Tomaso Erseghe

tomaso.erseghe@unipd.it

UNIVERSITÀ DEGLI STUDI DI PADOVA

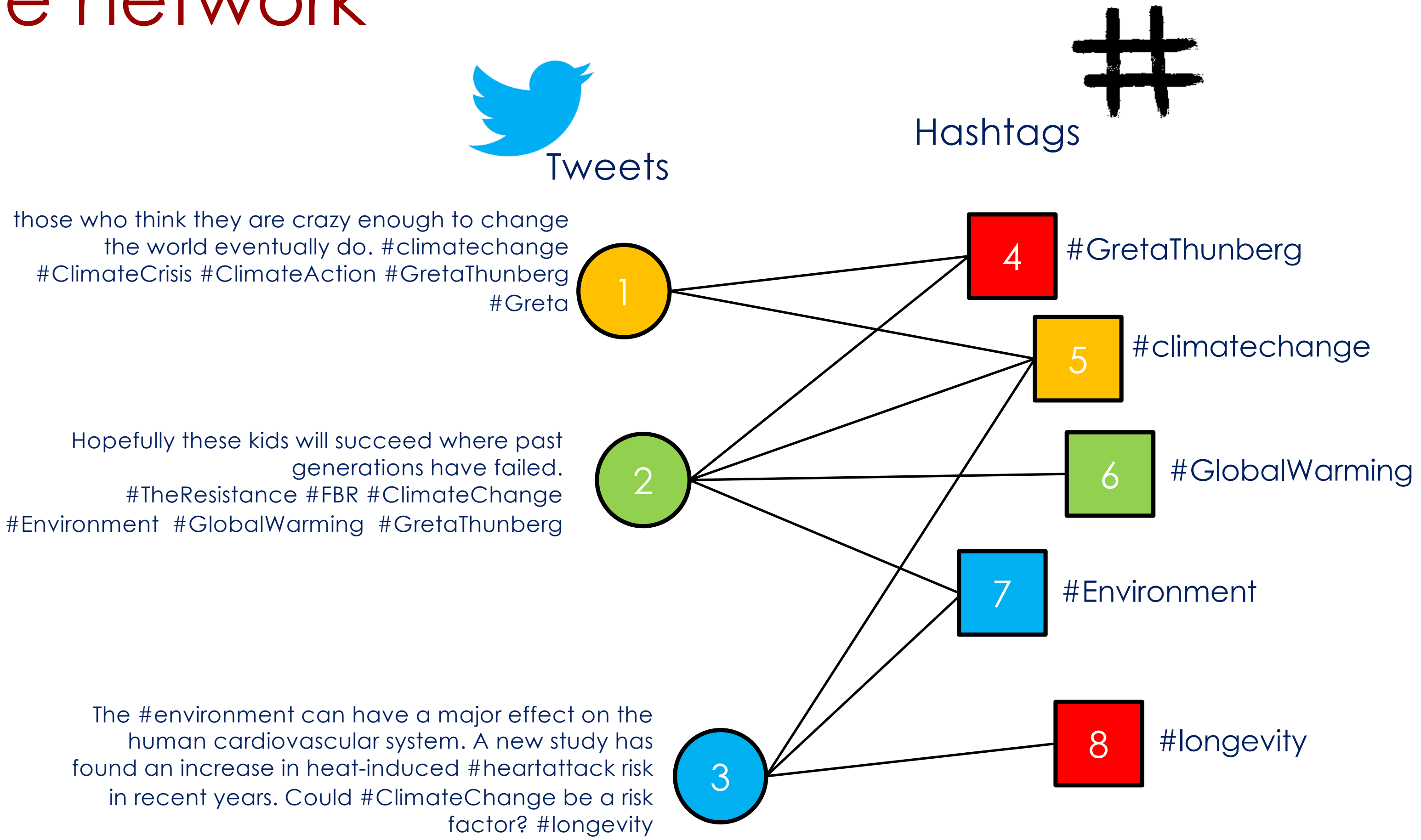
A method to map the linguistic markers of the social discourse onto its semantic network

T. Erseghe, L. Dzanko, L. Badia, C. Suitner

Rationale

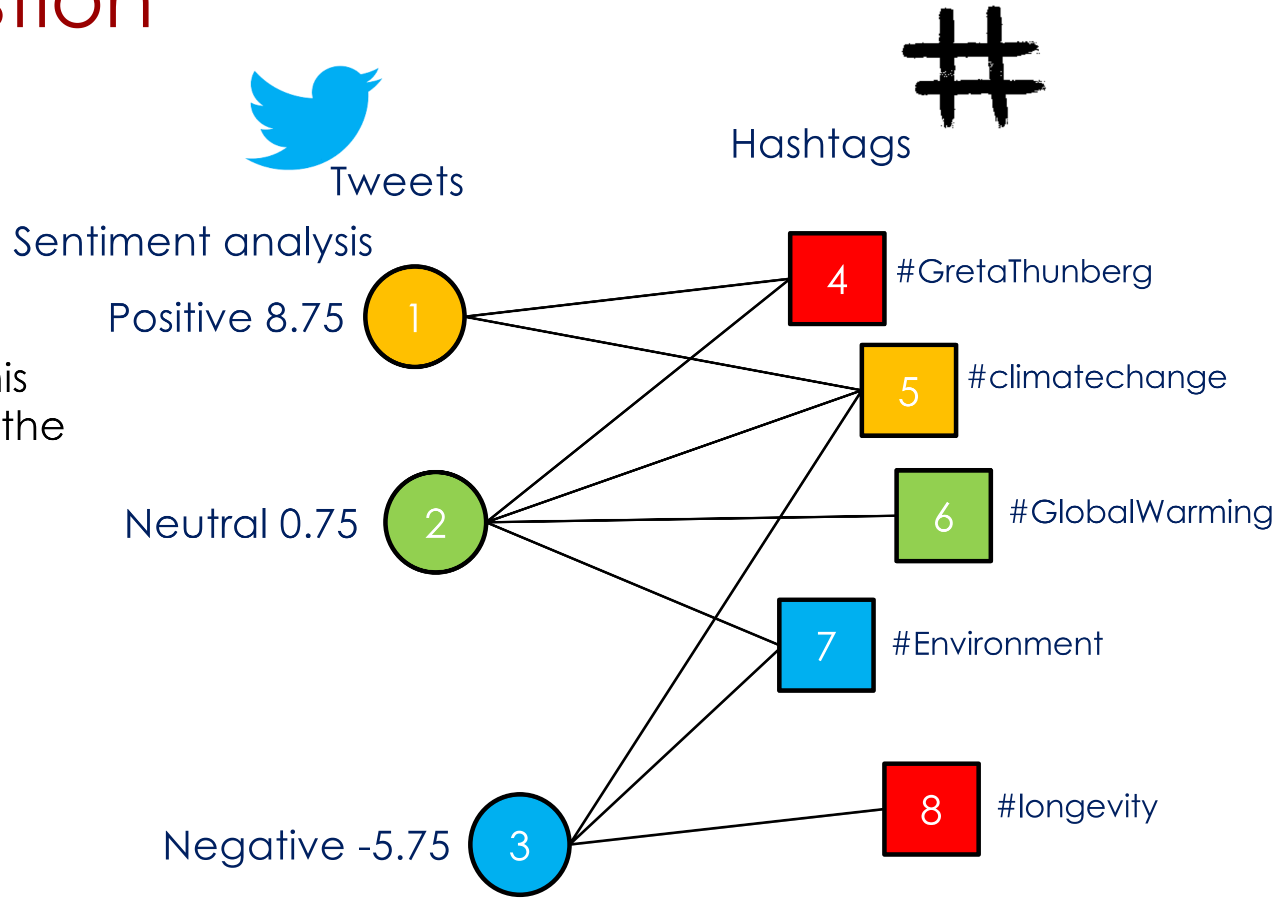
- Linguistic markers can be naturally associated to textual data
  - E.g., sentiment analysis of Tweets
- We would like to **project** this information onto the semantic network
  - i.e., onto the words appearing in Tweets
- We do it by exploiting **network science** tools

The network

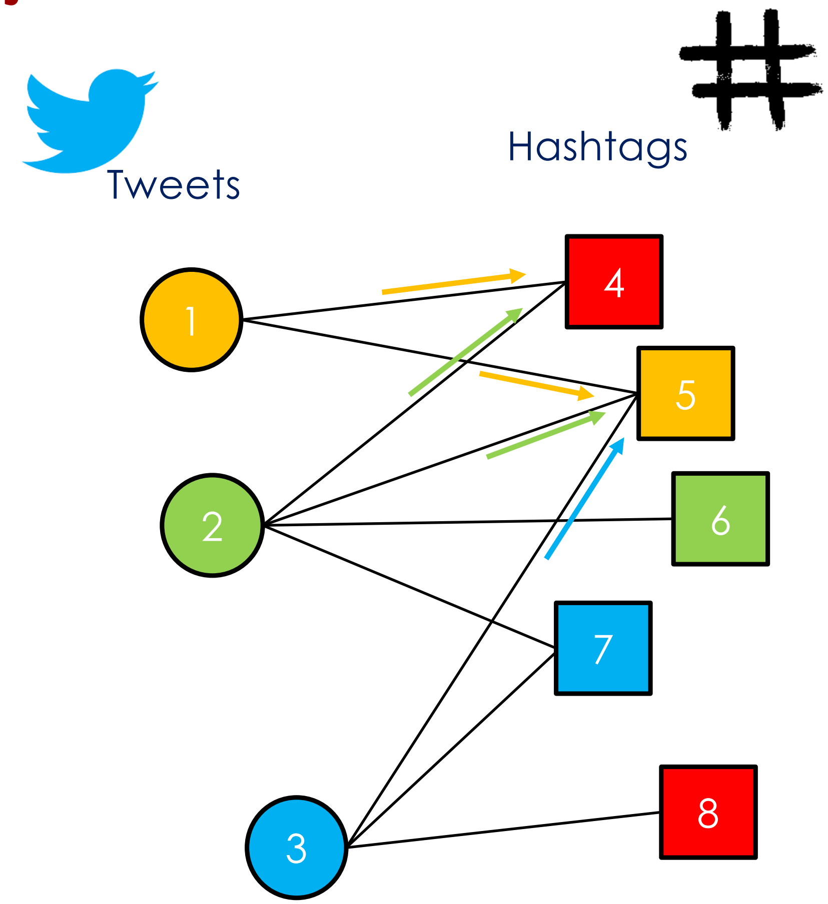

Tweets
Hashtags
those who think they are crazy enough to change the world eventually do. #climatechange #ClimateCrisis #ClimateAction #GretaThunberg #Greta
Hopefully these kids will succeed where past generations have failed. #TheResistance #FBR #ClimateChange #Environment #GlobalWarming #GretaThunberg
The #environment can have a major effect on the human cardiovascular system. A new study has found an increase in heat-induced #heartattack risk in recent years. Could #ClimateChange be a risk factor? #longevity
1
2
3
4
#GretaThunberg
5
#climatechange
6
#GlobalWarming
7
#Environment
8
#longevity

The question

How to **project** this information onto the hastags?!?!?

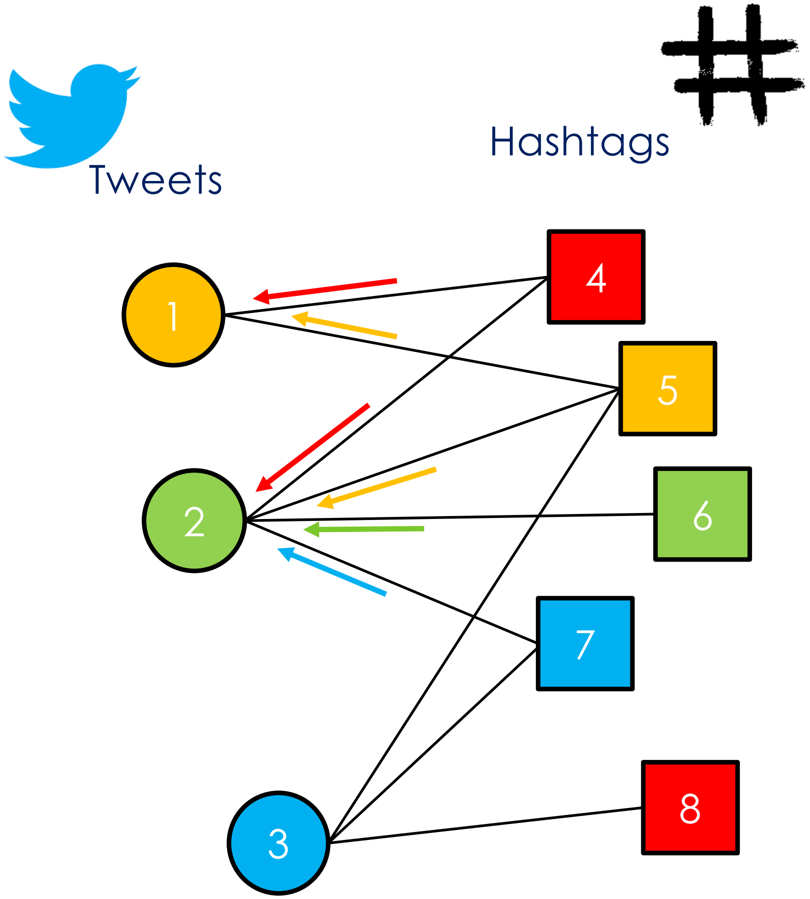


Basic idea: one-step projection



- Each hashtag captures the **average sentiment value** of the tweets it appears in

Improvement: PLMP projection



- Each hashtag captures the **average sentiment value** of the tweets it appears in
- Each tweet captures the **average sentiment** of the hashtags it contains

we iterate the two steps until convergence

PLMP insights

- Idea
  - Project markers from tweets to hashtags/words and viceversa, until convergence
- Many variants can be identified, here the most promising
- Similar to **PageRank**

In-isolation markers

$$\underbrace{\begin{bmatrix} \boldsymbol{m}_w \\ \boldsymbol{m}_t \end{bmatrix}}_{\boldsymbol{m}} = \alpha \underbrace{\begin{bmatrix} \mathbf{0} & \boldsymbol{B}_1 \\ \boldsymbol{B}_2 & \mathbf{0} \end{bmatrix}}_{\boldsymbol{M}} \begin{bmatrix} \boldsymbol{m}_w \\ \boldsymbol{m}_t \end{bmatrix} + (1-\alpha) \underbrace{\begin{bmatrix} \tilde{\boldsymbol{m}}_w \\ \tilde{\boldsymbol{m}}_t \end{bmatrix}}_{\boldsymbol{q}}$$

- But here matrices are **row-normalized** (need a specific proof for convergence)
- Statistically **more reliable** than one-step agency projection (same generalization of degree → pagerank)
- Result is **uncorrelated** with PageRank (i.e., independent on the centrality of words)

Test case

- #MeToo versus #FridaysForFuture calls to action
- Markers: Agency and affiliation

Agency & affiliation

Collective efficacy → agency

Agency = perception that an individual is able to contribute to/a group can collectively reach a social change, believing that the actions can contribute to a broader change

Typically associated with **action verbs**: do, change, make

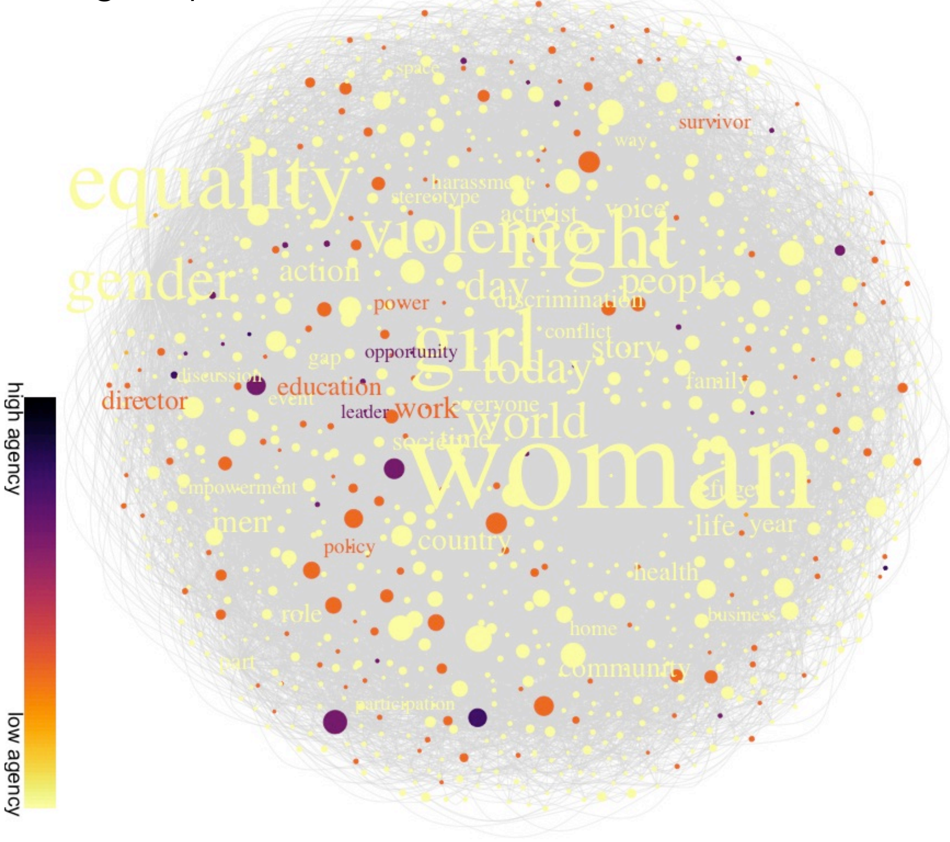
Social identity → affiliation

Affiliation = associating with the topic or consider it important, perception to belong to a group

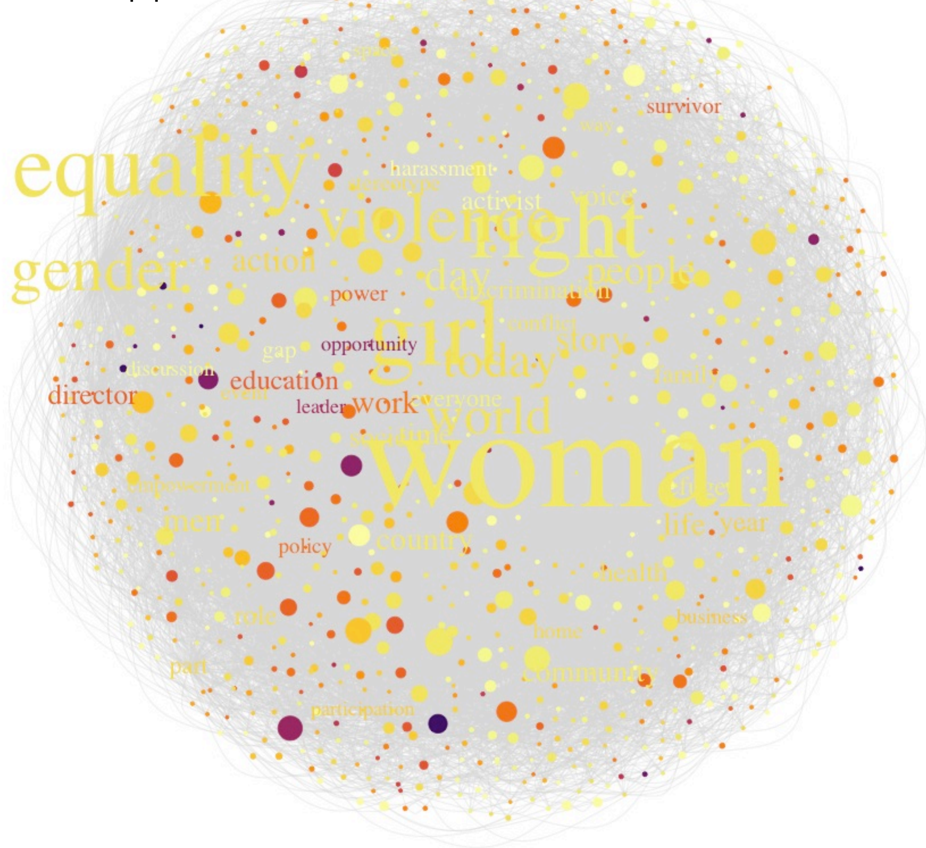
Typically associated with **pronouns**: we, us

What we get – A comparison

Agency in isolation



PLMP approach



Test case (2)

- Agency and affiliation increase in average

#MeToo	pre	post	variation
agency	1.67	1.83	+9.7%
affiliation	3.33	3.70	+10.9%

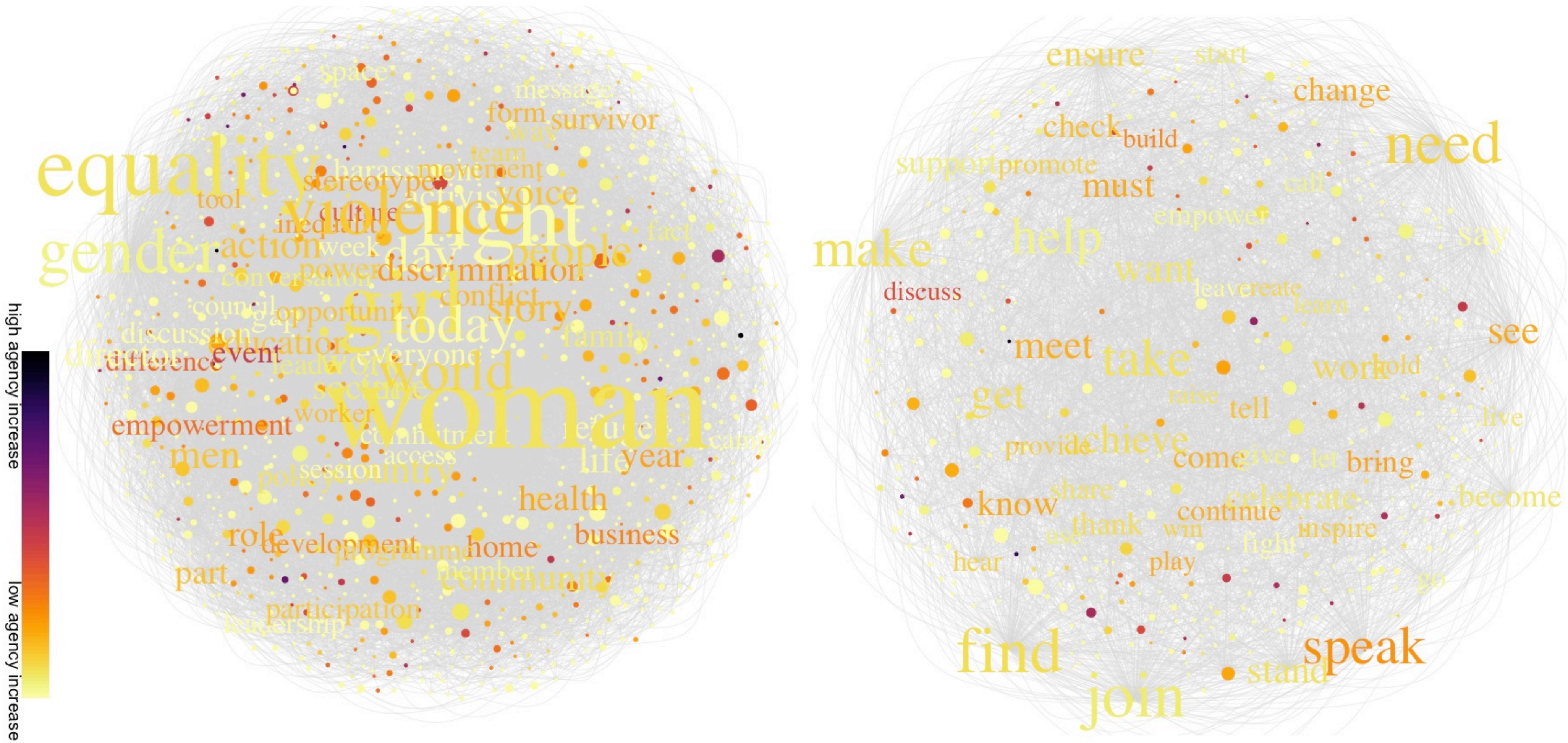
#FridaysForFuture	pre	post	variation
agency	1.56	1.65	+6.1%
affiliation	2.09	2.29	+9.5%

- We measure increase before→after the call to action
- Prestige measure

$$p = \frac{m_{\text{post}} - m_{\text{pre}}}{m_{\text{post}} + m_{\text{pre}}}$$

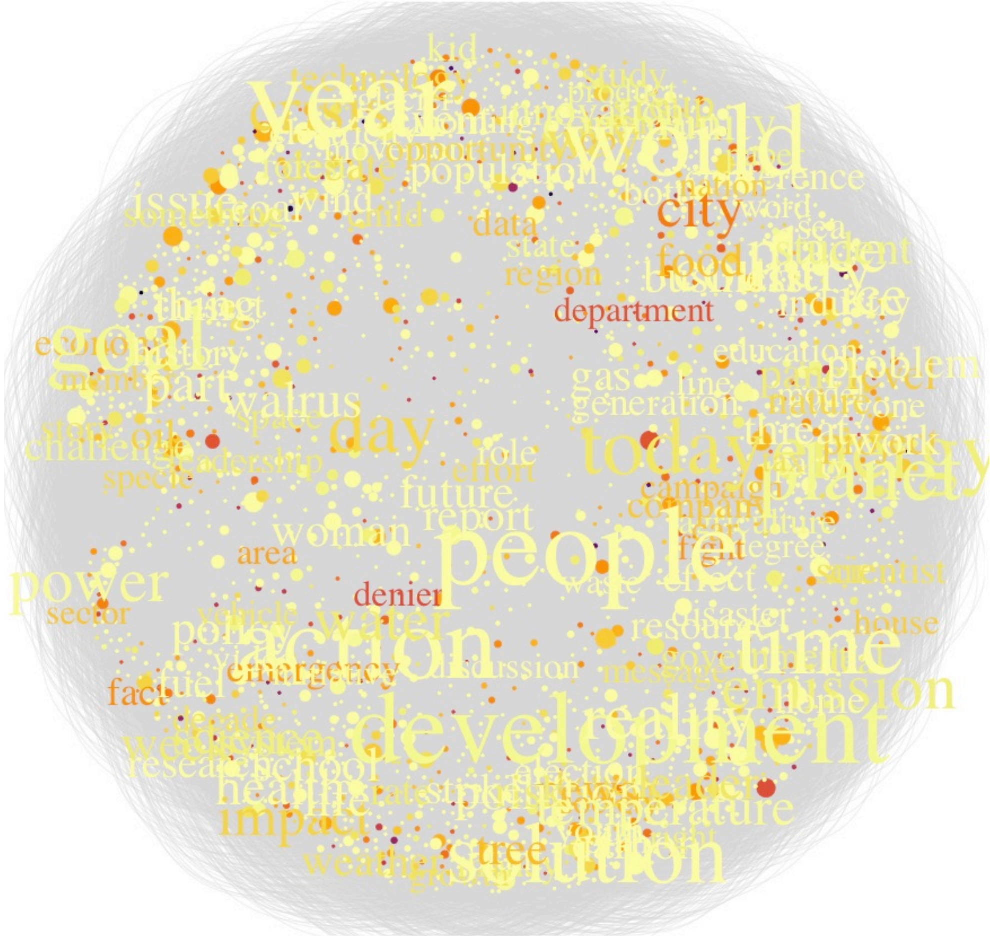
Agency increase in #MeToo

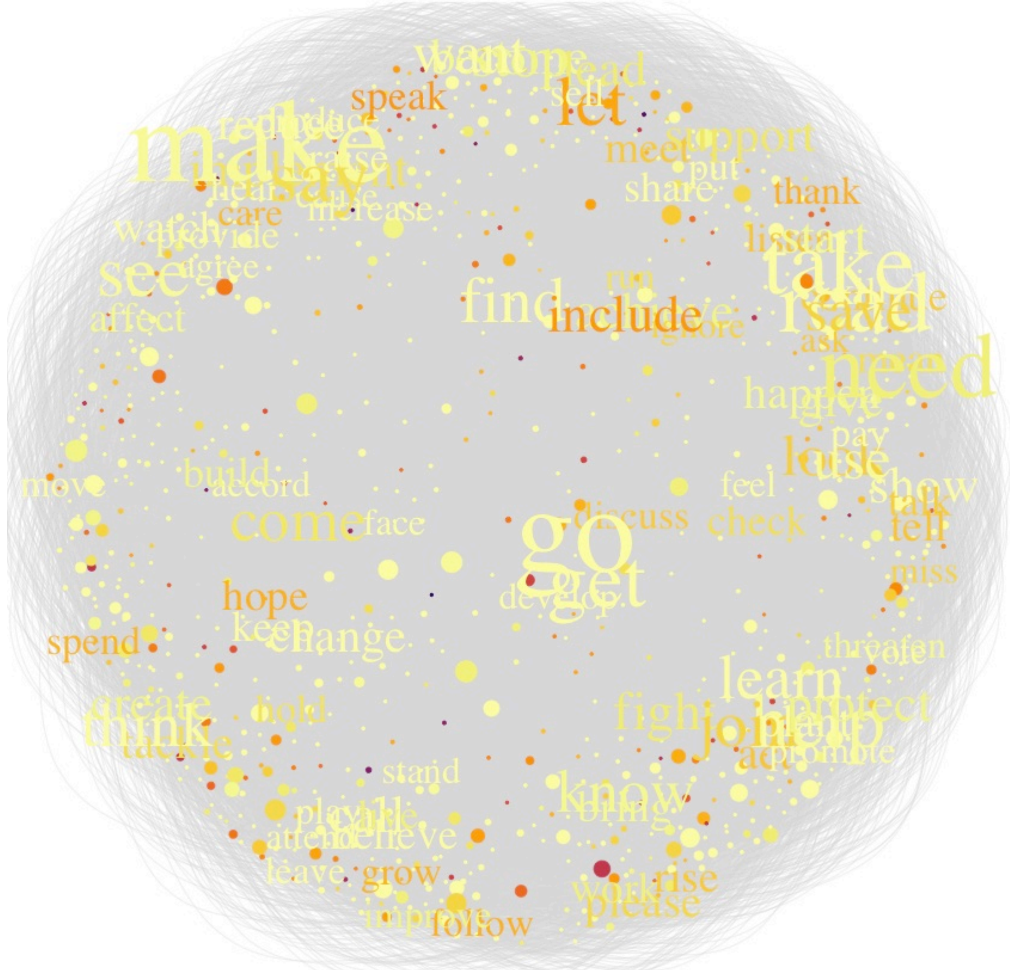
2017 → 2018



Agency increase in #FridaysForFuture

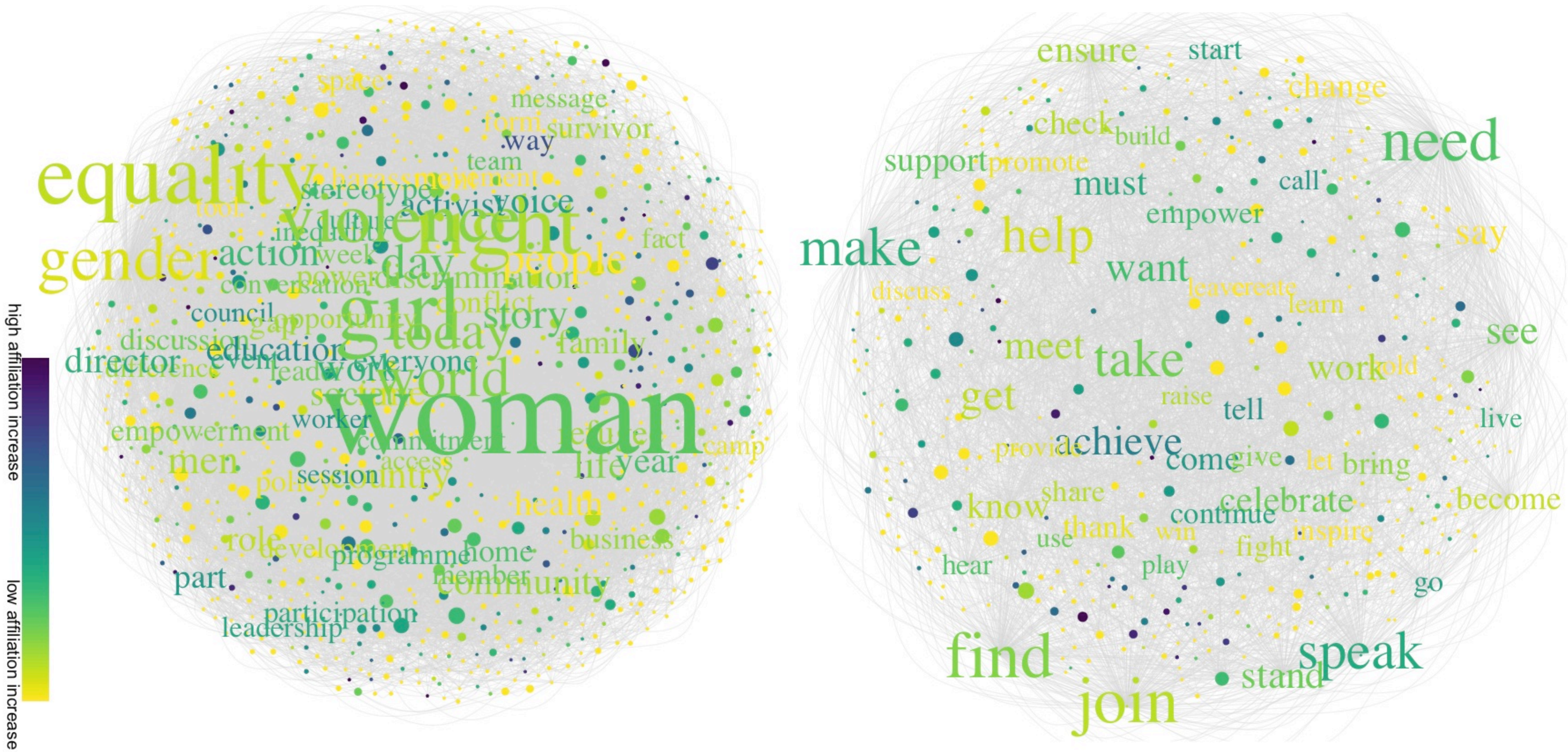
2018 → 2019





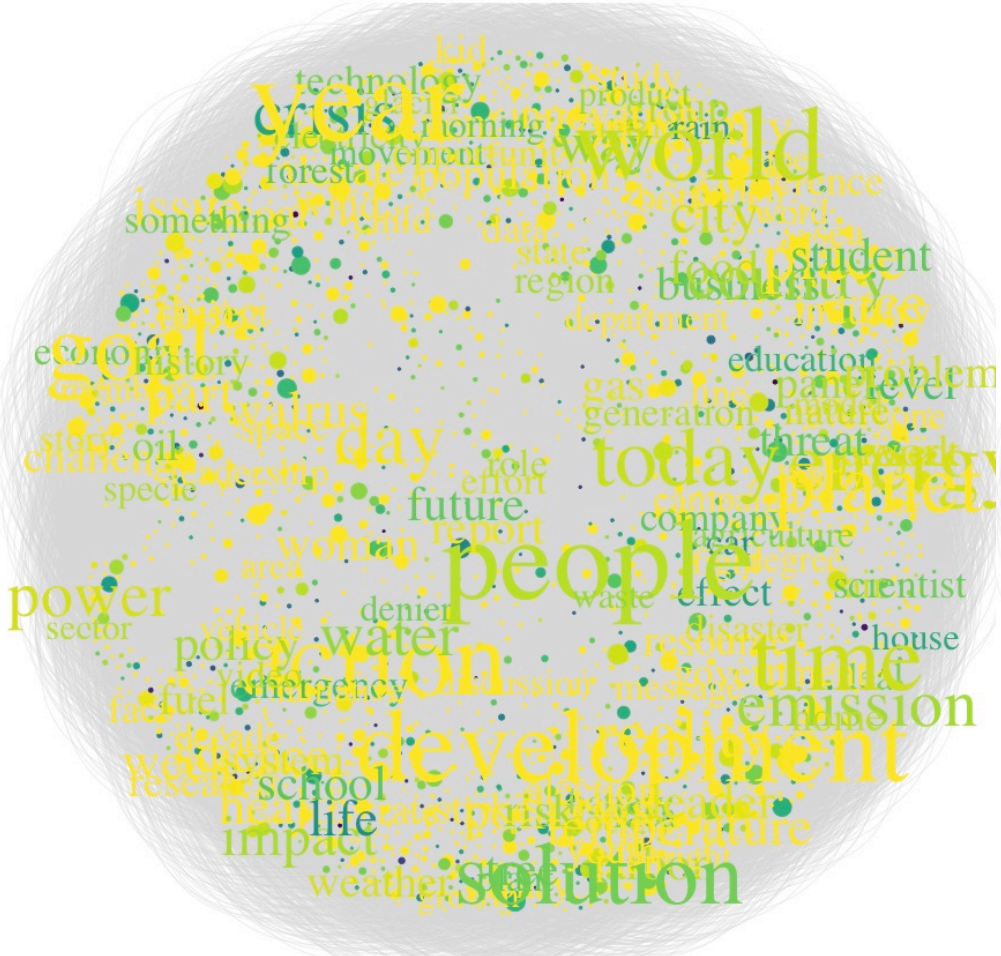
Affiliation increase in #MeToo

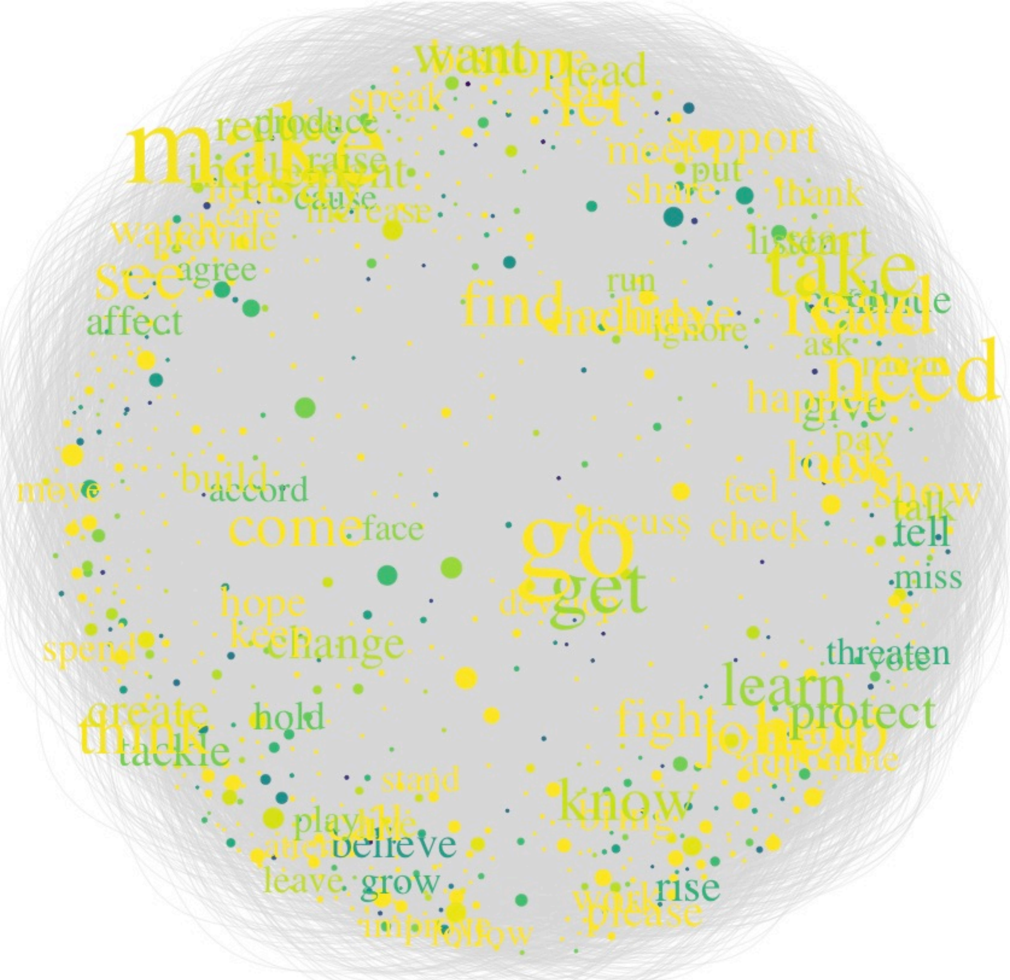
2017 → 2018



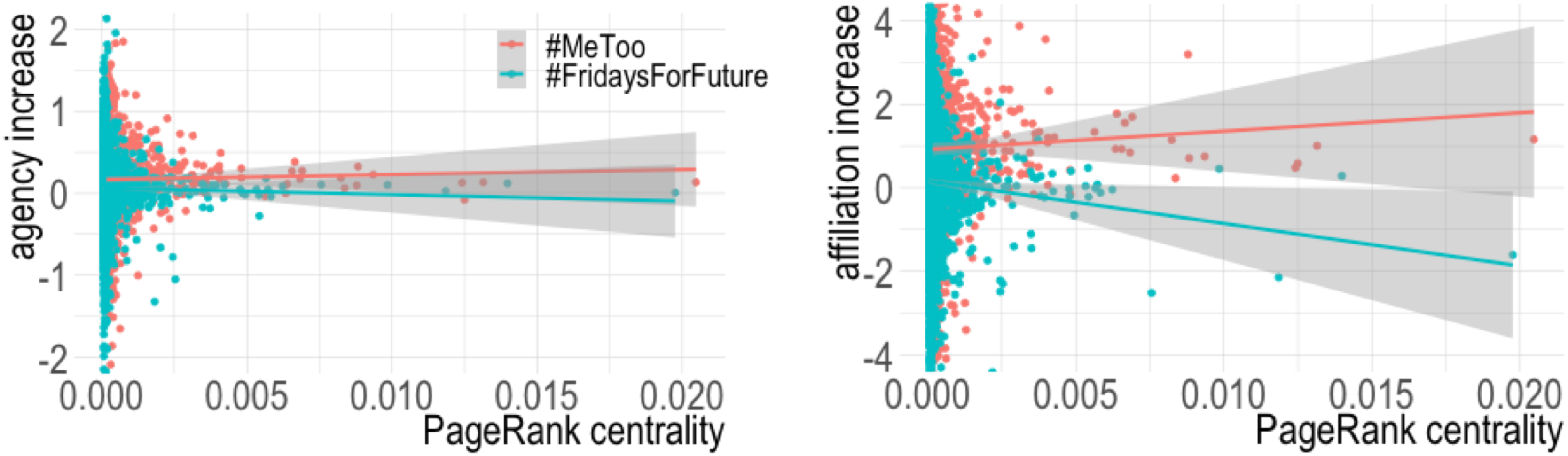
Affiliation increase in #FridaysForFuture

2018 → 2019





Relation with PageRank centrality



- Affiliation and agency **grow faster** in #MeToo (especially for high-ranked words)
- Statistically **meaningful** effect (mixed full-factorial linear model)

Conclusions

- PLMP is able to efficiently assign socio-psichological markers to words
- PLMP is a robust approach
- PLMP can capture structural semantic differences, e.g., in calls to action
  - #FridaysForFuture appears as a **sparser discourse** (less focused discussion)

    Planet in #FridaysForFuture in **not** agentic, as it appears in mixed tweets
  - #MeToo is much more **focused** (focused discussion)

    Woman in #MeToo is agentic as it only appears in agentic tweets
- Worth applying to
  - different contexts (e.g., political elections)
  - similar contexts (e.g., human right as in #MeToo, scientific matter as in #FridaysForFuture)