Queen Mary University of London

institute IMDEA networks

香港科技大學 THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Exploring online manifestations of real-world inequalities on the Nextdoor Social Network

Waleed Iqbal

Coseners-MSN 2023

Published in AAAI ICWSM 2023, in Collaboration with Vahid Ghafouri, Gareth Tyson, Guillermo Suarez-Tangil, Ignacio Castro

Can we identify

rich & poor?

Can we **identify**
rich & **poor**
from their
online **text**?

Nextdoor: within neighbourhood interactions

How Nextdoor works?

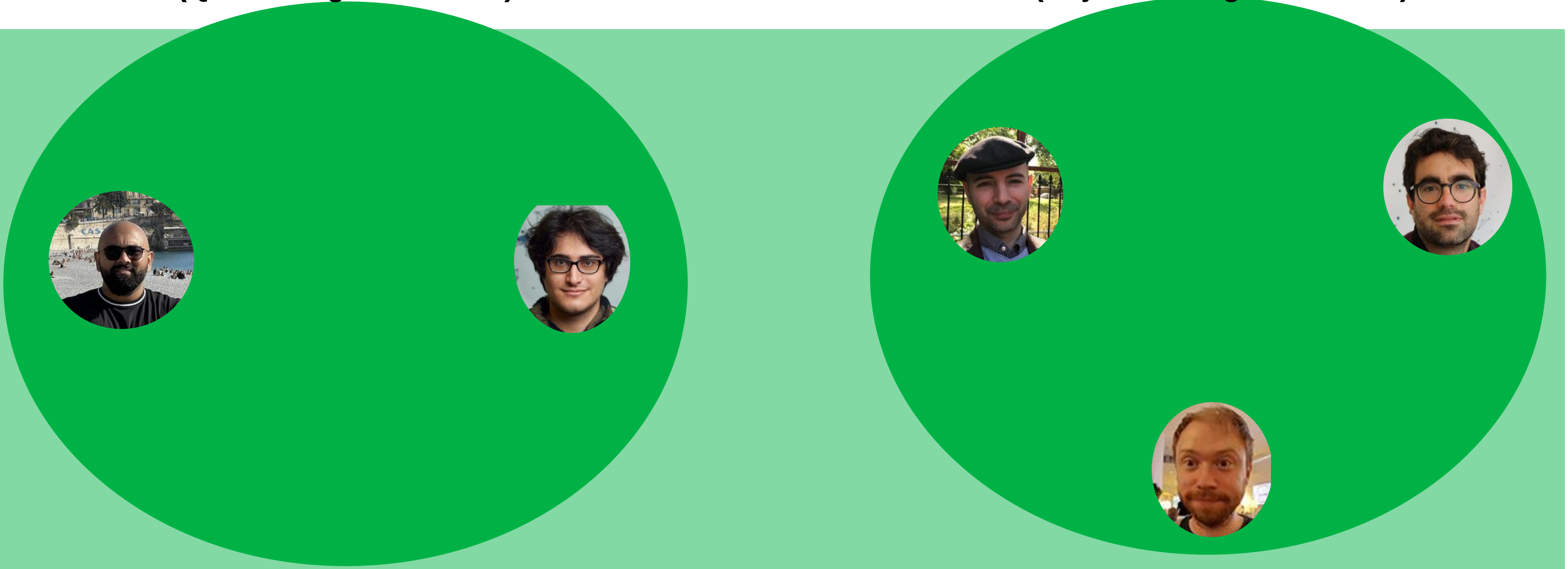
Mile End
(QMUL neighbourhood)

Bethnal Green
(Adjacent neighbourhood)

How Nextdoor works?

Mile End
(QMUL neighbourhood)

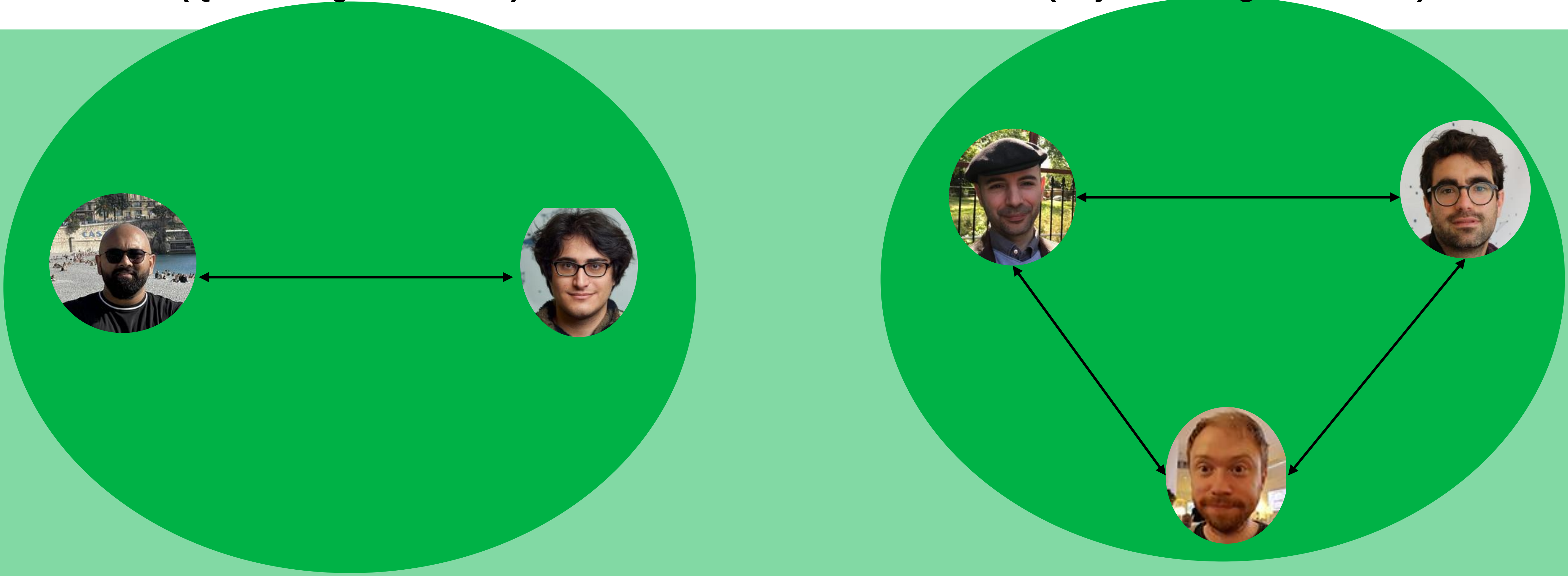
Bethnal Green
(Adjacent neighbourhood)



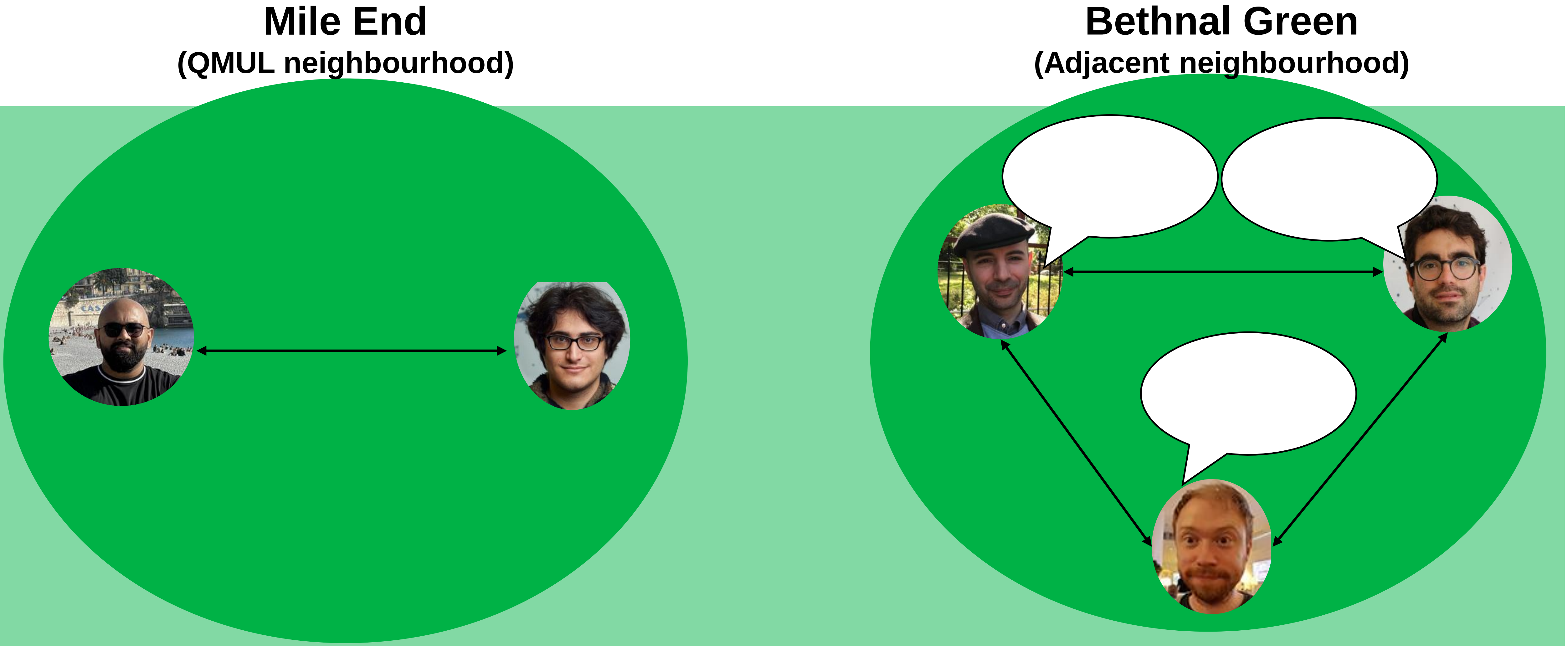
How Nextdoor works?

Mile End
(QMUL neighbourhood)

Bethnal Green
(Adjacent neighbourhood)



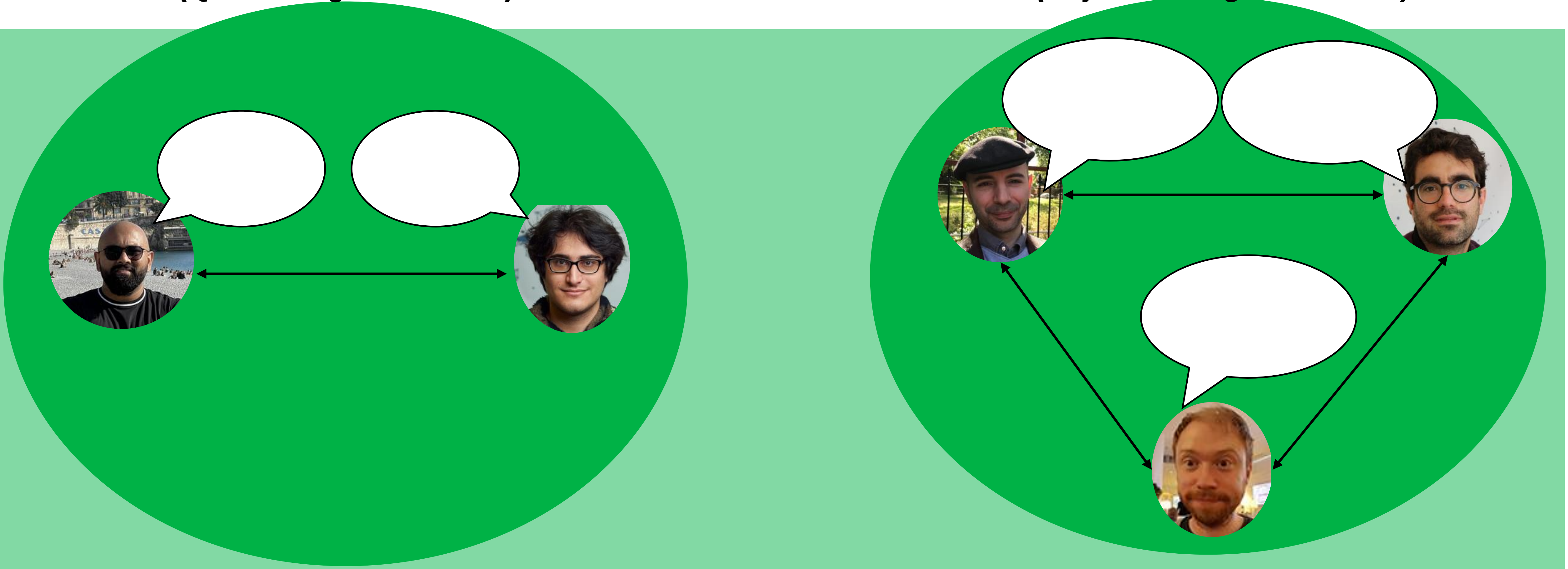
How Nextdoor works?



How Nextdoor works?

Mile End
(QMUL neighbourhood)

Bethnal Green
(Adjacent neighbourhood)



Nextdoor data

Attributes	USA	UK	Total
Posts	2,201,051	351,894	2,602,045
Neighborhoods	64,283	3,325	67,608
Cities	5,849	10	5,859
zip code(USA)/LSOA(UK)	30872	2512	33284
Comments	17,421,050	2,246,814	19,667,864
Neighbors	6,6480,730	1,744,948	68,225,678

lowest geo-units

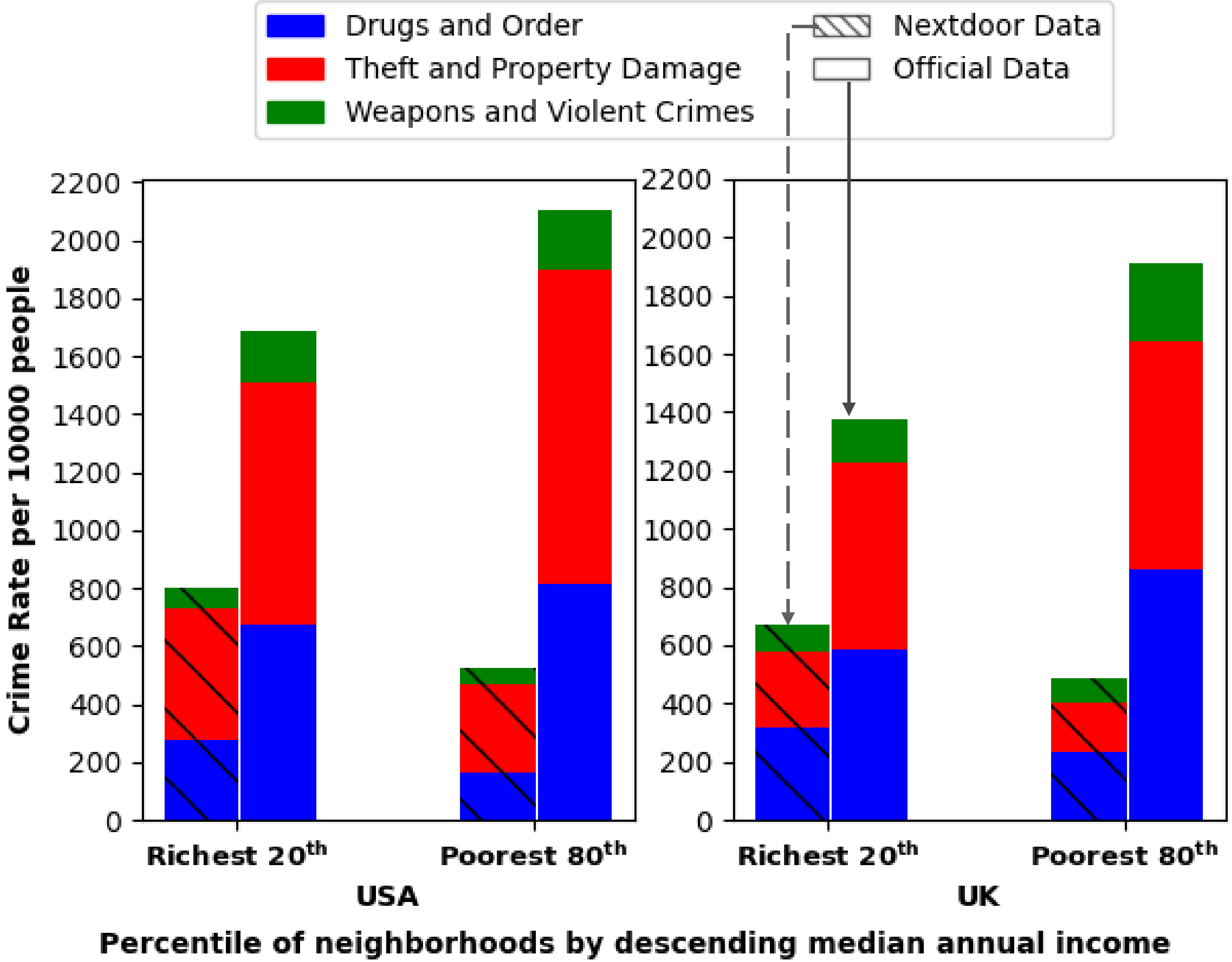
Methodology

- Neighbourhood→ geolocation→ Official statistics
  - Median income
  - Crime
- Identify crime related conversations: Semantic Search S-BERT
- Sentiment: VADER

Does the online text of **rich** & **poor** **differ**?

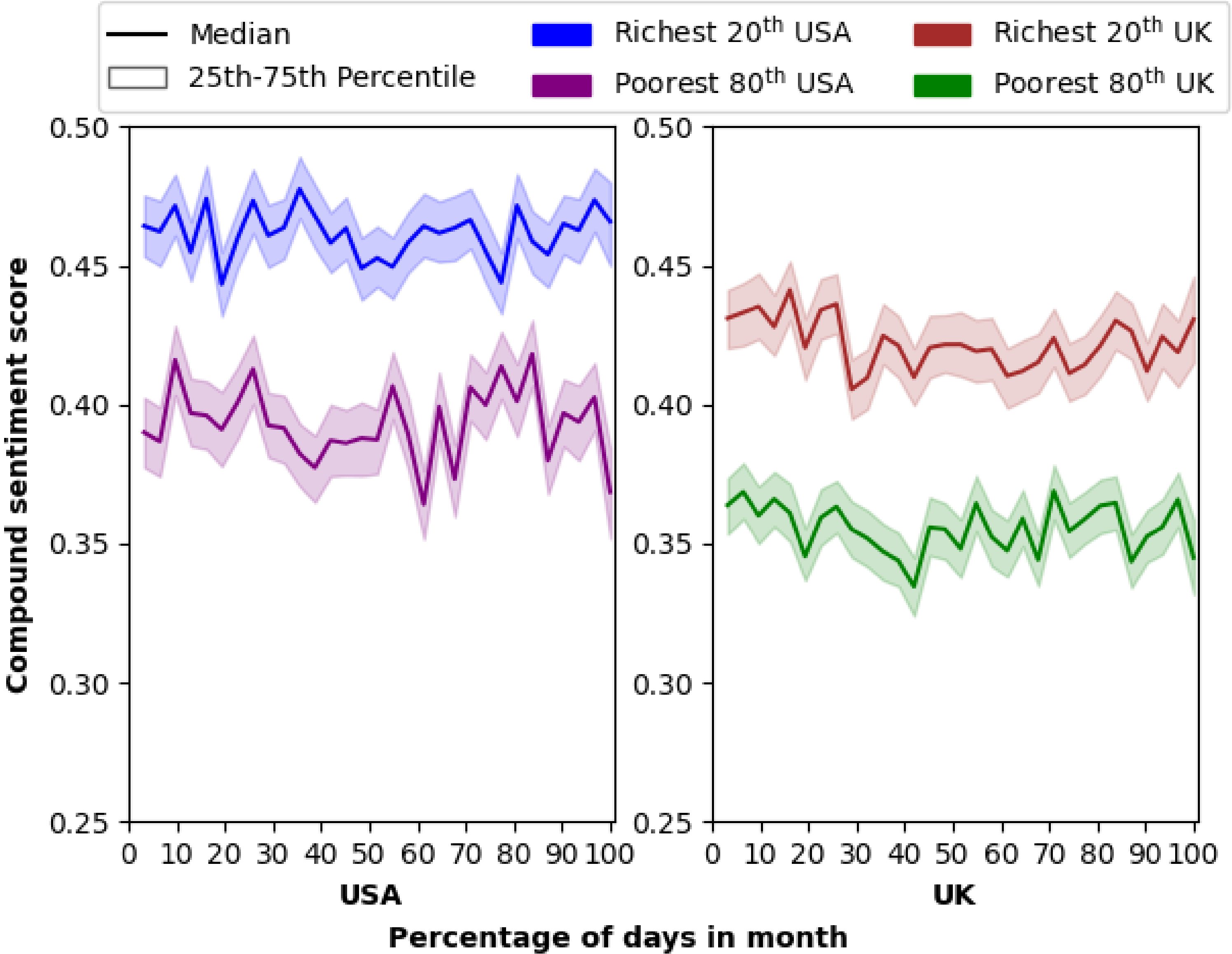
Who talks more about crime?

Richer neighbourhoods talk more about crime



Will richer have more negative sentiment?

More positive sentiment in richer neighbourhoods

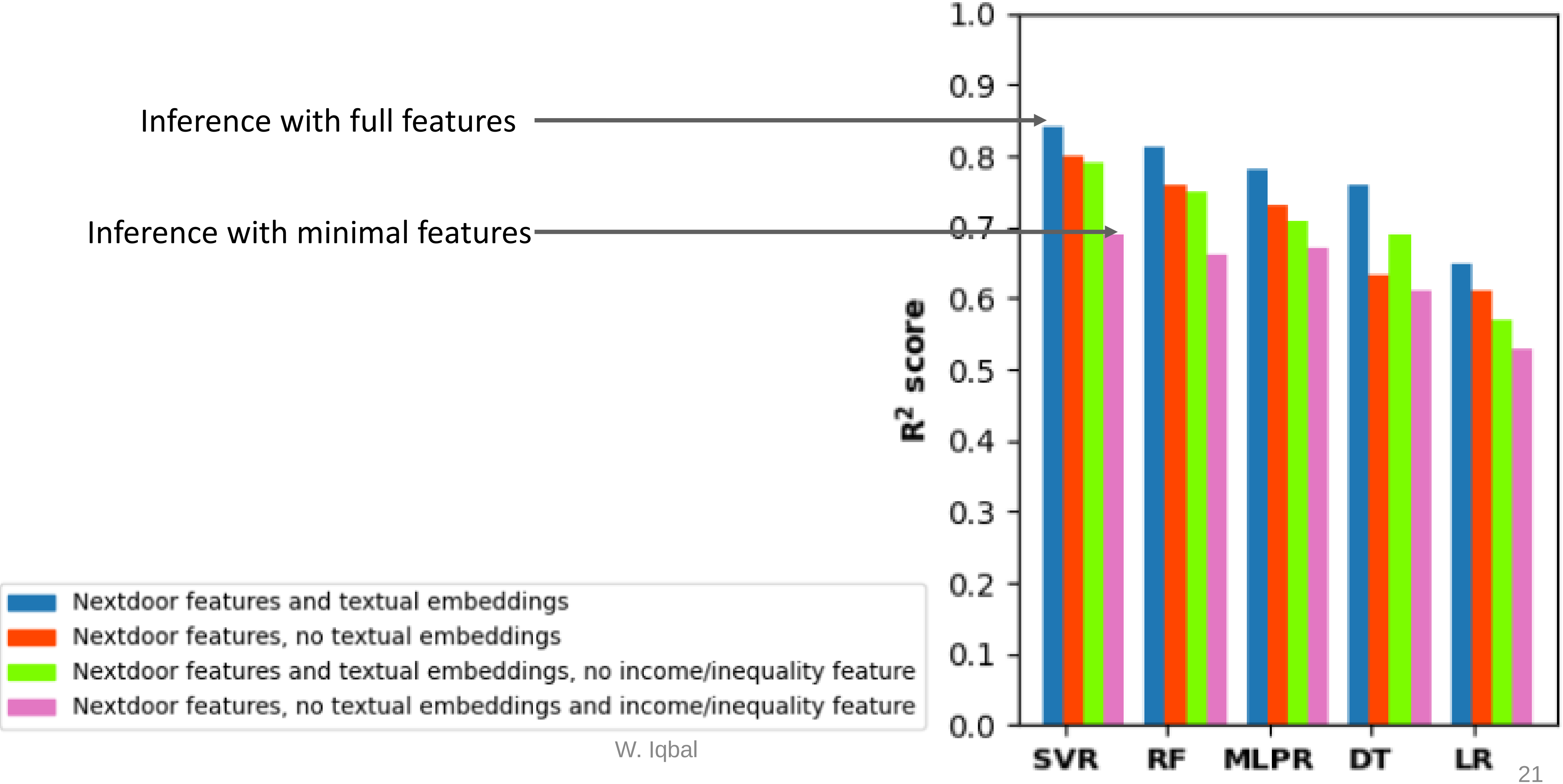


Can we **infer neighbourhood's income** by the **text** posted **online**?

Can we infer who is rich/poor just from the text?

- Embeddings: dimensionality reduction (768→5)
- Discussed-to-official crime ratio
- Other features
- Multiple common ML models

We can infer the level of income from the text



Politics and **income**

(Preliminary Results)

Datasets

- Twitter → Politician tweets
  - 10.1 Million from UK
  - 2.2 Million from USA

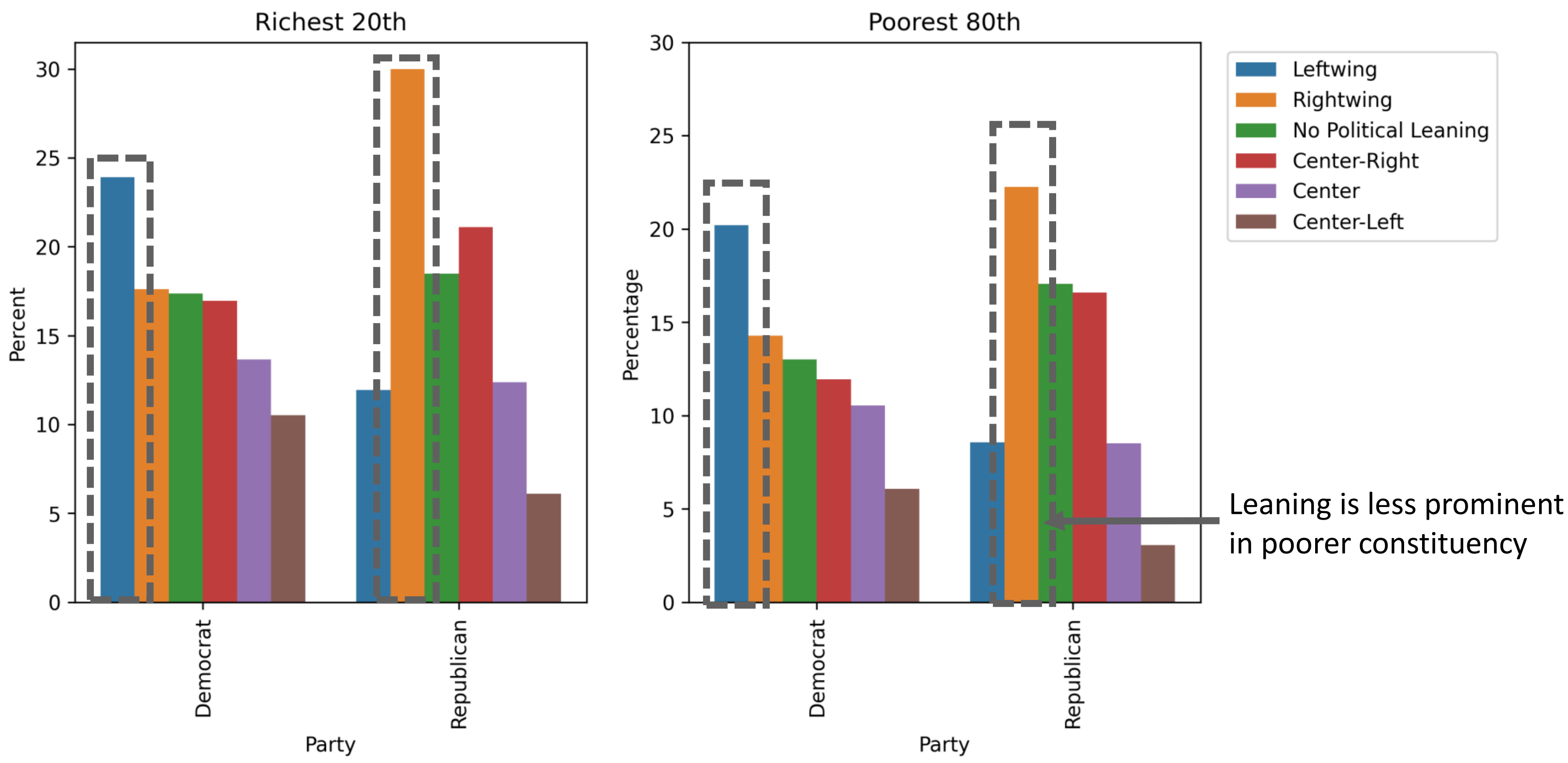
- Nextdoor → Neighbourhood posts
  - 4.5 Million from UK
  - 24.3 Million from USA

Methodology

- Neighbourhood→ constituency→ Official statistics
  - Median income

- Political leaning in tweets → ChatGPT, manual annotation
  - Fine-tuned BERT
    Setfit Model
  - Few shot Learning

Politician agrees with party ideology



USA (Twitter)

Conclusion

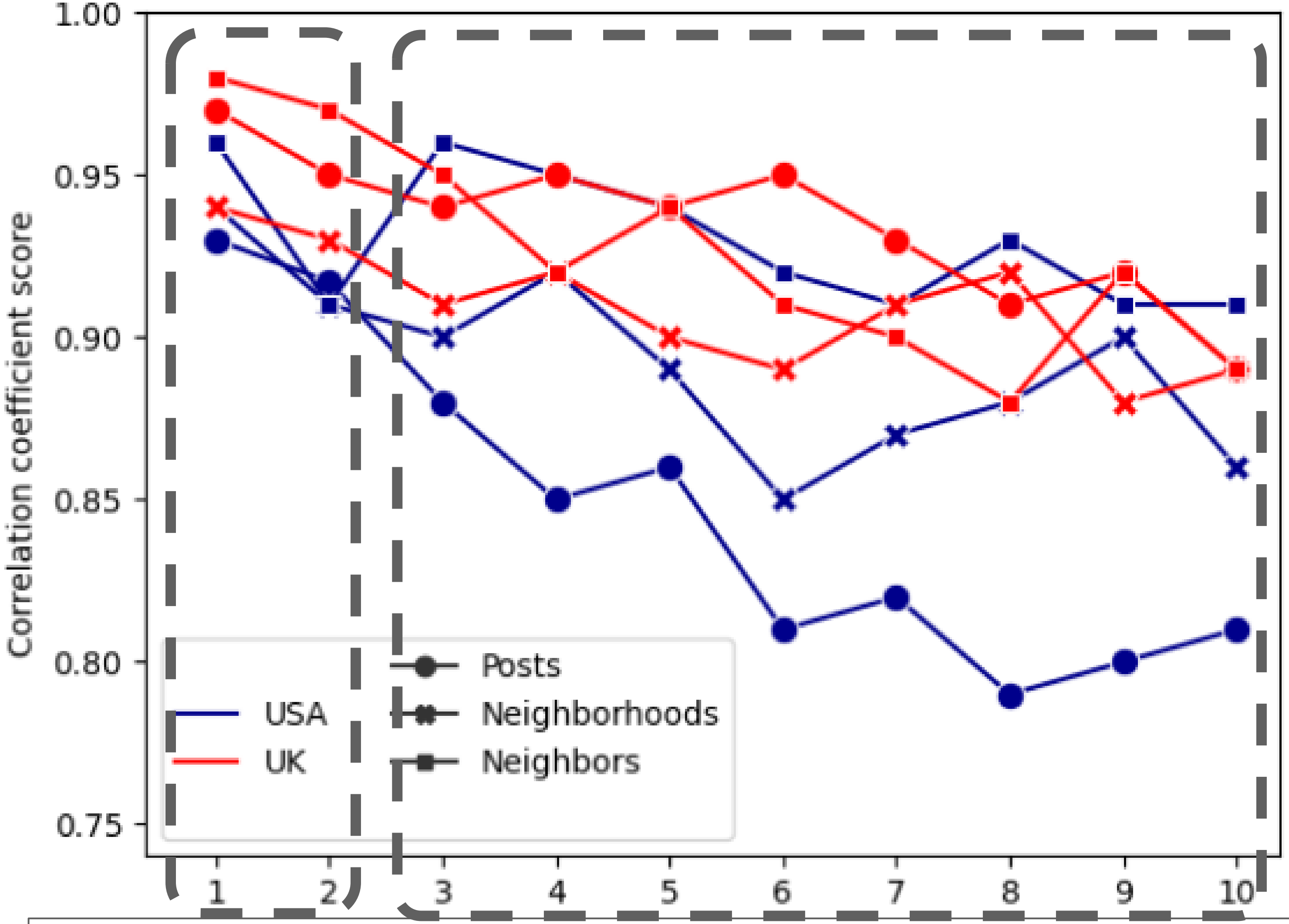
- Rich and poor neighbourhoods have distinct online text.
- User generated posts can predict neighbourhood's income.

Future work

- What about politics?
- What about inequality?
- Can we generalise?
- We are also in talk to collaborate with Nextdoor.

Backup Slides

How representative the data is?



20% Richest

80% poorest

Inequality across neighborhoods

Vicinity of a neighborhood:

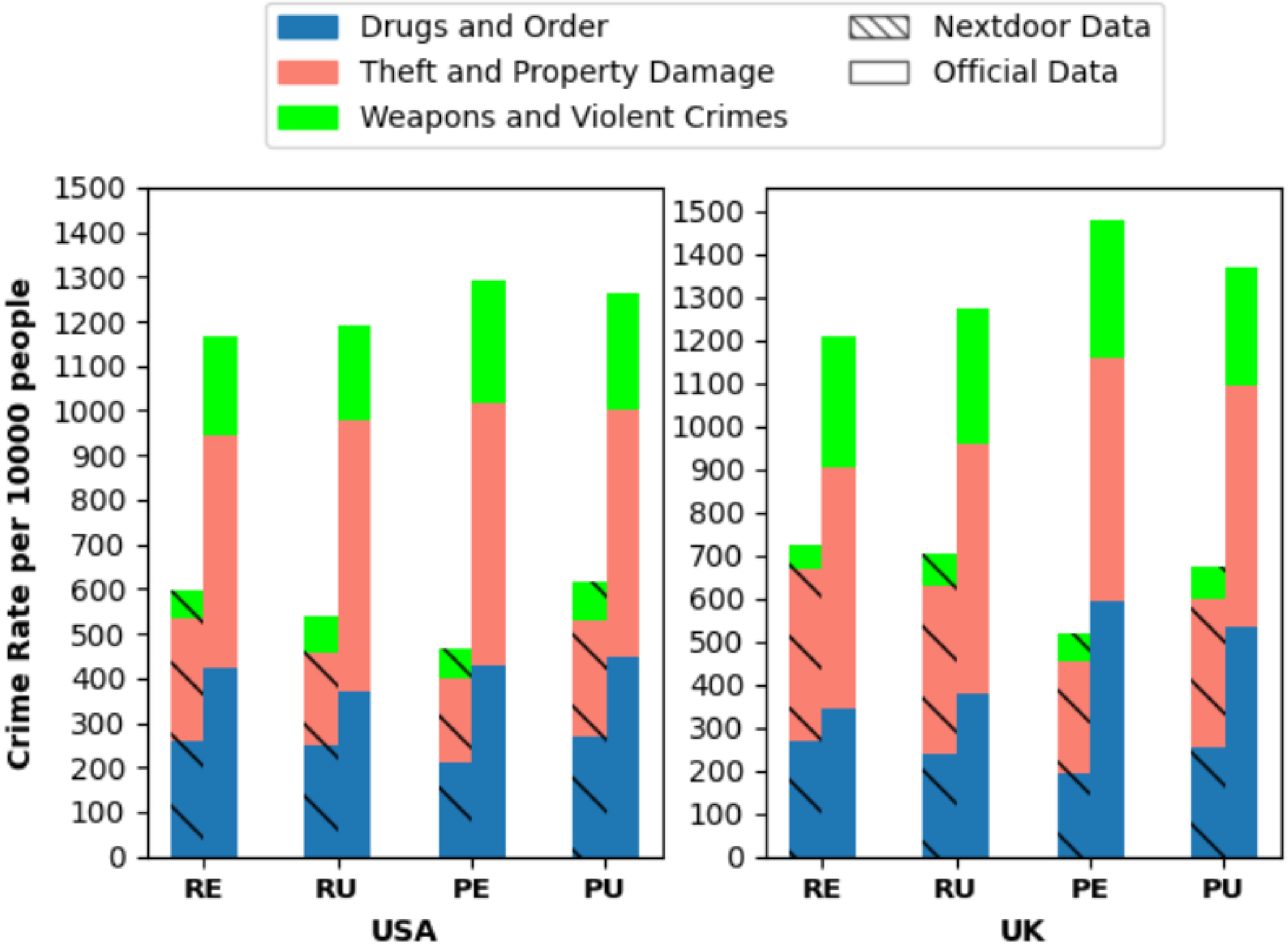
Neighborhoods within 25 (USA) and 3 (UK) miles

Atkinson Index for each neighborhood:

1 → inequality

0 → equality

Inequality matters for poorer neighborhoods



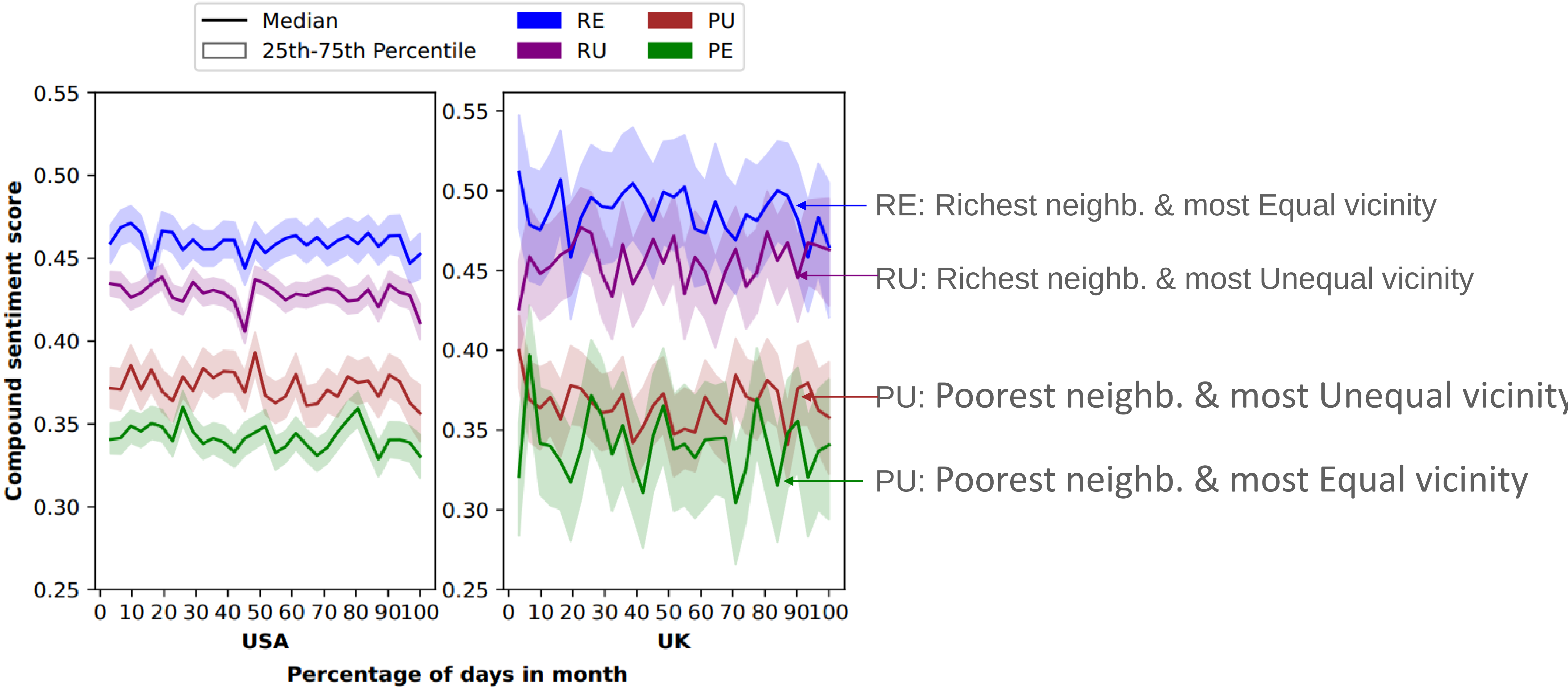
RE: Richest neighb. & most Equal vicinity

RU: Richest neighb. & most Unequal vicinity

PE: Poorest neighb. & most Equal vicinity

PU: Poorest neighb. & most Unequal vicinity

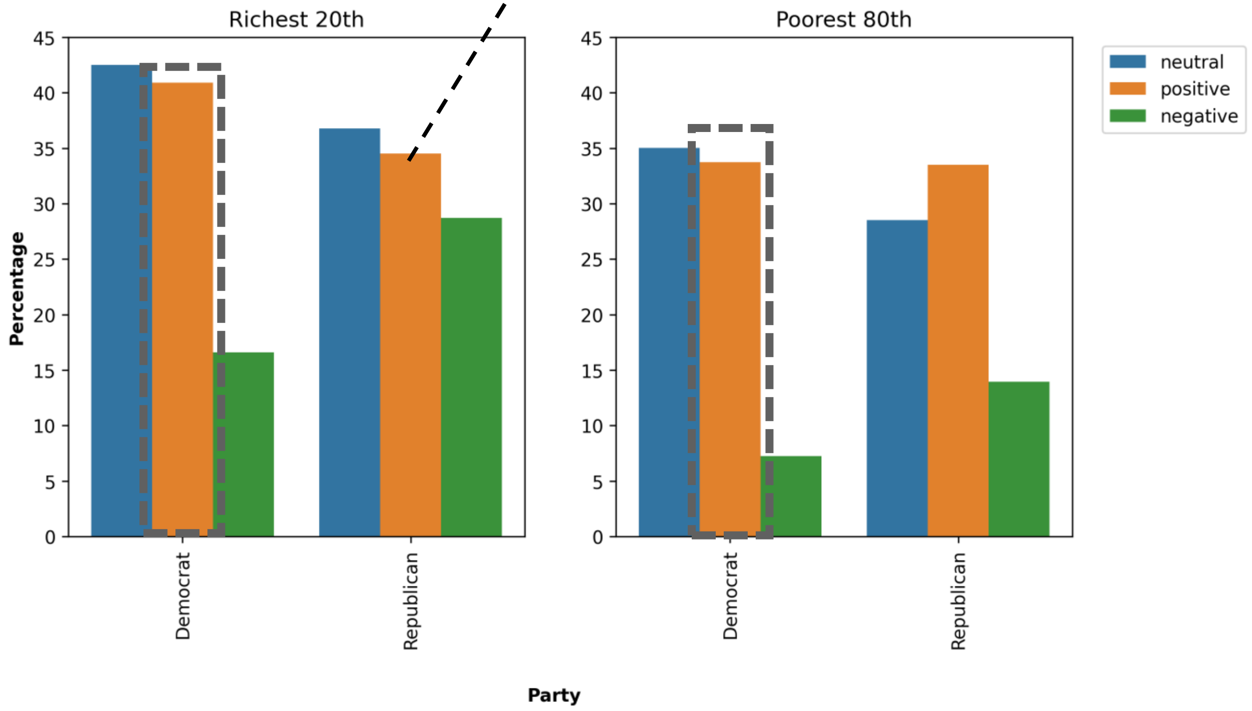
Inequality matters for poorer neighborhoods



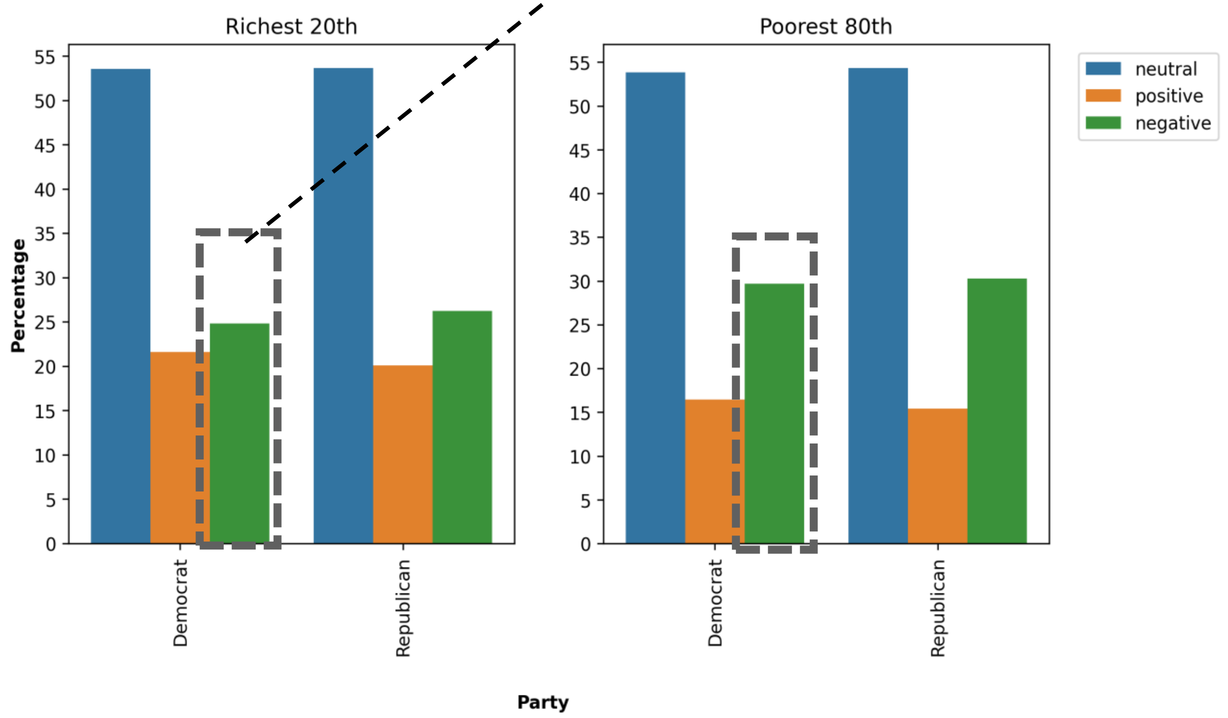
Richer neighborhoods have more positive stance

Politicians have higher positive stance.

ND users have negative stance on political posts



Twitter



Nextdoor