

Social networks of lexical innovation

Investigating the diffusion of neologisms on Twitter

Quirin Würschinger
LMU Munich

6th March 2020

Todo list

technical innovation, example like <i>blockchain</i>	2
code / plot: mark subsets in plots	6
decide: either overall network status or last stage	15

Contents

1 Introduction	2
2 Modeling the conventionalization of lexical innovations	2
3 Investigation the conventionalization of lexical innovations empirically	3
4 Going beyond frequency	3
5 Data	4
6 Method: social network analysis	4
7 Frequency: diffusion trajectories	4
7.1 Sample selection	4
7.1.1 General sample	4
7.1.2 Case studies	5
8 Diffusion over time	6
8.1 No diffusion: <i>microflat</i>	6
8.2 Successful diffusion	6
8.2.1 <i>ghosting</i>	6
8.2.2 <i>upcycling</i>	10
8.2.3 Increasing diffusion: <i>lituuation</i>	11

8.3	Unsuccessful diffusion	14
8.3.1	<i>alt-left</i>	14
8.3.2	<i>solopreneur</i>	15
9	Overall degrees of diffusion	15
9.1	Networks	15
9.2	Usage intensity	18
9.3	Networks and usage intensity	18
10	Frequency vs. networks	19

1 Introduction

- **Social media:** has changed the way we communitate.
 - It has changed the social fabric of our society (elections, press vs. ‘influencers’) and the sociolinguistic dynamics of how we communicate (fake news)
 - It has also changed the language system and the way the language system changes. Much as **cultural innovations:** like XXX ‘go viral’, new digital modes of communication also affect the way **linguistic innovations:** spread.
 - real viruses like Corona seem to spread similarly as linguistic innovations ‘go viral’ (editorial staff, 2020)
- This opens up new research questions and new ways to tackle previous questions in sociolinguistics. (sociolinguistics → computational sociolinguistics)
 - new data
 - new methods: social network analysis
 - research questions
 - How do new words spread?
 - Which factors influence their spread?

technical innovation, example like *blockchain*

2 Modeling the conventionalization of lexical innovations

- Research question: how do new words spread in the speech community?
- previous perspectives
 - **structural:** language system, lexicalization, institutionalization, word-formation processes etc. Bauer, 1983; Lipka, 2005
 - **cognitive:** (Schmid, 2008)
 - **sociolinguistic:** S-curves (Labov, 2007; Milroy, 1992)
- current framework: based on the EC-Model (Schmid, 2020)

- spread across usage contexts
- spread across speakers

3 Investigation the conventionalization of lexical innovations empirically

- Previous work has produced some important insights.
- I focus on the sociolinguistic dimension of lexical innovation in this paper.
- Previous empirical approaches have been limited in studying this because of the lack of information regarding the sociolinguistic dynamics of the spread of new words: how many speakers are affected? how are they interacting?

Overview of previous approaches

- traditional corpora (Elsen, 2004)
- web corpora Renouf, Kehoe and Banerjee, 2006; Kerremans, Stegmayr and Schmid, 2012
 - linguistic creativity and innovation happen there
 - big amounts of data
 - * big neologism samples
 - * big corpora (low-frequency nature of neologisms)
 - more informal sources
- social media corpora Grieve, Nini and Guo, 2016; Eisenstein, O'Connor, Smith and Xing, 2014
- hotbed
- driving force
- social network information
 - users
 - community characteristics
 - influencers

4 Going beyond frequency

- frequency
- corpus-as-input & corpus-as-output, usage intensity (Stefanowitsch & Flach, 2017)
- sociolinguistic information
 - number of users
 - social network characteristics

- influencers

5 Data

- sample
 - basis: bottom-up selection by NeoCrawler (Kerremans & Prokic, 2018, 2)
 - extension
 - * quite stable: not topical
 - * reasonably successful: e.g. technical innovations like *blockchain*
 - * sociolinguistically interesting: e.g. political terms such as *covfefe*
- corpus
 - longitudinal: retrospective
 - big data
 - social network information

6 Method: social network analysis

- basis for networks: interactions between users
 - mentions
 - retweets
- anatomy of a tweet
- network structure
 - nodes: users
 - edges: interactions

7 Frequency: diffusion trajectories

usage intensity serves as a baseline

7.1 Sample selection

7.1.1 General sample

- clustering: the words can be clustered in these groups (Kerremans, 2015)
- distinguishing between stable and unstable usage: [coefficient of variation](#)
- distinguishing between degree of success: [cumulative usage intensity](#)
 - no success
 - limited

- advanced
- S-curves
 - we don't expect S-curve trajectories for **topical**: neologisms because of variable conceptual salience (cf. Nini, Corradini, Guo and Grieve, 2017)
 - for stable neologisms we might expect S-curves [model testing for S-curves](#)
 - * according to sociolinguistic theory we expect certain sociolinguistic dynamics in their spread
 - * in the following sections we will employ social network analysis to empirically test these longstanding hypotheses

7.1.2 Case studies

- selection criteria
 - stable
 - successful
 - sociolinguistically marked vs. unmarked
 - comparable in frequency
- selection
 - advanced conventionalization: *shareable*
 - limited conventionalization:
 - * *alt-right*
 - * *alt-left*

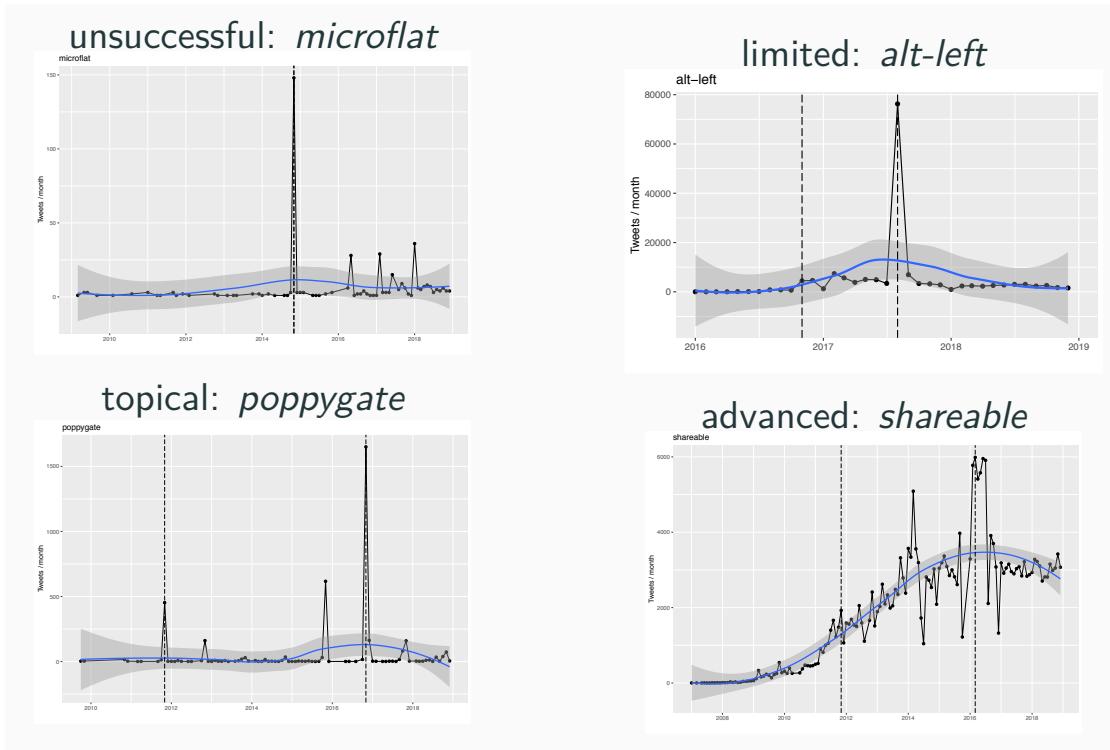


Figure 1: Usage intensity for case studies

Subsetting

I subset four stages in the diffusion to zoom in on different phases of the diffusion process

- first: the first 1,000 attestations
- mean
- max
- last

code /
plot:
mark sub-
sets in
plots

I will go beyond frequency and look into the sociolinguistic dynamics more closely

- sociolinguistic dynamics of diffusion over time
- sociolinguistic conventionality status of neologism

8 Diffusion over time

8.1 No diffusion: *microflat*

8.2 Successful diffusion

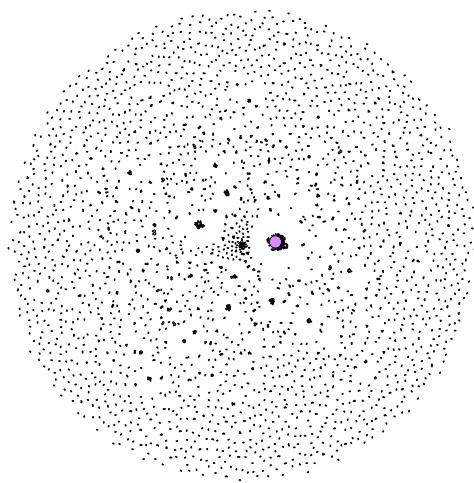
8.2.1 *ghosting*

- background of *shareable*

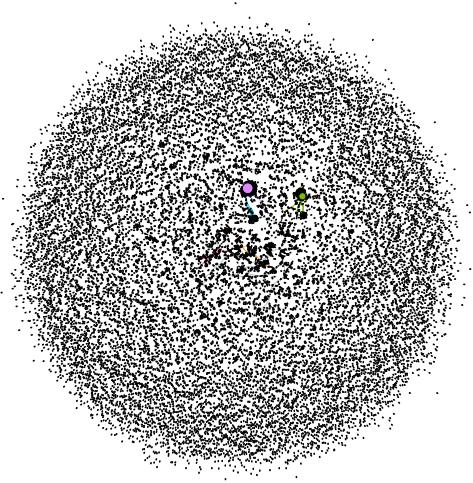
- linguistic
 - * form: It's
 - * meaning: broad, diverse scope; significant and stable increase in ‘semantic carrying capacity’
- diffusion history: first attestations, common uses, etc.
- template: S-curve
- shape corresponds quite nicely

Figure 2: Social network of diffusion for *ghosting* over time.

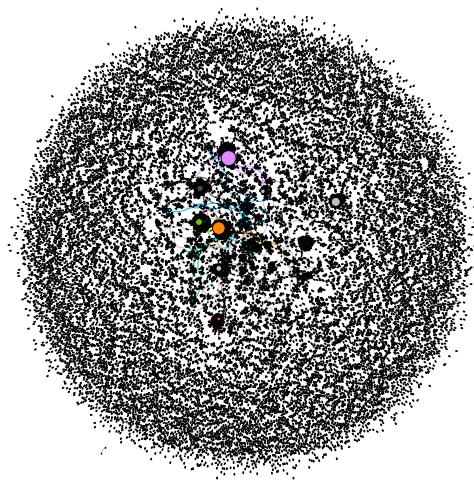
(a) First stage



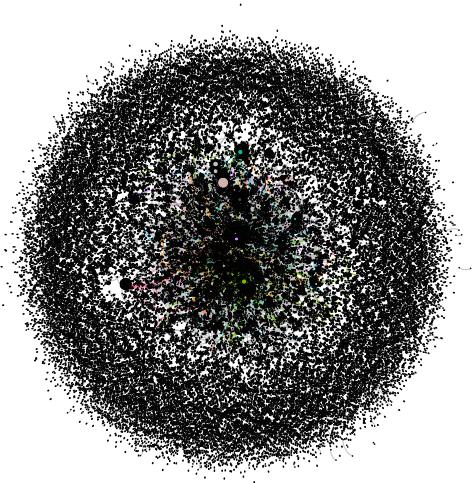
(b) Second stage



(c) Third stage



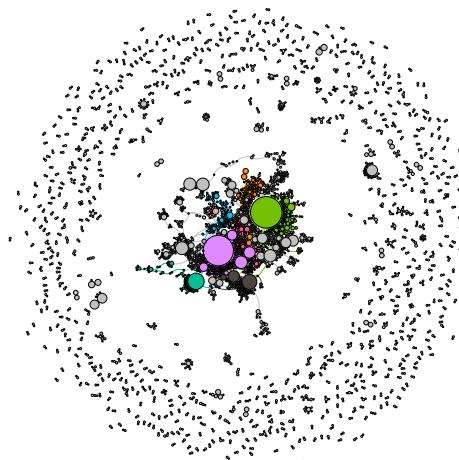
(d) Fourth stage



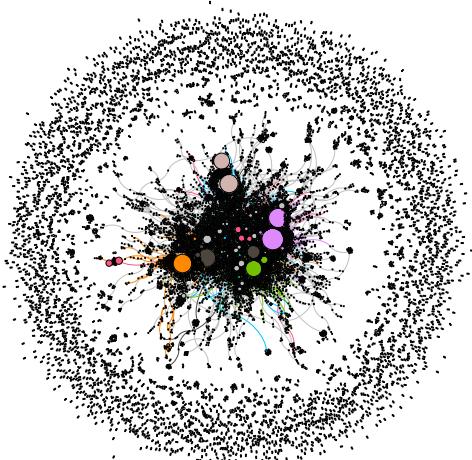
8.2.2 *upcycling*

Figure 4: Social network of diffusion for *upcycling* over time.

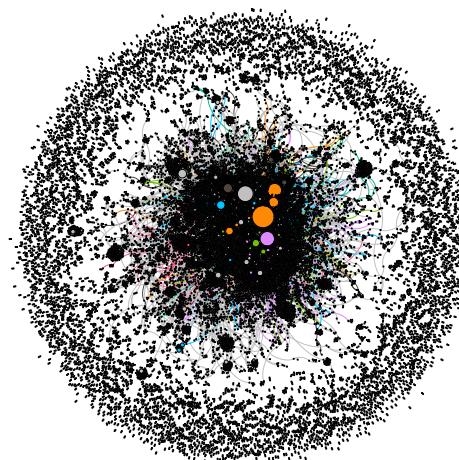
(a) First stage



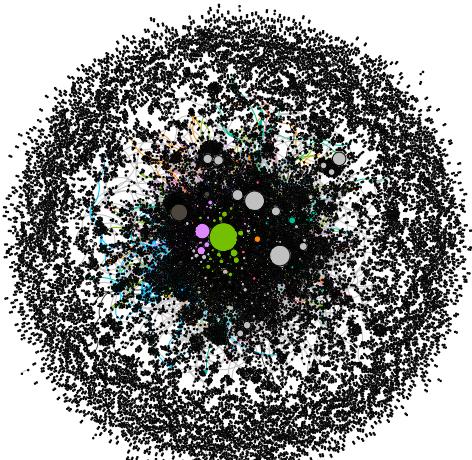
(b) Second stage



(c) Third stage



(d) Fourth stage



8.2.3 Increasing diffusion: *lituation*

Figure 6: Usage frequency of *lituation* over time.

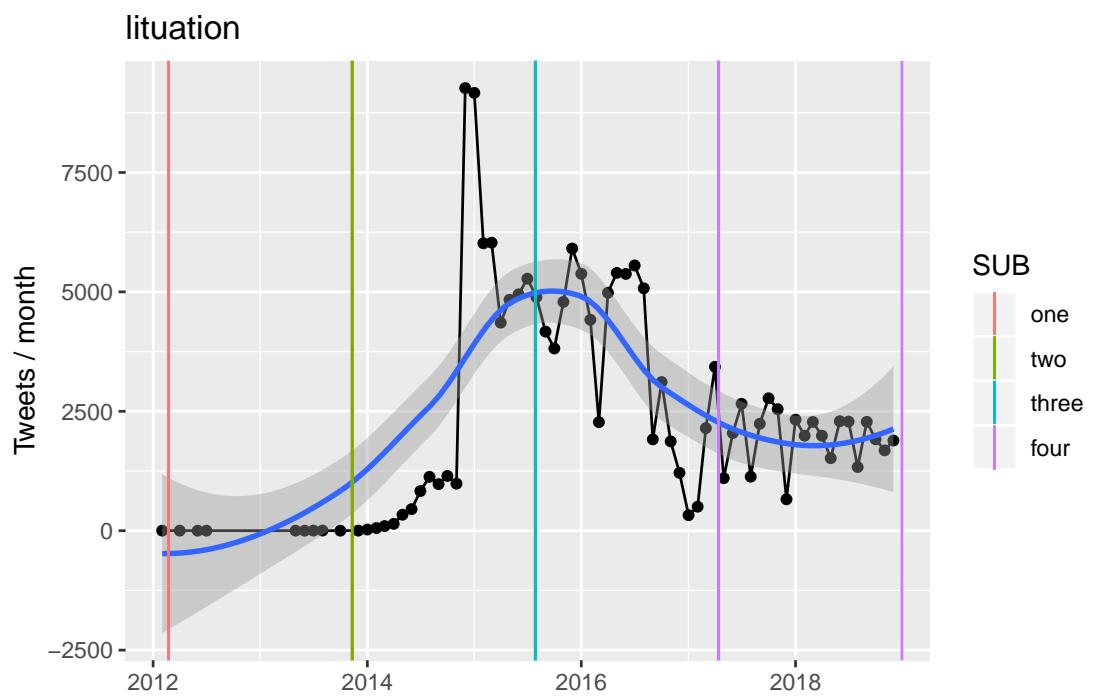
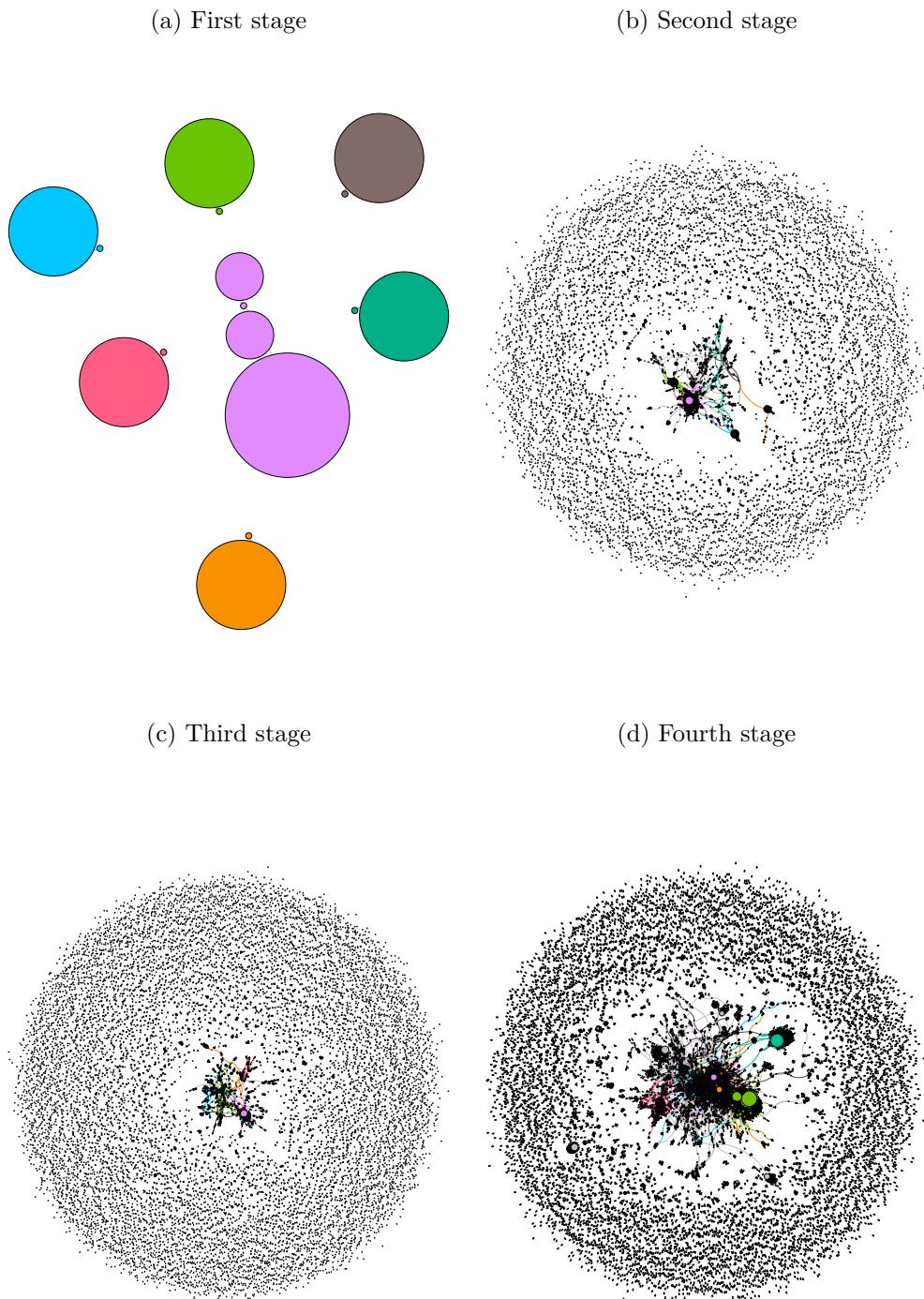


Figure 7: Social network of diffusion for *lituation* over time.



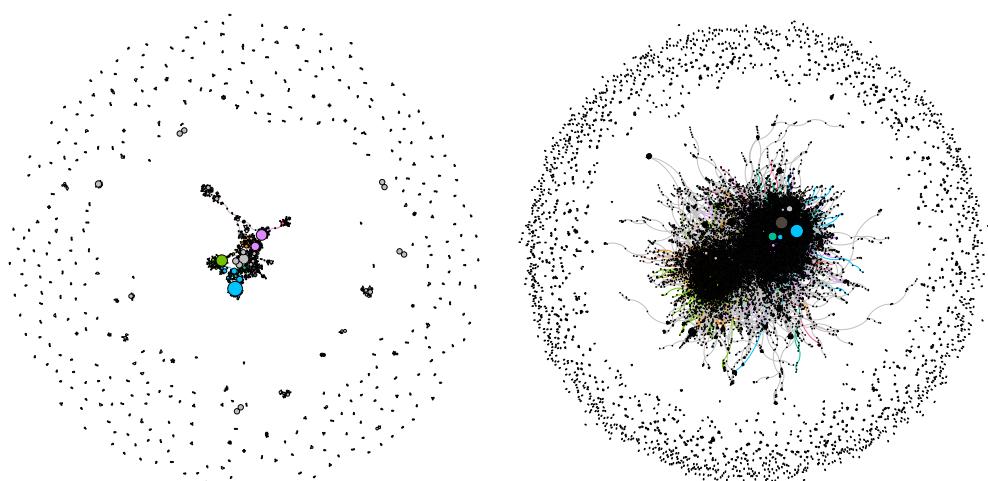
8.3 Unsuccessful diffusion

8.3.1 *alt-left*

Figure 9: Social network of diffusion for *alt-left* over time.

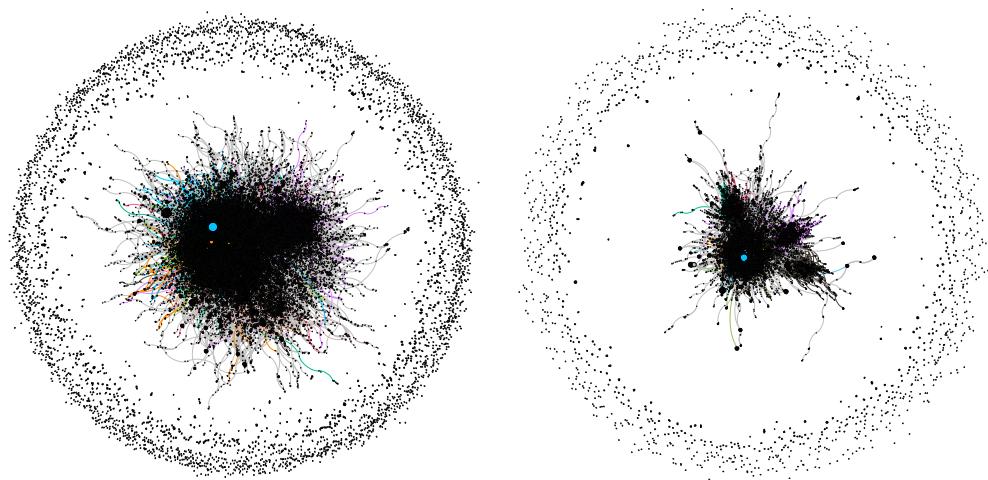
(a) First stage

(b) Second stage



(c) Third stage

(d) Fourth stage



8.3.2 *solopreneur*

Figure 11: Social network of diffusion for *solopreneur* over time.

9 Overall degrees of diffusion

9.1 Networks

Networks
last stage

decide:
either
overall
network
status or
last stage

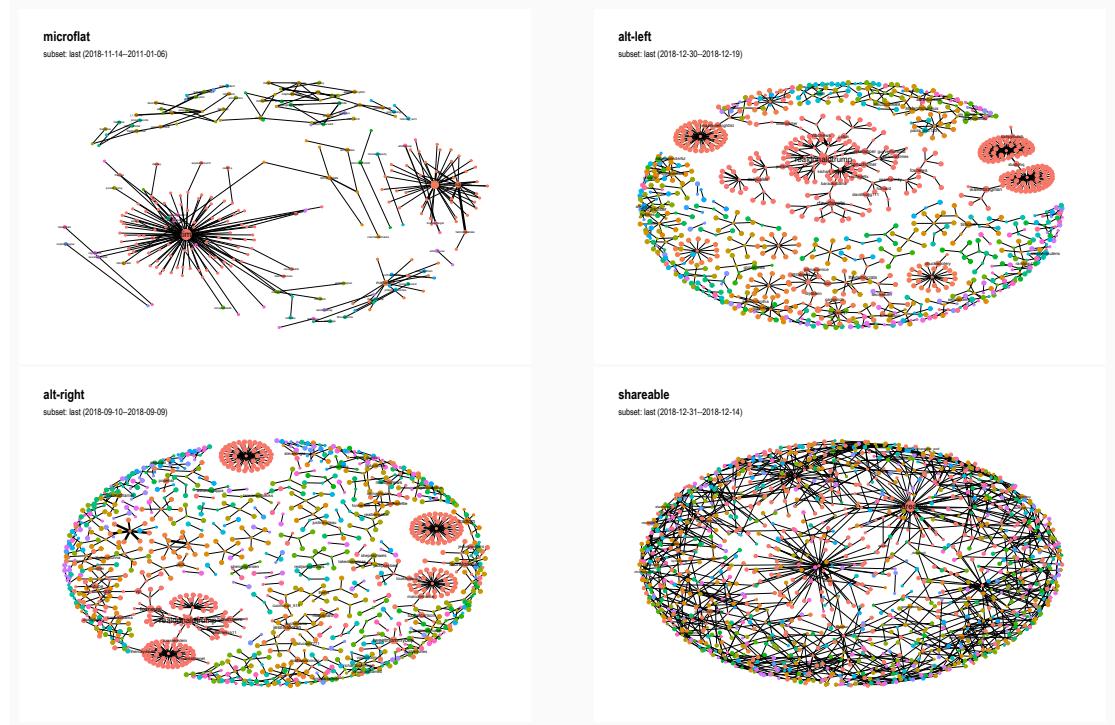


Figure 12: Network status for case study lemmas

whole period

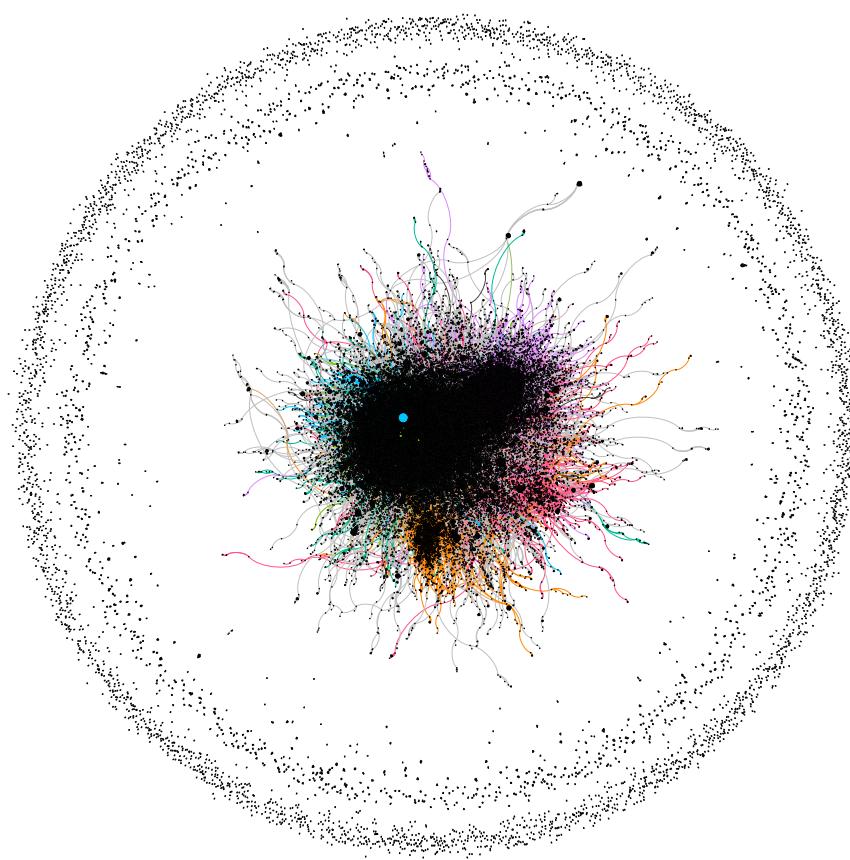


Figure 13: Network for interaction across the whole period for *alt-left*

Metrics

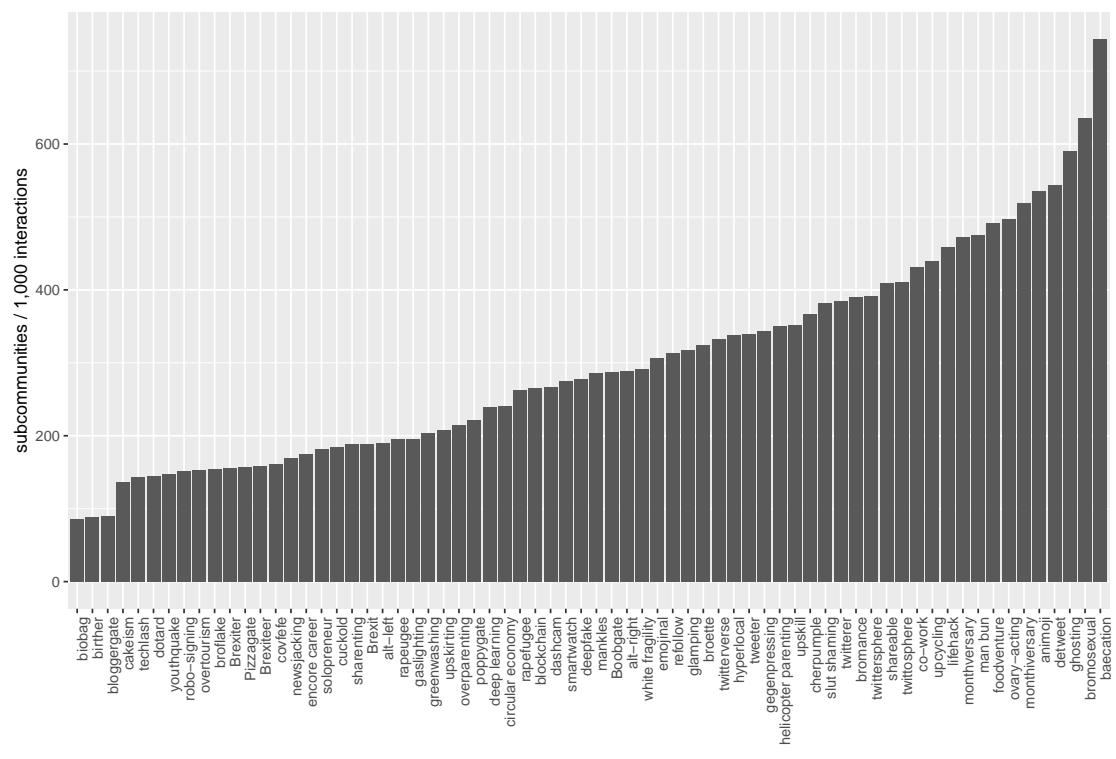


Figure 14: Network information for all lemmas

9.2 Usage intensity

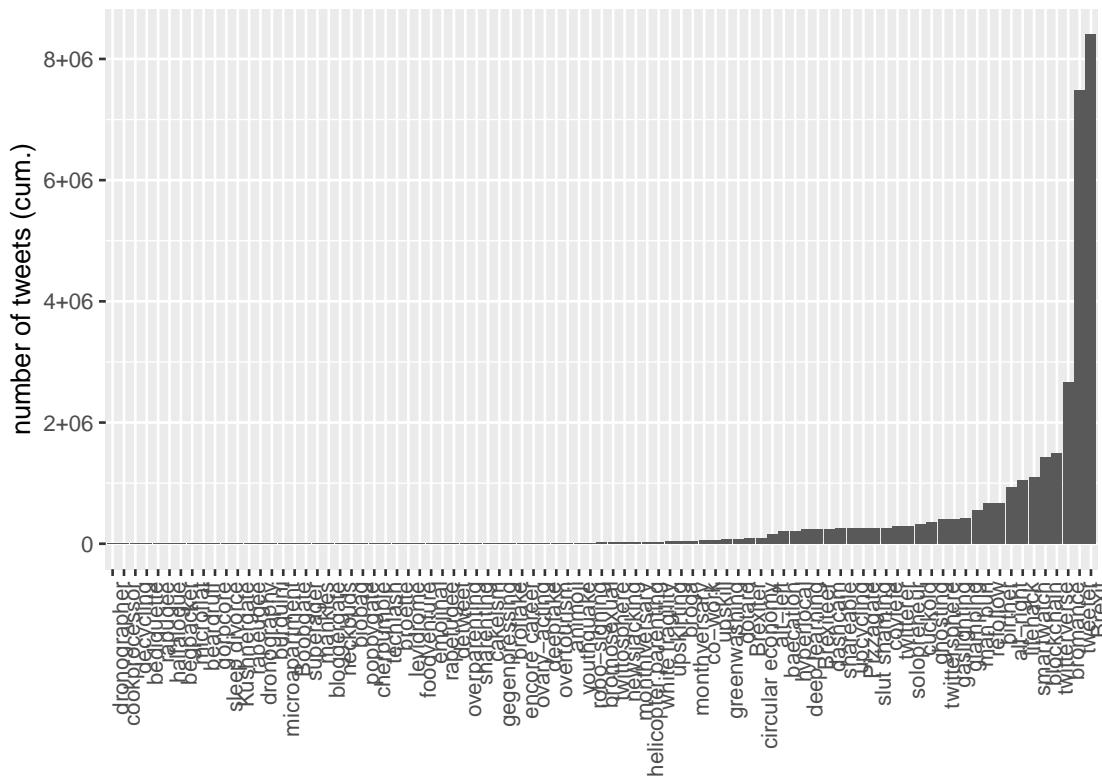


Figure 15: Cumulated frequency counts for all lemmas

9.3 Networks and usage intensity

Grouping and interpretation

- advanced
- topical
- little dispersion
 - political camps:
 - * propaganda: *alt-right*, *alt-left*, *covfefe*, *birther*
 - * Brexit terms: *Brexiteer*, *Brexiter*, *Brexit*
 - technical

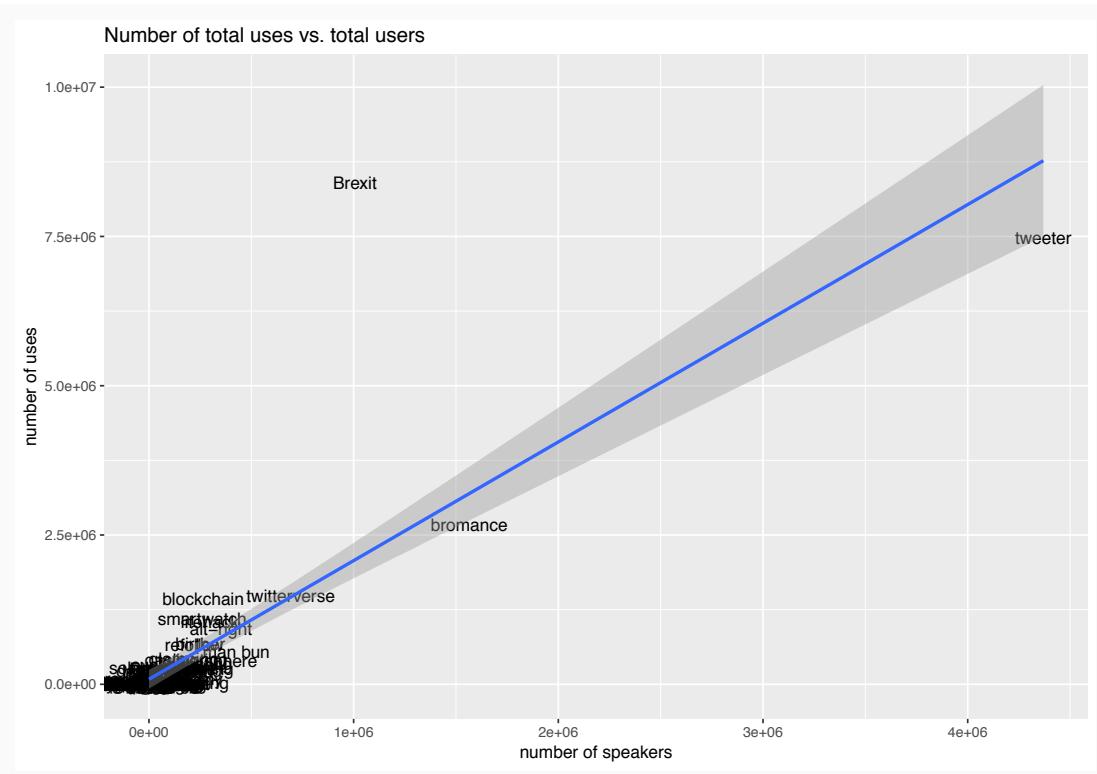


Figure 16: Users vs. usage

10 Frequency vs. networks

References

- Bauer, L. (1983). English word-formation. Cambridge: Cambridge University Press.

editorial staff, E. (2020). Diseases show to interact and spread like internet memes. Retrieved from <https://eandt.theiet.org/content/articles/2020/02/diseases-shown-to-interact-and-spread-like-internet-memes/>

Eisenstein, J., O'Connor, B., Smith, N. A. & Xing, E. P. (2014). Diffusion of lexical change in social media. *PLOS ONE*, 9(11), 1–13.

Elsen, H. (2004). Neologismen. Tübingen: Narr.

Grieve, J., Nini, A. & Guo, D. (2016). Analyzing lexical emergence in Modern American English online. *English Language and Linguistics*, (21), 99–127.

Kerremans, D. (2015). A web of new words. Frankfurt a. M.: Lang.

Kerremans, D. & Prokic, J. (2018). Mining the web for new words: semi-automatic neologism identification with the NeoCrawler. *Anglia*, (136), 239–268.

Kerremans, D., Stegmayr, S. & Schmid, H.-J. (2012). The NeoCrawler: Identifying and retrieving neologisms from the internet and monitoring ongoing change. (pp. 59–96). Berlin: Mouton de Gruyter.

- Labov, W. (2007). Transmission and diffusion. *Language*, 83(2), 344–387.
- Lipka, L. (2005). Lexicalization and institutionalization: revisited and extended. *SKASE Journal of Theoretical Linguistics*, 2(2), 40–42.
- Milroy, J. (1992). *Linguistic variation and change: On the historical sociolinguistics of english*. Oxford: Blackwell.
- Nini, A., Corradini, C., Guo, D. & Grieve, J. (2017). The application of growth curve modeling for the analysis of diachronic corpora. *Language Dynamics and Change*, 7(1), 102–125.
- Renouf, A., Kehoe, A. & Banerjee, J. (2006). Webcorp: an integrated system for web text search. Rodopi.
- Schmid, H.-J. (2008). New words in the mind: concept-formation and entrenchment of neologisms.
- Schmid, H.-J. (2020). *The dynamics of the linguistic system. Usage, conventionalization, and entrenchment*. Oxford: Oxford University Press.
- Stefanowitsch, A. & Flach, S. (2017). The corpus-based perspective on entrenchment. In H.-J. Schmid (Ed.), *Entrenchment and the psychology of language learning: how we reorganize and adapt linguistic knowledge* (pp. 101–128). Boston, USA: American Psychology Association and de Gruyter Mouton.