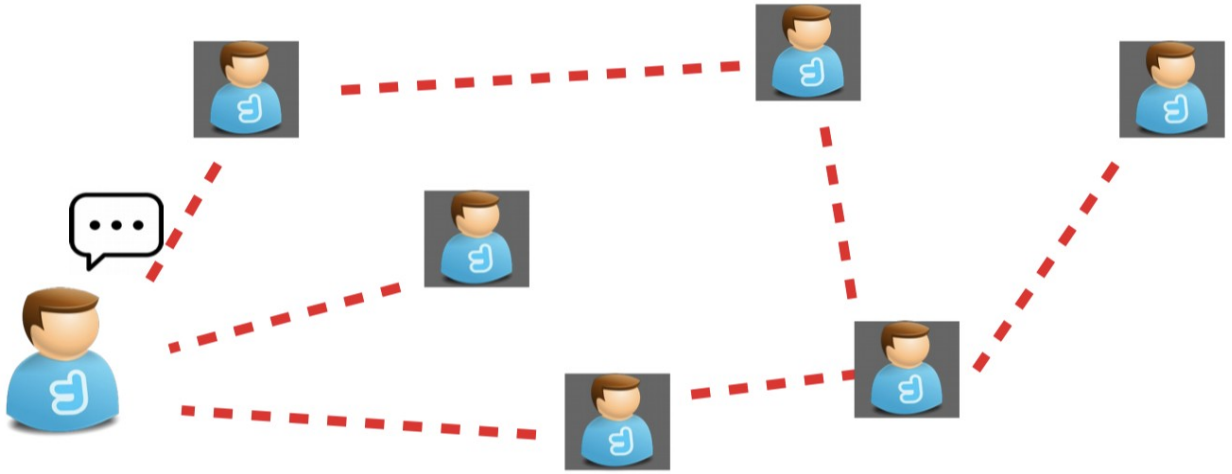
Presentation plan

1. What are the state-of-the-art computational approaches in your field, and what fundamental social science questions do they allow to answer?

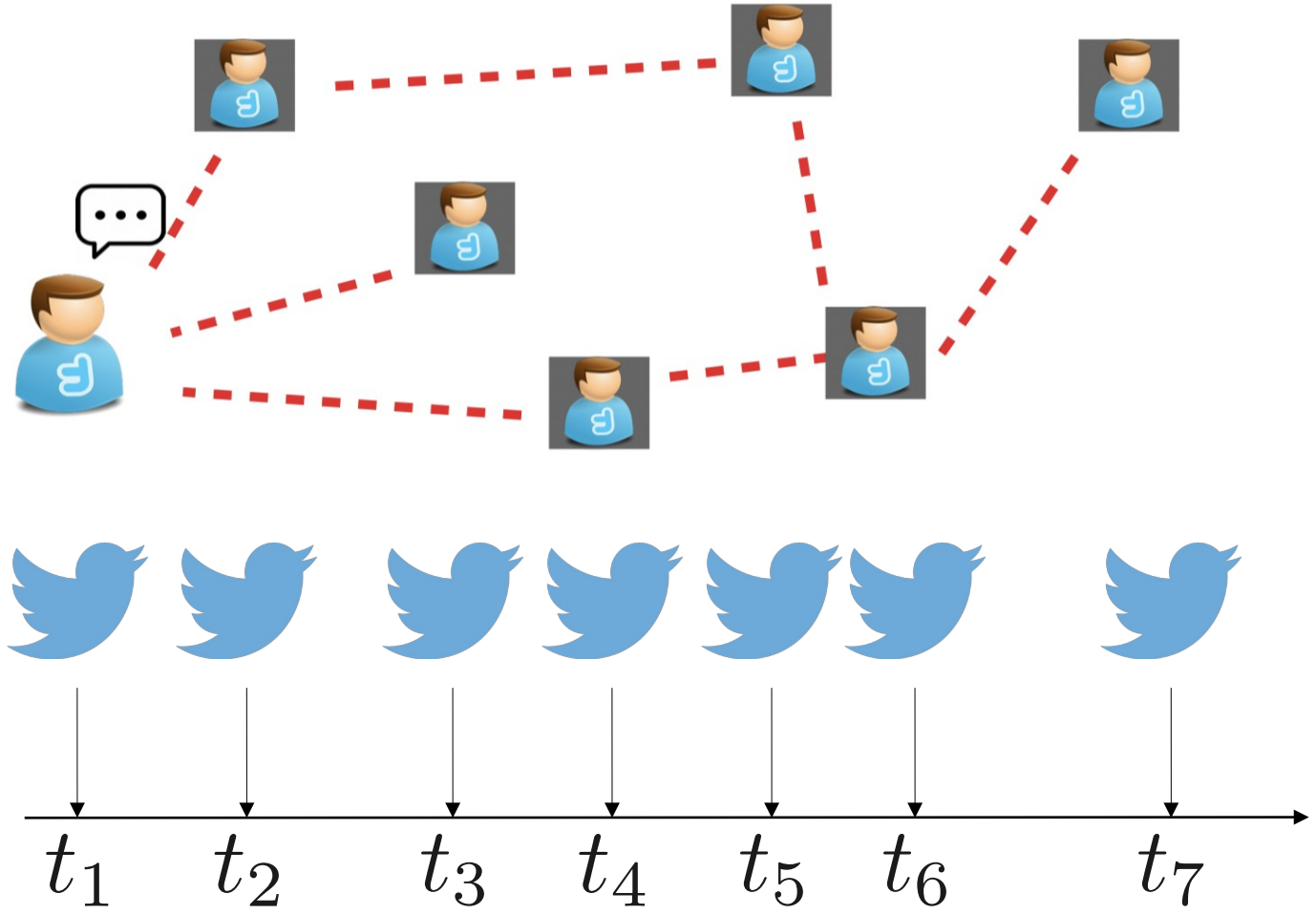
2. Where are the main challenges to developing computational social science capabilities in Australia, from your perspective?

3. What is the best way forward as a field in the current research funding landscape? Should we be unifying or diversifying our research efforts? How do you see the future of interdisciplinary collaborations?

Reshare cascades in online social platforms



Reshare cascades in online social platforms



Reshare cascades: a collection of time stamped reshare events (retweets) of an online post (tweet).

Hawkes process – definition

$$\lambda(t|\mathcal{H}_t) = \mu(t) + \sum_{i:t>T_i} \phi(t - T_i)$$

base intensity (exogenous)

self-excitation (endogenous)

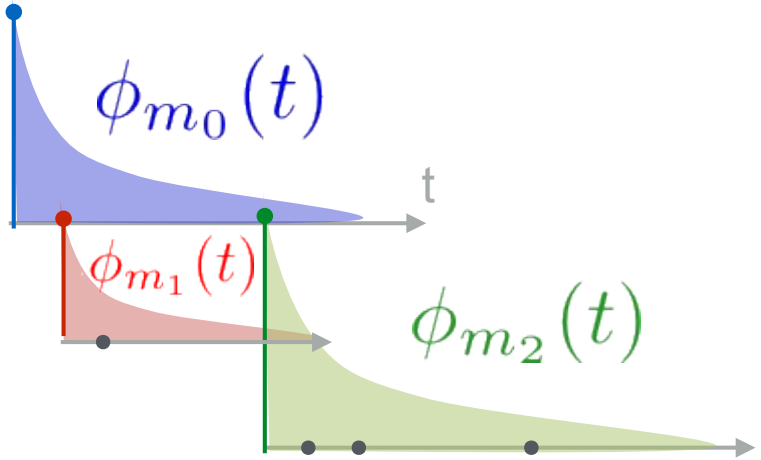
$$\phi_m(t - T_i) = \kappa m^{\beta} (t - T_i)^{-(1+\theta)}$$

the rate of 'daughter' events

content virality

user influence

memory



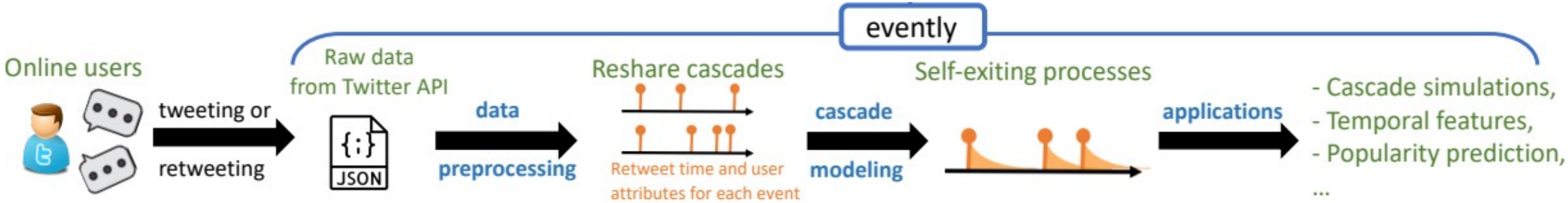
S. Mishra, M.-A. Rizoiu, and L. Xie, "Feature Driven and Point Process Approaches for Popularity Prediction, " in Proc. International Conference on Information and Knowledge Management (CIKM '16), Indianapolis, USA, p. 1069-1078, 2016.

evently – a point-process toolbox

Maintainer: Quyu Kong
quyu.kong@uts.edu.au
https://github.com/behavioral-ds/evently

Designed with an emphasis on online information diffusion modeling.

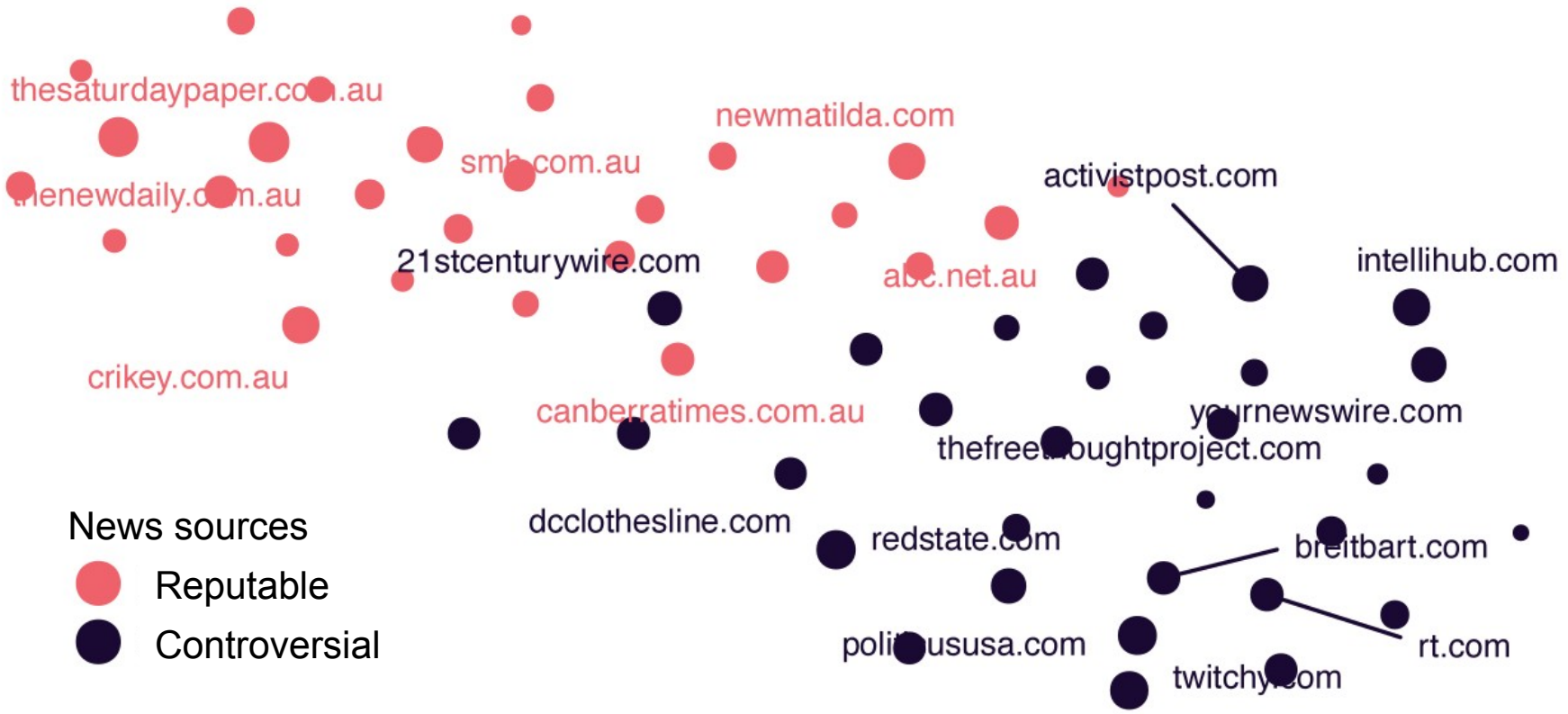
- Data processing tools for Twitter
- Popularity measures on online information diffusions
- Diffusion embeddings for online users and online content



Kong, Q., Ram, R., & Rizoiu, M.-A. (2021). evently: A Toolkit for Analyzing Online Users via Reshare Cascade Modeling. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining (pp. 1097–1100). New York, NY, USA: ACM. https://doi.org/10.1145/3437963.3441708

(1) Separating controversial from reputable

[Kong et al 2020]



Reputable and controversial sources are separable based solely on how their information spreads

Detect controversial news without content analysis

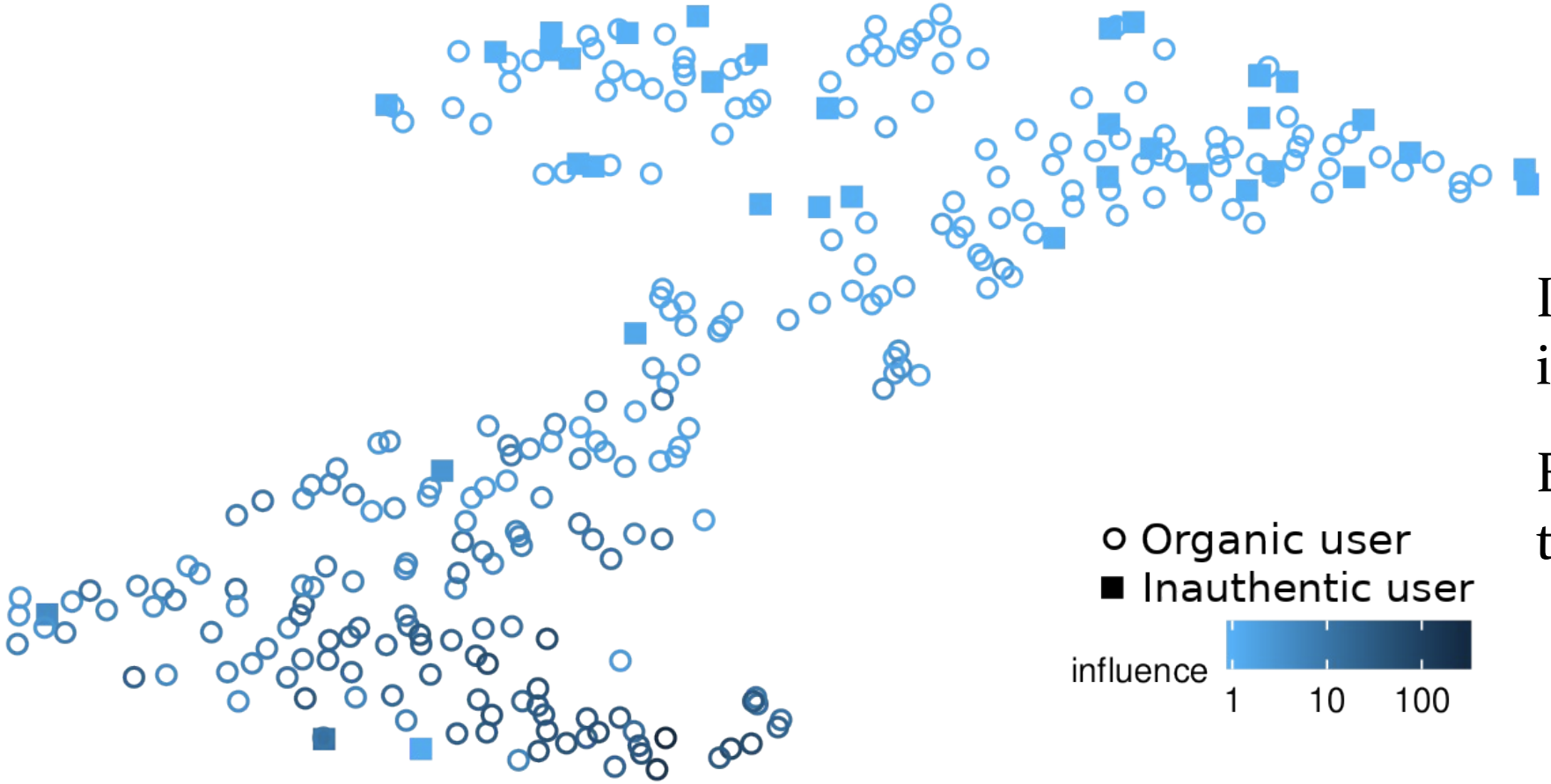
birdspotter

Maintainer: Rohit Ram
rohit.ram@uts.edu.au
https://github.com/behavioral-ds/BirdSpotter

- An **easy-to-use** Twitter user analysis tool
  - Designed for political scientists, sociologists and, data-practitioners
  - Models and labels the attributes of Twitter users
  - Prepackaged a state-of-the-art offline bot detector
  - Quantifies influence via a Tweet Dynamics System
- The **intuitive visualization birdspotter.ml** for dataset exploration and narrative construction.
- A **versatile Twitter user classifier,** trainable.

Ram, R., Kong, Q., & Rizoiu, M.-A. (2021). Birdspotter: A Tool for Analyzing and Labeling Twitter Users. In Proceedings of the 14th ACM International Conference on Web Search and Data Mining (pp. 918–921). New York, NY, USA: ACM.

(2) Identify influential inauthentic users



Identify users engaged in influence operations

Estimate their impact on the wider community

Presentation plan

1. What are the state-of-the-art computational approaches in your field, and what fundamental social science questions do they allow to answer?

2. Where are the main challenges to developing computational social science capabilities in Australia, from your perspective?

3. What is the best way forward as a field in the current research funding landscape? Should we be unifying or diversifying our research efforts? How do you see the future of interdisciplinary collaborations?

Main challenges for computational social science capabilities in Australia

Funding – nowhere near the amount of funding available in the US

Previously, stakeholders simply reached out to acquire the US/Europe/Japan solution without using local talent – changing in recent years

Working interdisciplinary – research is siloed into Schools, and Faculties, and FOR codes

Collaboration of the online social platforms – no data is given, no support, not even acknowledgment that CSS exists

Presentation plan

1. What are the state-of-the-art computational approaches in your field, and what fundamental social science questions do they allow to answer?

2. Where are the main challenges to developing computational social science capabilities in Australia, from your perspective?

3. What is the best way forward as a field in the current research funding landscape? Should we be unifying or diversifying our research efforts? How do you see the future of interdisciplinary collaborations?

The way forward?

Funding dedicated to interdisciplinary research – cross-/inter-disciplinary seed grants

Adopt a problem-oriented approach, rather than a solution-oriented approach – **what can we solve together?**, rather than **what can my tool solve?**

Who do I know who has this set of skills that I don't, but I need for this project?

Create discussion venues, without getting bogged into yet another committees – workshops with stated outcomes?

The future – research is a tool, a pathway; solving complex societal issues will likely require knowledge across fields; center projects on real requirements (preferably with funding)