



# *Bridging Nations:* Quantifying the Influence of Multilinguals in the European Twitter Network

#### Julia Mendelsohn

● juliamendelsohn.github.io
▶ @jmendelsohn2
➡ juliame@umich.edu



#### **David Jurgens**

⊕ jurgens.people.si.umich.edu
❤ @david\_jurgens
➡ jurgens@umich.edu







#### Ceren Budak

Cbudak.com

- 🕑 @cerenbudak
- 🔀 cbudak@umich.edu



Information shared on social media can spread quickly across the globe, crossing linguistic and national borders.







#### But how?





We use network analysis and causal inference to quantify the role of individual multilingual Twitter users in information exchange across countries and languages

## Quantifying the role of multilinguals

Structural Role

**Communication Influence** 

To what extent do multilingual users act as bridges?

ഹ്റ

How does having a multilingual friend impact one's sharing behavior?







How would X's structural role and communication influence be different if they were monolingual?

We use causal inference to isolate the effect of multilingualism.

#### 1: Quantifying the structural role of multilinguals





Unit	Users from <b>C<sub>1</sub></b> who post in <b>L<sub>1</sub></b>		
Treatment	Posting in $L_1^{and} L_2^{c}$		
Outcome	Betweenness centrality (log)		



#### 2: Sharing links across language boundaries

Is a monolingual more likely to share **domains** from another language if they have a multilingual friend?

Unit	Monolinguals from $C_1$ who use $L_1$
Treatment	Having a multilingual friend who posts in L <sub>1</sub> and L <sub>2</sub>
Outcome	Sharing a <b>domain</b> associated with L <sub>2</sub>



#### 3: Spreading hashtags across language boundaries



Is a monolingual more likely to share **hashtags** from another language if they have a multilingual friend?

Unit	Monolinguals from <b>C<sub>1</sub></b> who use <b>L<sub>1</sub></b>
Treatment	Having a multilingual friend who posts in L <sub>1</sub> and L <sub>2</sub>
Outcome	Sharing a <b>hashtag</b> associated with L <sub>2</sub>



#### Constructing networks for pairs of countries

- Undirected network of mutual mentions from Decahose ('12-'19)
- Location inference to create network subsets containing all nodes and edges from C<sub>1</sub> UC<sub>2</sub>
- **C<sub>1</sub>**, **C<sub>2</sub>** must have single, distinct official/dominant languages
  - Switzerland excluded due to multiple languages
  - O Germany and Austria cannot be a pair together



#### Constructing networks for pairs of countries

- Undirected network of mutual mentions from Decahose ('12-'19)
- Location inference to create network subsets containing all nodes and edges from C<sub>1</sub> UC<sub>2</sub>
- **C<sub>1</sub>**, **C<sub>2</sub>** must have single, distinct official/dominant languages
  - Switzerland excluded due to multiple languages
  - O Germany and Austria cannot be a pair together



#### Constructing networks for pairs of countries

- Undirected network of mutual mentions from Decahose ('12-'19)
- Location inference to create network subsets containing all nodes and edges from C<sub>1</sub> U C<sub>2</sub>
- C<sub>1</sub>, C<sub>2</sub> must have single, distinct official/dominant languages
  - Switzerland excluded due to multiple languages
  - O Germany and Austria cannot be a pair together



## Classifying monolinguals and multilinguals

- Determine language usage based on individual tweets
  - Decahose 10% sample from Feb 2018 June 2020, no retweets
  - Twitter's automatic language identification
  - Minimum threshold of 5 total tweets
- A multilingual user of L<sub>1</sub> and L<sub>2</sub> has at least 10% of their tweets in each language

### Study 1: Estimating effects on betweenness

For all valid ( $C_1$ ,  $C_2$ ). If  $\geq 100 C_1$  users in post in  $L_1 \& L_2$ :



#### Effect of multilinguality on betweenness centrality

• For majority of pairs (C<sub>1</sub>, C<sub>2</sub>): posting in L<sub>1</sub> & L<sub>2</sub> increases betweenness centrality.



Change in betweenness (ATT) However, 1.0 0.73 \*\*\* 0.59 \*\*\* effects vary 0.42 \*\*\* 0.5 widely across 0.0 Polish/German Turkish/German Polish/German bilinguals bilinguals bilinguals country pairs in Germany in Poland in Germany





• We include retweets to determine top domains/hashtags by language and users' sharing of domains/hashtags

### Identifying hashtags by language



Hashtags shared in all tweets from all European countries

Hashtags shared in all tweets in language L<sub>1</sub> Filter hashtags in the top 200 overall or with count  $\leq$  10 Log-odds to get ≤ 2k hashtags most associated with L<sub>1</sub>

- Trending hashtags change fast, so top hashtags calculated for 14 day intervals.
- A user shares a hashtag from language L if they share a top
   L hashtag from interval *i* during interval *i* or *i*+1

### Examples of hashtags and domains

German	Portuguese	Turkish	Polish	English
cdu	fcporto	çağlarertuğrul	pis	oddoneout
spd	todosportugal	sustunuz	konwencjapis	remain
merkel	capricórnio	pazartesi	topmodel	eastenders
klimaschutz	aquário	cumartesi	thevoiceofpoland	liarjohnson
noafd	sportingcp	burcuözberk	kaczyński	ncfc
tagesschau.de	publico.pt	tele1.com.tr	wpolityce.pl	manchestereveningnews.co.uk
faz.net	record.pt	haber.sol.org.tr	niezalezna.pl	whounfollowedme.org
spiegel.de	maisfutebol.iol.pt	diken.com.tr	dorzeczy.pl	theneweuropean.co.uk

## Study 2: Estimating effects on domain use

For all valid  $(C_1, C_2)$ . If  $\geq 100 C_1$  monolinguals have  $(L_1, L_2)$  multilingual neighbor:



## Study 3: Estimating effects on hashtag use

For all valid  $(C_1, C_2)$ . If  $\geq 100 C_1$  monolinguals have  $(L_1, L_2)$  multilingual neighbor:



## Multilingual friends often influence sharing



Multilingual treatment has significantly positive effect on information sharing for nearly half of country pairs (p < 0.05)

Effect of having a multilingual friend on sharing another language's domains

The effect of multilingual users varies widely across country pairs



Effect of having a multilingual friend on sharing another language's hashtags



#### How can we explain this variation?

#### Geography

## Migration

#### Language







## We explore how these factors correlate with the role of multilingual Twitter users in information exchange

Note: while the previous causal inference studies *quantify* the effect of multilinguals compared to monolinguals, this next analysis *compares* the effect of multilinguals across country pairs

#### Regression of causal effects on country/language factors

- Dependent variables for **C**<sub>1</sub> in country pair (**C**<sub>1</sub>, **C**<sub>2</sub>):
  - Potential outcome difference of posting in L<sub>1</sub> & L<sub>2</sub> on betweenness
  - Risk ratio of having  $(L_1, L_2)$  multilingual friend on  $L_2$  domain sharing
  - Risk ratio of having (L<sub>1</sub>, L<sub>2</sub>) multilingual friend on L<sub>2</sub> hashtag sharing
- Independent variables:
  - Total pop. of C<sub>1</sub> & C<sub>2</sub>, ratio between populations (World Bank, 2018)
  - % foreign-born pop. of C<sub>1</sub> & C<sub>2</sub> (World Bank, 2015)
  - % of pop. who are migrants from C<sub>1</sub> to C<sub>2</sub> (OECD, 2018)
  - % of pop. who are migrants from C<sub>2</sub> to C<sub>1</sub> (OECD, 2018)
  - Distance (km) between population centers of C<sub>1</sub> & C<sub>2</sub> (CEPII GeoDist)
  - Linguistic closeness between L<sub>1</sub> and L<sub>2</sub> based on genetic relationship

## Measuring linguistic closeness

Level 0: no established relationship

• (German, Turkish), (Spanish, Hungarian)



Does not consider similarities due to language contact (e.g. lexical borrowing)

#### Multilinguals play greater role in info exchange when C<sub>1</sub> & C<sub>2</sub> have higher foreign-born population

Multilinguals may be	Dependent variable: causal effects of $L_1/L_2$ multilingualism in country $C_1$				
more likely to connect int'l offline networks		multilingual betweenness	multilingual friend domain usage	multilingual friend hashtag usage	
	$\begin{array}{ccc} \% \ C_1 & { m foreign-born} \\ \% \ C_2 & { m foreign-born} \end{array}$	$0.005^{***}$ 0.001	0.006 0.026**	0.016 0.067***	
	$\%$ migrants $C_1$ to $C_2$	$3.445^{*}$	55.517***	-0.130	
Hashtags and domains in	$\%$ migrants $C_2$ to $C_1$	$31.237^{***}$	96.209***	58.780***	
L may be more	distance	0.00001	$0.0004^{***}$	0.0003***	
internationally relevant	linguistic closeness	$-0.032^{***}$	0.080	$-0.118^{**}$	
,	Observations	172	219	219	
	$\mathbb{R}^2$	0.663	0.398	0.345	

Note:

- - - ---

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

# Multilinguals play greater role in info exchange when there is more migration between $C_1 \& C_2$

German-Turkish users have higher betweenness than German-Norwegian users. Their German friends are more likely to share L<sub>2</sub> links & hashtags.

Migration from C<sub>2</sub> to C<sub>1</sub> is more highly associated with role of (L<sub>1</sub>, L<sub>2</sub>) multilinguals in C<sub>1</sub> than migration from C<sub>1</sub> to C<sub>2</sub>

Dependent variable: causal effects of  $L_1/L_2$  multilingualism in country  $C_1$ multilingual multilingual friend multilingual friend betweenness domain usage hashtag usage  $0.005^{***}$ 0.006 0.016  $\% C_1$  foreign-born 0.067\*\*\*  $\% C_2$  foreign-born 0.001  $0.026^{**}$ % migrants  $C_1$  to  $C_2$  $3.445^{*}$ 55.517\*\*\* -0.130% migrants  $C_2$  to  $C_1$ 31.237\*\*\* 96.209\*\*\* 58.780\*\*\*  $0.0004^{***}$ 0.0003\*\*\* distance 0.00001 $-0.032^{***}$  $-0.118^{**}$ linguistic closeness 0.080Observations 172 219219  $\mathbb{R}^2$ 0.6630.3980.345Note: \*p<0.1; \*\*p<0.05; \*\*\*p<0.01

26

# Multilinguals have greater communication influence when $C_1 \& C_2$ are further apart

People may rely more on individual multilinguals to access info from afar, while info from nearby may be spread by other means



_ op on a on o o				
	multilingual betweenness	multilingual friend domain usage	multilingual friend hashtag usage	
$\% C_1$ foreign-born	0.005***	0.006	0.016	
$\% C_2$ foreign-born	0.001	0.026**	0.067***	
% migrants $C_1$ to $C_2$	$3.445^{*}$	55.517***	-0.130	
$\%$ migrants $C_2$ to $C_1$	31.237***	96.209***	58.780***	
distance	0.00001	0.0004***	0.0003***	
linguistic closeness	$-0.032^{***}$	0.080	$-0.118^{**}$	
Observations	172	219	219	
$\mathbb{R}^2$	0.663	0.398	0.345	

Dependent variable: causal effects of  $L_1/L_2$  multilingualism in country  $C_2$ 

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

# Multilinguals have greater influence when $L_1 \& L_2$ are less closely related

Users of similar languages may not need to rely on individual multilinguals to share info across languages



	multilingual betweenness	multilingual friend domain usage	multilingual friend hashtag usage
$\% C_1$ foreign-born	0.005***	0.006	0.016
$\% C_2$ foreign-born	0.001	0.026**	0.067***
% migrants $C_1$ to $C_2$	$3.445^{*}$	55.517***	-0.130
$\%$ migrants $C_2$ to $C_1$	$31.237^{***}$	96.209***	58.780***
distance	0.00001	$0.0004^{***}$	0.0003***
linguistic closeness	$-0.032^{***}$	0.080	$-0.118^{**}$
Observations	172	219	219
$\mathbb{R}^2$	0.663	0.398	0.345
Note:		*p<0.1; *	**p<0.05; ***p<0.01

Dependent variable: causal effects of  $L_1/L_2$  multilingualism in country  $C_1$ 

Exception: domains Maybe language is less relevant because of multilingual sites? Or because domains are closely tied to countries/regions?

# Multilinguals have greater influence when $L_1 \& L_2$ are less closely related

Users of similar languages may not need to rely on individual multilinguals to share info across languages



	multilingual betweenness	multilingual friend domain usage	multilingual friend hashtag usage
$\% C_1$ foreign-born	0.005***	0.006	0.016
$\% C_2$ foreign-born	0.001	0.026**	0.067***
% migrants $C_1$ to $C_2$	$3.445^{*}$	55.517***	-0.130
% migrants $C_2$ to $C_1$	$31.237^{***}$	96.209***	58.780***
distance	0.00001	$0.0004^{***}$	0.0003***
linguistic closeness	$-0.032^{***}$	0.080	$-0.118^{**}$
Observations	172	219	219
$\mathbb{R}^2$	0.663	0.398	0.345
37		*	

Dependent variable: causal effects of  $L_1/L_2$  multilingualism in country  $C_1$ 

Note:

\*p<0.1; \*\*p<0.05; \*\*\*p<0.01

### Next Steps

- Robustness checks for data collection and causal inference methods
- How does the effect of multilingual friends depend on topic?
  - Do multilingual friends matter more for localized issues (e.g. politics), while topics like pop culture and sports may circulate more easily via other mechanisms?
  - Can address possible topical confound for language-specific hashtags discussed earlier
- Explore multilinguality as continuous variable & "implicit" multilinguality
- Disentangle country and language assumptions to study:
  - Users in highly multilingual countries (e.g. Switzerland)
  - Users of minority languages (e.g. Basque)
  - Code-mixing and linguistic fluidity within tweets

#### Conclusion

- To understand how information spreads around the world, we need to understand how it crosses languages and countries.
- We quantify the structural role and communication influence of multilinguals in 3 studies of European Twitter
- Multilinguals play important role in information exchange
- Effects vary across country pairs, and are associated with linguistic, geographic, and migration relationships
- This work is ongoing, and we would love any suggestions!











#### Thank you! *Merci vielmal!*

#### Julia Mendelsohn

⊕ juliamendelsohn.github.io
♥ @jmendelsohn2
➡ juliame@umich.edu



#### **David Jurgens**

● jurgens.people.si.umich.edu
● @david\_jurgens
> jurgens@umich.edu



#### 



#### Ceren Budak

Cbudak.com

- У @cerenbudak
- 🔀 cbudak@umich.edu

