

SOBIGDATA ^{PPP}
PREPARATORY PHASE PROJECT

FIR Future
Artificial
Intelligence
Research

The ego network of words

Kilian Ollivier, **Chiara Boldrini**,
Andrea Passarella, Marco Conti

IIT-CNR, Pisa

iit ISTITUTO
DI INFORMATICA
E TELEMATICA
CNR

CNR Consiglio Nazionale
delle Ricerche

Our goal

Use **data science** to uncover **cognitive constraints** in human behavior **beyond social interactions**.

- **Identify** and **describe** the **cognitive constraints** affecting the human brain in everyday activities, and more specifically in the **production** of spoken and written **language**.
- Leverage **digital traces** of human behavior, including social media posts and interview transcriptions, using cutting-edge techniques in **unsupervised classification** and **NLP**.
- Inspiration from **existing research** on cognitive constraints in **social interactions**, utilizing an **individual-centered model** to describe **cognitive effort distribution**.



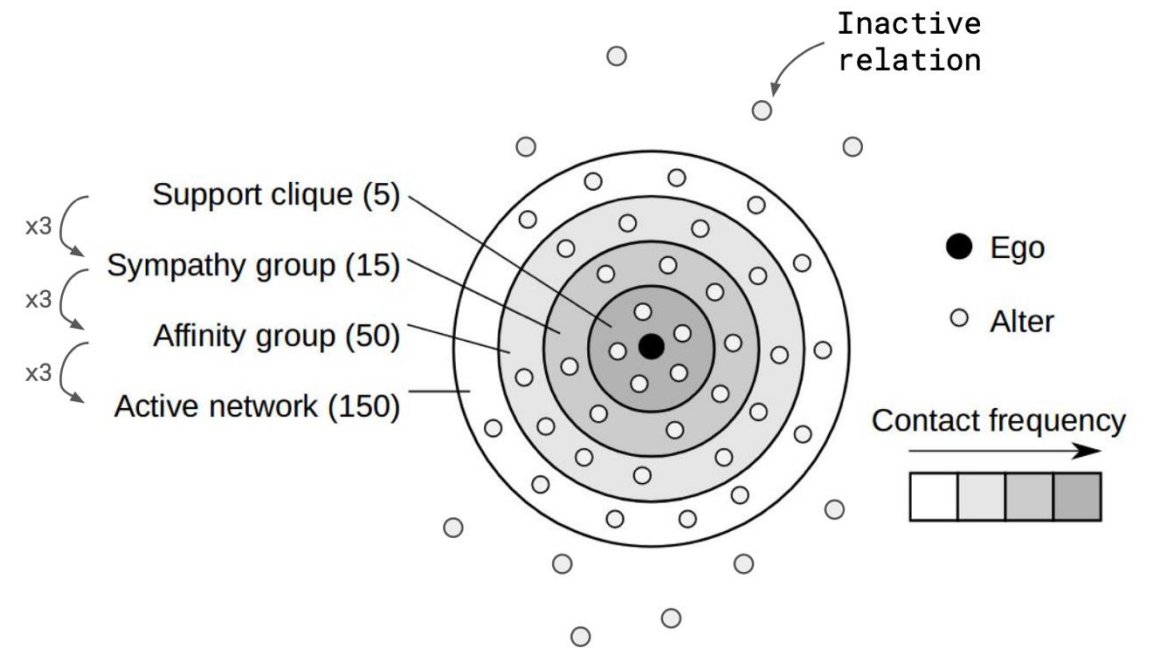
The ego network model: cognitive constraints in social relations

Micro view of social relations centered on an individual (ego) within a social network.

Recurrent pattern of **5-15-50-150** with a stable **scaling ratio** of around **3**.

- **Support clique:** 5 most reliable relatives and friends.
- **Active network:** approximately 150 alters, contacted at least once a year.

Patterns holds true across both real-life and online social networks.

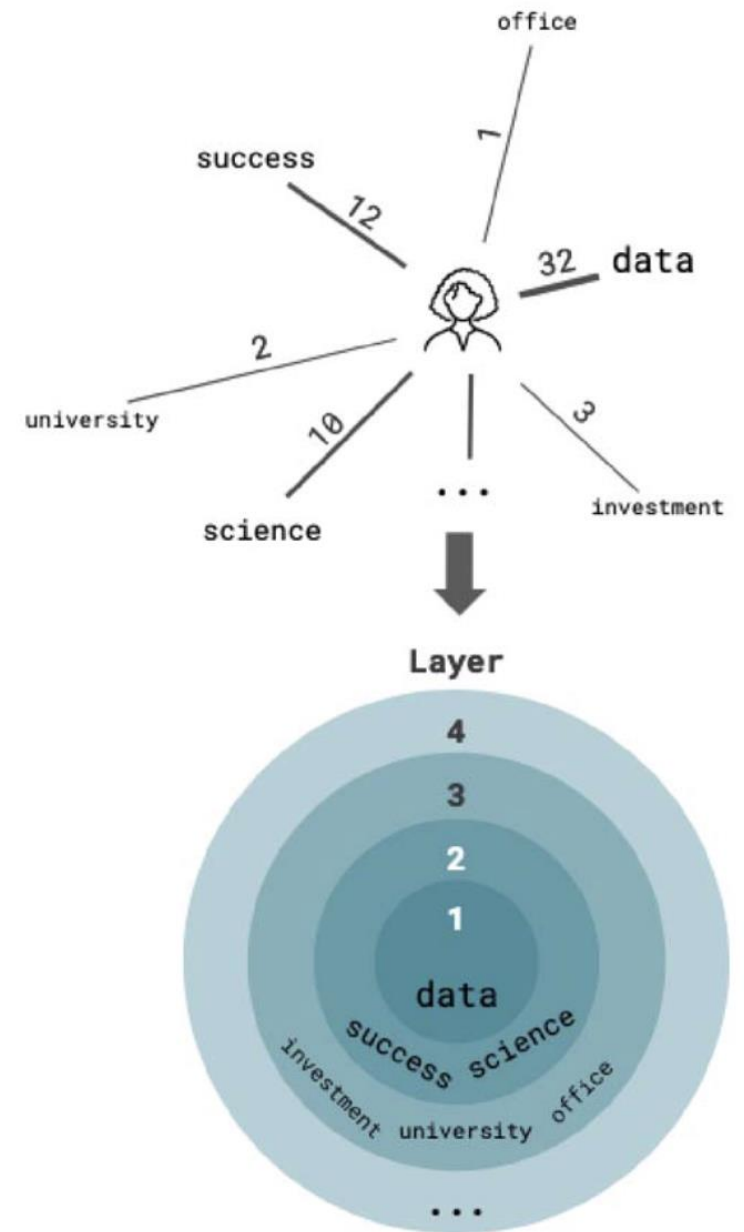


Dunbar, R. I. (1998). The social brain hypothesis. Evolutionary Anthropology: Issues, News, and Reviews: Issues, News, and Reviews, 6(5), 178-190.

Focus on language production

Can individual word frequency distributions be arranged in circles?

If so, this forms a new kind of **ego network**—this time, built from **words**.



Research questions

1. What are the structural properties of ego networks of words ?

Ollivier, K., Boldrini, C., Passarella, A., & Conti, M. (2020, October). Structural invariants in individuals language use: The “ego network” of words. In International Conference on Social Informatics (pp. 267-282). Cham: Springer International Publishing.

2. How words are semantically organized in this layered model ?

Ollivier, K., Boldrini, C., Passarella, A., & Conti, M. (2022). Structural invariants and semantic fingerprints in the “ego network” of words. Plos one, 17(11), e0277182.

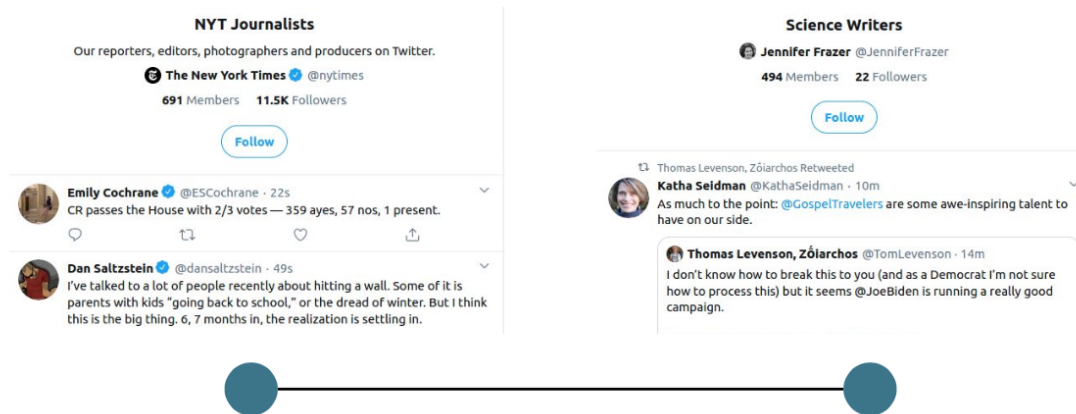
3. Does an “active” part of the model naturally emerge ?

Ollivier, K., Boldrini, C., Passarella, A., & Conti, M. (2024). Unveiling Cognitive Constraints in Language Production: Extracting and Validating the Active Ego Network of Words. IEEE Transactions on Computational Social Systems.



Core datasets

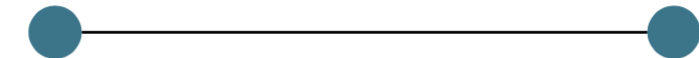
4 sets of tweets timelines extracted from 4 distinct set of users



2 preexisting Twitter lists of **New York Times Journalists** and **Science Writers**

Sample of users
having tweeted
with the hashtag
#MondayMotivation
on 16/01/2020

Sample of users
having tweeted in
english from UK
on 11/02/2020



2 sets of **randomly chosen users**
(after bot filtering)

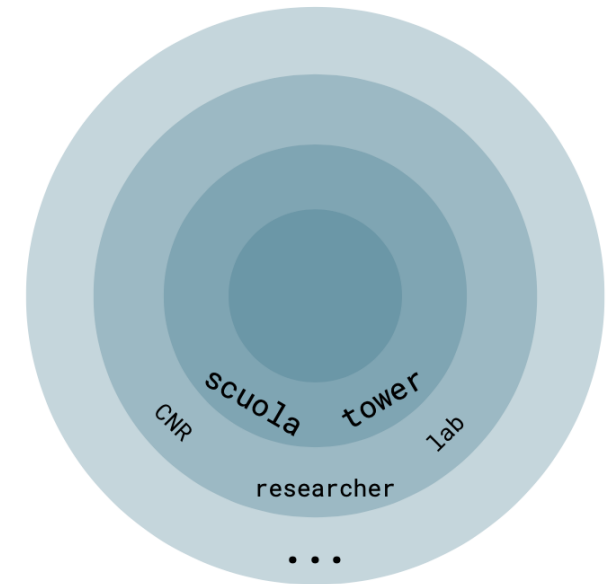
Building the ego network of words

Word	Freq.
pisa	32
scuola	12
tower	10
cnr	3
researcher	2
lab	2

MeanShift
clustering

Word	Freq.	Cluster
pisa	32	1
scuola	12	2
tower	10	2
cnr	3	3
researcher	2	3
lab	2	3

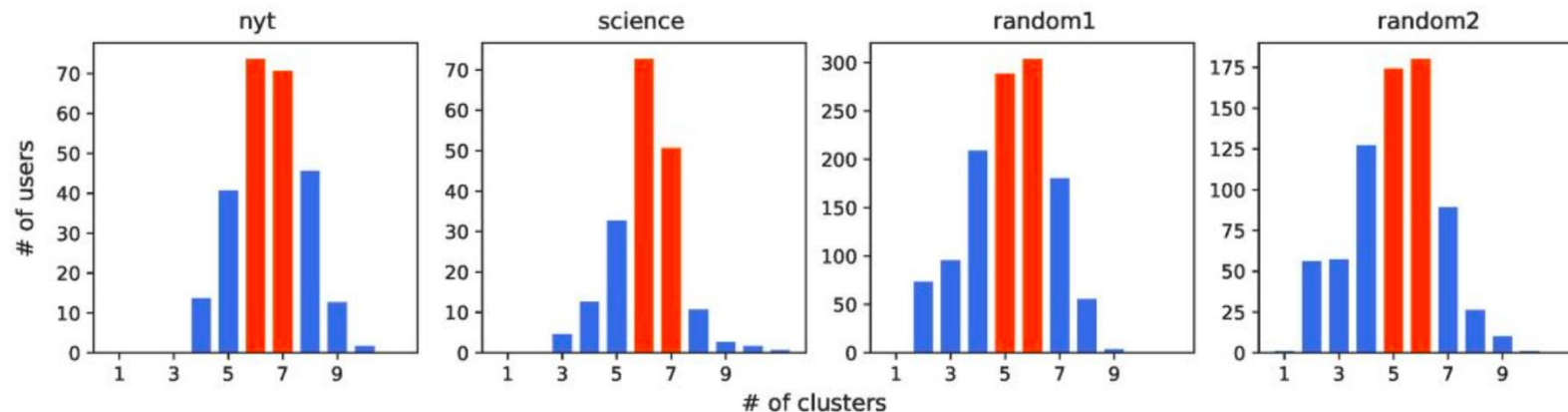
Layered
structure



Words with a similar frequency
of use are grouped together

RQ1: take-home messages

- The words can usually be grouped into **5 to 7 layers**.
 - *4 to 5 in the social ego networks*
- Layers size **approximately doubles** when moving from layer i to layer $i + 1$, with the only exception of the first layer.
 - *Scaling ratio of 3 in social ego networks*

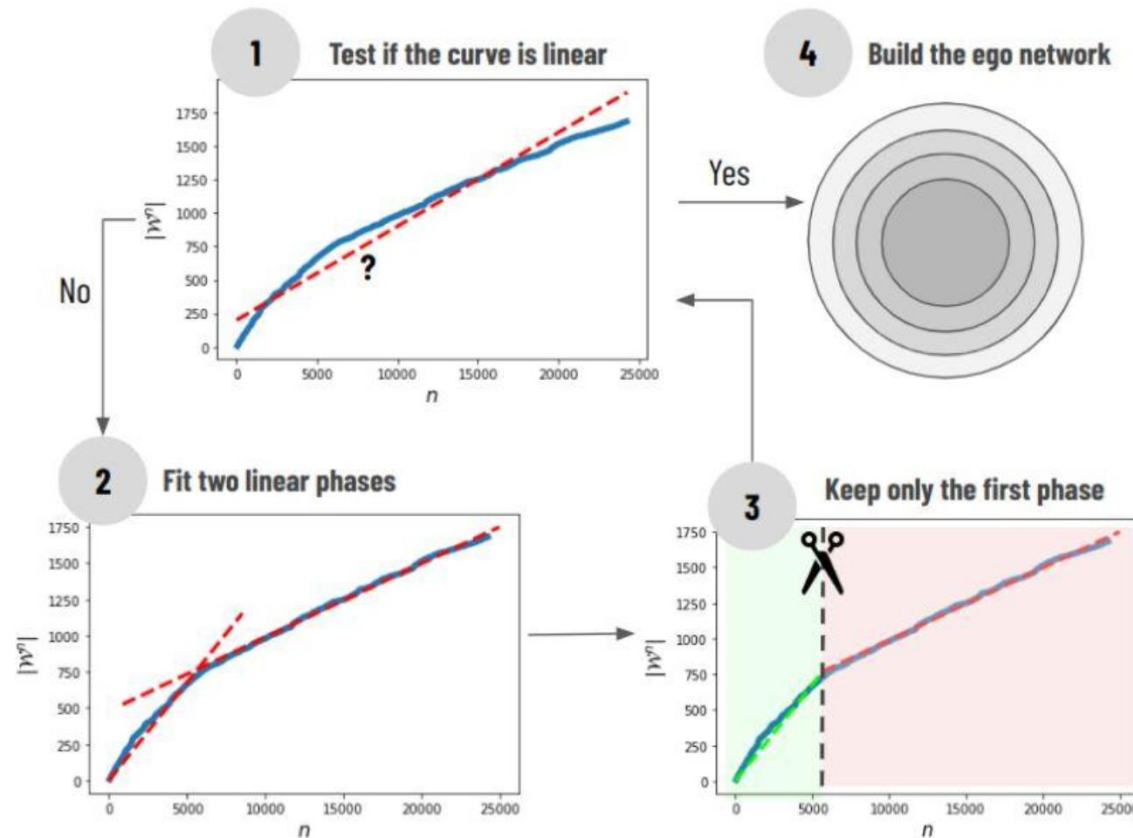


RQ2: take-home messages

- Ring #1 generates proportionally more topics compared to other rings.
- Semantic profile of ring #1 is the most unique and semantically distant from others.
- Important topics in ring #1 are also important in other rings, indicating a predictive role of the innermost ring for the entire ego network.
- **The semantic profile of the innermost ring serves as the semantic fingerprint of the entire ego network of words.**



RQ3: algorithm to extract the “active” ego networks



1 Linearity test.

2 If the curve is not linear, we find the best model fit with two linear parts.

3 Keep only the part of the curve which fits the first linear part of the model, then come back to 1).

4 Build the ego network of words base on the remaining part (the active part).

Why important? The “active” part of social ego networks was grounded in a shared understanding of how human social interactions work. **Now we have a quantitative tool**

Conclusions & future work

- **Take-home messages**
 - An active ego network exists and can be extracted with a data driven methodology.
 - The extracted ego network exhibits structural invariants exhibits several structural invariants,
 - With a semantic perspective, we can focus on the innermost ring, which plays a role of a semantic fingerprint of the whole ego network of words.
- **Next: Explore application scope of ego networks of words**
 - Reddit/Discord: Groups of users with a specific interest in a topic.
 - Other languages or family of languages
 - LLM's ego networks of words
- **Next: Use the ego network of words as a social microscope**
 - Investigate linguistic production as a function of age and socio-economic category
 - Cross-study with social ego network
 - Study structural and semantic features characteristics of a specific intimacy level

