Twitter Twitter on the Wall, which University's the Fairest of Them All?

Exploring brands' social perception on social media using Big Data

Big Data in Psychology, Trier 7-9 June 2018

UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

Biella Marco
Zogmaister Cristina
Ceolato Stefania
Parozzi Elena

Goals

- **Method**[1] to explore the perception of different targets
  - Institutions (eg. Universities, Brands…)
  - Persons (eg. Political Figures, Influencers, Opinion Leaders…)
  - Social phenomena
- **Index** to quantify the perception of each target along certain dimensions
  - Dimensions depend on the nature of the targets
- Map the **perceptual space** where all targets lay

1 this method is a modifyied version of the social perceptio score used by Culotta & Cutler (Culotta, A., & Cutler, J. 2016. *Mining Brand Perceptions from Twitter Social Networks*. Marketing Science 35(3):343-362)

Crafting

How to **assess** and **quantify** perception?

Theoretical and Methodological Framework

- Similar accounts attract **similar** followers
- Accounts closely related to a given dimension are **exemplar** accounts, they are prototypical of that dimension

If

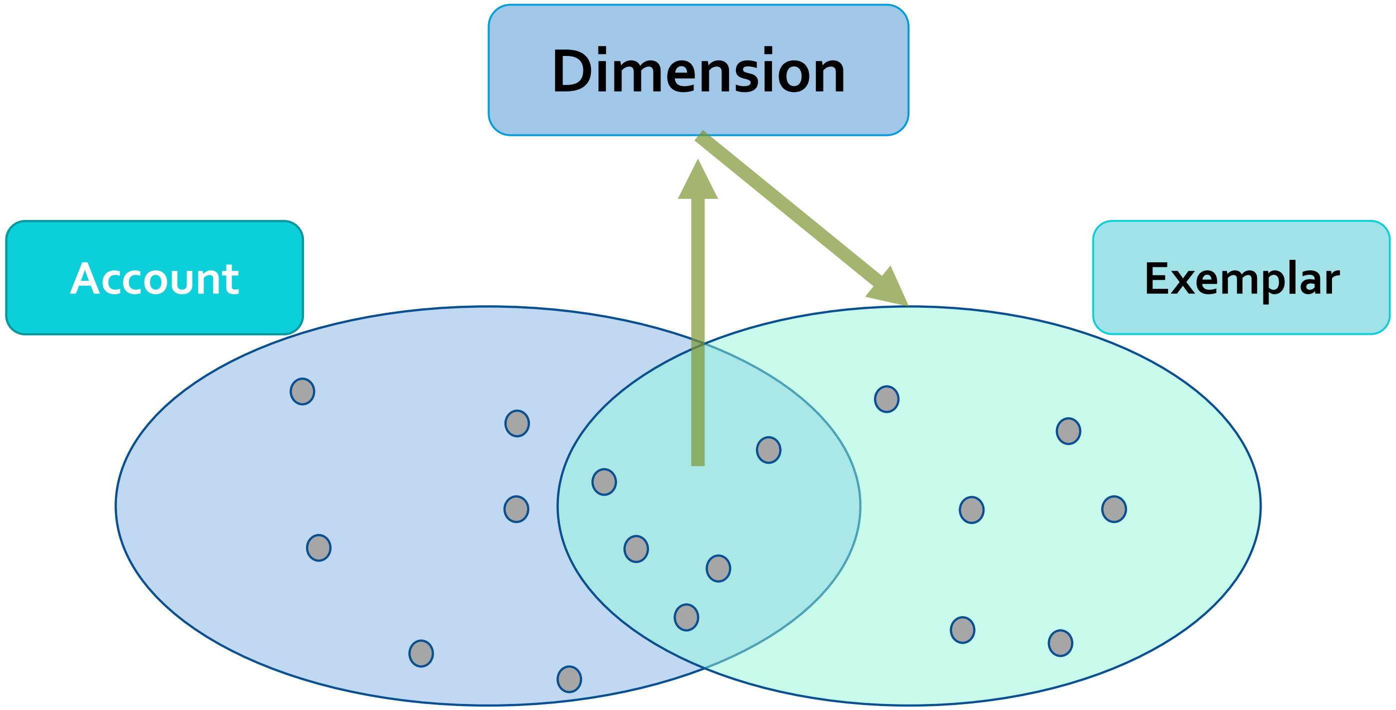
- A follower is interested in a given **dimension** (eg. Research) it follows accounts related to that dimension
- If a follower follows a **target** account, that account has some characteristics related to the dimension

Therefore

- The more the follower a **target** account shares with **exemplar** accounts, the stronger the bond between the **dimension** and the target

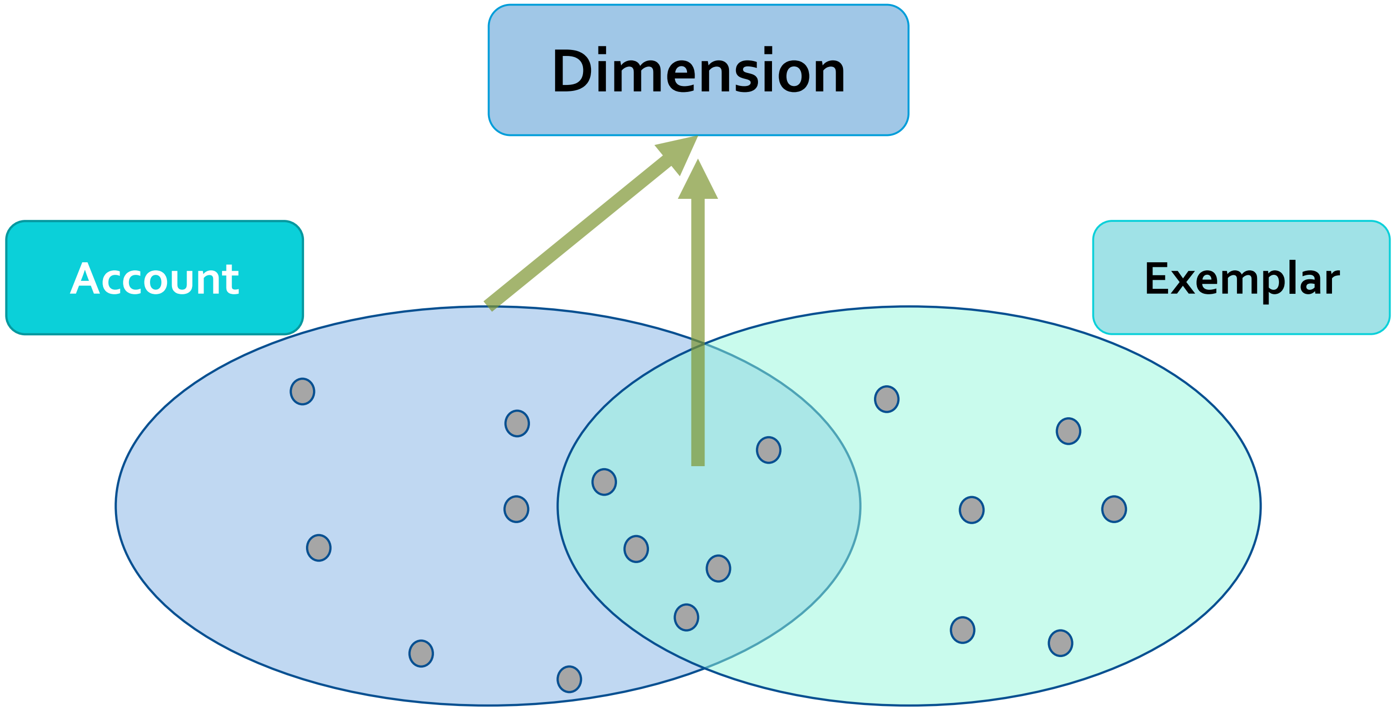
Theoretical and Methodological Framework

- Similar accounts attract **similar** followers
- Accounts closely related to a given dimension are **exemplar** accounts, they are prototypical of that dimension



Theoretical and Methodological Framework

- Similar accounts attract **similar** followers
- Accounts closely related to a given dimension are **exemplar** accounts, they are prototypical of that dimension



Building Dimensions
- Choosing exemplars

- Culotta & Cutler 2016 selected **exemplar** accounts that appeared in the first lists retrieved using the name of the **dimension** as a query keywords
  - «Objective» selection
  - Rigid dimensions
- In the present research we selected **exemplar** accounts
  - Each author came up with a given list of exemplars
  - Lists were merged
  - Each account in the merged list was re-classified in one dimension by independent raters

Building Dimensions
- Choosing exemplars

- Pros
  - Choose wisely (based on research questions)
  - Flexibility
  - Ad-hoc dimensions
    - *Silicon-valleyness*
    - *Left-winginess, right-winginess*
  - Crafting your lenses
    - Social phenomena
    - Competitors
- Cons
  - Different opinions (Fleiss' kappa = .69)
  - Subjective selection (theory informed, limited inferences)

University's Dimensions

Research

- ERC_Research
- Esa
- EU_H2020
- StampaCnr
- istat_it

Teaching

- edizionimulino
- CorsiEcm_Info
- alpha_test
- Educaform
- eidosco
- ISTUD_IT

Communication

- nature
- Focus_it
- NatGeoTvItalia
- RaiCultura
- AnsaScienza

Tech&Innovation

- timwcap
- CorInnovazione
- tag_school
- TalentGardenit
- ImpactHubMilano

Employability

- 24job
- cliclavoro
- AdeccoItalia
- AlmaLaurea
- cgilnazionale
- inail_gov

Economics*

- CorrierEconomia
- BorsaItalianaIT
- CNNMoney
- ansa_economia
- bancaditalia

Church*

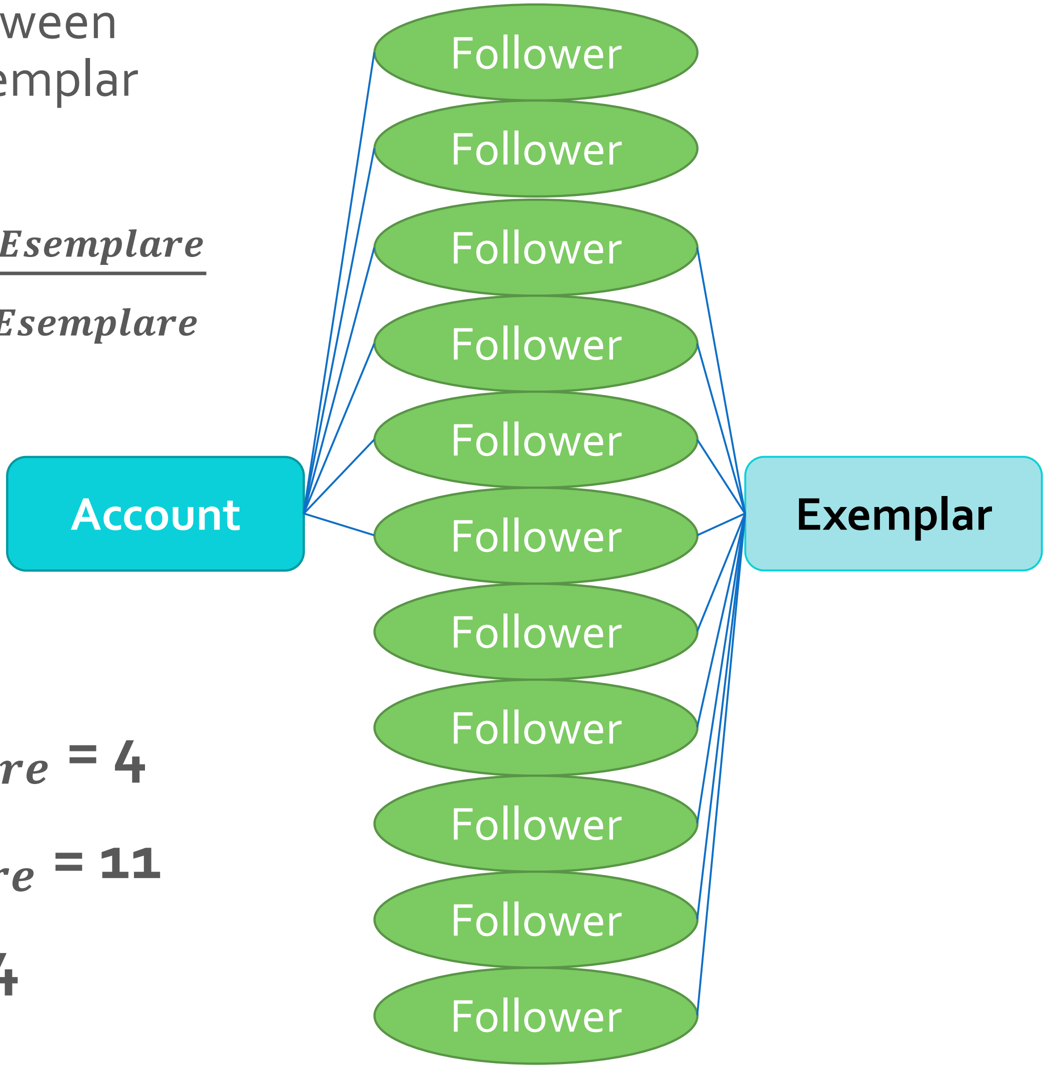
- Avvenire_NEI
- fam_cristiana
- Chiediloaloro
- CaritasItaliana
- AzioneCattolica

*for validation purposes

Similarity
-Jaccard Similarity Coefficient

Shared followers between
target account and exemplar

$$J(A, E) = \frac{F_{Account} \cap F_{Esemplare}}{F_{Account} \cup F_{Esemplare}}$$



$F_{Account} \cap F_{Esemplare}$ = 4

$F_{Account} \cup F_{Esemplare}$ = 11

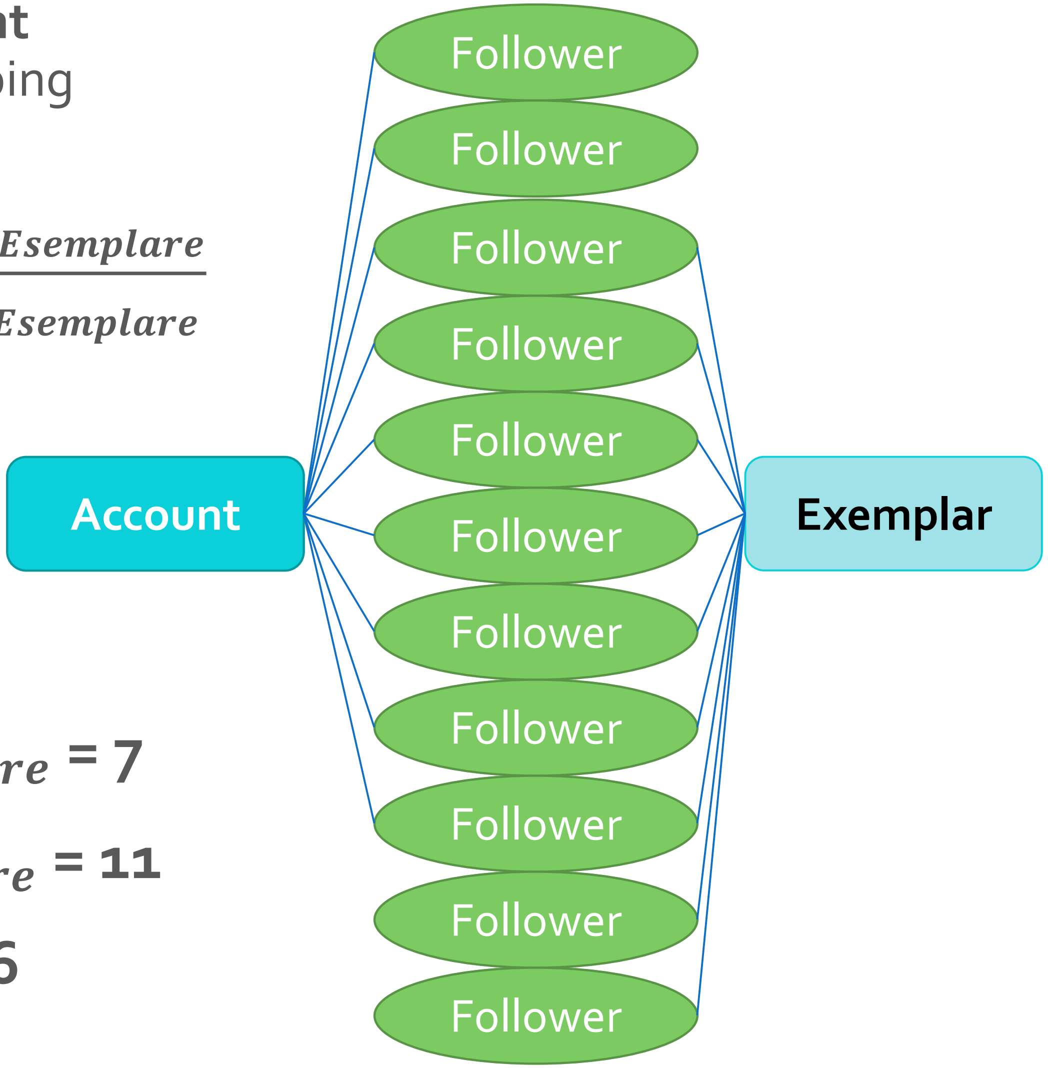
$J(A, E)$ = .364

Similarity

-Jaccard Similarity Coefficient

Greater coefficient
for higher overlapping

$$J(A,E) = \frac{F_{Account} \cap F_{Esemplare}}{F_{Account} \cup F_{Esemplare}}$$



$F_{Account} \cap F_{Esemplare}$ = 7

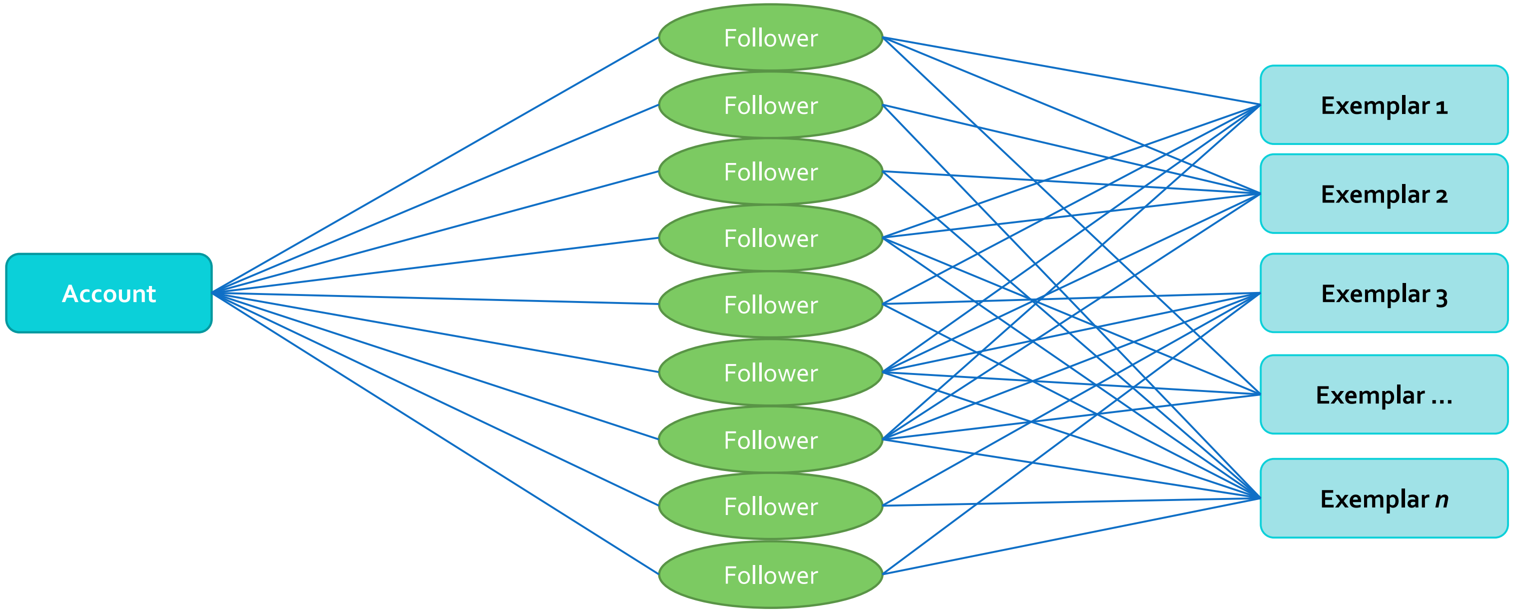
$F_{Account} \cup F_{Esemplare}$ = 11

$J(A,E)$ = .636

Scoring

-Social Perception Score[1]

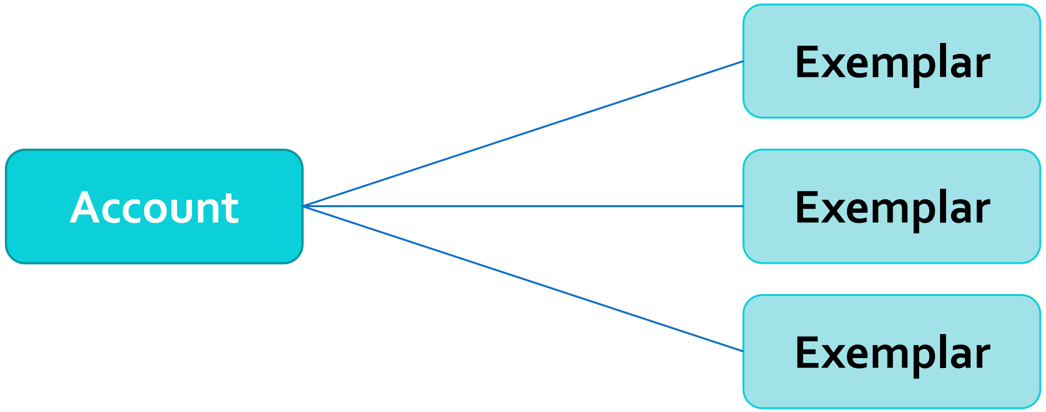
$$SPS(A,D) = \frac{\sum_{E_i \in D} \left(\frac{1}{F_{E_i}}\right) J(A,E_i)}{\sum_{E_i \in D} \left(\frac{1}{F_{E_i}}\right)}$$



From similarity to scoring

- **Shared followers** between target account and exemplars

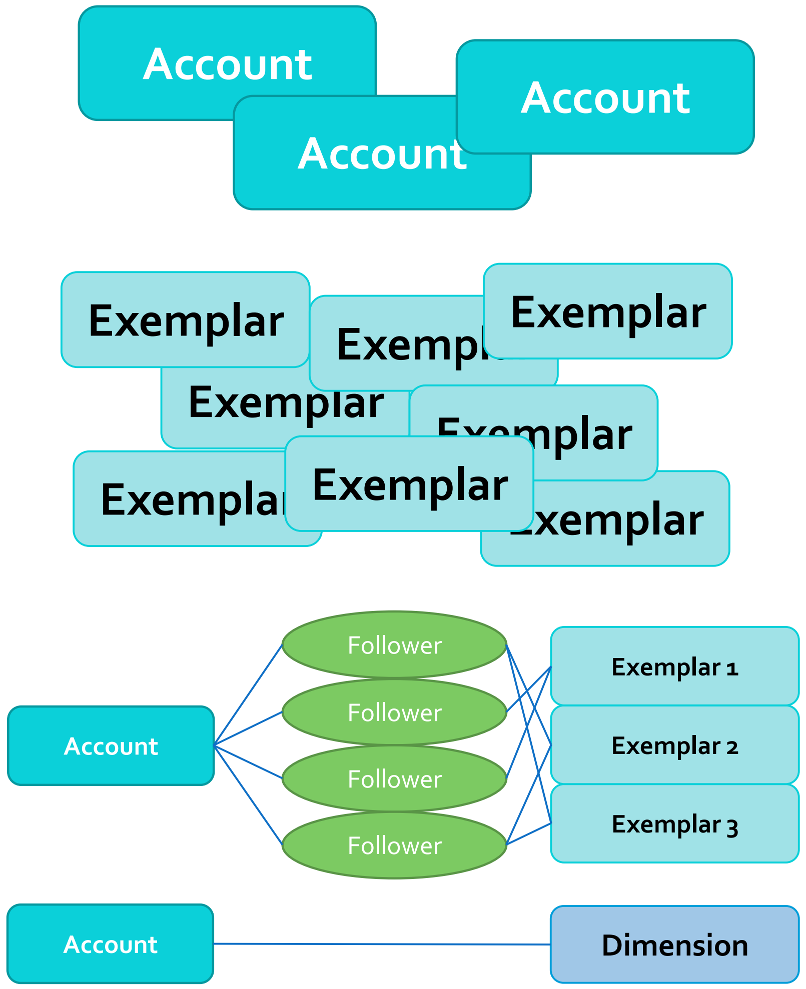
- **Aggregating** the similarities between target and each exemplars



- **Scoring** the target along the dimension

Method
-Scoring Process

- Target account
  - Competitors
  - Institutions
  - Social phenomena
- Dimensions
  - Exemplars
    - Top players
    - Opinion leaders
    - Influencers
- Similarity
  - Shared followers
  - Jaccard Similarity Coefficient
- Scoring
  - Social Perception Score

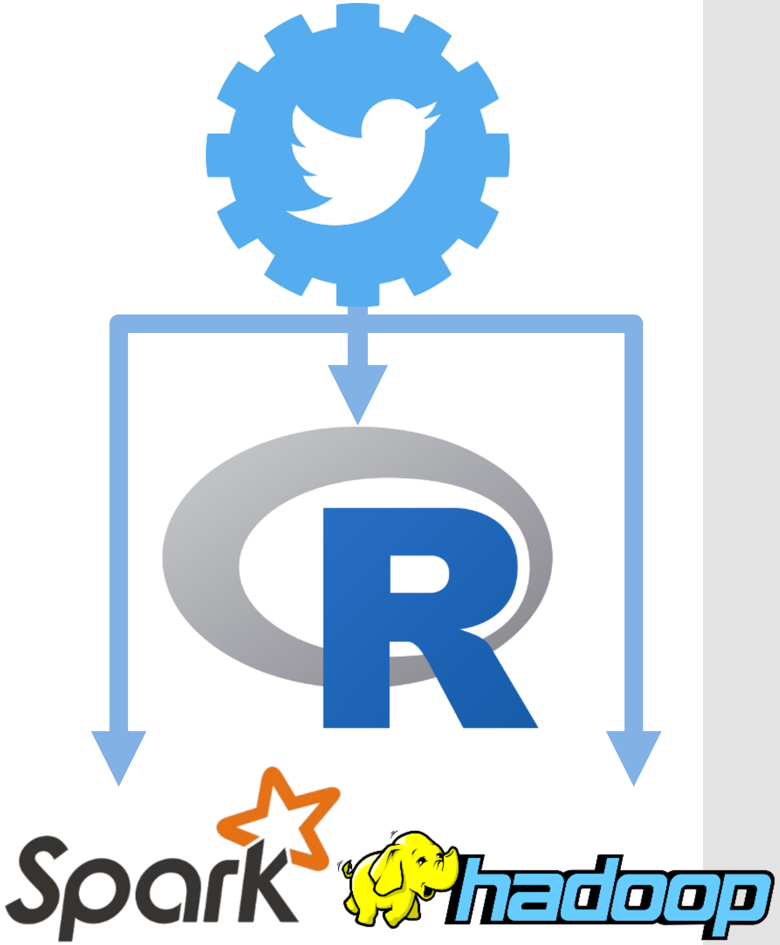


Validating

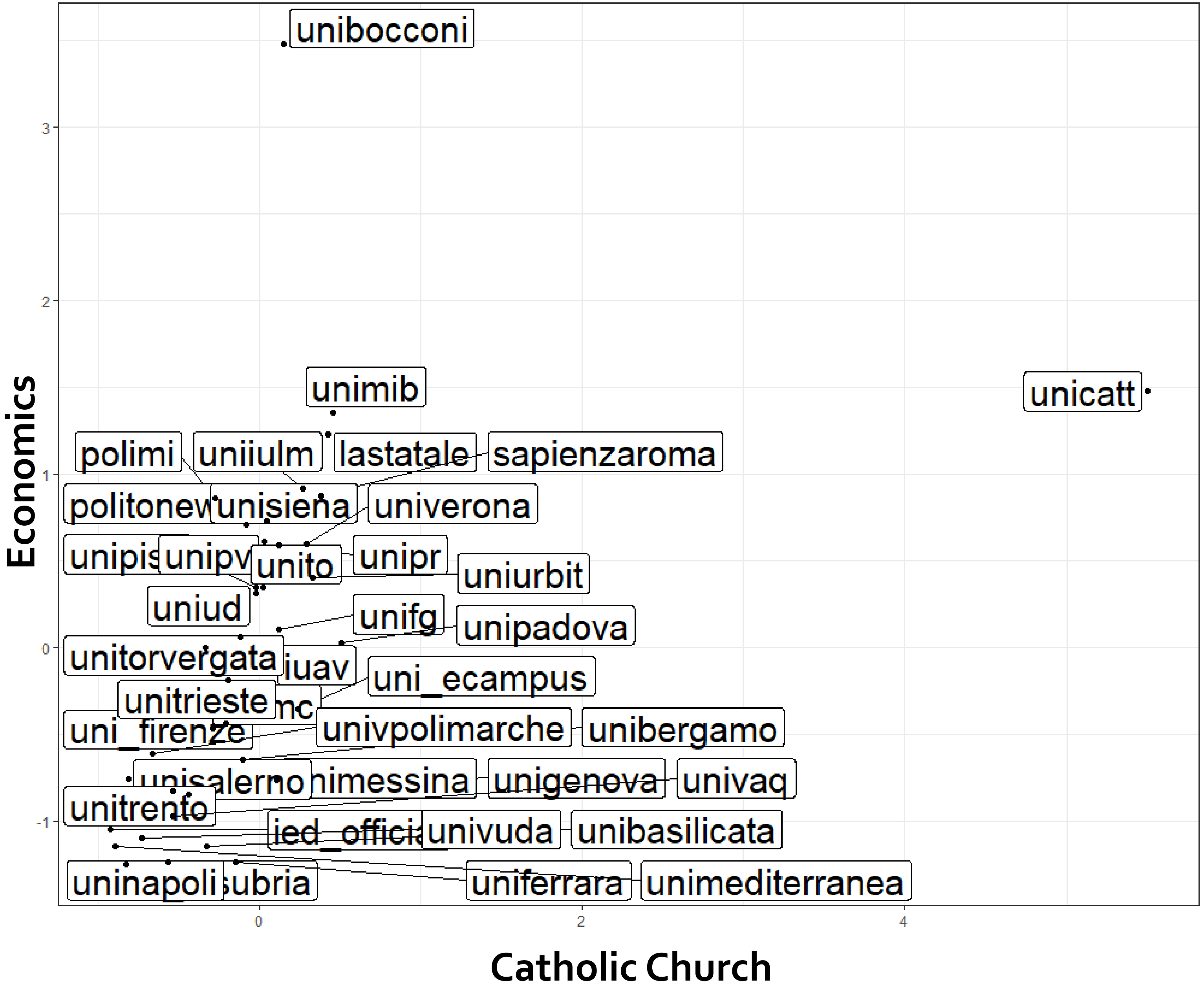
Is the **score** meaningful?

Data Analysis

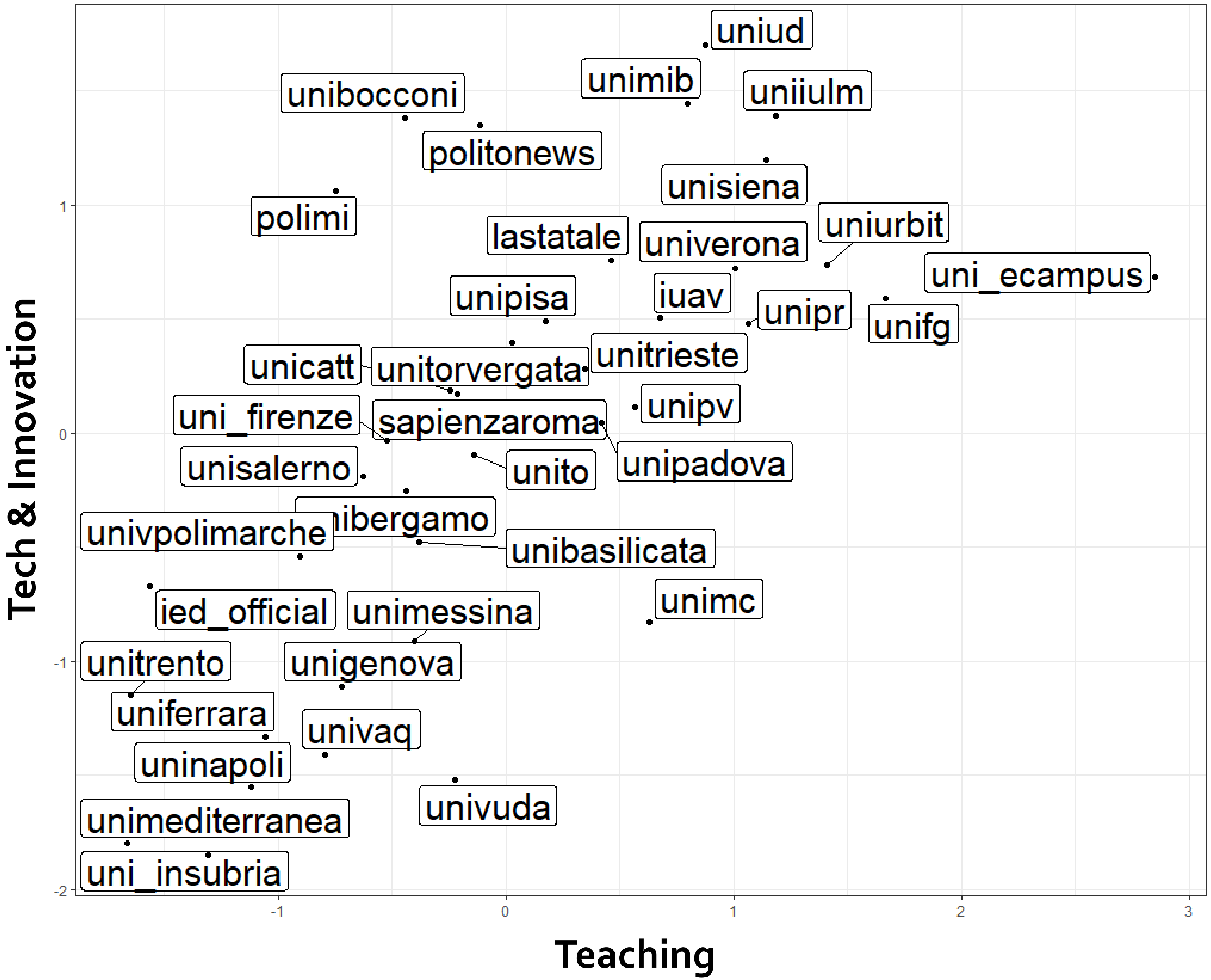
- Web scraping
  - Follower (targets & exemplars)
  - Twitter API
  - One txt for each account
- Analysis
  - Computing scores (base, set operations)
  - Explorations (base, ggplot2)
  - Network data (igraph)
- MapReduce philosophy
  - Parallel, distributed, scalable computing
  - Spark, Hadoop, ecc…



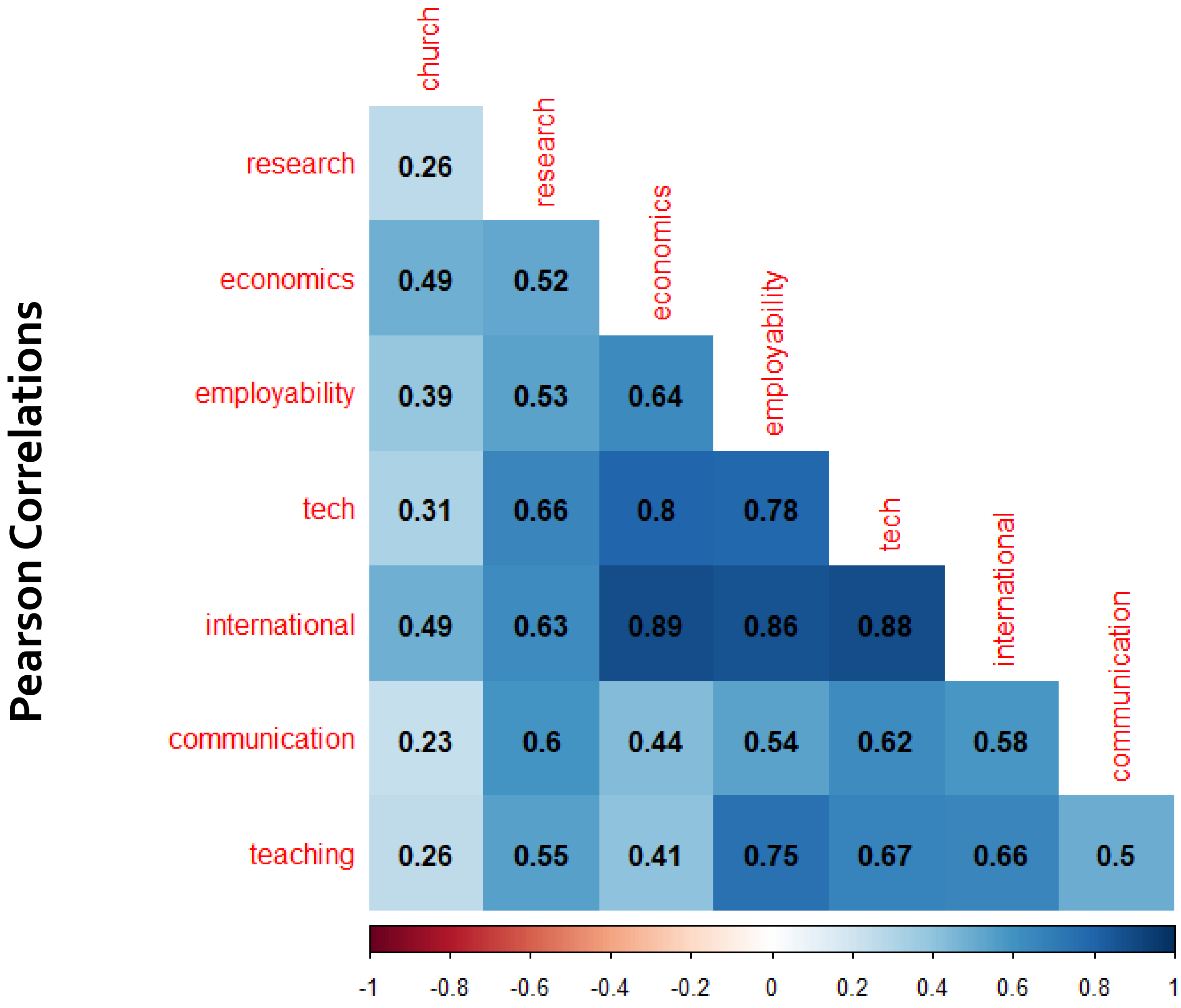
Validation
- Meaning and Expectations



Validation
- Meaning and Expectations



Validation
- Meaning and Expectations

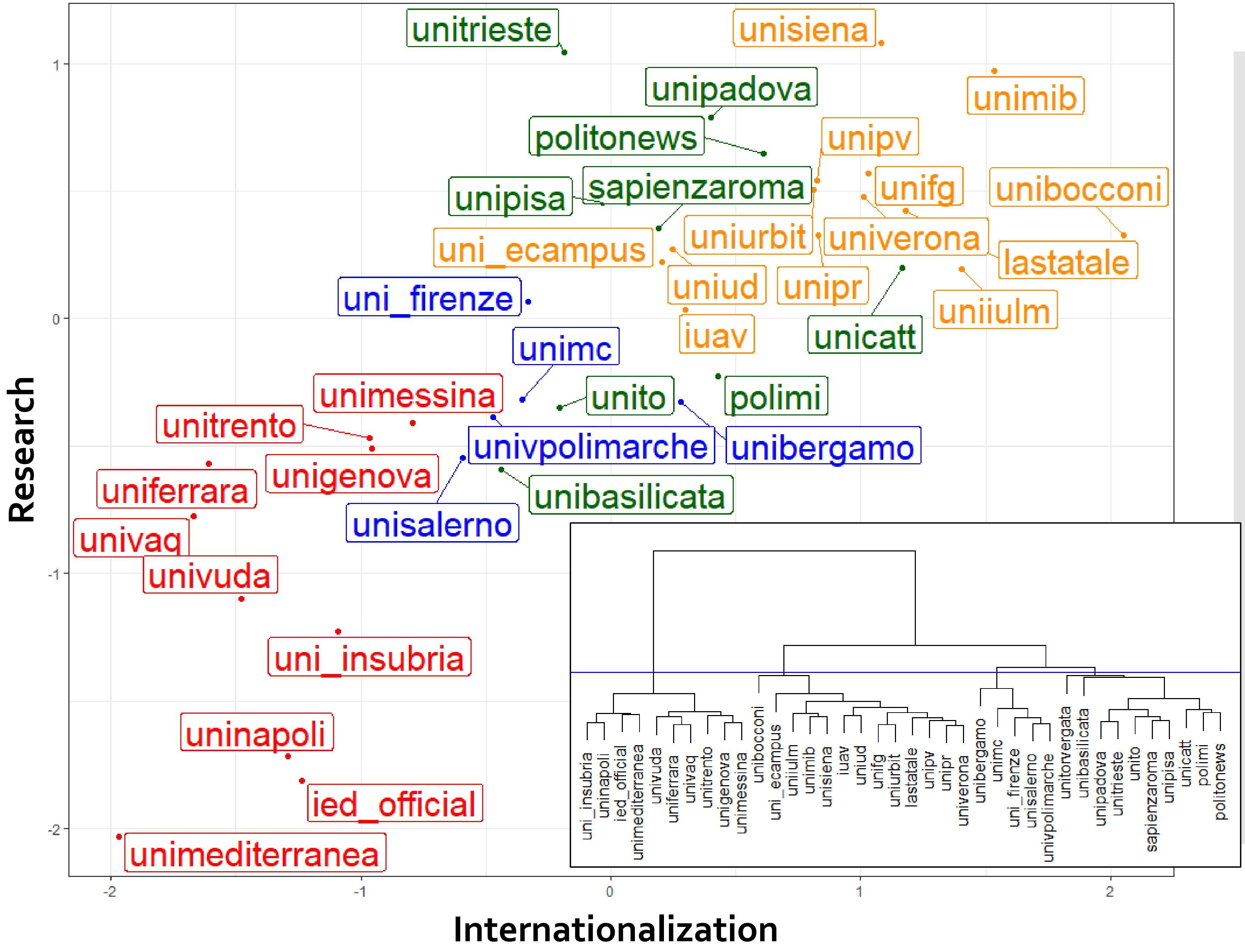


Validation
- External Coherence

Rank **Il Sole 24 Ore**, italian economics journal

Dimensione	**Correlation$_{Kendall}$**
Research	.22
Economy	.72
Internationalization	.44
Employability	.11
Innovation	.61
Communication	.44
Technology	.77

Cluster Analysis
- 4 clusters

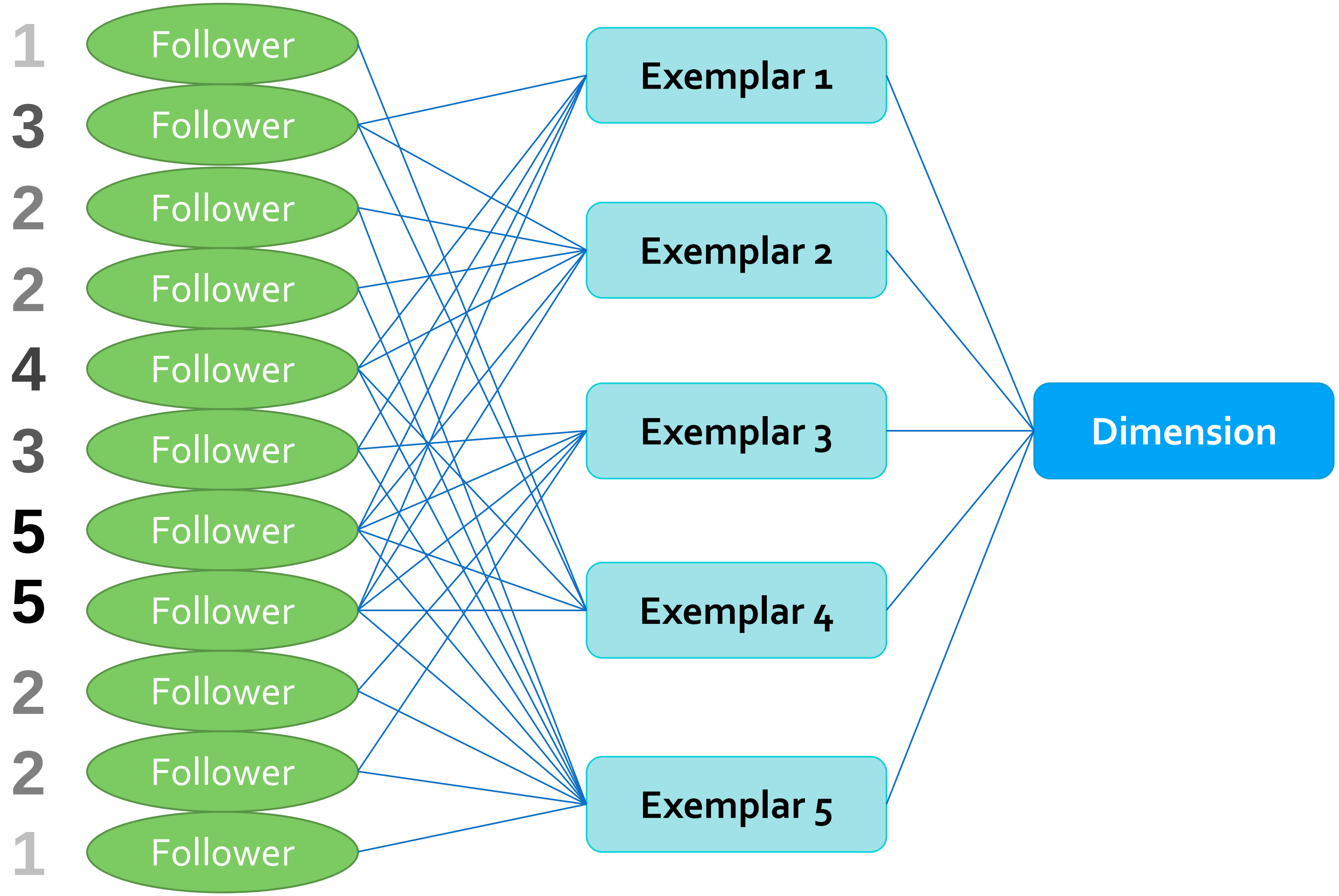


Explaining

What drives each **target's** score?

Explaining

- Followers' impact



Follower's Impact

- Most Valuable Followers

Research	Employability	International	Tech&Innovat	Economics	Teaching
2349879146	3944889083	176540776	123356739	1914625608	1935716222
945982915	74395114	63796828	22083	244162538	461961209
1467599203	2240525990	17467699	291096724	325522575	787322
2989340277	3033476780	1398654900	494357346	3417395211	96938901
364343909	1134455509	42417984	79719617	425384503	113582262
122888810	113582262	586941998	82060480	464662192	342614745
2511974010	261645765	17038869	56326516	238611868	419988263
254100283	707661	202151035	74090914	267083256	100035152
108979965	375546791	2260150651	419951447	335490496	1139245314
22083	410077401	258790331	523451543	593540542	1464732296
…	…	…	…	…	…

Follower’s Impact
- Further Analysis

- Who are the **followers** that drives **target’s** perception?
  - Understanding social phenomena
  - Spot communities
  - Market segmentation (targeting new customers)

- Impact Estimation
  - Given a followers community (eg. Followers of a specific target), estimate their impact on other dimension
  - Refine dimensions based on the followers

Thank you!

- Co-authors:

  Cristina, Stefania e Elena



- Contacts:
  - m.biella6@campus.unimib.it - @marcobie
  - cristina.zogmaister@unimib.it - @czogmaister
  - s.ceolato@campus.unimib.it - @stregattooo
  - e.parozzi@campus.unimib.it - @ElenaParozzi