

Euskara eta Gazteak Sare Sozialetan: Erabiltzaile gazte eta helduen edukiei buruzko azterketa

Joseba Fernandez de Landa eta Rodrigo Agerri
HiTZ Center - Ixa, University of the Basque Country UPV/EHU
{joseba.fernandezdelanda, rodrigo.agerri}@ehu.eus

Asmoa

- **Gazteak** eta **euskara** aztertu **sare sozialetan**.
 - Gazteen errealitate ezezagunera hurbilpena.
 - Erronka berrietara euskara nola egokitzen ari den ezagutu.
- Twitterren gazteak **zertaz** eta **zeinekin** aritzen diren azaldu.

Ekarpenak

- Ikerketa sozialerako bide berriak irekitzea.
 - Datu-iturri berrien ustiaketa.
 - Teknika berrien proposamen eta aplikazioa.
 - Gizarte eta konputazio zientziak bateratu.

Datuak

- **Euskal txiolariak identifikatu** Twitterren
 - Umap.eus
 - **7.977** erabiltzaile euskaldun.
- **Datu jasoketa** Twitterretik
 - *tweepy*
 - Erabiltzaile euskaldunen eduki osoa.
 - Bakoitzetik 3.200 txio gehienez.
- **Jasotako datuak**
 - 5,2 M txio pertsonal
 - 3,3 M euskaraz (% 61)
 - 5,5 M birtxio
 - 2,9 M euskaraz (% 53)

Metodoak

Adinaren iragarpena

- **Gazte / heldu identifikatzailea**
 - Idazteko eraren arabera sailkatu txioak: formal / informal
 - Ikasketa Automatikoko sailkatzailea: *Ixa-pipes*
 - Txio informalen kontzentrazioa gazte/heldu den erabakitzeke

Textu eta harremanen analisisa

- **a) Txiolarien ohiko gaiak identifikatu**
 - Txio pertsonalen testuan oinarriturik gaiak azalderatu
 - *Topic-modeling* teknika: *LDA* algoritmoa
- **b) Txiolarien harremanak azalderatu**
 - Birtxioetan oinarrituta harremanen grafoa sortu
 - Azpi-taldeak sortzeko *Modularity* algoritmoa

Bistaraketa

Gazte / heldu banaketa

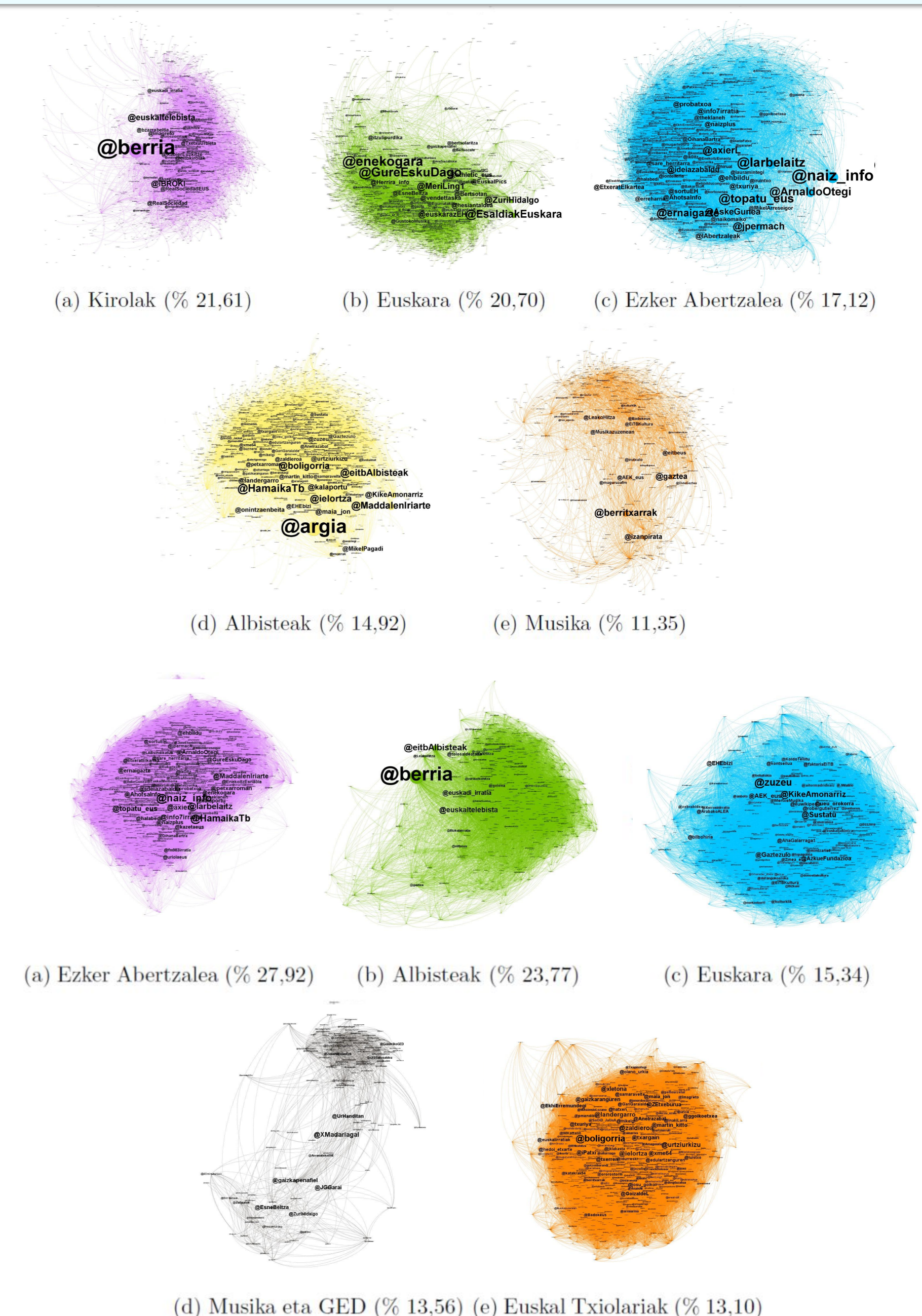
- **Gazteak:**
 - **1.579** erabiltzaile
 - **0,5 M** txio euskaraz (a)
 - **0,4 M** birtxio euskaraz (b)

- **Helduak:**
 - **5.508** erabiltzaile
 - **2,6 M** txio euskaraz (a)
 - **2,4 M** birtxio euskaraz (b)

a) Gaiak



b) Harremanak



Ondorioak

- **Euskal txiolari gazteak zertaz eta zeinekin.**
 - Gertukoekin komunikatzeko eta egunerokotasuneko gertakariak adierazteko erabiltzen dute.
 - Harremanak euskal herriko gai politikoekin eta aisialdiarekin lotzen dira.
- **Euskara presente dago Twitterren:** testuinguru berrietara moldatzeko gai da.
 - 6 milioi txio ia 8.000 erabiltzailearenak.
 - Komunikabide eta politikarekin dago erlazionatua.
 - Euskaldunen gaiei buruz aritzeko erabiltzen da gehienbat.