

Social networks of lexical innovation

Investigating the diffusion of neologisms on Twitter

Quirin Würschinger
LMU Munich

14th March 2020

Todo list

impact of social media social fabric: elections, press, influencers how we communicate: fake news language system: blends, acronyms, hashtags	3
---	---

Contents

1 Introduction	3
2 Theoretical background	4
2.1 Modelling the conventionalization of lexical innovations	4
2.1.1 Research perspectives	4
2.1.2 Current framework: the EC-Model (Schmid 2020)	4
2.1.3 The S-curve model	5
2.2 Investigating lexical innovation empirically	6
2.3 Goal: investigating the sociolinguistic dynamics	6
3 Data and methods	7
3.1 Data	7
3.2 Method: social network analysis	7
4 Sample	8
4.1 Neologism candidates	8
4.2 Case studies	8
5 Frequency	9
5.1 Overall degree of diffusion	10
5.2 Temporal dynamics	10
5.2.1 Coefficient of variation	10
5.3 Speaker counts	11

6 Diffusion over time	11
6.1 Advanced: <i>ghosting</i>	11
6.2 Advanced: <i>upcycling</i>	13
6.3 Increasing diffusion: <i>lituation</i>	15
6.4 Centralized use: <i>alt-left</i>	17
6.5 Narrowing: <i>solopreneur</i>	19
6.6 Overall diachronic development for case studies	20
7 Full sample	20
7.1 Diachronic development	20
7.2 Biggest changes	20
8 Networks vs. frequency	20
8.1 Plots	20
8.2 Correlation	20
8.3 Grouping and interpretation	20
9 Conclusion	22

1 Introduction

The Corona virus has recently spread with shocking speed and has tragically affected the lives of people around the world. Its fatal consequences have demonstrated the devastating power of exponential contagion. A leading research group of biologists analysing the spread of Corona have shown that the virus spreads supra-exponentially and they have likened its dynamics to the diffusion of cultural and linguistic innovations such as internet memes. (editorial staff, 2020) Does this confirm the popular perception that certain cultural and linguistic innovations ‘go viral’, after all?

Societies constantly evolve, new products and practices emerge and speakers continually invent and adopt new words which diffuse through social networks of communicative interaction. Influential models in sociolinguistics like the S-curve model (J. Milroy, 1992) and its compatibility with economic models (Rogers, 1962) suggest commonalities between the diffusion of cultural and linguistic innovations. These models assume that the diffusion of linguistic innovations across social networks follows universal trajectories and that rates of diffusion depend on sociolinguistic dynamics such as network density and the presence or absence of weak ties (Granovetter, 1977).

Unlike research on biological and cultural diffusion processes, sociolinguistic research has only fairly recently been provided with datasets that are suitable for large-scale, data-based approaches and that contain the necessary information to study these phenomena empirically using social network analyses.

The technological innovation of social media platforms like Twitter and the increasing availability of linguistic data from these sources for academic research allow for large-scale usage-based investigations of authentic language use.

The informal nature and the sheer size of these data have promoted significant advances in the study of new words, as neologisms are often propagated via social media channels and as their inherently low frequency of occurrence requires large corpora. In contrast to modern web corpora, which share these features, the availability of metadata about speakers and their interactions has been particularly important for the advancement of sociolinguistic research.

For example, recent sociolinguistic work on lexical innovation has used social media data to gain insights about general trajectories (Nini, Corradini, Guo & Grieve, 2017) and geographical patterns (Eisenstein, O’Connor, Smith & Xing, 2014; Grieve, 2017, 2018) of diffusion and about factors that influence whether new words spread successfully (Grieve, forthcoming).

new methods: social networks

The study of social networks has had a long research tradition in sociolinguistics and has shaped current models of diffusion (e.g. J. Milroy and Milroy, 1985). However, large-scale empirical corpus studies using social network analysis have only recently become feasible in the field of sociolinguistics.

promising in other fields: agent-based modeling also promising for lexical innovation
not been done enough I will do it

This paper presents an empirical approach to investigate these phenomena by study-

impact
of social
media so-
cial fabric:
elections,
press, in-
fluencers
how we
commu-
nicate:
fake news
language
system:
blends, ac-

ing the trajectory and sociolinguistic dynamics of how new words diffuse across social networks on Twitter.

Section X will present the theoretical framework for modelling the diffusion of new words ... and operationalization Section X will indicate the data and methods I use for my empirical investigation Section

2 Theoretical background

2.1 Modelling the conventionalization of lexical innovations

Research question: How do new words diffuse and become conventional lexical items in a language system?

2.1.1 Research perspectives

A substantial body of linguistic research has tackled this question from different **perspectives**. (Schmid, 2016, p. 16)

From a **structural** perspective, main areas of interest include which word-formation processes are involved in forming new words, whether they are formally and semantically transparent, whether they show variation and change in the process of lexicalization and which status the resulting neologisms have in the language system (institutionalization). (e.g. Bauer, 1983; Lipka, 2005)

Cognitive perspectives focus on how individuals process and store lexical innovations. Speakers generally use new words when they experience a communicative need to talk about entities or practices that cannot be expressed by their language's inventory of conventional words yet. In order for neologisms to successfully diffuse, speakers need to successfully negotiate their meaning (co-semiosis) in discourse, others need to adopt the behaviour of using these words (co-adaption). Continued exposure and use of new words can then lead to the entrenchment of new words in the mental lexicon of speakers. (Schmid, 2008)

Sociolinguistic perspectives transcend the level of the individual to study the diffusion of new words across speakers. The diffusion of lexical innovations is commonly seen as successful when the majority of the speech community has accepted a new word as a conventional lexical unit that can and is being used in communicative practice.

2.1.2 Current framework: the EC-Model (Schmid 2020)

The EC-Model integrates both structural, cognitive and sociolinguistic perspectives on language structure, variation and change.

Entrenchment: individuals

Conventionalization: society, sociolinguistics → focus of this paper

Usualization

Diffusion

definition

2.1.3 The S-curve model

S-curve models of linguistic change (Labov, 2007; J. Milroy, 1992; Nevalainen, 2015) assume universal sociolinguistic dynamics for the diffusion of linguistic innovations.

- The **trajectory** of spread is expected to follow an S-curve shape, with low rates of diffusion in early stages, followed by a period of accelerating spread with a tipping point at the mid point in the diffusion curve after which diffusion slows down and the curve flattens towards the end of the diffusion process.
- These temporal trajectories are assumed to correspond to the **sociolinguistic dynamics** of which individuals and groups interact with each other and adopt the target innovation.
 - In the **first stage** of slow diffusion only a small number of early adopters take up the innovative words. The individuals who use the new word typically form dense networks connected by strong ties. The structure of tight-knit communities of potentially like-minded individuals of similar backgrounds facilitates the successful negotiation of meaning (co-semiosis) of new words. High rates of interactions in these communities leads to high rates of exposure for individuals, which fosters co-adaption, entrenchment and usualization of new words in these communities.
 - In cases of successful diffusion the initial stages are followed by an **acceleration in spread** when new words increasingly reach speakers outside these tight-knit communities via weak ties (Granovetter, 1977). Rates of diffusion increase substantially when speakers that are not part of the initial group of early adopters start to accomodate the new words, allowing the innovations to reach a broader spectrum of the speech community.
 - In **later stages**, rates of diffusion slow down again as the majority of the speech community has already adopted the new words while smaller pockets of speakers remain reluctant to take up the new words.
- S-curve models have mainly been applied to the **linguistic domains** of phonology and syntax. Fundamental differences between lexemes and linguistic items on other levels such as phonemes and grammatical constructions might affect whether the validity and reliability of such models for *lexical* innovation.
 - For example, grammatical constructions such as the *going to* future used to express a speaker's future intention serve to fulfill relatively abstract communicative needs that remain stable over time. By contrast, on the lexical level, linguistic innovations are typically tied to concrete cultural referents such as products and practices whose conceptual relevance is much more volatile over time. For example, many lexical innovations such as *millennium bug* denoting the fear of a computer crash at the beginning of the new millennium can show high rates of diffusion and become entrenched and conventional among the majority of the speech community. Without continual conceptual relevance in public discourse, however, these words fail to pass on to the next generation.

of speakers. S-curves are commonly assumed to be found when linguistic innovations compete for ‘**semantic carrying capacity**’ (Nini et al., 2017), however, in many if not most cases of *lexical* innovation the conceptual carrying capacity is far from stable over time which represents a critical deviation from the traditional assumptions behind S-curves in language change.

- However, the generally strong theoretical and empirical basis of the S-curve model for language innovation and change, also from studies on the diffusion of cultural innovations (Rogers, 1962), and the precise formulation of the sociolinguistic dynamics underlying different phases of diffusion still make it an attractive **blueprint** for the empirical study of the sociolinguistic diffusion of lexical innovations.

2.2 Investigating lexical innovation 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 & Banerjee, 2006; Kerremans, Stegmayr & Schmid, 2012; Lemnitzer, n.d.; Gérard, 2017; Cartier, 2017)
 - 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 et al., 2014
- hotbed
- driving force
- social network information
 - users
 - community characteristics
 - influencers

2.3 Goal: investigating the sociolinguistic dynamics

- going beyond frequency

- sociolinguistic dynamics
 - number of users
 - social network characteristics
 - influencers

3 Data and methods

3.1 Data

- corpus
 - longitudinal: retrospective, starting from early attestations
 - big data
 - social network information
- sample
 - basis: bottom-up selection by NeoCrawler (Kerremans & Prokic, 2018, 2)
 - covering candidates from clusters found in earlier empirical work (Kerremans, 2015)
 - * no diffusion
 - * topical
 - * recurrent
 - * advanced
 - extension
 - * reasonably successful: e.g. technical innovations like *blockchain*
 - * sociolinguistically interesting: e.g. political terms such as *covfefe*

3.2 Method: social network analysis

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

4 Sample

4.1 Neologism candidates

- previous categorization (Kerremans, 2015)
 - no diffusion
 - topical
 - recurrent
 - advanced
- dimensions
 - overall degree of diffusion (synchronic): successful vs. unsuccessful: **usage frequency, degree centralization**
 - * no success
 - * limited
 - * advanced
 - temporal dynamics of diffusion (diachronic)
 - stability: stable vs. topical **coefficient of variation**
 - trend
 - diffusion: increasing degree of diffusion
 - centralization: decreasing degree of diffusion
 - speed

4.2 Case studies

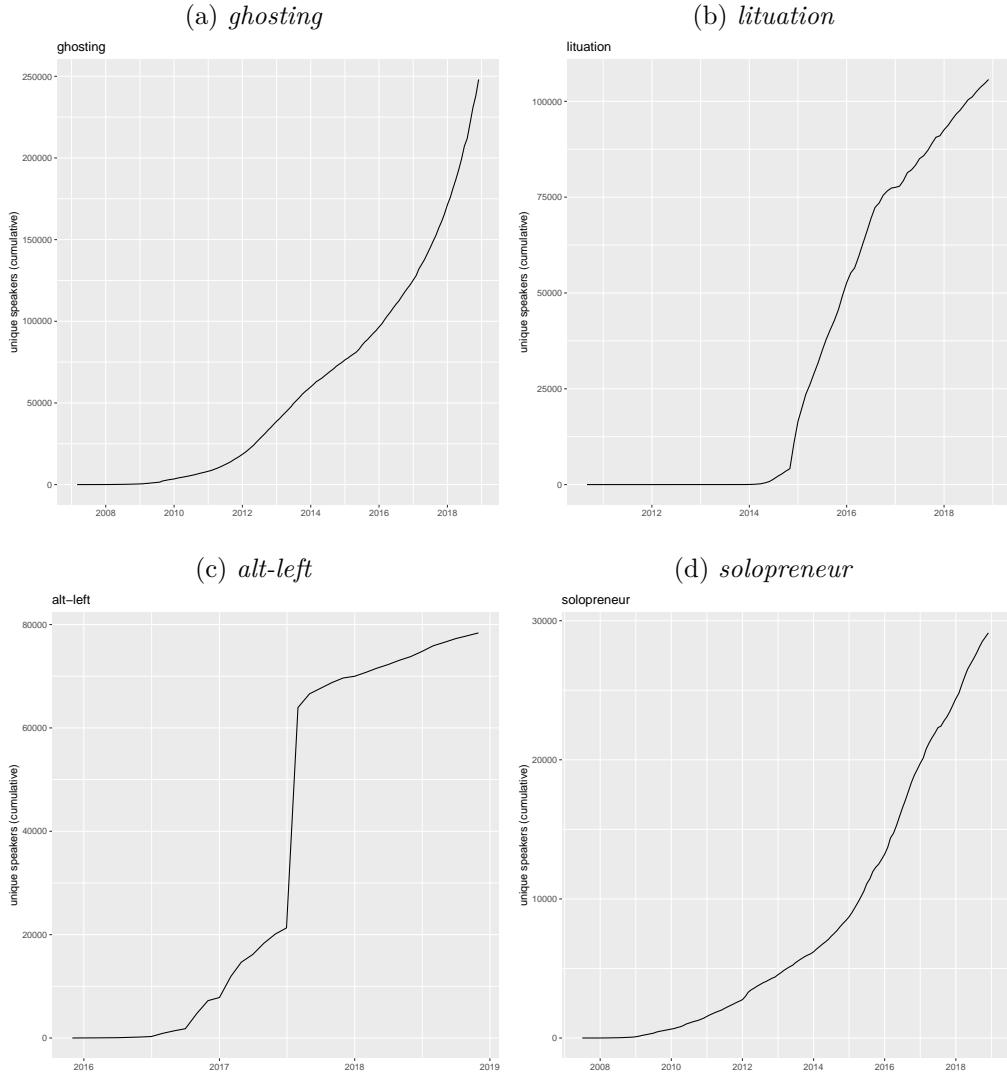
- criteria
 - covering clusters of neologism candidates
 - frequency counts comparable
- cases (Kerremans, 2015)
 - no diffusion: *microflat*
 - limited
 - * topical: *poppygate*
 - * centralized: *alt-left*
 - * decreasing: *solopreneur*
 - advanced diffusion:
 - * advanced: *upcycling*
 - * increasing: *hyperlocal*

5 Frequency

- Frequency is usually taken as the measure to approximate the degree of conventionality and entrenchment of linguistic entities.
- There are a number of underlying assumptions.
 - speech community: frequency = conventionality
 - * high freq. = majority of speakers show entrenchment
 - individuals: exposure → entrenchment
- problems
 - temporal dynamics (e.g. *millenium bug*)
 - output != input (Stefanowitsch & Flach, 2017)
 - high freq != many speakers
 - many speakers != ‘majority of the speech community’

5.1 Overall degree of diffusion

Figure 1: Cumulative counts for unique speakers that have used the target neologisms.



5.2 Temporal dynamics

5.2.1 Coefficient of variation

- most volatile candidates: *poppygate*, *burkini* etc.
- will be excluded from in-depth analyses

5.3 Speaker counts

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

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

6 Diffusion over time

6.1 Advanced: *ghosting*

Figure 3: Usage frequency for *ghosting*.

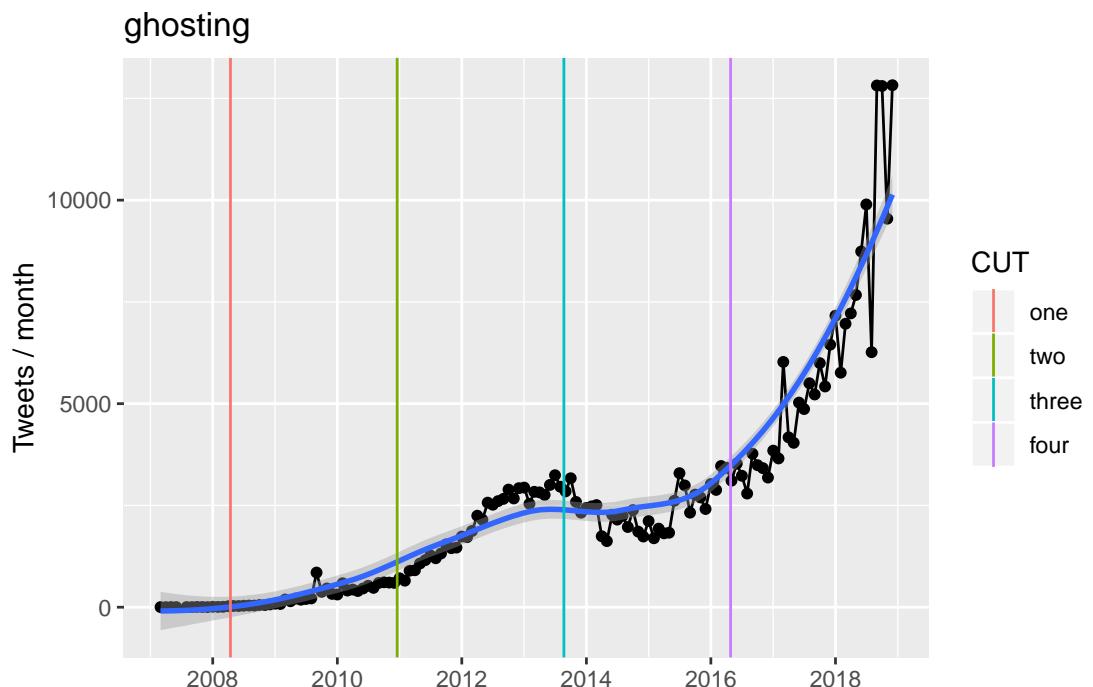
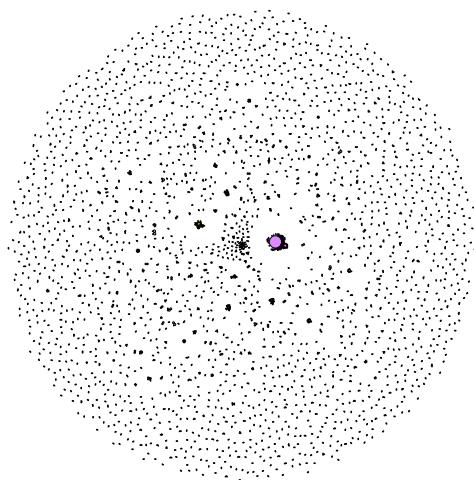
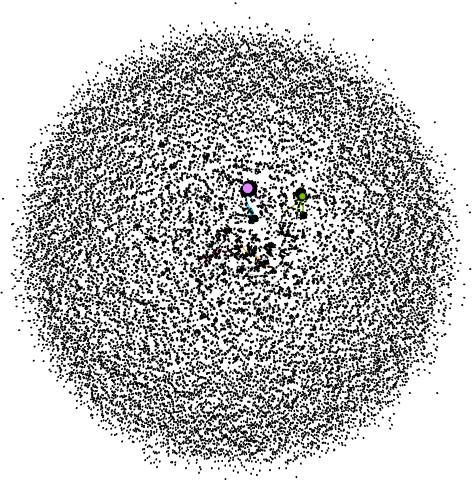


Figure 4: Social network of diffusion for *ghosting* over time.

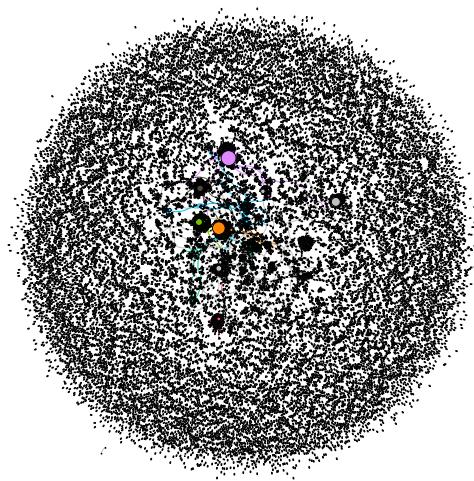
(a) First stage



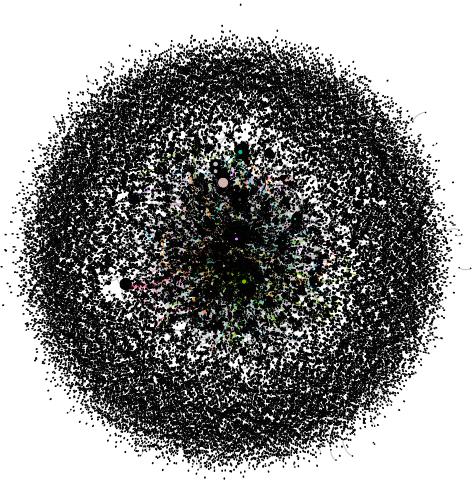
(b) Second stage



(c) Third stage



(d) Fourth stage



6.2 Advanced: *upcycling*

Figure 6: Usage frequency for *upcycling*.

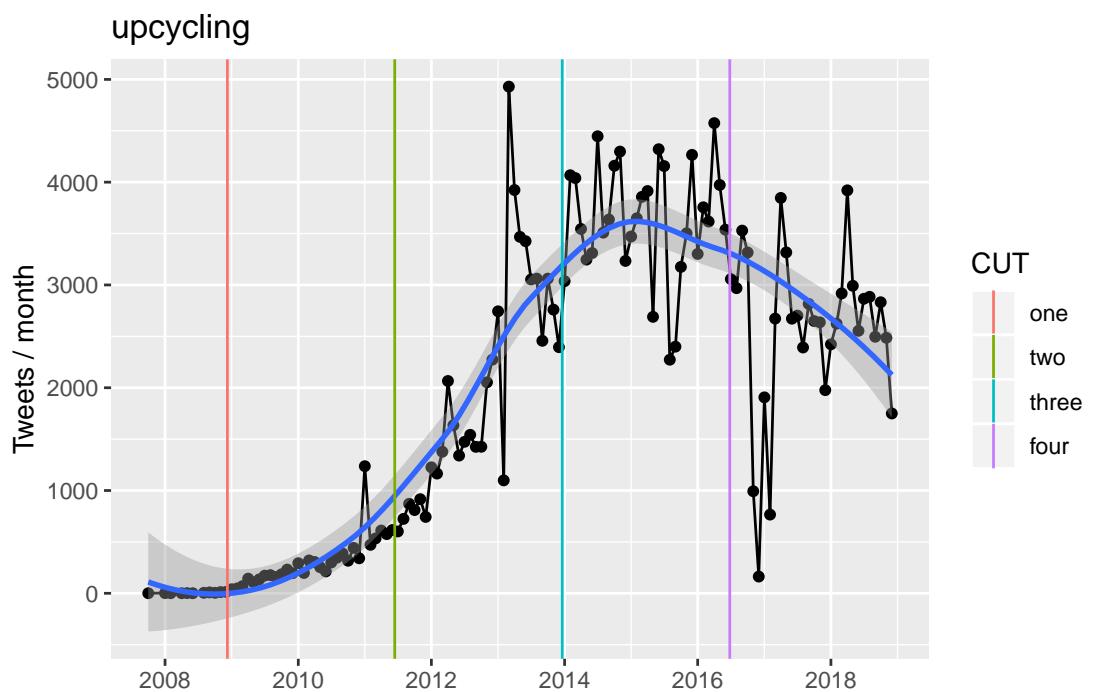
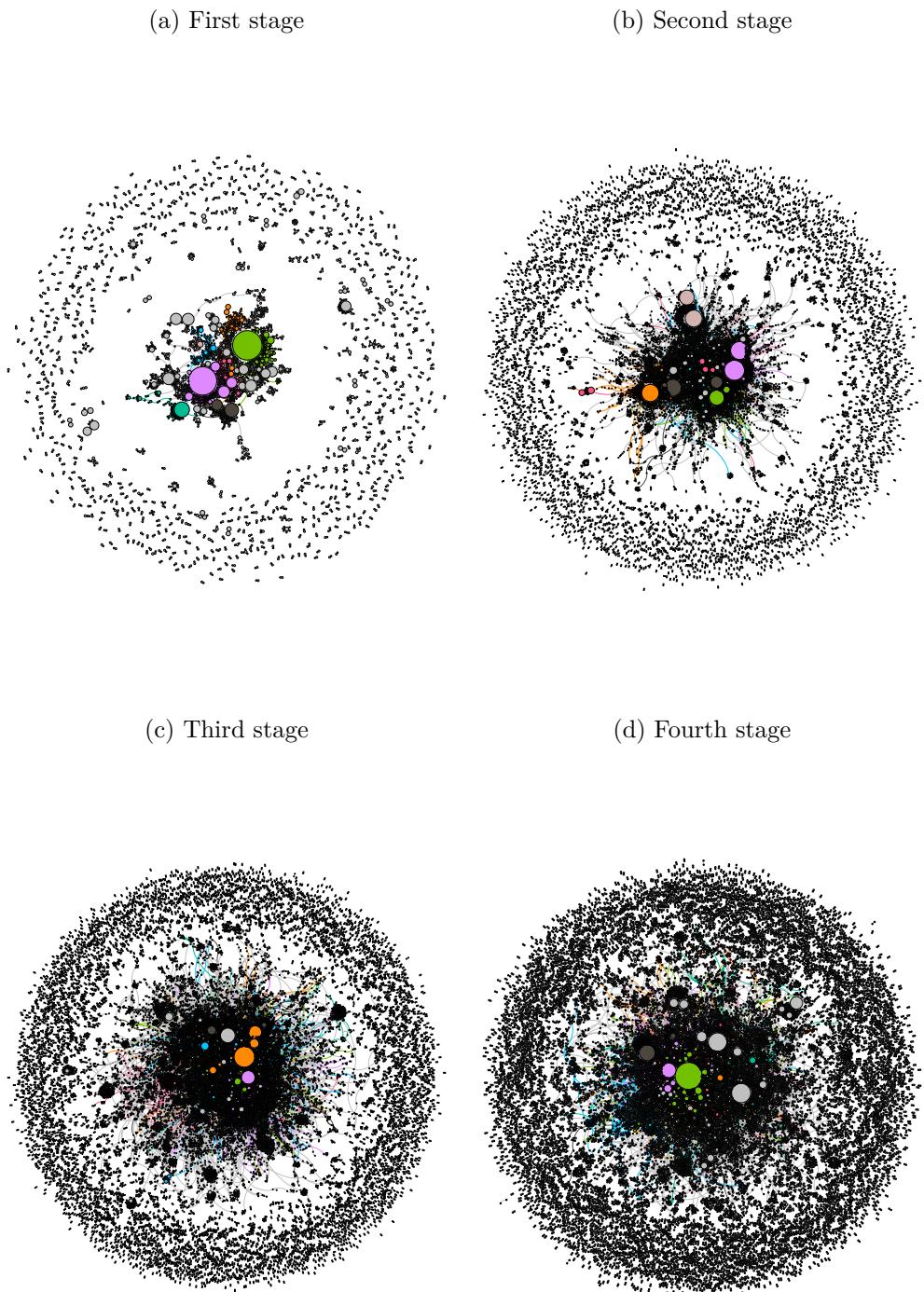


Figure 7: Social network of diffusion for *upcycling* over time.



6.3 Increasing diffusion: *lituation*

Figure 9: Usage frequency of *lituation* over time.

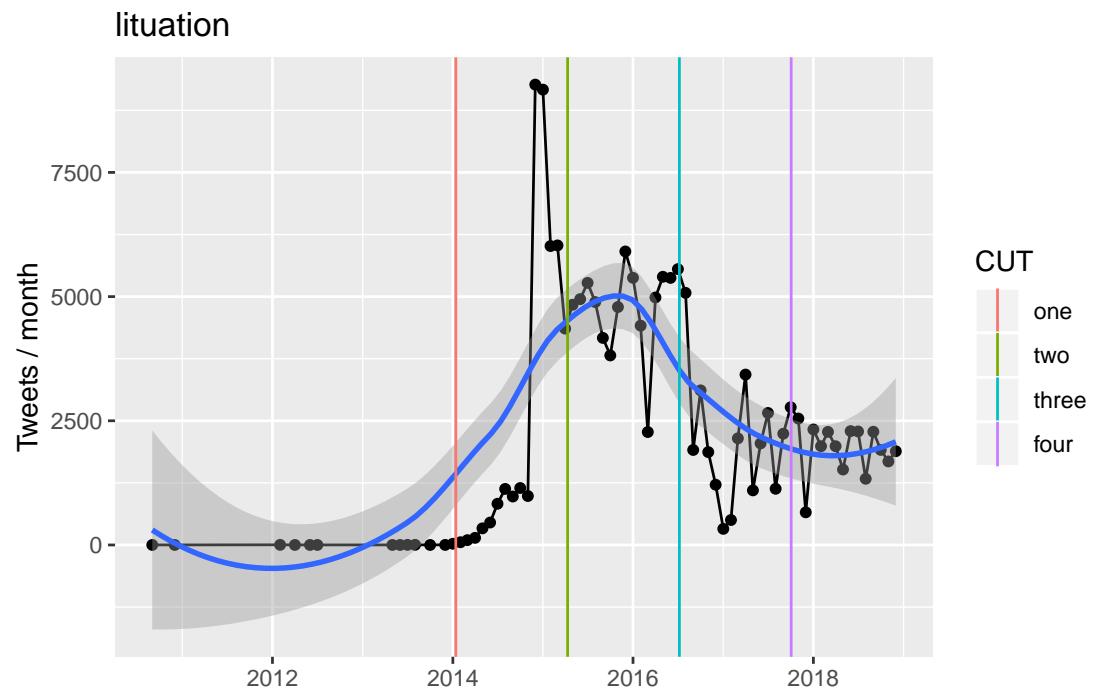
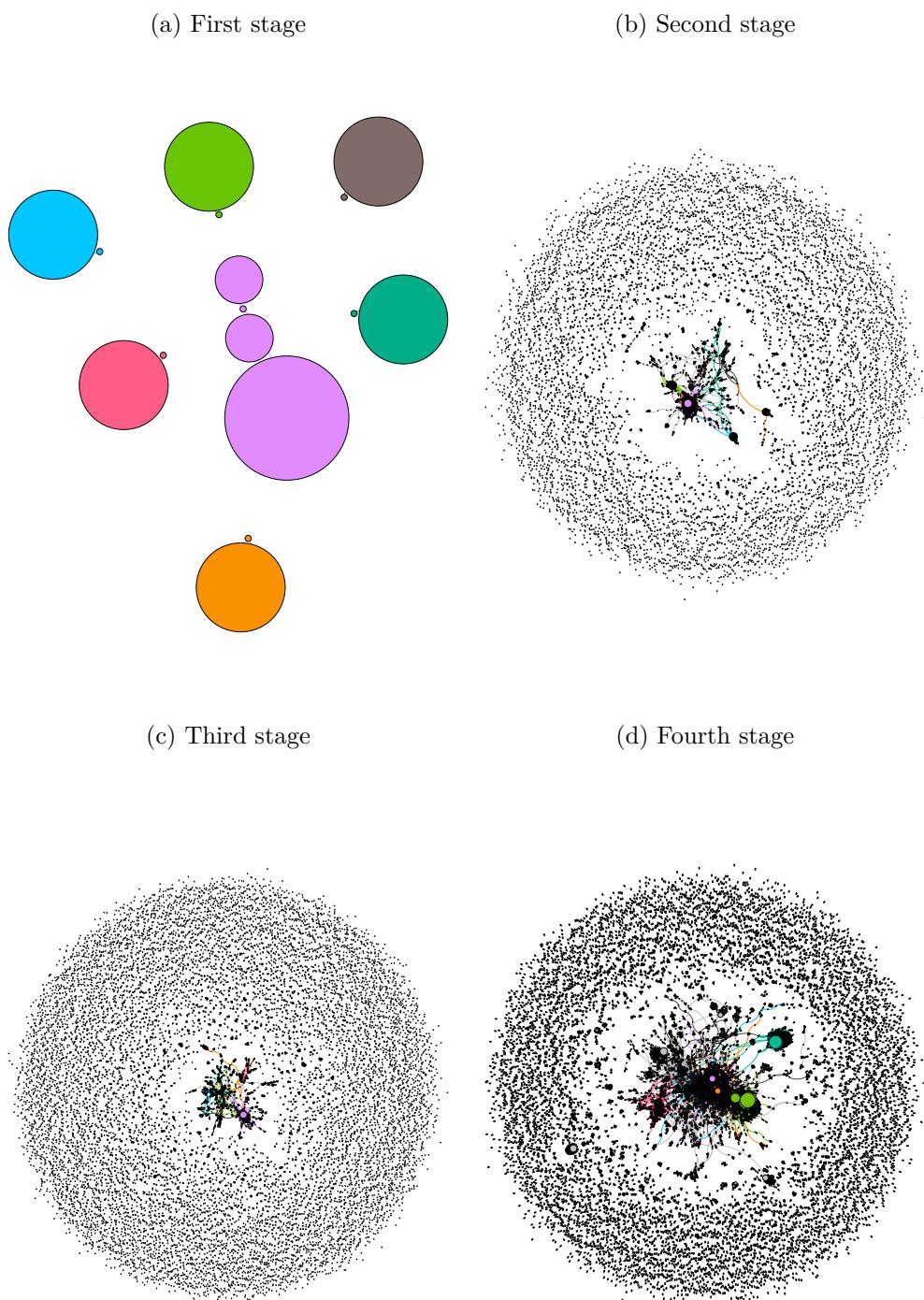


Figure 10: Social network of diffusion for *lituation* over time.



6.4 Centralized use: *alt-left*

Figure 12: Usage frequency of *alt-left* over time.

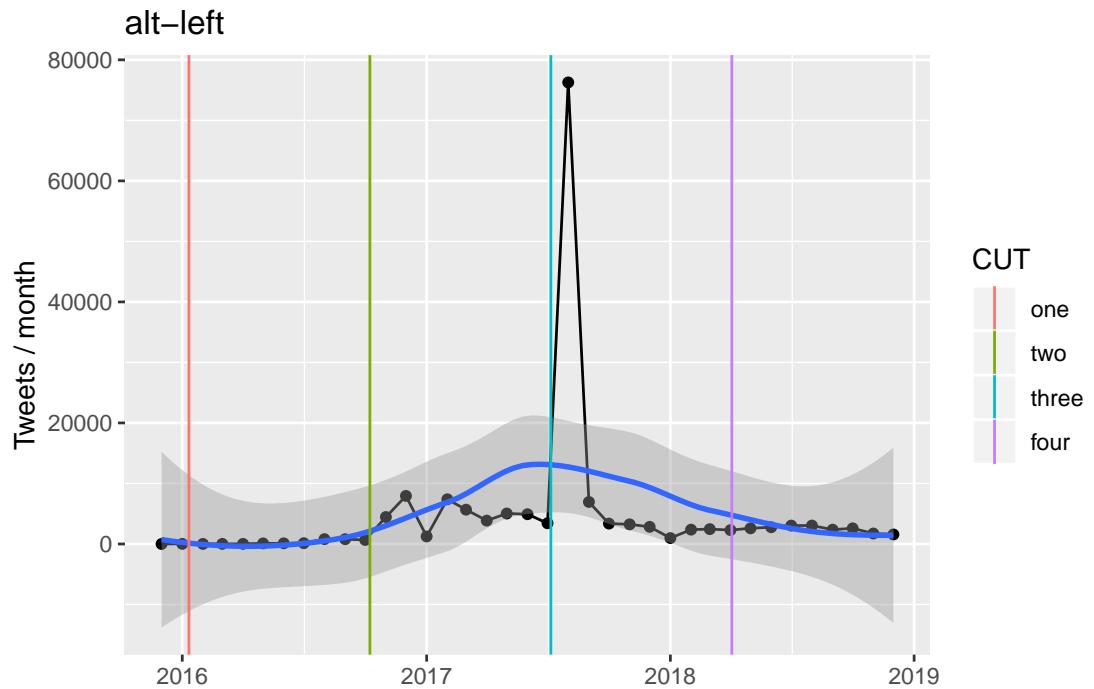
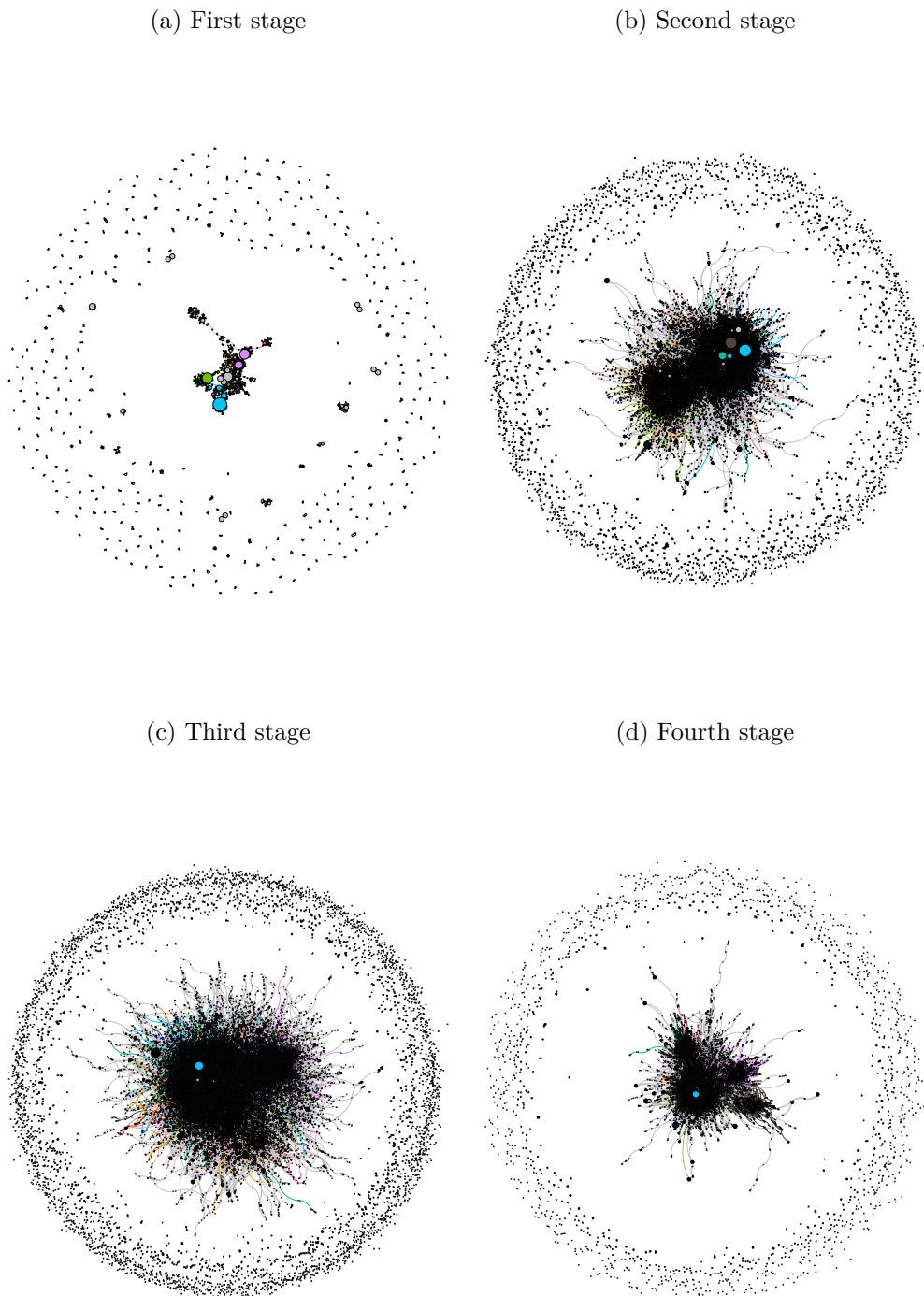


Figure 13: Social network of diffusion for *alt-left* over time.



6.5 Narrowing: *solo**preneur*

Figure 15: Usage frequency of *solo**preneur* over time.

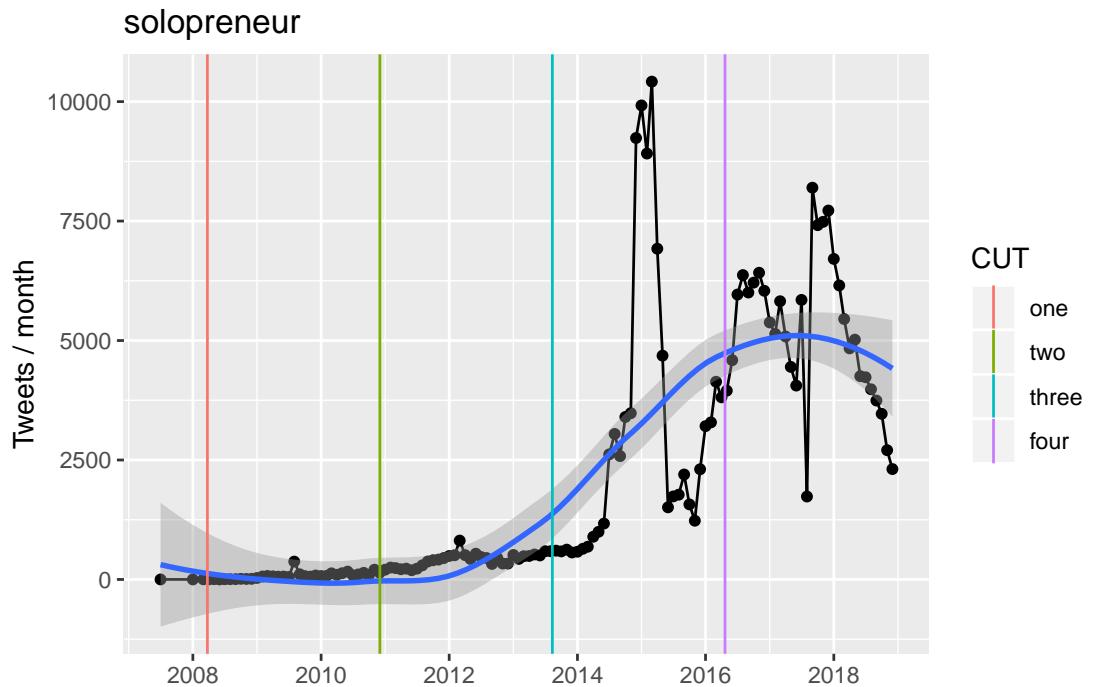
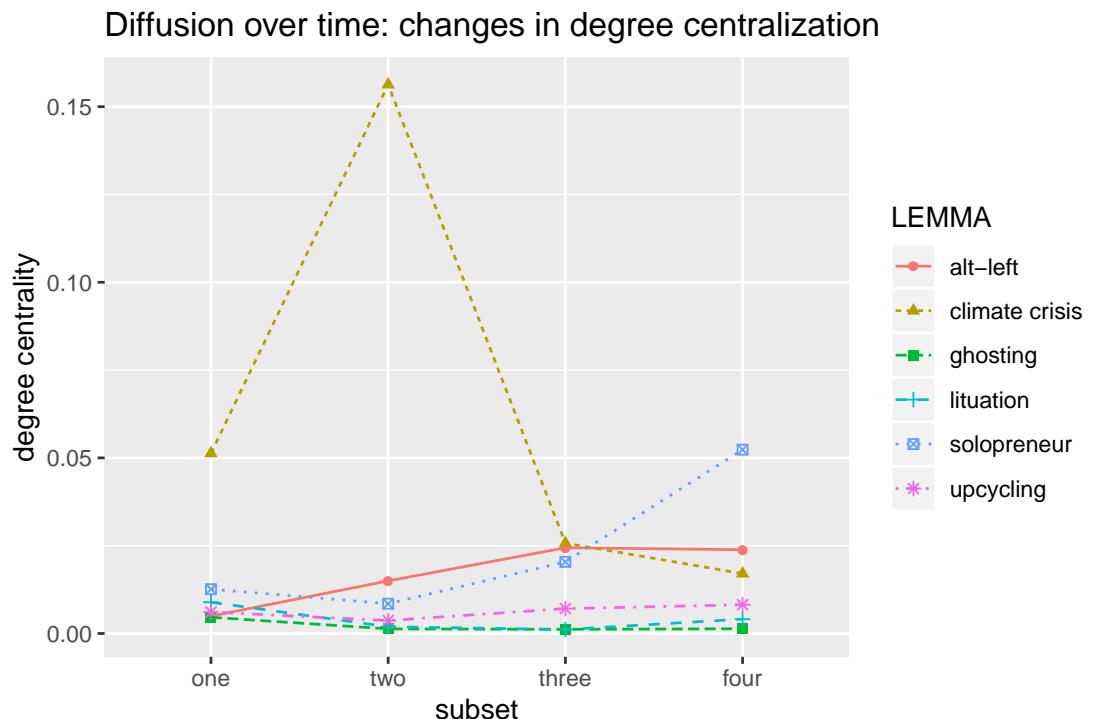


Figure 16: Social network of diffusion for *solo**preneur* over time.

6.6 Overall diachronic development for case studies

Figure 17: Degree centralization over time for case study words.



7 Full sample

7.1 Diachronic development

7.2 Biggest changes

- increasing diffusion
- centralization

8 Networks vs. frequency

8.1 Plots

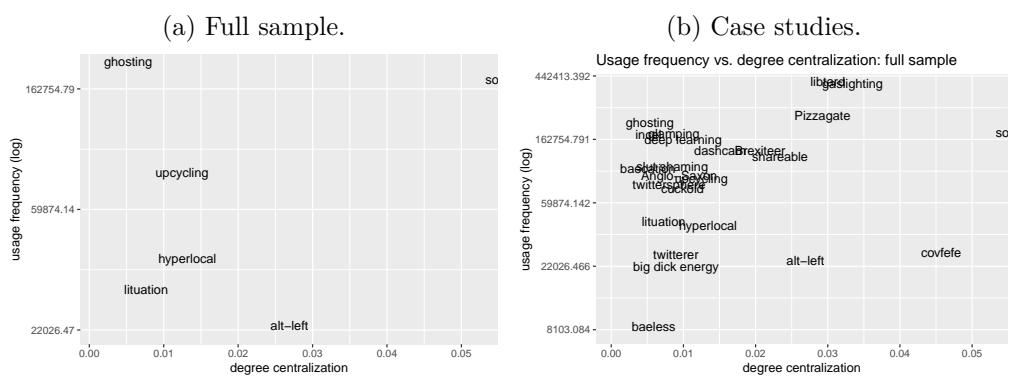
8.2 Correlation

8.3 Grouping and interpretation

Divergences between frequency and network analysis

- advanced

Figure 18: Degree centralization over time for the full sample.



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

9 Conclusion

- going beyond frequency is important

References

- Bauer, L. (1983). English word-formation. Cambridge: Cambridge University Press.
- Cartier, E. (2017). Neovelle, a web platform for neologism tracking. In *Proceedings of the software demonstrations of the 15th conference of the European chapter of the Association for Computational Linguistics* (pp. 95–98). Valencia, Spain: Association for Computational Linguistics.
- 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.
- Gérard, C. (2017). The logoscope: semi-automatic tool for detecting and documenting the contexts of french new words.
- Granovetter, M. S. (1977). The strength of weak ties. In S. Leinhardt (Ed.), *Social networks* (pp. 347–367). Academic Press.
- Grieve, J. (2017). *Geographical patterns of lexical innovation*. Workshop 'Diffusion of Lexical Innovations', LMU Munich.
- Grieve, J. (2018). Mapping emerging words in new york city.
- Grieve, J. (forthcoming). Natural selection in the modern English lexicon. *English Language and Applied Linguistics*.
- 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.
- Lemnitzer, L. (n.d.). Wortwarte. Retrieved January 14, 2018, from <http://www.wortwarte.de/>
- 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.
- Milroy, J. & Milroy, L. (1985). Linguistic change, social network and speaker innovation. *Journal of Linguistics*, 21(2), 339–384.
- Nevalainen, T. (2015). Descriptive adequacy of the s-curve model in diachronic studies of language change. VARIENG.
- 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.
- Rogers, E. M. (1962). *Diffusion of innovations*. New York: Free Press of Glencoe.
- Schmid, H.-J. (2008). New words in the mind: concept-formation and entrenchment of neologisms.
- Schmid, H.-J. (2016). *English morphology and word-formation: An introduction* (2nd ed.). Berlin: Erich Schmidt Verlag.
- 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.