Tweeting about exams
Social Media Discussion of British School Exams

Tom Sutch, Nicole Klir & James Keirstead

Section One

Background

Cambridge Assessment & British school exams

- GCSEs and A Levels in Britain
  - 16 - 18 year olds
  - Mainly taken over May-June
  - Increasing media coverage of exam-related tweets

- Cambridge Assessment
  - An international exams group
  - 8 million learners
  - Over 170 countries
  - 3 exam boards

- Research questions
  - To establish the extent of exam related tweeting
  - To determine patterns over time
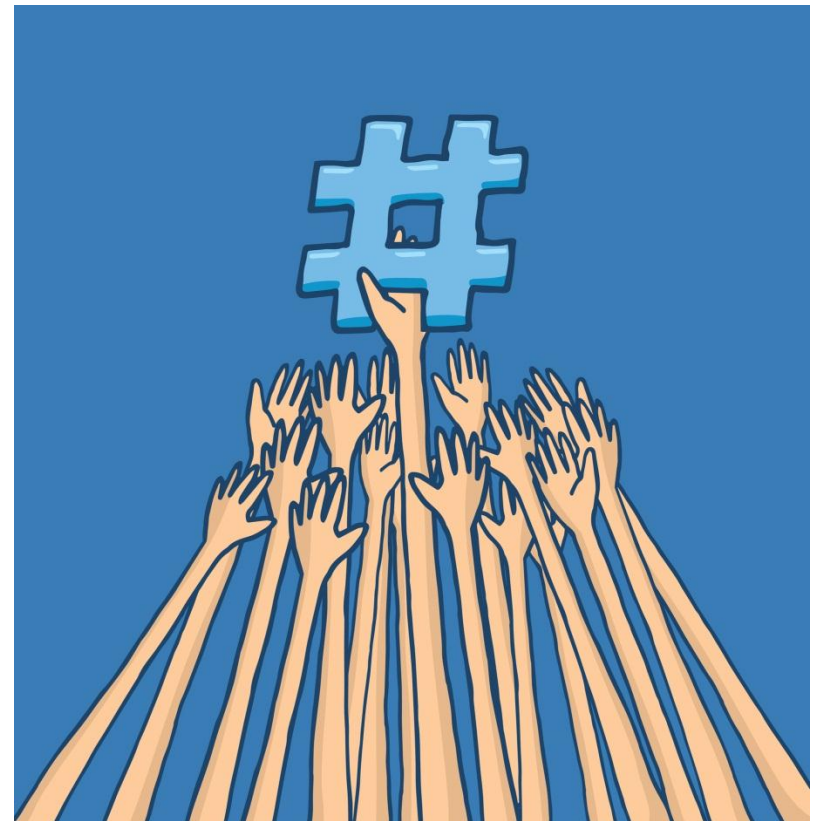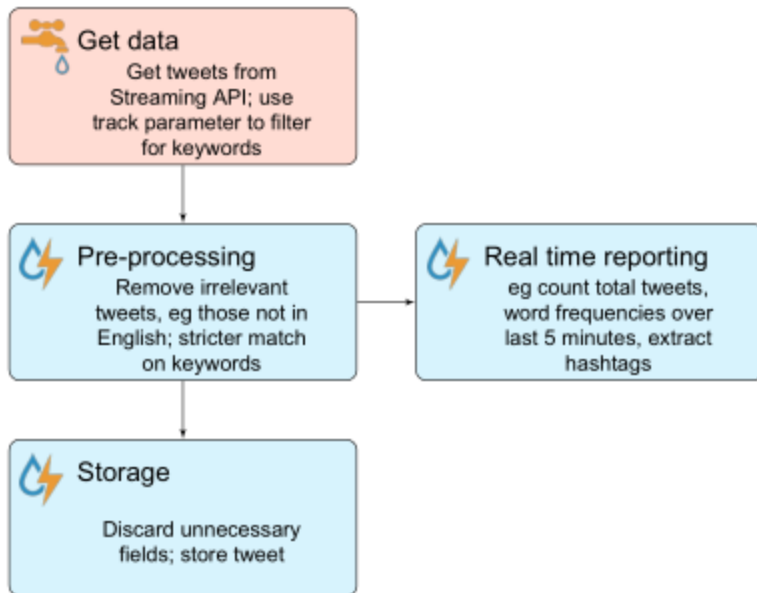 - To investigate the topics being discussed and sentiments expressed

Cambridge Assessment

Section Two

Data

Data

Collected from 14 May to 14 July 2016
- Exam session plus an additional period to give context

Get data
Get tweets from Streaming API; use track parameter to filter for keywords

Pre-processing
Remove irrelevant tweets, eg those not in English; stricter match on keywords

Real time reporting
eg count total tweets, word frequencies over last 5 minutes, extract hashtags

Storage
Discard unnecessary fields; store tweet

Twitter streaming API + Apache Storm
→ stored data for later processing
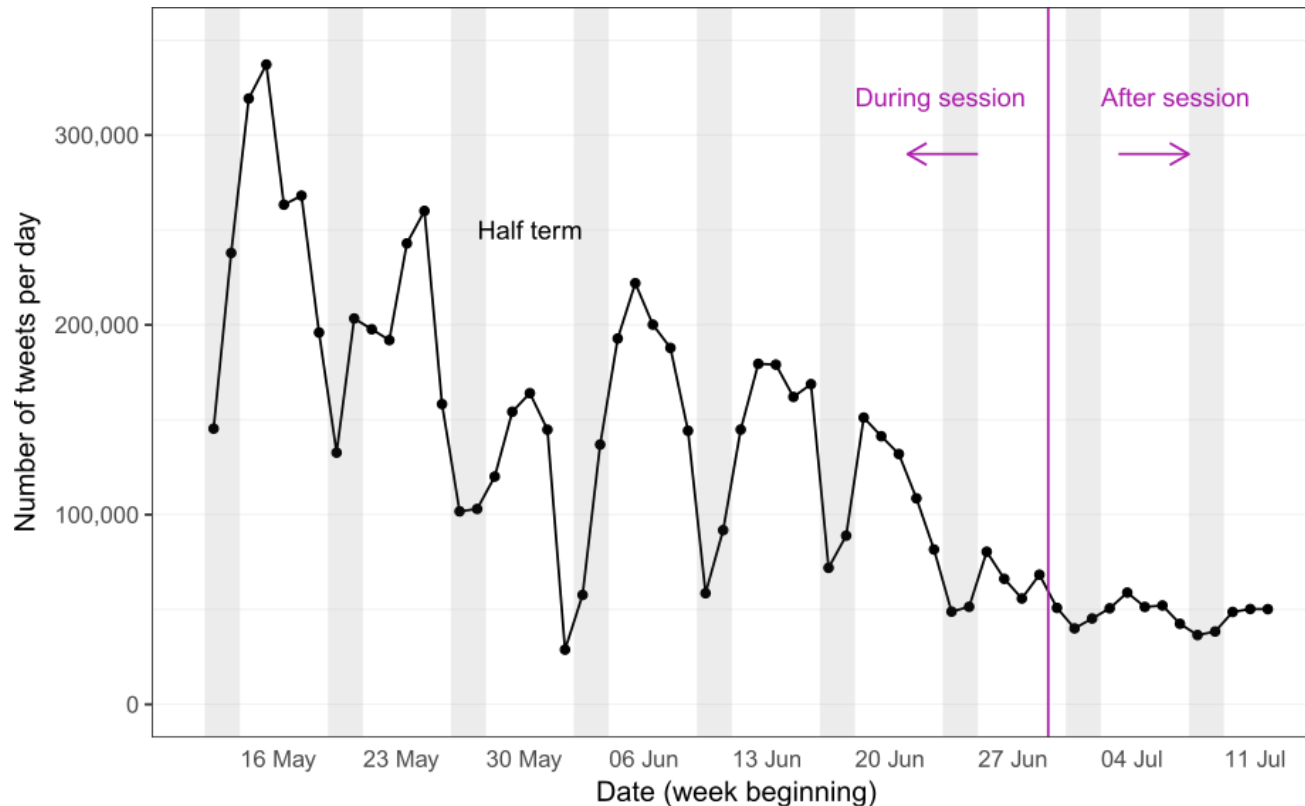
Cambridge Assessment

Section Three

Results

Twitter activity over time

High-level overview:

- 6.44 million tweets collected in total

- Volume declines over the session

- Weekly trend seen within the session

Cambridge Assessment

Twitter activity over time

Weekly trends:

- Higher numbers of tweets before and after exams

- Peaks the evening before an exam

- Fewer tweets over the weekend

- In-session tweets vs. background noise post-session

Cambridge Assessment
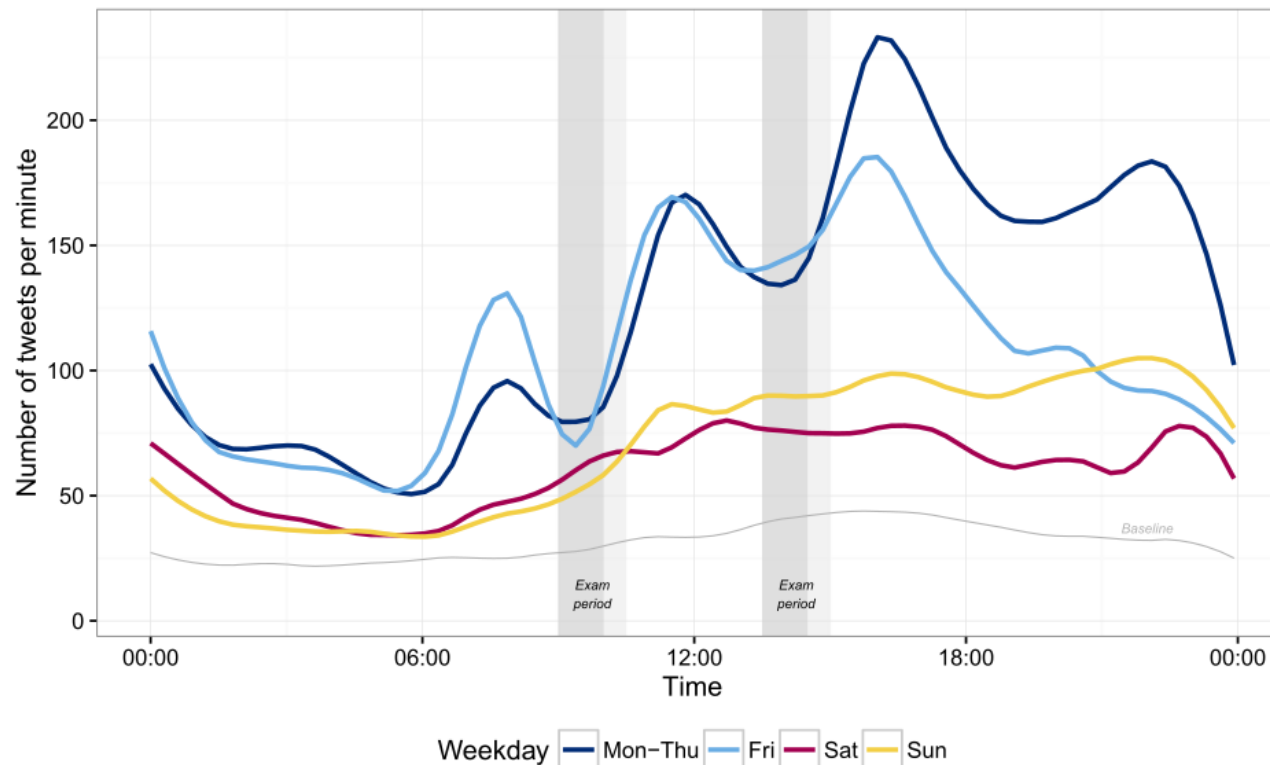
Tweet content - over the exam period

Overview of the corpus of 6.4 million tweets:

- Word frequencies: Top 20 most frequently occurring words included the words **tomorrow**, **luck, day** and **final**
- Basic sentiment analysis using emoji as a proxy for sentiment
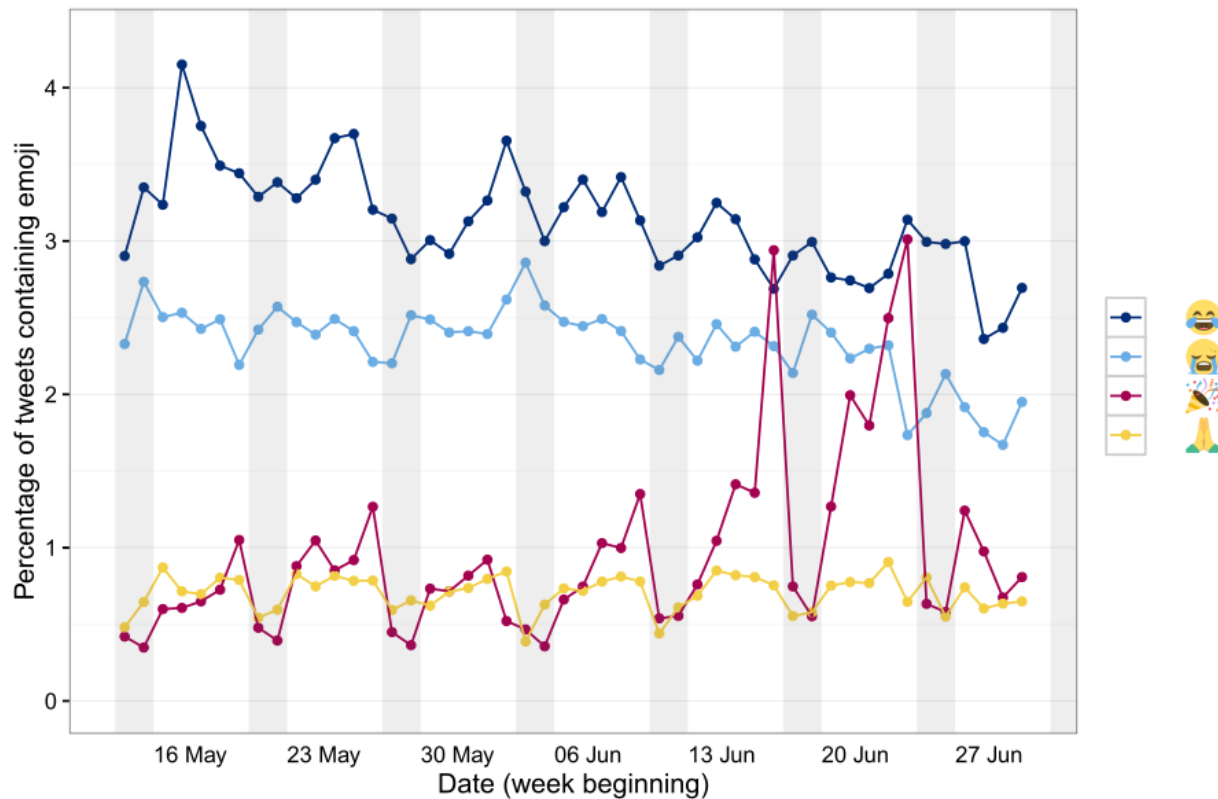
Tweet content: GCSE Mathematics

Examine one exam for content and sentiment:

- Filtered dataset for tweets sent on 25-26 May and keyword *maths*
   - Word and emoji counts show a clear pattern in tweet content
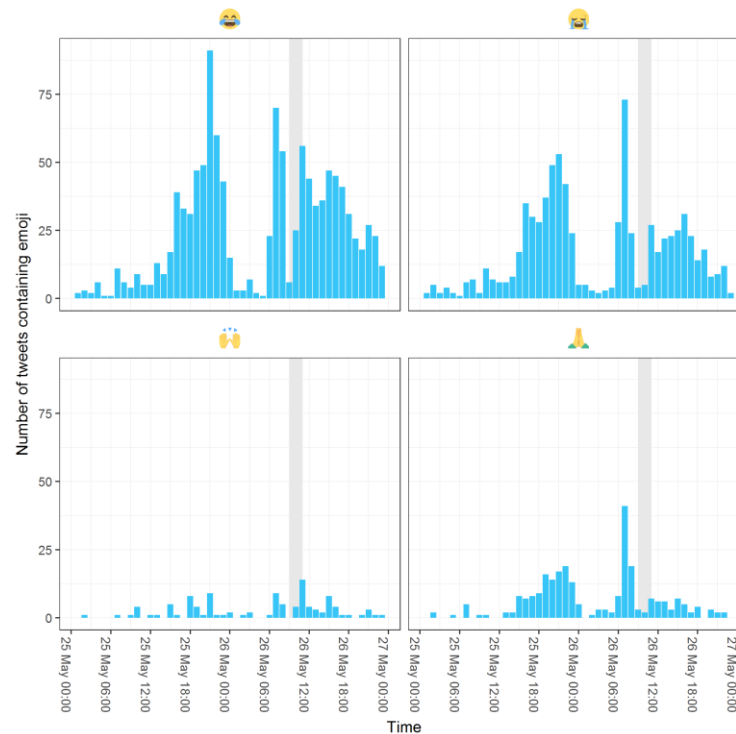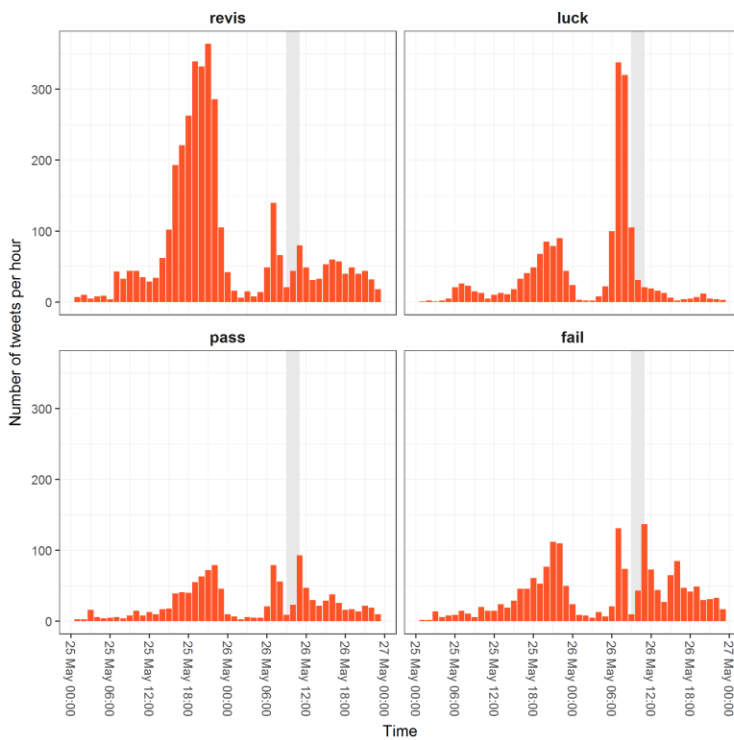
Section Four

Discussion

Discussion & further work

- Tweet volumes decline over the exam period

- Students more likely to tweet about particular exams the day before and just after the exam

- Phases in tweeting:
  - Revision : luck : post-exam analysis
  - Mirrored by use of emoji

- We're not looking to minimise or influence this discussion

Further work:

- Examine tweet content in more detail
  - Link responses to particular question types or contexts
  - Image/video analysis

Cambridge Assessment
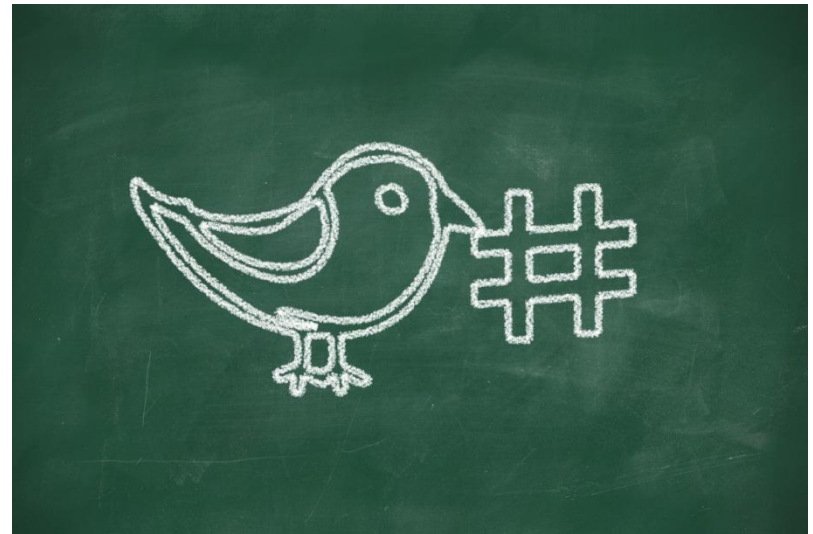
Limitations and concerns

Ethics:
- Young people in a stressful situation
- May have different understandings of privacy when using social media
- May not feel consented to research
- Potential traceability and loss of confidentiality
- Detection of illicit/reportable activities

We discarded unnecessary data (e.g. real name and location) before storage and carried out high-level numerical analysis.

Representativeness:
- *Sample:* Twitter users unlikely to be a random subset of candidates
- *Bias:* Candidates more likely to tweet if they have strong opinions or may not be expressing their true feelings

We cannot draw inferences beyond the data we have, however Twitter data is a useful source as it shapes media coverage of exams and therefore has a potential influence on public attitudes.

Cambridge Assessment

Questions

Sutch, T., and Klir, N. (2017)
Tweeting about exams: Investigating the use of social media over the summer 2016 session
Research Matters: A Cambridge Assessment publication, 23, 2–9
tinyurl.com/research-matters-23