AR

AR-EN

EN

TL

Sociolinguistic Analysis of Twitter in Multilingual Societies

Suin Kim, Ingmar Weber, Li Wei, Alice Oh

KAIST, QCRI, BBK

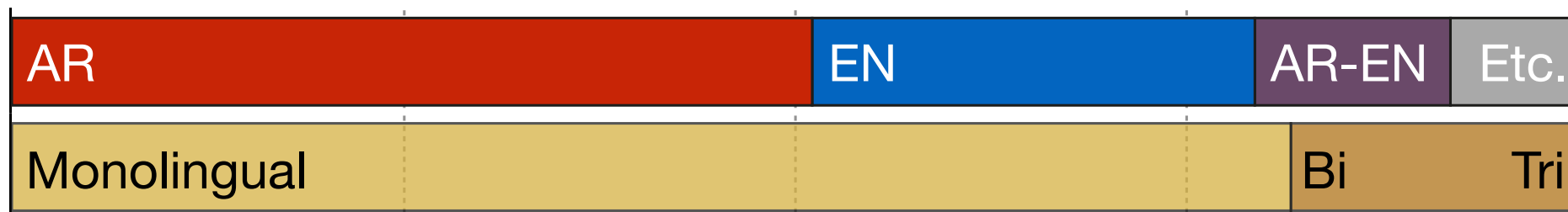# Modern Society is Multilingual

- In the multilingual societies…

  - The language we speak is an integral part of our culture

  - A spoken language can be a political statement with wide-ranging implications

  - The dynamics of language choice could explain the prosperity or demise of local languages

- We study the phenomena of multilingual societies and the role that bilinguals play in them, questioning

  1. *Potential segregation*: **Do monolinguals cluster together?**

  2. *Social status*: **Is there a "High Class" language?**

  3. *Social capital*: **Do bilinguals act as a bridge? If so, how?**

  4. *Language preservation*: **Will language converge to single lingua franca?**

  5. *Code switching*: **Do bilinguals use different language for different stories?**
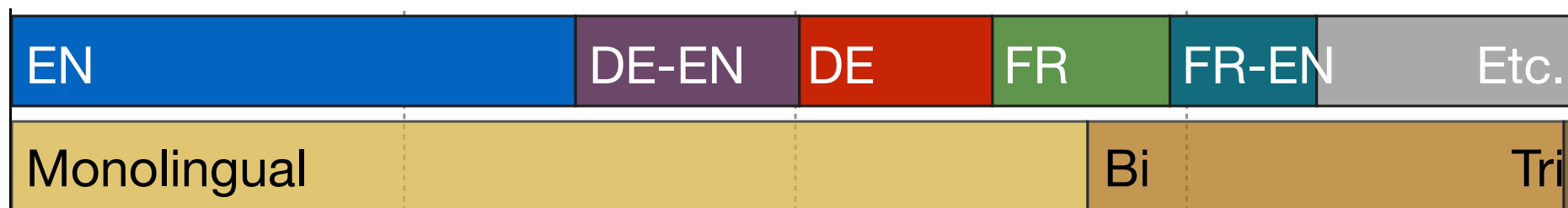
Twitter Data

- We captured users with geographic information in their profile

- We crawled all available tweets and detected languages, classifying each user into language categories
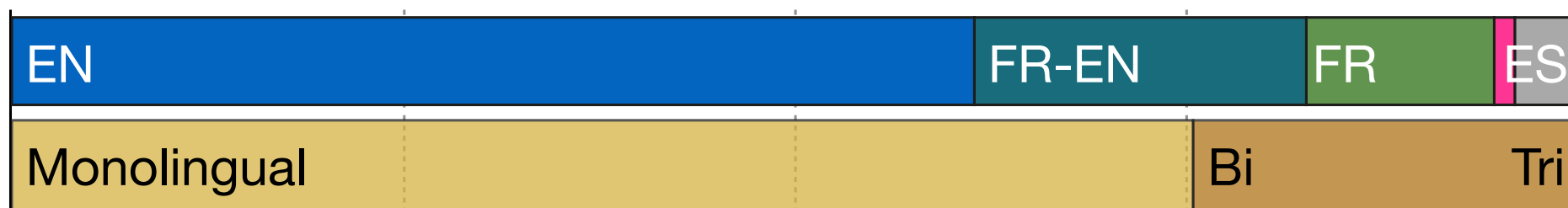
Qatar

| AR | EN | AR-EN | Etc. |

| Monolingual | Bi | Tri |

Switzerland

| EN | DE-EN | DE | FR | FR-EN | Etc. |

| Monolingual | Bi | Tri |

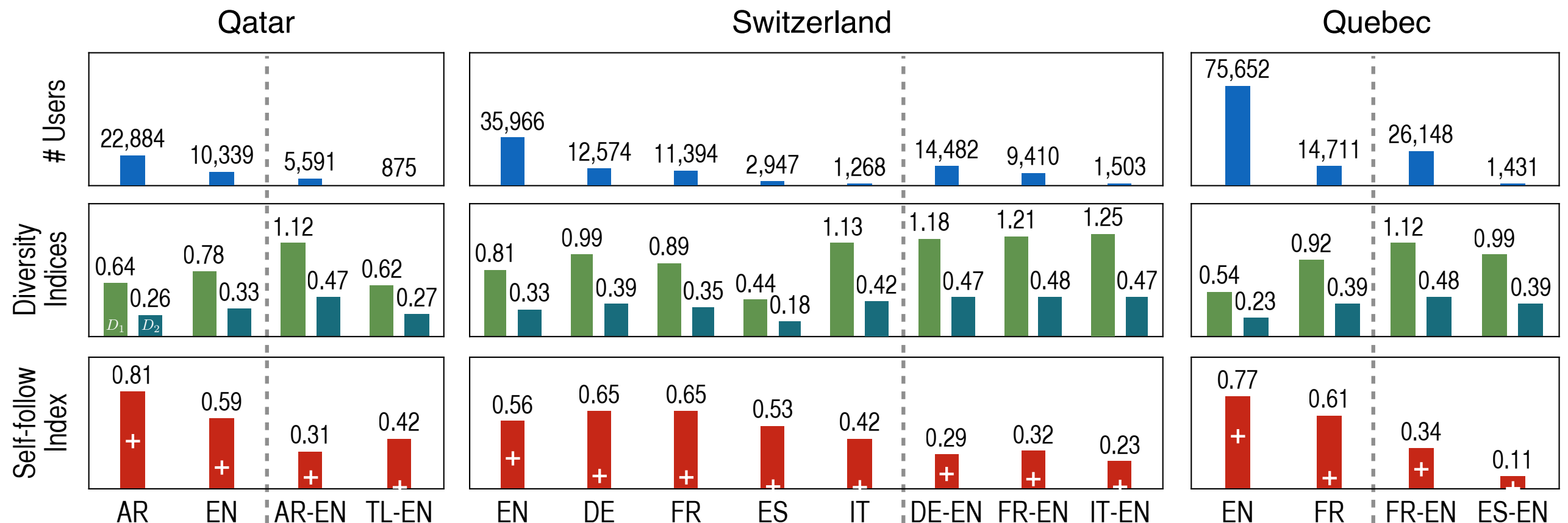Quebec

| EN | FR-EN | FR | ES |

| Monolingual | Bi | Tri |

Q1: Do Monolinguals Cluster Together?

- **Diversity indices**: How users make diverse connection in language speaking groups?

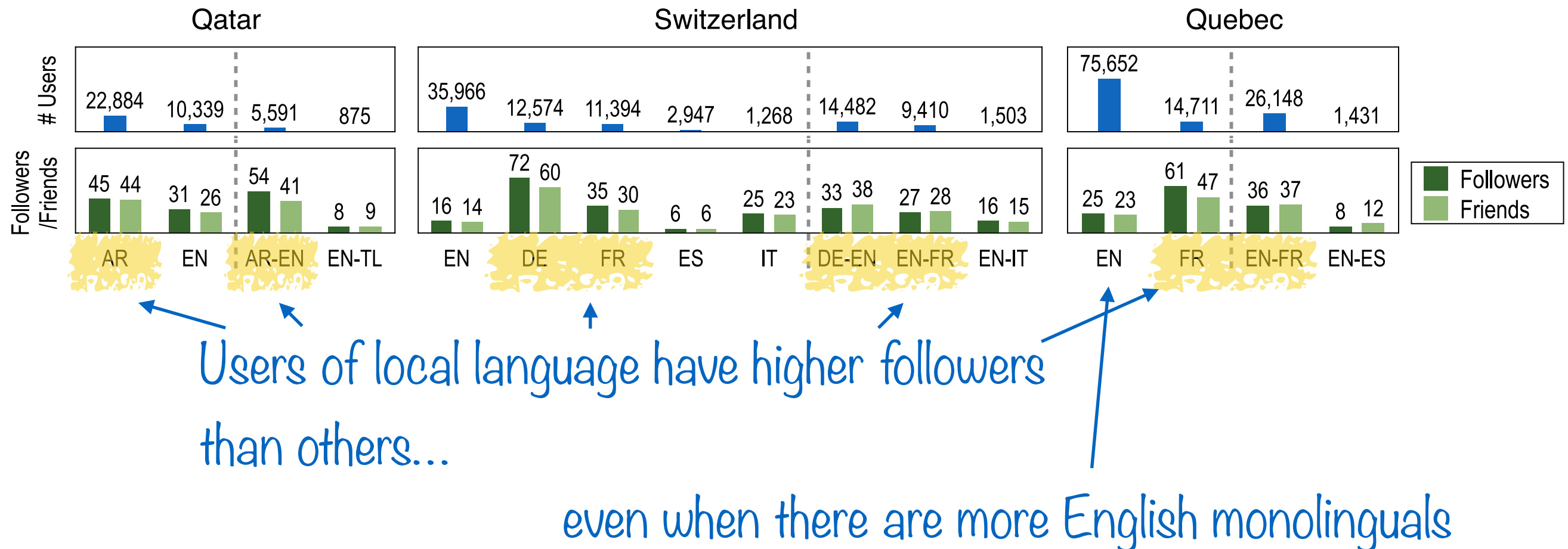- **Self-follow index**: How users tend to make homogeneous connections?

Monolinguals have lower diversity index and higher self-follow index

# A1: YES.

Q2: Is There a More Influential Language?

- Studies have shown that tweets from a user with a high in-degree are more likely to be retweeted (Cha 2010, Suh 2010)

Qatar — # Users: AR 22,884; EN 10,339; AR-EN 5,591; EN-TL 875

Followers/Friends: AR 45/44; EN 31/26; AR-EN 54/41; EN-TL 8/9

Switzerland — # Users: EN 35,966; DE 12,574; FR 11,394; ES 2,947; IT 1,268; DE-EN 14,482; EN-FR 9,410; EN-IT 1,503

Followers/Friends: EN 16/14; DE 72/60; FR 35/30; ES 6/6; IT 25/23; DE-EN 33/38; EN-FR 27/28; EN-IT 16/15

Quebec — # Users: EN 75,652; FR 14,711; EN-FR 26,148; EN-ES 1,431

Followers/Friends: EN 25/23; FR 61/47; EN-FR 36/37; EN-ES 8/12

Users of local language have higher followers than others…

even when there are more English monolinguals

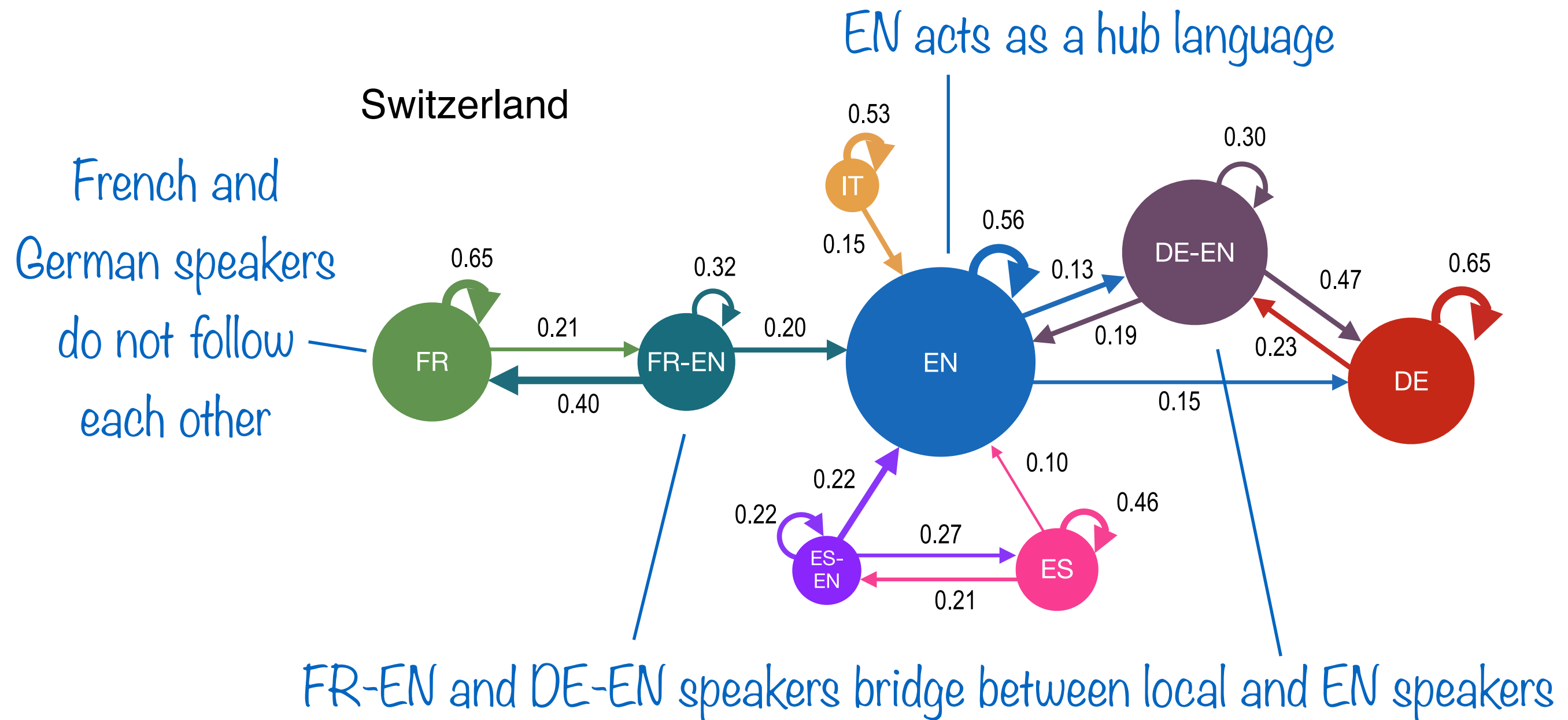- Tweeting in the local language exert higher influence within the regional network

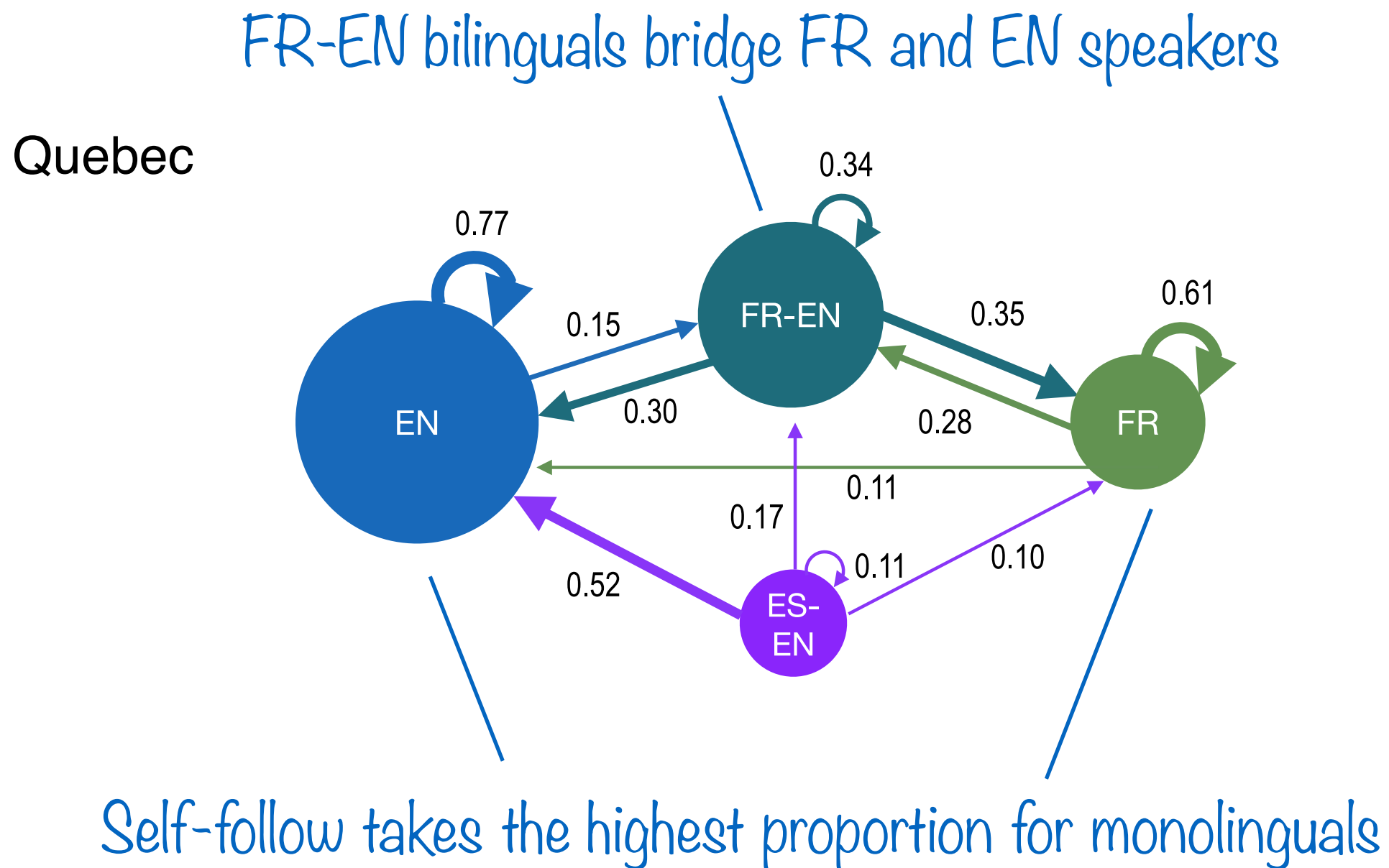**A2: YES (LOCAL LANGUAGES).**

Q3: Do Bilinguals Act as Bridge? If So, How?
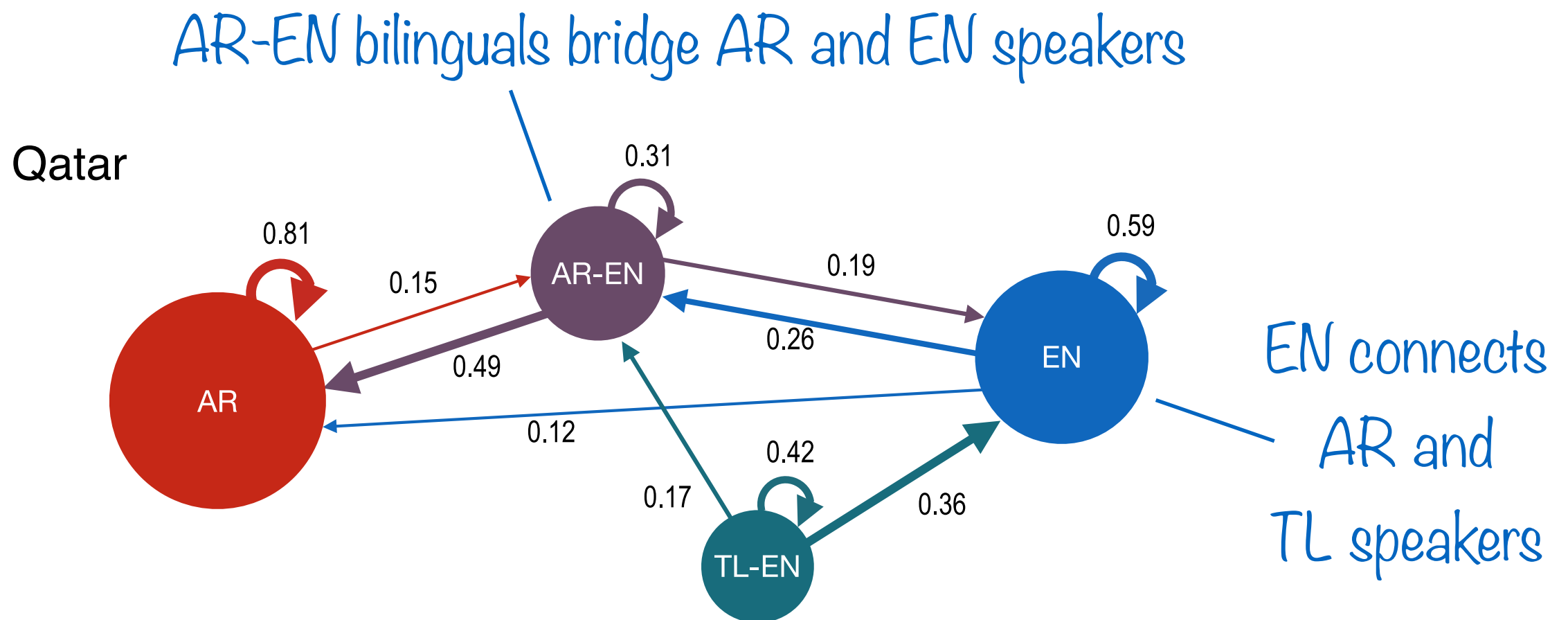
Per-user averaged ratio of followings
Ratio < 0.10 not shown

Q3: Do Bilinguals Act as Bridge? If So, How?

EN acts as a hub language

Switzerland

French and German speakers do not follow each other

FR-EN and DE-EN speakers bridge between local and EN speakers

Q3: Do Bilinguals Act as Bridge? If So, How?

FR-EN bilinguals bridge FR and EN speakers

Quebec

Self-follow takes the highest proportion for monolinguals

Q3: Do Bilinguals Act as Bridge? If So, How?

- English acts as a hub language

- Bilingual group bridges two monolingual groups

- In-group following takes the largest proportion for monolingual groups

- Monolingual users do not follow other monolingual users

Switzerland
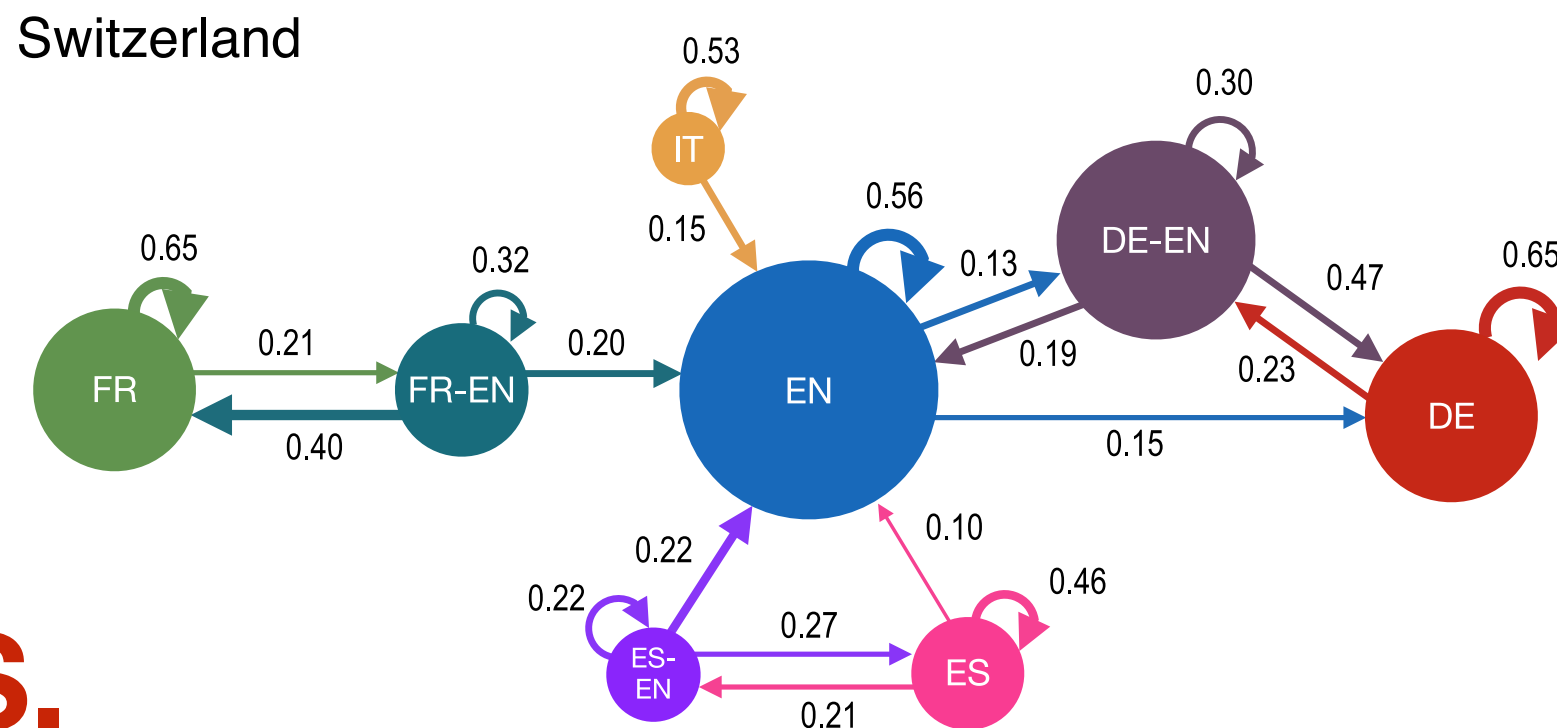
A3: YES.

Q4: Will Languages Converge to Single Lingua Franca?

- How should multilingual users choose their language?
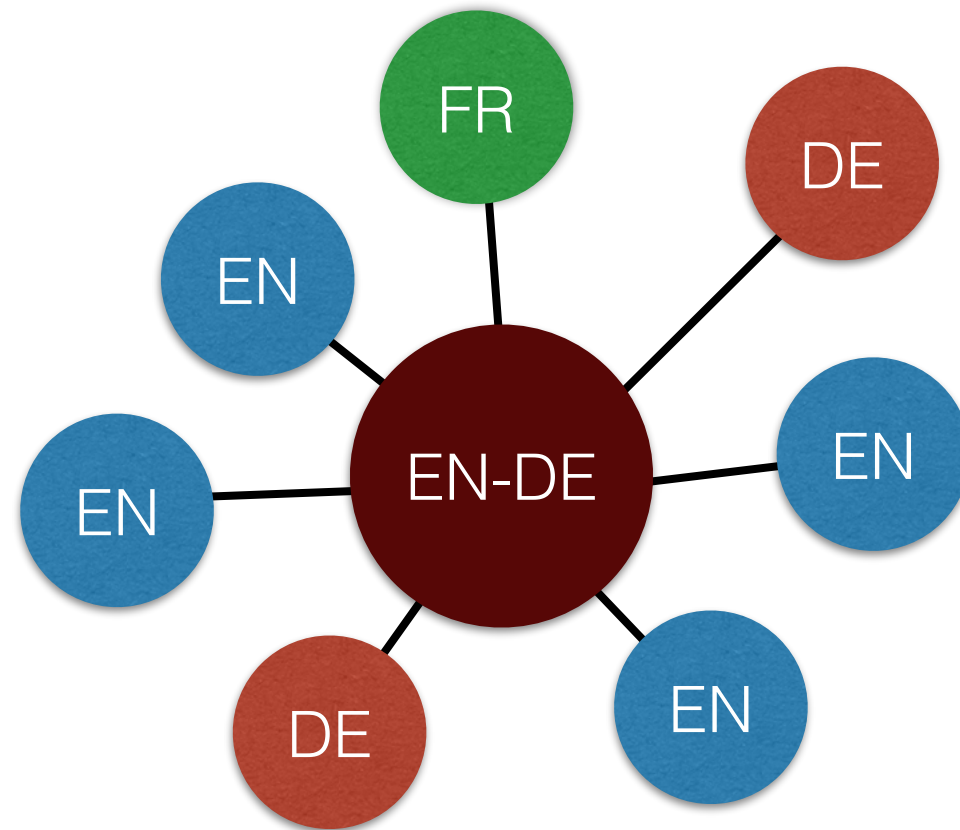
A game-theoretic approach to "maximise the audience"
*will force German-English speakers to tweet in English*

- Will it lead to a global convergence to a single lingua franca?

- What is the threshold that users change their language?

Q4: Will Languages Converge to Single Lingua Franca?

Bilingual users "mimic" the language mix observed along their followers

A non-steep, centered language mix indicates the language convergence is not being accelerated

A4: NO.

Bilingual user with no Arabic friends speaking 50% Arabic in all the tweets
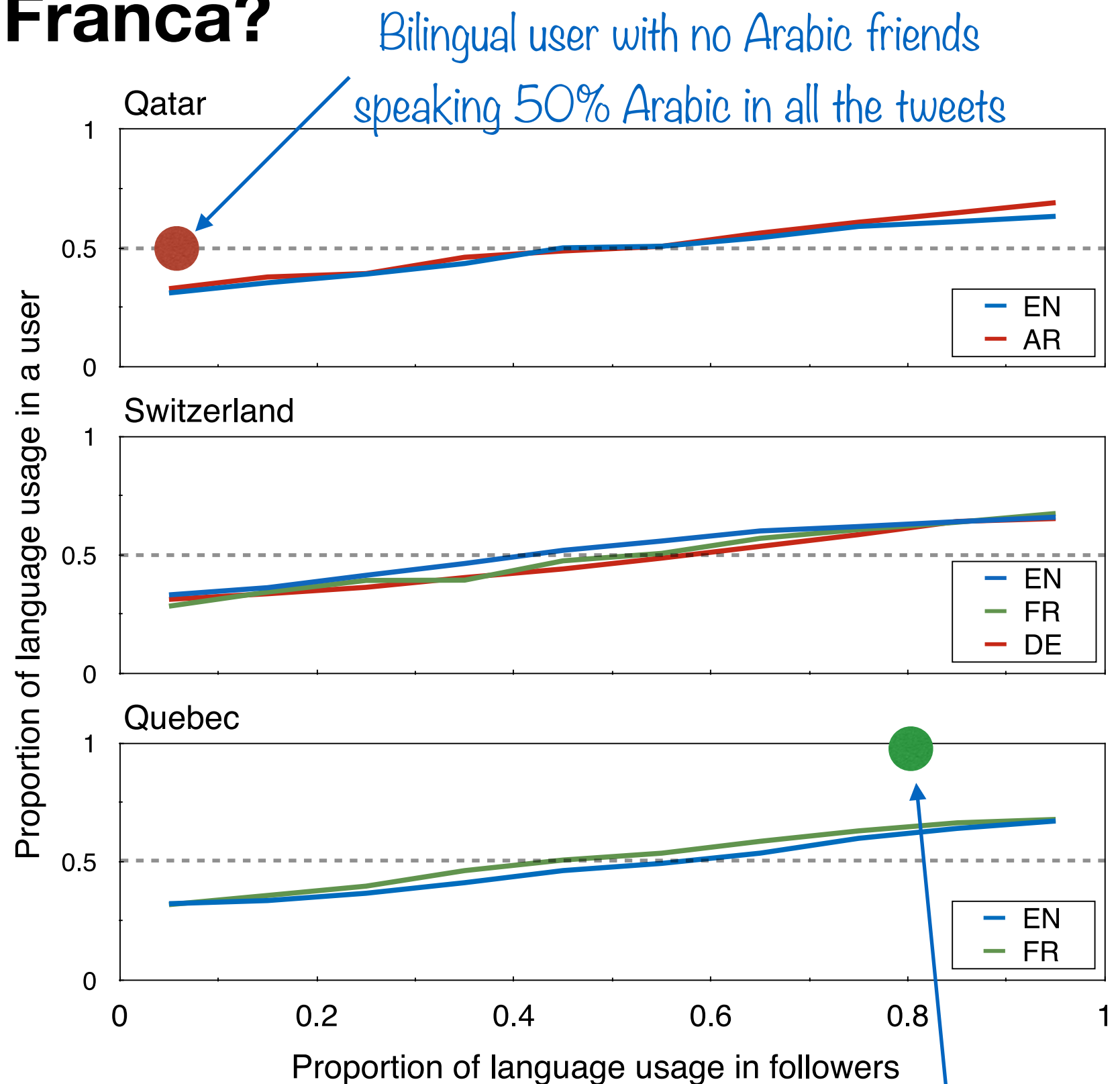
Bilingual user with 80% French friends speaking French all the time

Q5: Do Bilinguals Use Different Language for Different Stories or Topics?

- Use probabilistic topic model (LDA) on tweets from bilinguals

- Grouped tweets having hashtags with same meaning but different languages: (#switzerland - #suisse, #qatar - #قطر)

Bilingual users post informational and political tweets in local language

events, tourism, photography, leisure-related tweets in English for the non-local audience

QATAR, AR-EN BILINGUAL		
#	Label	Top words
#قطر	development government emir/god gcc countries	national doha development vision government national qa foundation hamad bin god sheikh tamim emir doha egypt kuwait bahrain uae
#qatar	national day sports event photography recruitment	day national doha gcc about today doha volleyball football uae photo instagram katara instagood love job please send cv recruitment org

SWITZERLAND, EN-FR BILINGUAL		
#suisse	politics news/radio recruitment ski	more country politics no federal ch news tar info thank you radio job manager head senior engineer ski weather romandie snow rentals
#switzerland	scenery greeting party/club wine/fashion	lake sun sky sunset beautiful night love like time good show morning club dj party enjoy welcome house fashion valais basel wine beautiful

12

Summary

- A large-scale computational analysis of language use and network characteristics in multilingual societies

- Findings

  1. **Monolinguals cluster together**, while bilinguals do not

  2. Users speaking **local language have more influence**

  3. Language-mixing ratio of bilingual users **mirrors the mix of their followership**

  4. **Bilinguals bridge between monolinguals** with **English as a hub**

  5. Bilinguals express informative/political/**debatable topics in a local language**, while posting event/tour/**enjoyment topics on the English**