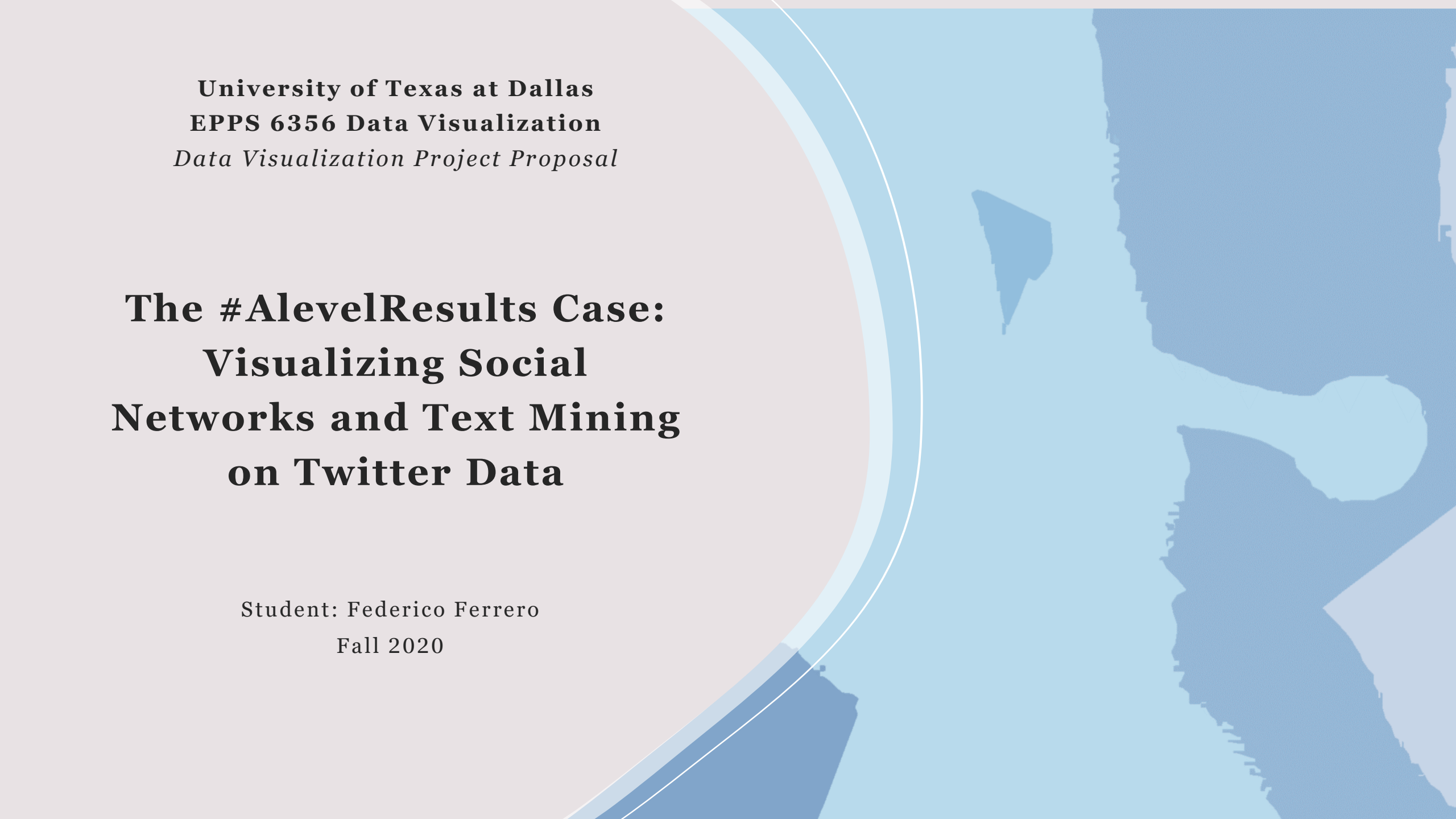
University of Texas at Dallas
EPPS 6356 Data Visualization
Data Visualization Project Proposal

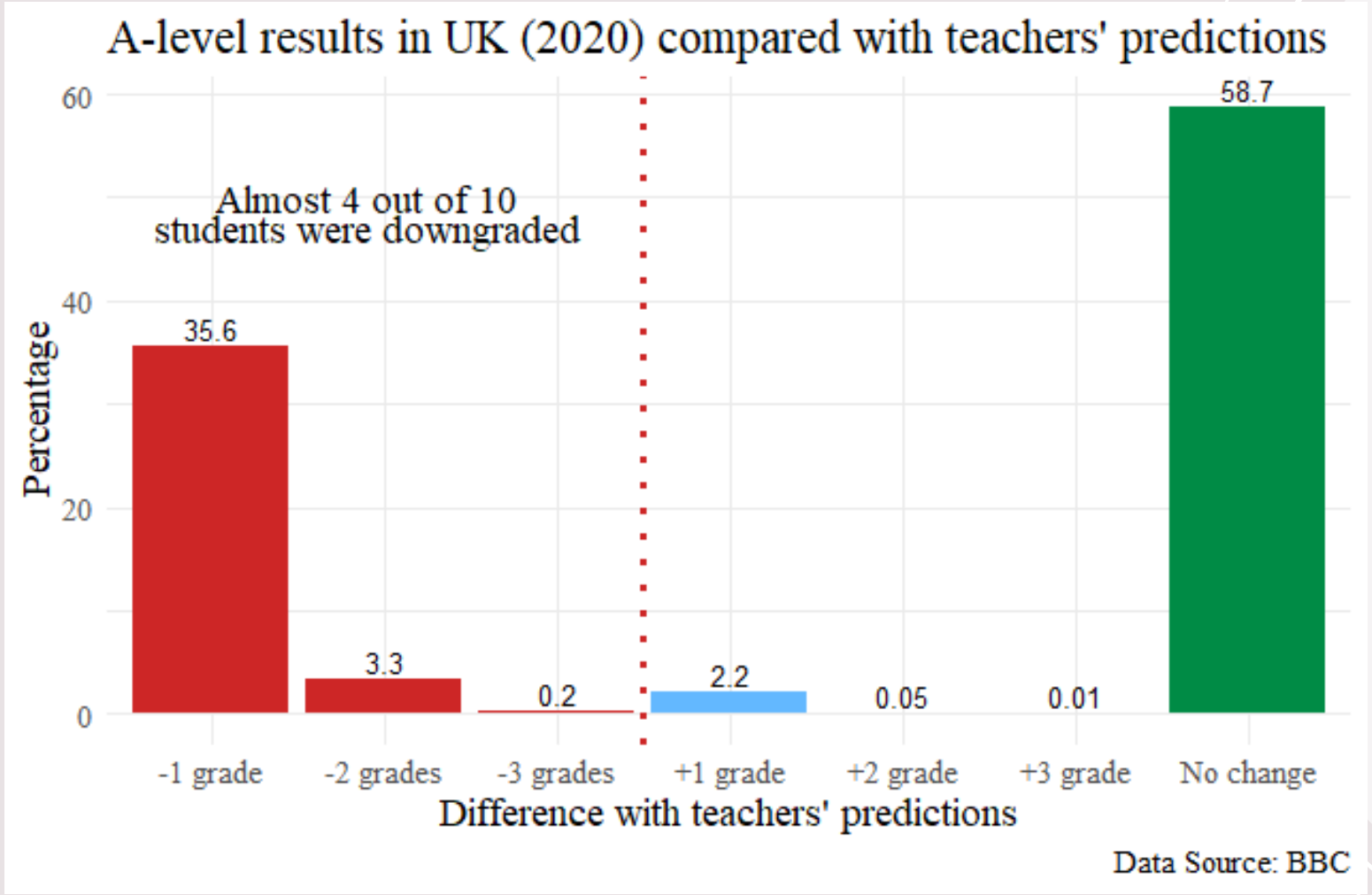
The #AlevelResults Case: Visualizing Social Networks and Text Mining on Twitter Data

Student: Federico Ferrero
Fall 2020



Context of the research questions: The *#AlevelResults* case

- **A-level exams**: exams that are taken annually in the UK by young people to determine which university they can aspire to.
- **COVID-19 pandemic**.
- **High school teachers were asked** (1) to give a predicted grade for their pupils and then (2) rank them within their class.
- Posteriorly, (3) an **algorithm** (named "Ofqual") was expected to adjust and standardize results at the national level.



Results by 9/13/2020

Ofqual's algorithm **reduced 39.1%** of the estimated grades and **increased 2.26%** compared to the teachers' predictions.

```r
# Barplot A-level exam general results
area.color <- c("firebrick3", "firebrick3", "firebrick3", "springgreen4", "steelblue1",
                "steelblue1", "steelblue1")
barplot<-ggplot(data=mydata, aes(x=Difference, y=Percentage)) +
  geom_bar(stat="identity", width=0.9, fill=area.color)+
  geom_text(aes(label=Percentage), vjust=-0.3, size=3.5)+
  labs(title="A-level results in UK (2020) compared with teachers' predictions",
       caption= "Source: BBC",
       x="Difference with teachers' predictions", y = "Percentage")+
  theme_minimal() +
  geom_vline(xintercept = 3.5, linetype="dotted",
             color = "firebrick3", size=1.1)+
  annotate("text", x = 1.9, y = 50, label = "Almost 4 out of 10", family="serif", size=5) +
  annotate("text", x = 1.9, y = 47, label = "students were downgraded",  family="serif", size=5) +
  theme(text=element_text(size=14,  family="serif"))
barplot
```

Context of the research questions: The *#AlevelResults* case

- Results favored students who were enrolled in better schools even if their grades were worse, compared to those in schools with historically worse performance.
- **Private schools** vs **Public schools**.
- Impact on **future educational life**: conditions access to higher education institutions & limits opportunities for social mobility.

Media response...

WIRED BACKCHANNEL BUSINESS CULTURE GEAR IDEAS SCIENCE SECURITY SIGN IN

An Algorithm Determined UK Students' Grades. Chaos Ensued

This year's A-Levels, the high-stakes exams taken in high school, were canceled due to the pandemic. The alternative only exacerbated existing inequities.

BBC Sign in News Sport Reel Worklife Travel

NEWS

Home US Election Coronavirus Video World US & Canada UK Business

N. Ireland N. Ireland Politics

A-levels: Over a third of NI results lowe than estimated grades

13 August 2020

Autumn term chaos feared over exam resits and appeals

Teachers warn that new pupils may be neglected as schools and colleges take action in wake of the A-level algorithm debacle

dents, teachers and parents protest in London against do

A-levels: Anger over 'unfair' results this year

By Sean Coughlan, Katherine Sellgren, Judith Burns
BBC News family and education

13 August 2020

Coronavirus pandemic

I'his year's A-level results are a fiasco – but the system was already broken

Dan Davies

Coronavirus has shown the fragility of grading pupils and exposed problems with the entire examination process

A-levels: 'Dreams ruined by an algorithm'

13 August 2020 Share

Subscribe

The Guardian

News Opinion Sport Culture Lifestyle

A-levels

Do the maths: why England's A-level grading system is unfair

The Guardian

Culture Lifestyle

A-levels

Punishment by statistics': the father who foresaw A-level algorithm flaws

On Twitter trending topic...

#AlevelResults

Why study the case:

Public opinion has spoken out about ***results*** *considered* ***biased*** *coming from* ***evaluations*** *driven by* ***algorithms*** *in* ***educational settings***.

General objective

To explore data from Twitter with the hashtag #AlevelResults in order to produce visualizations about the social network and about the content of the published tweets.

Data

78,839 tweets have been downloaded **from** 8/14/2020 **to** 9/2/2020 (when discussions dissolves) using as a criterion the hashtag **#AlevelResults**.

Variables:

- *Id_str*
- *From_user*
- *Text*
- *Created_at*
- *Time*
- *In_reply_to_user_id_str*
- *In_reply_to_screen_name*
- *From_user_id_str*
- *In_reply_to_status_id_str*
- *Profile_image_url*
- *User_followers_count*
- *User_friends_count*
- *User_location*

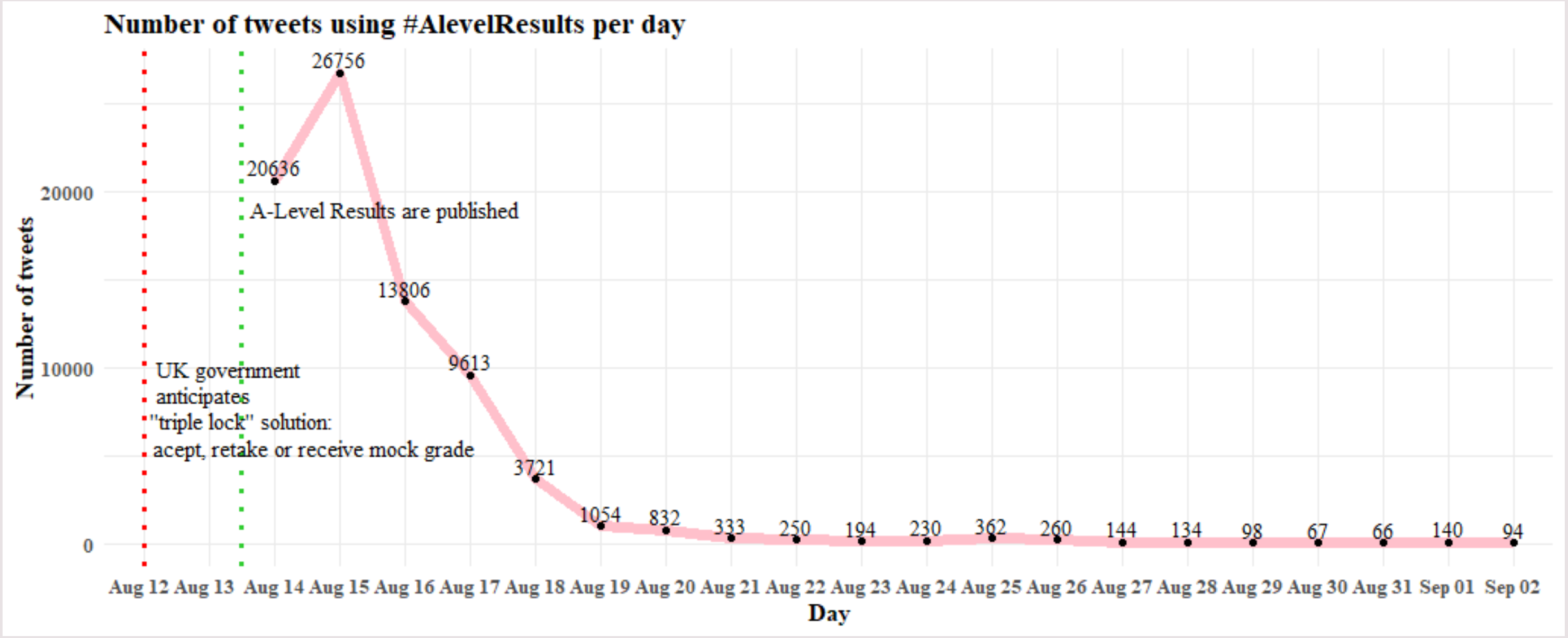
Visualization method(s)

General charts to initially explore data (ggplot2)

Social Network Analysis (igraph)

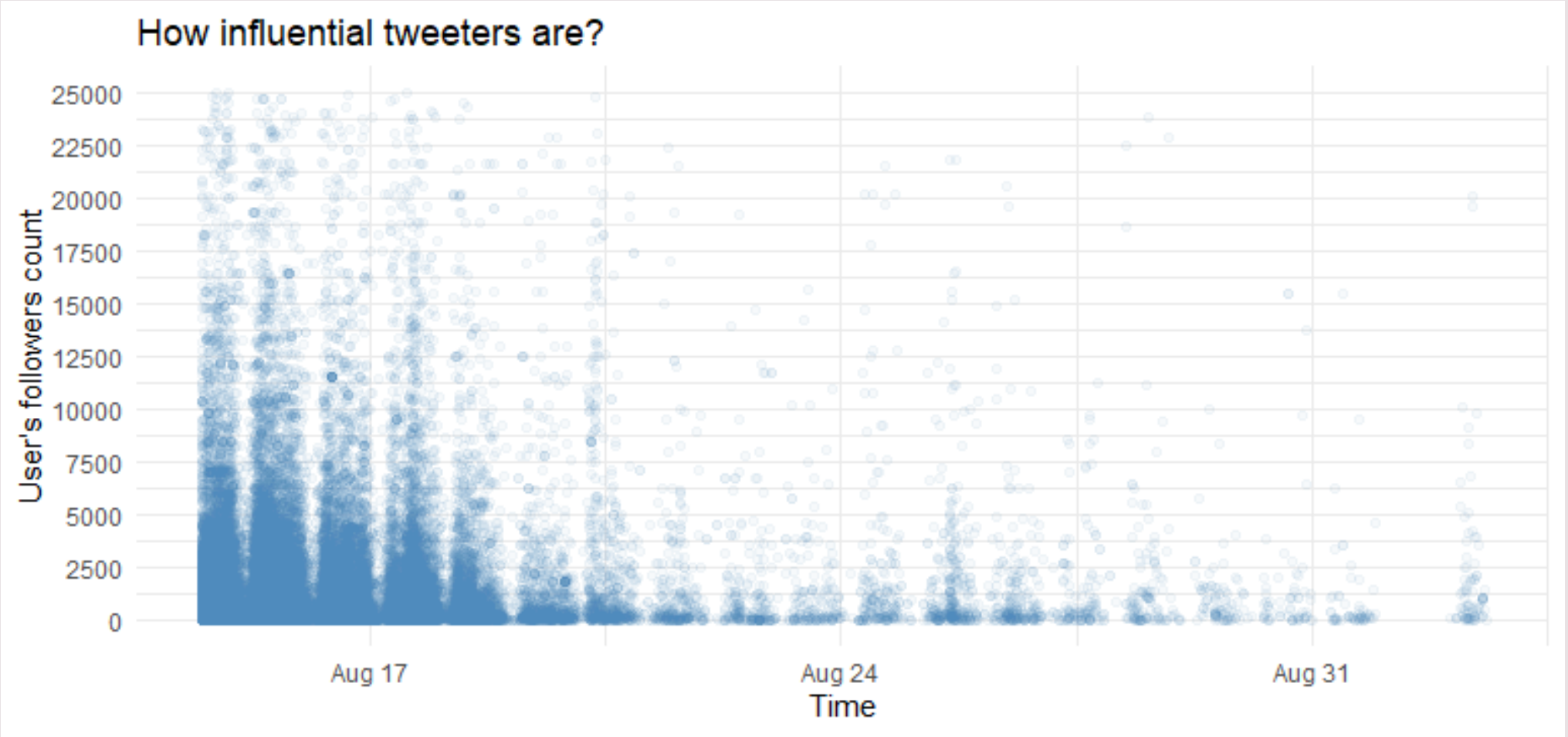
Text Mining (Sentiment Analysis)

How has the reaction around #AlevelResults on Twitter developed in **time**?



```r
# Number of tweets per day: line chart
mydata3 <- read.delim("/Users/feder/Desktop/tweets_per_day.txt")
head(mydata3)
ggplot(data=mydata3, aes(x=Day, y=Number.of.tweets, group=1)) +
  geom_line(color="pink", size=2.5)+
  geom_text(aes(label=Number.of.tweets), vjust=-0.3, size=3.8, family = "serif")+
  geom_point()+
  labs(title="Number of tweets using #AlevelResults per day",
       x="Day", y = "Number of tweets")+
  scale_color_brewer(palette="Paired")+
  geom_vline(xintercept = 1, linetype="dotted",
             color = "red", size=1.2)+
  geom_vline(xintercept = 2.5, linetype="dotted",
             color = "limegreen", size=1.2)+
  annotate("text", x = 4.7, y = 19000, label = "A-Level Results are published", family = "serif") +
  annotate("text", x = 2.3, y = 10000, label = "UK government", family = "serif") +
  annotate("text", x = 1.9, y = 8500, label = "anticipates", family = "serif") +
  annotate("text", x = 2.5, y = 7000, label = "''triple lock'' solution:", family = "serif") +
  annotate("text", x = 3.6, y = 5500, label = "acept, retake or receive mock grade", family = "serif") +
  theme_minimal()+
  theme(text=element_text(size=12, family = "serif", face = "bold"))
```

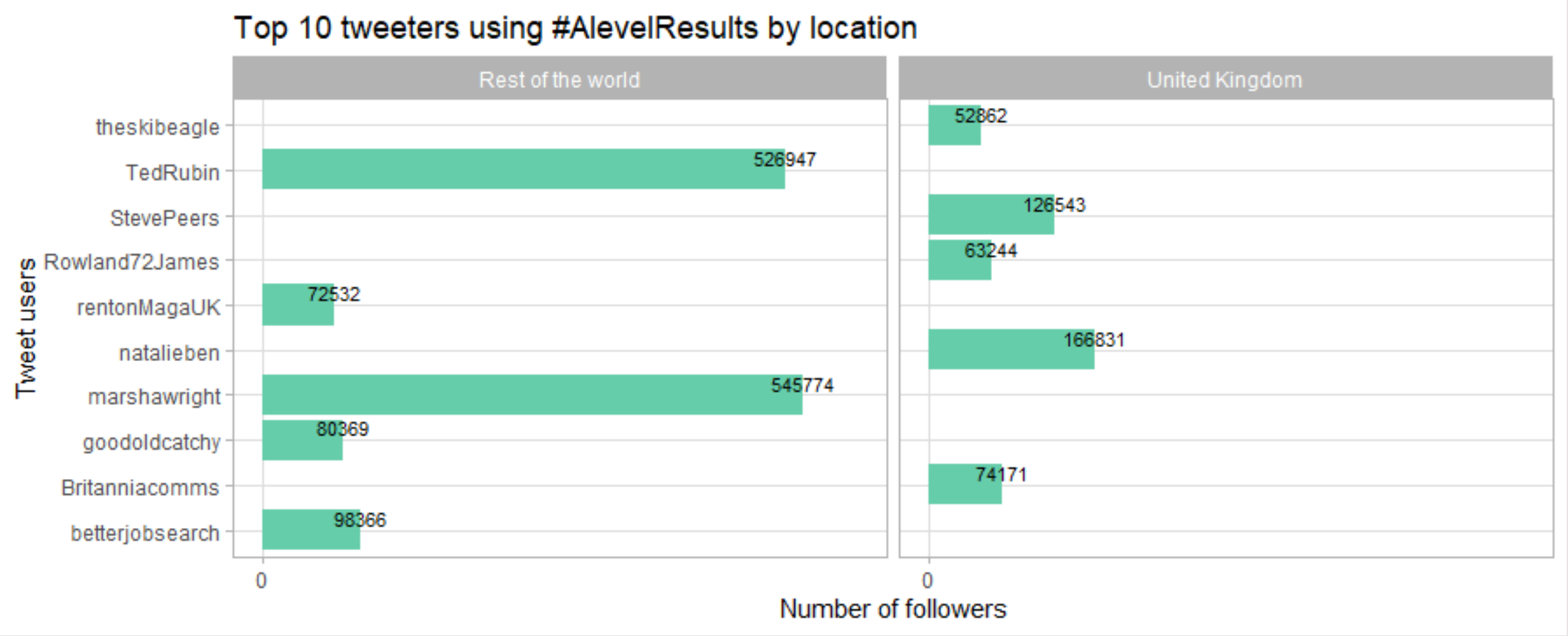
How has the reaction around #AlevelResults on Twitter developed in **time**? (considering number of followers)



```r
# check if time is continous variable
class(complete_tweets$time)

# scatterplot
ggplot(aes(x=time, y=user_followers_count), data=complete_tweets) +
  geom_point(alpha=0.05, color= 'steelblue')+
    labs(title="How influential tweeters are?",
         x="Time", y = "User's followers count")+
  scale_y_continuous(limit=c(1,25000), breaks = seq(0,25000,2500)) +
  theme(text=element_text(size=14,  family="Garamond")) +
  theme_minimal()
```

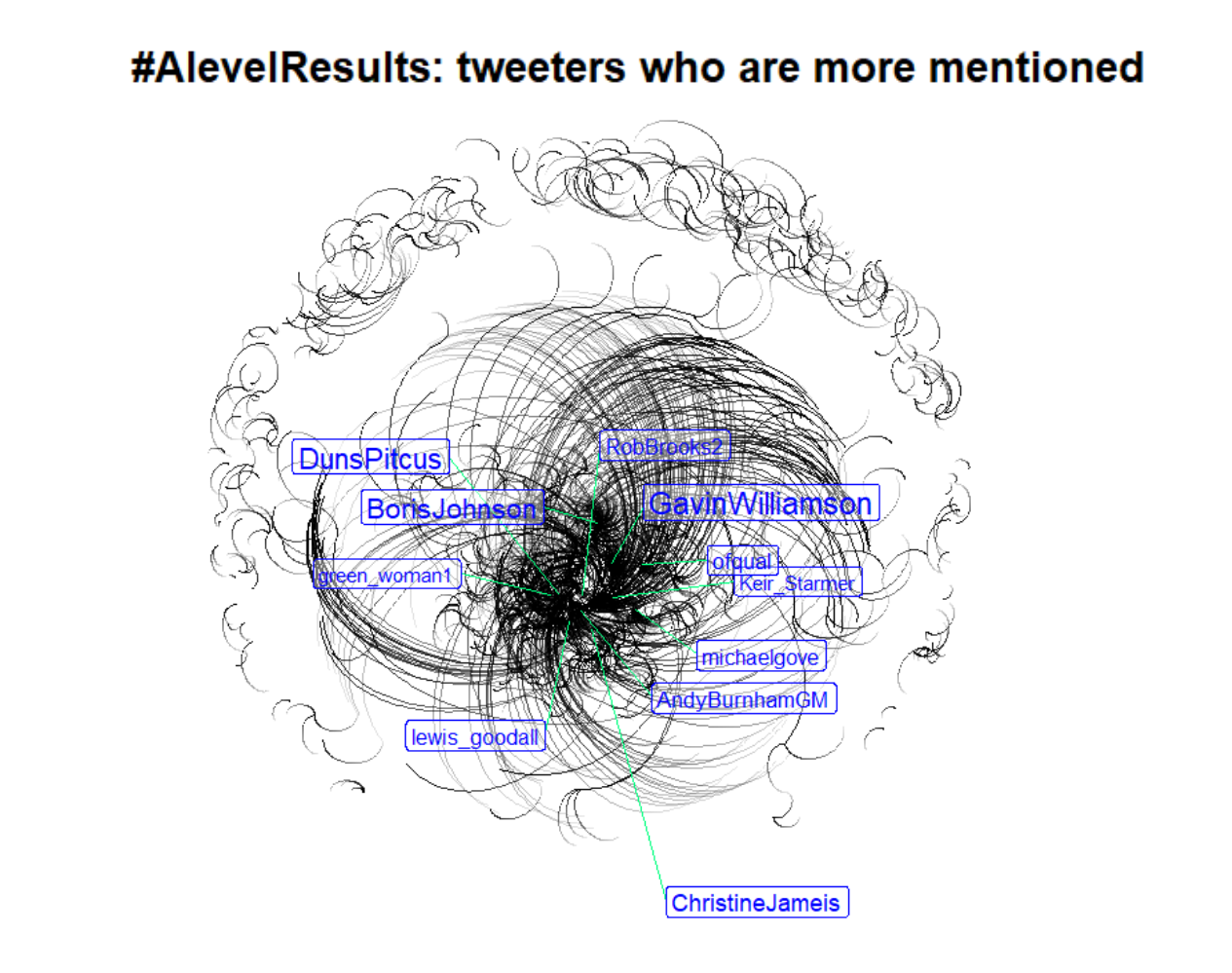
Who are the **top users** involved in the discussion around #AlevelResults on Twitter?
From which **geographic** locations did these users participate?



```
barplot2<-ggplot(data=mydata2, aes(x=number_of_followers, y=from_user, xaxt="n")) +
  geom_bar(stat="identity", width=0.9, fill="aquamarine3") +
  geom_text(aes(label=number_of_followers), vjust=-0.3, size=3) +
  labs(title="Top 10 tweeters using #AlevelResults by location",
  x="Number of followers", y = "Tweet users") +
  theme(text=element_text(size=12,family="Serif")) +
  theme_light() +
  scale_x_continuous(limit=c(0,600000), breaks=seq(0,600000,600001)) +
  facet_wrap(~location, ncol=2)

barplot2
```

What is the **influential network** or sub-community that most participates in?
Who are the **most interpellated** ones in the discussion?



```
g <- read.delim("/Users/feder/Desktop/BASE14.txt")

# create data graph
g <- graph.data.frame(g, directed= T)

# remove loops in the igraph
g <- simplify(g, remove.multiple = F, remove.loops = T)

# check number of edges and nodes
E(g)# edges= 2312
V(g)# vertex= 49111

# defining labels and degrees
V(g)$node_label <- unname(ifelse(degree(g)[V(g)] > 20, names(V(g)), ""))
V(g)$node_size <- unname(ifelse(degree(g)[V(g)] > 20, degree(g), 0))

# People who are more mentioned is highlighted / layout: kk
ggraph(g, layout = "kk") +
        geom_edge_arc(edge_width=0.1, aes(alpha=..index..)) +
        geom_node_label(aes(label=node_label, size=node_size),
                        label.size=0, fill="#ffffff66", segment.colour="springgreen",
                        color="blue", repel=TRUE, family="Apple Garamond") +
        coord_fixed() +
        scale_size_area(trans="sqrt") +
        labs(title="#AlevelResults: tweeters who are more mentioned") +
        theme_graph(base_family="Apple Garamond") +
        theme(legend.position="none")
```

What is the **network** (or subcommunity) that **participates** the **most**?

S Styles @DunsPitcus · 15 ago.

Imagine if Leicester City to were stripped of their 2025-16 Premier League Trophy with honours being awarded to Manchester Utd instead because an algorithm decided that as LCFC didn't finish in the top ten the previous decade their position wd be downgraded to 7th. **#AlevelResults**

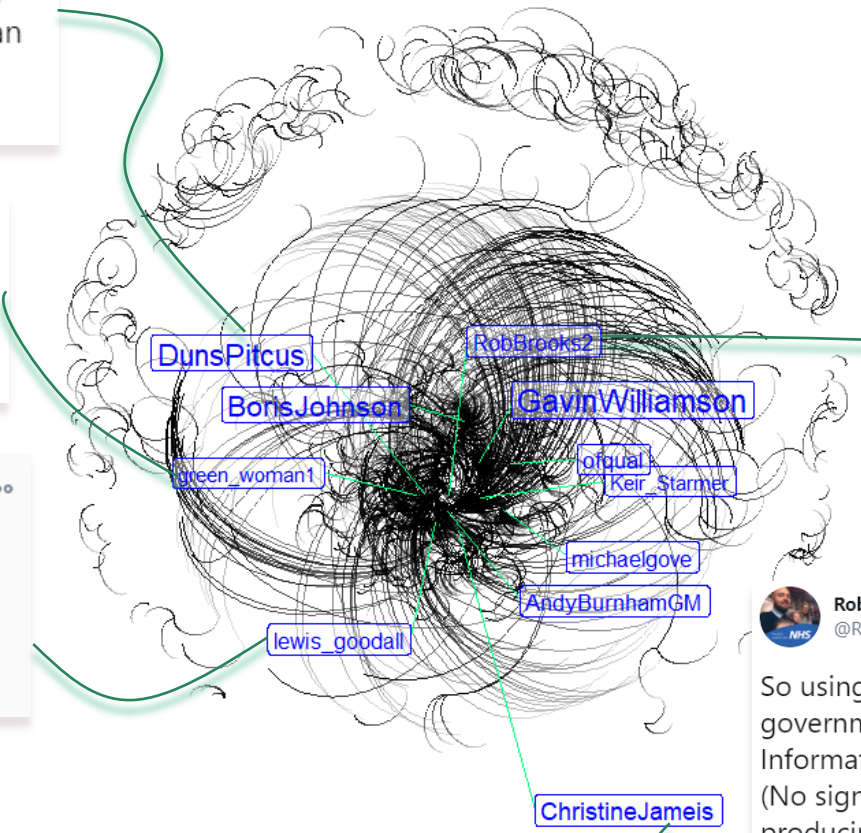
Green World @green_woman1 · 17 ago.

Oh I think young people in #England know exactly what's going on. Imagine not having the sense to keep future electors happy. ##ToryShambles #ToryChaos #ExamResults **#AlevelResults** #GavinWilliamsonResign #JohnsonHasFailed

Lewis Goodall @lewis_goodall · Aug 17

BREAK: Gavin Williamson and Ofqual confirm that the govt will abandon standardisation and use teacher assessment for GCSEs and A-levels in England.

Hundreds of thousands of students just had their exams results upgraded.



Rob Brooks @RobBrooks2

So using my own algorithm I have graded three governments #AlevelResults -
Information Technology = U
(No sign of the track and trace app they were producing for their final exam)

#FreePalestine #FreeAssange #BDS #Yemen #Kashmir @... · Aug 15

"Eton gets 80% slashed off its rates bill because it has charitable status, while most state schools must pay 100%." Daily Mirror. #classwar **#AlevelResults**

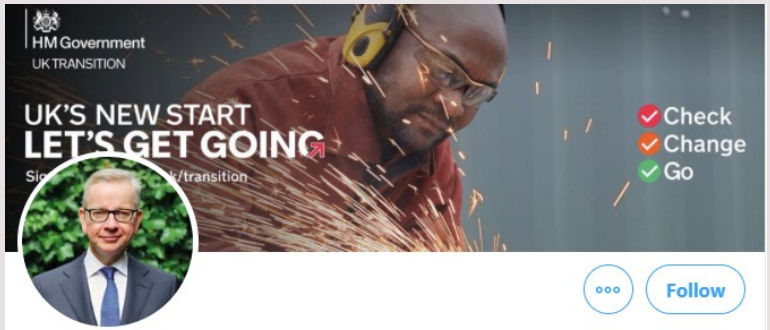
Who are the **most interpellated** ones in the discussion?

Boris Johnson
@BorisJohnson

Prime Minister of the United Kingdom and @Conservatives leader. Member of Parliament for Uxbridge and South Ruislip.

United Kingdom · gov.uk/coronavirus · Joined April 2015

458 Following **3.1M** Followers

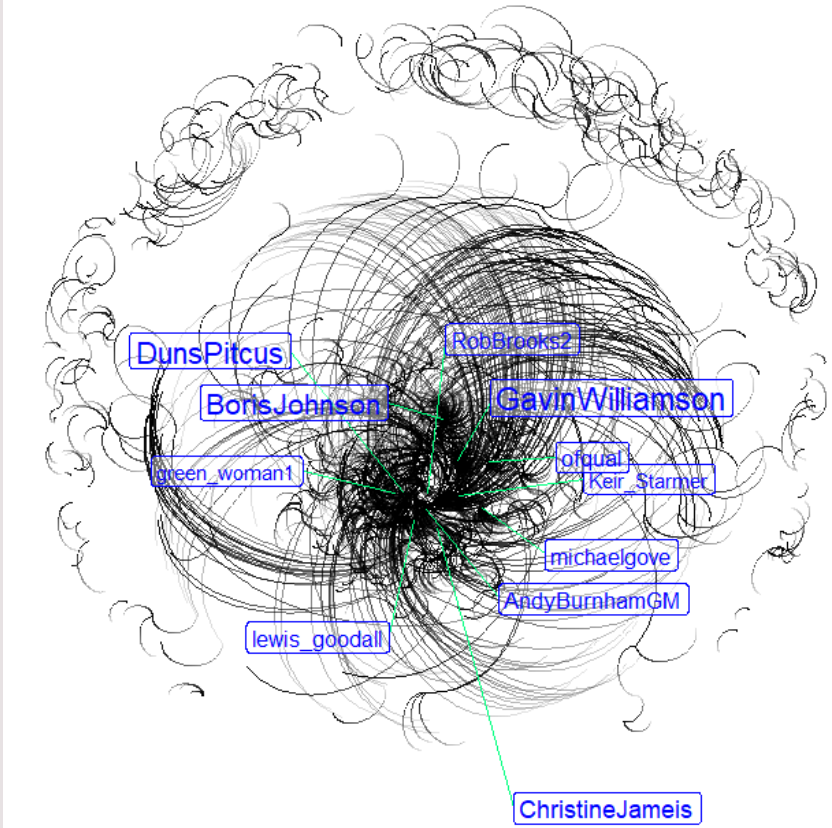


Michael Gove
@michaelgove

MP for Surrey Heath. Chancellor of the Duchy of Lancaster @CabinetOfficeUK

michaelgove.com · Joined June 2016

2,331 Following **207.2K** Followers



Andy Burnham
@AndyBurnhamGM

Labour Mayor of Greater Manchester

Leigh, Greater Manchester · greatermanchester-ca.gov.uk/themayor · Joined May 2010

2,362 Following **393.1K** Followers

Gavin Williamson
@GavinWilliamson

Conservative Member of Parliament for South Staffordshire and Education Secretary. For any casework issues please email gavin@gavinwilliamson.org

South Staffordshire · Joined September 2011

521 Following **86.7K** Followers

Ofqual
@ofqual

The regulator of qualifications, exams and assessments in England. If you have a question for us, please email public.enquiries@ofqual.gov.uk

Coventry, UK · gov.uk/ofqual · Joined January 2009

524 Following **37.1K** Followers

Keir Starmer
@Keir_Starmer

Leader of the Labour Party. Member of Parliament for Holborn and St Pancras. Former Director of Public Prosecutions.

London · keirstarmer.com · Joined April 2014

412 Following **1M** Followers

What **feelings** are associated with the texts published in the tweets?

Sentiment Analysis: tweets that mention @BorisJohnson

Sentiment Analysis: tweets that mention @GavinWilliamson

Sentiment Analysis: tweets that mention @ofqual

```r
# TOKENIZATION: split text = it separes one by one the words from our database and indicates
# in which lines they are
library(tidytext)
# remove punctuation, convert to lowercase, add id for each tweet!
boris <- boris %>%
  dplyr::select(text) %>%
  unnest_tokens(word, text)
boris

# remove stop words
data(stop_words)
boris  <- boris %>%
  anti_join(stop_words)
boris

# count words
boris %>%
  count(word, sort = TRUE)

# wordcloud: most common words in tweets that mention @BorisJohnson
library(wordcloud)
boris %>%
  anti_join(stop_words) %>%
  count(word) %>%
  with(wordcloud(word, n, max.words = 100))

# generate a wordcloud showing sentiment analysis
library("reshape2")
boris %>%
  inner_join(get_sentiments("bing")) %>%
  count(word, sentiment, sort = TRUE) %>%
  acast(word ~ sentiment, value.var = "n", fill = 0) %>%
  comparison.cloud(colors = c("indianred3", "chartreuse4"),
                   max.words = 100)
```

Explore the #AlevelResults Database!: Shiny App

https://federico-jf.github.io/Data-Visualization-UTD-EPPS-6356/

Conclusions

- ###
- ###
- ###
- ###

References

Abdelsadek, Y., Chelghoum, K., Herrmann, F., Kacem, I., & Otjacques, B. (2018). Community extraction and visualization in social networks applied to Twitter. Information Sciences, 424, 204-223.

Bruns, A. (2012). How Long is a Tweet? Mapping Dynamic Conversation Networks on Twitter Using Gawk and Gephi. Information, Communication & Society, 15:9, 1323-1351, DOI: 10.1080/1369118X.2011.635214

Ediger, D. et al., (2010). Massive Social Network Analysis: Mining Twitter for Social Good. 39th International Conference on Parallel Processing, San Diego, CA, 583-593, DOI: 10.1109/ICPP.2010.66.

Kolaczyk, E. D., & Csárdi, G. (2014). Statistical Analysis of Network Data with R. New York: Springer.

Silge, J., & Robinson, D. (2017). Text mining with R: A tidy approach. New York: O'Reilly Media.

Journalistic References about "A Level Exams" Case

Alex Hern (14 August 2020). Do the maths: why England's A-level grading system is unfair. The Guardian. Retrieved from: https://amp.theguardian.com/education/2020/aug/14/do-the-maths-why-englands-a-level-grading-system-is-unfair?__twitter_impression=true

Amit Katwala. (15 August 2020). An Algorithm Determined UK Students' Grades. Chaos Ensued. Wired. https://www.wired.com/story/an-algorithm-determined-uk-students-grades-chaos-ensued/

Dan Davies (15 August 2020). This year's A-level results are a fiasco – but the system was already broken. The Guardia. Retrieved from: https://www.theguardian.com/commentisfree/2020/aug/15/a-level-results-system-ofqual-england-exam-marking

Donna Ferguson and Michael Savage (16 August 2020) Autumn term chaos feared over exam resits and appeals. The Guardian. Retrieved from: https://www.theguardian.com/education/2020/aug/16/autumn-term-chaos-feared-over-exam-resits-and-appeals

Heather Stewart and Kate Proctor (17 August 2020)Penny Mordaunt becomes first minister to voice concerns about A-level crisis. The Guardian. Retrieved from: https://www.theguardian.com/education/2020/aug/17/delaying-a-year-wont-be-an-option-for-many-a-level-pupils-says-penny-mordaunt

Helen Pidd (14 August 2020). 'Punishment by statistics': the father who foresaw A-level algorithm flaws. The Guardian. Retrieved from: https://amp.theguardian.com/education/2020/aug/14/punishment-by-statistics-the-father-who-foresaw-a-level-algorithm-flaws?__twitter_impression=true

No specified author (13 August 2020). A-levels: 'Dreams ruined by an algorithm'. BBC News. Retrieved from: https://www.bbc.com/news/uk-northern-ireland-53767773

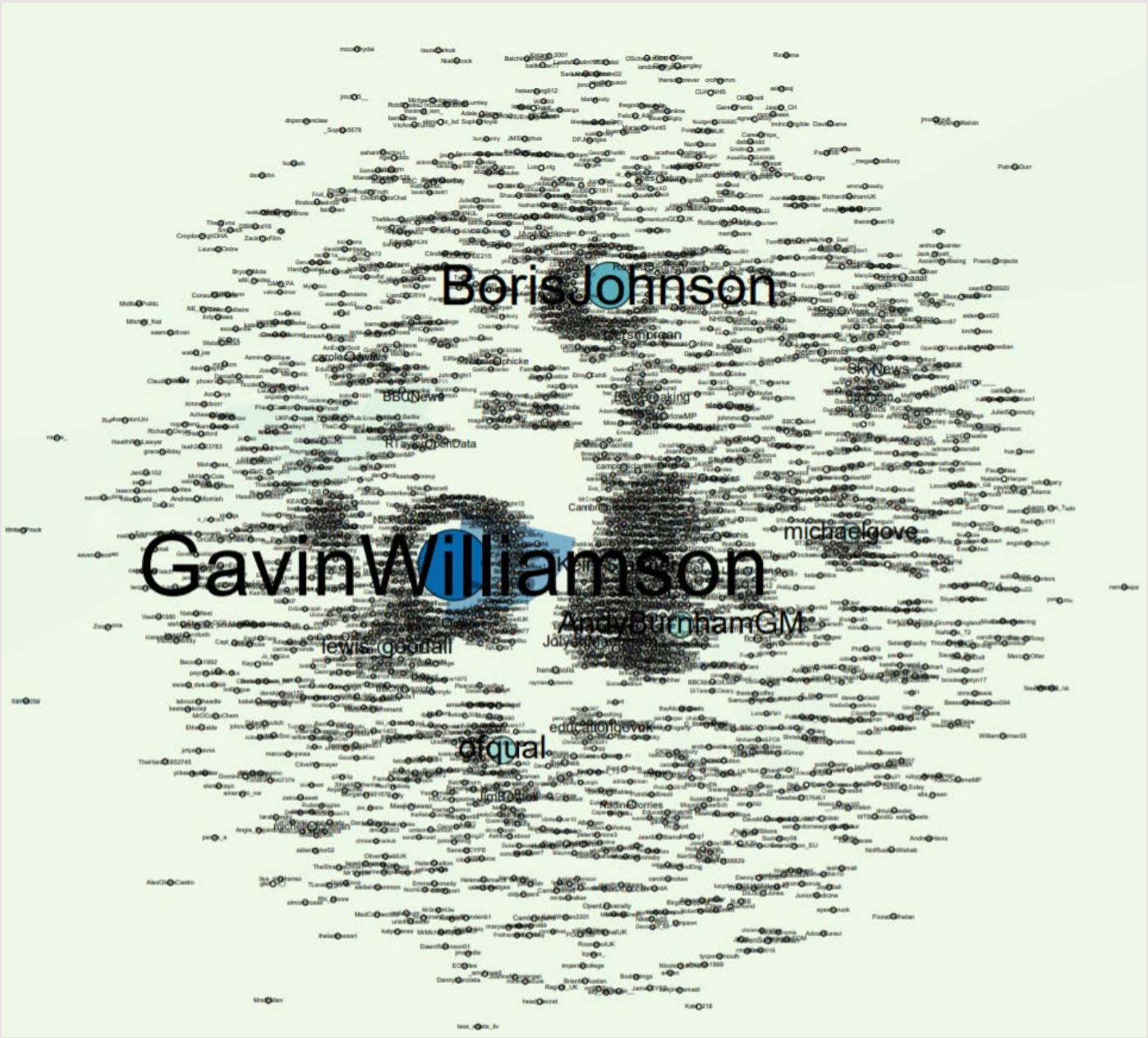
No specified author (13 August 2020). A-levels: Over a third of NI results lower than estimated grades. BBC News. Retrieved from: https://www.bbc.com/news/uk-northern-ireland-53757767

Ofqual (2020). Awarding GCSE, AS, A level, advanced extension awards and extended project qualifications in summer 2020: interim report. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/909368/6656-1_Awarding_GCSE__AS__A_level__advanced_extension_awards_and_extended_project_qualifications_in_summer_2020_-_interim_report.pdf

Sean Coughlan, Katherine Sellgren, Judith Burns (13 August 2020) A-levels: Anger over 'unfair' results this year. BBC News. Retrieved from: https://www.bbc.com/news/education-53759832

Thank you

Rachael!
@Rachael_Davis56
@ofqual downgrading me from AAA to ACC because of where I live #AlevelResults
Traducir Tweet
STOP
BEING
POOR
4:08 a. m. · 13 ago. 2020 · Twitter for Android


BorisJohnson
GavinWilliamson
michaelgove
AndyBurnhamGM
lewis_goodall
ofqual
educationgovuk
BBCNews
SkyNews