Vol 5, No. 1, 2020

Journal of
language | works

SEIZ Matters: Modelling the spread of concepts on Twitter

Jonathan Hvithamar Rystrøm
School of Communication and Cognition, University of Aarhus
jhvithamar@gmail.com

Abstract

How do different concepts spread on social media? This question is becoming increasingly important as much of our time, discussion, and news consumption move online. This paper investigates the use of two models from epidemiology, namely the classical SI-model and the sociology-inspired SEIZ-model, to model and understand this phenomenon. I study the spread of two concepts during the 2019 Danish national election, klimatosse (climate fool) and Paludan on Twitter, both of which played key roles in the election season and had epidemic qualities in their usage throughout social media. I find that although both models can provide decent fits of the data, the SEIZ-model outperforms the SI-model by a wide margin. Furthermore, the parameters can be interpreted to provide a deeper understanding of the two phenomena and how they spread.

Keywords: contagion, SEIZ, Twitter, epidemiology

1. Introduction

Social media plays an increasingly vital role in our society. From its humble origins as a place to connect with friends, it has become a catalyst of revolution, playing a critical role in the Arab Spring (Howard et al., 2011). However, it also has a more sinister side, e.g. when bots spread mountains of misinformation during the Brexit campaign and during the 2016 US presidential election (Gorodnichenko et al., 2018; Howard & Kollanyi, 2016). Combined with hyper-targeted advertisement, the campaigns were able to shape public opinion in a way never seen before. This had led researchers to expand the classical two-step flow model of media (Katz, 1957) to a three-step flow model, where social media is seen as a middle step between mass media and interpersonal communication (Jensen, 2009; Zeller & Hermida, 2015). Understanding how opinions are formed and how different concepts originate and spread on social media is crucial for the future of our democracy. This paper aims to investigate a central piece of the puzzle: The spread of concepts on twitter.

Using the term ‘concept’ I refer to a word or a small set of words that describe the same phenomena. Concepts could be feelings such as “happiness” which can be described by a set of words such as happiness, joy and well-being. The notion of concept is thus able to encapsulate more than single words without being on the level of discourses (Parker, 1990).

Empirical evidence is central to solve the puzzle of how concepts spread. The discourses surrounding The Danish Election Cycle of 2019 created multiple interesting concepts. This paper will investigate two of these.

The first is the concept “klimatosse” (climate fool). When the Danish People’s Party suffered a resounding defeat in the election for the European Parliament, politician Pia Kjærsgaard partly blamed the results on “klimatosser” (-r indicates plural). The term became widely popular mainly with proponents of the green revolution and was recently named Word of the Year by the Danish Language Council (Vestergaard, 2019).

The second concept is part of the discourse surrounding Rasmus Paludan. Paludan is a far-right politician, whose party Stram Kurs ran for the national election. He got much media coverage by doing protests and publicly burning the Quran, which led to intense discussions on social media. For this analysis I have chosen to treat Paludan/Stram Kurs as a single concept, namely as the symbol for a previously taboo; a right-extremist, anti-politically correct movement. In much of the public discourse Paludan/Stram Kurs as a symbol was more predominant than the ordinary discussions revolving politicians and political parties.

This paper investigates and models the spread of these two concepts. To do so, necessitates an overview of the toolbox of previous research on the spread of concepts. The following sections provides such an overview.

1.1 Psycholinguistic theories of word spread

Language shapes our thoughts (Sapir, 1929; Whorf, 1940). From low-level perception of color (Winawer et al., 2007), space (Majid et al., 2004) and time (Casasanto et al., 2004; Casasanto & Boroditsky, 2008) to abstract constructs such as sexism (Dayhoff, 1983) the influence of words is pervasive. Understanding how we use language is therefore paramount to understand everything from individual thought processes to the structure of society.

Language, however, does not develop in a vacuum. It is created and evolved through everyday conversations, interactions and media consumptions. Historically, the study of word adoption has focused more on the individual rather than entire networks. Much of the existing research on this has focused on the role language play in children’s development (Pinker, 1994; Trecca et al., 2018) and acquisition of cognitive skills more generally (Tomasello, 1995, 2001). The application of research on individual word adoption is therefore not useful to investigation of the spread of concepts on social media.

Another approach to study words is the notion of interactive alignment (Fusaroli & Tylén, 2016; Garrod & Pickering, 2009). Interactive alignment is the process in which interlocutors align on different levels such as phonetic, lexical or semantic. According to (Clark, 1996) this eases the individual cognitive load and explains why conversations are easier than monologues although it intuitively requires more processing. Although a lot of research on this topic has been on face-to-face conversations, there is also evidence that the same effects apply in chat conversations (Michel & Smith, 2017). However, the focus of these studies has still been limited to dyadic interactions. It is therefore an open question how alignment influences multiple interacting interlocutors, as one would see in a social media setting.

1.2 Social media as a data source

One of the reasons dyadic and individual language has been the focus of the field is the availability of data. Individuals participate in a myriad of different communicative situations with many different groups throughout their day. Getting reliable data from a network this complex has historically been an insurmountable task. However, with social media this has changed. Most social media platforms keep an accurate record of all interactions (Goldfarb & Tucker, 2011), which offers researchers golden opportunities to test their hypotheses on hitherto unimaginable scales (Schwartz et al., 2013). Nevertheless, with the lack of theoretical frameworks to guide researchers in classical social communication literature, we must turn towards other fields for inspiration.

1.3 Influence from epidemiology

Given the social nature of language, researchers have likened the spread of ideas to the spread of diseases (Bettencourt et al., 2006). This in turn makes it possible to expand the analyzes of language use with tools from the vast toolbox of already existing epidemiological models. The main idea behind these models is to model the spread of a disease as people transition between different compartments. In the simplest model, the SI-model, people can be either Susceptible (S), or Infected (I) (Kermack & McKendrick, 1991). The dynamics between these compartments can be described by a system of differential equations (ODEs). In relation to the study of spread of phenomena in the social context, this would imply that as soon as susceptible individuals interact with infected individuals, they are immediately infected and start spreading the phenomena.

In comparison to the complex nature of language usage, the SI-model is far too simple. People do not necessarily immediately adopt an idea; depending on the complexity and social inertia of the idea it might take repeated exposure and social approval to finally adopt the phenomenon (Centola & Macy, 2007). Therefore, researchers have expanded the framework with the so-called SEIZ-model (Bettencourt et al., 2006) which allows for an exposure delay and for people to get exposed to a phenomenon without becoming affected. I elaborate on the SEIZ-models and its application to Twitter data in the methods section.

1.4 Influence from sociology

Early work on the spread and adoption of ideas and behaviors was centered around how these become widely accepted (e.g., Gramsci & Hoare, 1971). Much of later theory is focused on the barriers and “stickiness” of the adoption. (Centola & Macy, 2007; Romero et al., 2011). Broadly speaking, the theory divides the behavior into two categories of contagion: Simple and complex. A simple contagion is a behavior or idea that only requires a single exposure to be adopted. This is much like the dynamics of the common cold, which is well captured by naïve epidemiological models such as the flu. Classical network research has mainly focused on which settings facilitate the spread of simple contagions (Granovetter, 1977). The main findings have been that so-called “small world”-networks, consisting of many weak ties, diffuse simple contagions effectively.

However, many behaviors and ideas require more than just simple exposure to be adopted, namely complex contagion. A complex contagion is a behavior that requires multiple exposures and social reinforcement to be adopted. This calls for an entirely different network topology where people have more redundant ties to facilitate social reinforcement. Robust evidence show that although complex contagions take more time to spread because of the redundant ties, their adoption is stickier and lasts for longer (Centola & Macy, 2007).

Romero, Meeder, & Kleinberg (2011) have worked on diffusion of ideas, focusing on the proliferation of political hashtags on Twitter. They found that controversial political hashtags were more persistent, meaning that repeated exposures seemed to continue to have additional impact. Other work has focused more on how the different network structures impact the diffusion of ideas (Centola, 2010, 2011; Centola & Macy, 2007).

Another significant work combining the approaches from epidemiology, sociology and big data is Jin, Dougherty, Saraf, Cao, & Ramakrishnan (2013). They investigate how the spread of news stories differ from the spread of rumors using the SEIZ-model. Their research outlines a method for numerically estimating the parameters of the SEIZ-model, which I will adopt in this paper.

Although the research shows potential to tackle questions of diffusion, it is nevertheless relevant to investigate the spread of concepts. The concepts I investigate differ from hashtags in the fact that hashtags only live on social media, whereas words and concepts are used in all sorts of mediums. How social media influences the proliferation of concepts is therefore still an open question. Twitter is especially interesting in the Danish context, as it is used predominately by public stakeholders in Denmark (Derczynski et al., 2019; Olof Larsson & Moe, 2013) compared to e.g. snapchat or Instagram which are used more widely and recreatively (Chung et al., 2017; Larsen & Kofoed, 2016).

Based on this theoretical research, the present paper aims to investigate how well epidemiological models can be used to study the spread of new concepts on Twitter. Two questions are of particular interest: (1) whether the spread of concepts on Twitter can be accurately modeled using epidemiological models; (2) whether the SEIZ-model, which is tailored for social-epidemics, will outperform the more naïve SI-model.

2. Models and Methods

2.1 Explanation of SI- and SEIZ-model

In the following, I describe how I make the SI-model and the SEIZ-model comparable and applicable to the Twitter context.

SI-model: As previously mentioned, the SI-model has two compartments; Susceptible (S) and Infected (I). In my Twitter-model I will define them following Jin, Dougherty, Saraf, Cao, & Ramakrishnan (2013) but with the following modifications:

- Susceptibles will be users who have not yet tweeted about the topic
- Infected will be users who have tweeted about the topic
- Upon contact with an infected individual, the susceptible user will become immediately infected thus tweeting at once
- Once a user becomes infected, she will stay that way.

SEIZ-model: The SEIZ-model, introduced by (Bettencourt et al., 2006), is an improvement on the SI-model with regards to modelling social behavior. The main improvement comes from introducing two additional compartments: Exposed (E) and Skeptics (Z). In my implementation, Exposed will be users who have seen a tweet but wait a certain amount of time (called the exposure delay) before they tweet about it. The Skeptics will be users who have seen a tweet about the topic but choose not to tweet about it. Thus, the SEIZ-model allows for external factors to affect the user's decision during the exposure delay as well as social inertia with the Skeptic compartment.

The model shows that there are multiple ways a user can become infected. In the simplest case, she can become an Infected by initial contact with an Adopter (i.e. user in the Infected compartment). This is the type of behavior captured by the SI-model. Otherwise, there are two ways users can become infected after being exposed: They can either become an Infected by repeat exposure or they can spontaneously become Infected after an incubation period. The incubation period helps model the influence of exogenous factors; a shorter incubation period therefore signifies a higher influence of non-Twitter related sources.

It is also important to note that once a user becomes a Skeptic, they will never tweet about the subject.

2.2 Fitting the models

To fit the model, it is necessary to find the best numerical approximation for the parameters of the differential. For this task, I use the programming language, R (R Core Team, 2019), and the FME-package (Soetaert & Petzoldt, 2010). However, because of limitations in the data I must make the following assumptions:

- As I do not know the total population (N), the model will fit it as a constant parameter. Theoretically, N is the total number of users that could be exposed to the concepts. This is difficult to assess because of the unknown topology of the Danish twitter-network and the proliferation of bots.
- Because both the *klimatosse*-concept and *Paludan*-concept originated outside of Twitter, the initial parameters are unknown. The model will therefore fit them as constant parameters as well.

With these additional modifications to the model proposed by Jin, Dougherty, Saraf, Cao, & Ramakrishnan (2013) I will fit the SI-model and SEIZ-model to my datasets, and compare them to the ground-truth using the L2-error to compare the models.

2.3 Data

It is essential for my data to encompass the following attributes; First, data needs to capture a new concept which has had a rapid adoption process. Both concepts investigated in this paper were born during the election cycle and were widely discussed, which should make them fit the criterion. To verify, I checked how the concepts evolved on Google Trends[1]. This is, however, just a rough estimate as the patterns of Google searches does not necessarily follow the patterns of tweets.

Second, the users should ideally belong to a well-defined network, as to mitigate the effects of foreign news coverage of the phenomenon. This paper focuses solely on Danish-speaking users on Twitter as this creates a relatively limited group. A Danish-speaking user on Twitter is a user who has tweeted at least once about the subject in Danish. I use the language detection package *cldr* (The Chromium Authors et al., 2013) to do the filtering.

Lastly, the data should be as complete as possible. To accomplish this, I use Twitters Premium Search API, which has the advantage of gathering all tweets for a given period in contrast to the Standard Search API. Furthermore, the Premium Search API gives access to historical records instead of the two-week limit imposed by the Standard Search API.

I investigate the spread of concepts using two different datasets:

- *Klimatosse*: The dataset consists of 4557 tweets from 2533 unique users mentioning variations of the word *klimatosse* (Climate Fool) from the two weeks after 2019/05/26 where Pia Kjærsgaard gave her speech (Christoffersen, 2019).
- *Paludan*: The dataset consists of 44965 tweets from 10338 unique users mentioning the set of words “Paludan” or “Stram Kurs” between 2019/04/14 and 2019/06/07 (the day after the national election). This period was determined using Google Trends data.[2]

Thus, the model can be tested against both a rapid short-term epidemic (klimatosse) and a longer, more expansive epidemic (Paludan).

2.4 Processing

To fit the parameters to the data I make a data frame with cumulative count of the number of unique users who have tweeted about the topic. This corresponds to the user transferring to the Infected compartment the first time they tweet. For the *klimatosse*-data I summarize on the 15-minute level and for the *Paludan*-data I summarize on the 1-hour level because of computational limitation in the FME-package (Soetaert & Petzoldt, 2010).

[1] https://trends.google.com/
[2] See appendix

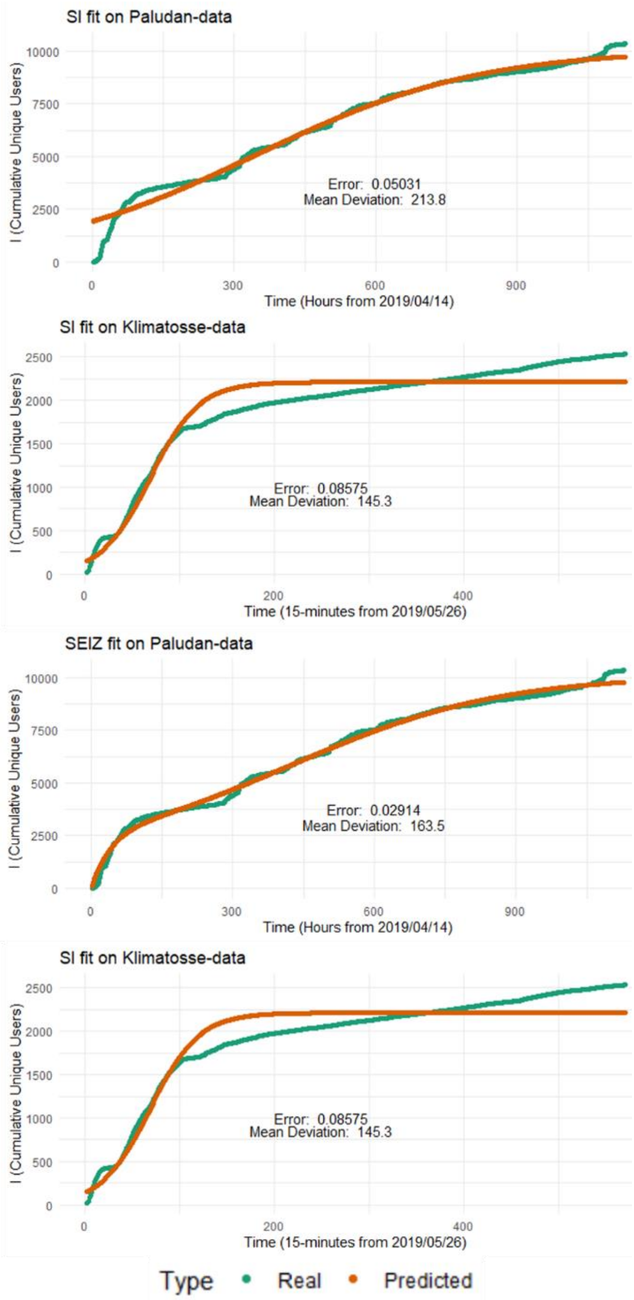


Figure 1: Fit of models to data

Figure 1 shows the parameters of the SEIZ-models, and the fit of the two models to data. The results clearly indicate the SEIZ-models outperform the SI-models, as they consistently have a lower error.

For the Paludan-data, the SEIZ-model has an error of 0.02914 and a mean deviation of 163.5 users per hour, which is better than the SI-model with an error of 0.05631 and a mean deviation of 213.8 users per hour.

For the klimatosse-data the SEIZ-model has an error of 0.0124 and a mean-deviation of just 15.2 users per 15-minute interval which is better than the SI-model, which has an error of 0.08575 and a mean deviation of 145.3 users per 15-minute interval.

3. Discussion

It is clear from the results that the epidemiological models seem to provide a good fit of the data. Furthermore, the SEIZ-models seem to clearly outperform the simpler SI-model indicating that the social adjustments to the SEIZ-model paid off.

The discussion will focus on the SEIZ-model, because it has richer, more meaningful parameters and provides a better explanation of the data. Figure 2 shows a graphical representation of the modelled diffusion.

Figure 2: Compartment dynamics

3.1 Klimatosse

As previously mentioned, the SEIZ-model fits the klimatosse-data extremely well with an almost nonexistent error. This means that I can provide an accurate description of the spread – given the assumptions and limitations. There are many interesting insights to be gathered from the model.

First, the initial burst during the first 24 hours seem to be driven mainly by people transferring from the Exposed compartment, as seen by the rapid decline in the green line of the left-hand plot in figure 2. There are two reasons for this transfer. The first reason is the relatively high number of people in the compartment initially (n=1562). In the real world, this can be interpreted as the number of people seeing the original news story from other sources such as the news, Facebook or from friends. The model, therefore, correctly picked up that the *klimatosse*-epidemic originated outside of Twitter. The second reason is the direct conversion from Susceptible to Infected. This indicates that we are dealing with a relatively simple contagion; as soon as a Susceptible sees a tweet reinforcing the news story, she tweets about it herself, without first spending time in the Exposed compartment. From a sociological perspective this could indicate that the use of *klimatosse* fits the hegemonic discourse (Gramsci & Hoare, 1971).

In the tweets, *klimatosse* is used either to mock Pia Kjærsgaard or to ironically proclaim how fanatic they, the pro-climate users, are about the climate. A minority uses it to show disdain for climate-change activists. Overall, the discourse is positive toward the green movement. This seems to follow surveys of the Danish population attitude toward climate change such as the one by Ipsos (2019), where 60% of respondents overall, and 72% of respondents between the ages of 18-24, believed that Denmark should lead the fight against climate change. It is therefore safe to say that the social inertia against participating in the *klimatosse*-epidemic is minimal.

Another closely related finding is the insignificant presence of Skeptics. The flat magenta line in the leftmost plot of figure 2 indicates there is almost no increase in Skeptics during the epidemic. This ties well into the simple dynamics of the *klimatosse*-epidemic; Because of the almost non-existing cost and potential in-group benefit of tweeting, most people felt inclined to tweet about the subject after exposure.

After the initial Exposed-driven uptick, the rest of the infected seem to come directly from the Susceptible compartment. The left plot of figure 2 shows, that the decrease in Susceptible (purple) corresponds to the increase in Infected (orange) and no other compartments are changing.

Coupled with the high amount of remaining Susceptibles at the end of the time-series, the model predicts this will continue for a while. However, when compared with the google trends data this data extraction period is based on, this does not seem to be the case. Rather there seem to be a drop-off already after one week – well within the data range. There are two reasons for this apparent disagreement between the data sources.

The first, and perhaps most obvious one, is the fact we are dealing with two separate mediums. One might not suspect users' tweeting patterns to be identical with their searching patterns. It is entirely possible that the *klimatosse*-phenomenon continued to live a life of its own as a twitter-native phenomenon. This would be supported in the model by the high degree of direct transmission between the Susceptible and Infected compartments. Had the Exposed compartment played a longer-lasting role as a middle step between Susceptible and Infected, the data could have been interpreted as the *klimatosse*-phenomenon being more endogenously borne.

The second reason highlights the descriptive limitation of epidemiological models: In being descriptive, the SEIZ-model implicitly assumes that the dataset is complete. If the dataset had included an additional week of tweets, plausibly with a continued levelling off, the parameters would have look rather different. This can also be related to the observation that the starting amount of Susceptibles – which I fitted as a constant parameter – might have been a lot lower without any influence on the fit of the Infected compartment. It is therefore not sound to extrapolate the patterns of the model.

3.2 Paludan

Despite the similarities between the two datasets, both exhibit good fits, with the SEIZ-model clearly outperforming the SI-model, there are some additional insights to be gained from the Paludan data. First and foremost, Skeptics play a far larger role. This helps dampening the diffusion rate to the Infected compartment after the initial outburst by absorbing more of the Susceptibles, as the average sentiment of the Paludan data (-0.42, n=44965) is significantly lower than the Klimatosse data (0.29, n=4557). An interpretation of this could be that *Paludan* is a more divisive concept. The sentiment scores were calculated using the Afinn-package (Nielsen, 2011). Using the yuen test from the WRS2 package (Mair & Wilcox, 2019) I found a significant trimmed mean difference of 0.44 (T=12.23, p = 0).

The divisiveness of Paludan might translate into social inertia. Whereas there is virtually no negative social cost of tweeting about klimatosser, Paludan might be too contentious of a topic. It is important to emphasize that being a Skeptic in the SEIZ-model does not imply anything about your opinions towards the subject; it simply means that you are aware but have not tweeted. This could help capture and quantify the potentially large amount of social media "lurkers" which has important implication (Gong et al., 2015).

3.3 Validity of the model

Although the results seem to fit the hypotheses, one might argue that the SEIZ-model is simply performing better because it has more parameters and therefore is susceptible to overfitting. Though a valid concern, this is not too worrying. Epidemiological models are not predictive tools, they are descriptive: We are more interested in understanding *how* the phenomenon happened than predicting how it will continue.

Nevertheless, it is still valid to question whether the model is describing signal or noise. There are a few arguments for signal. In the *klimatosse*-data where the model fits almost perfectly, the data seems smooth as well. The reason for the goodness of fit therefore seems to be the ability of SEIZ-models to capture both the initial outburst and the gradual decline in rate of change, rather than overfitting to noise. The SI-model, in contrast, seems to overshoot and level off too early.

In the *Paludan*-data both models also seem to provide a relatively smooth fit, although the data is quite noisy. The noise probably stems from the fact that the Paludan-phenomenon is a series of smaller epidemics. The reason for this is that Paludan was very reliant on channels outside of Twitter such as the general news-cycle, YouTube and word-of-mouth. Coupled with the sheer timespan of the data this allows for multiple sub-epidemics. The SEIZ-model can capture some of the exogeneous factors using the dynamics of the Z- and E-compartments, which might be the reasons for its relative success. However, it cannot model the sporadic outburst evoked by the complex interactions with the news cycle. One could imagine utilizing topic-modelling to categorize the data into smaller sub-epidemics which might provide an even more realistic fit. That is, however, outside the scope of this paper.

3.4 Further studies

There is still a lot to explore on the spread of concepts on social media. One highly current concept to explore is "samfundssind", introduced by Danish Prime Minister, Mette Frederiksen, in her televised address to the nation on March 11, 2020 following the coronavirus outbreak. Although known since 1992 (according to ordnet.dk) the word exemplifying the concept of an attitude that demonstrates that society takes precedence over narrow self-interest, has since been adopted by various commentators from both classical media and social media. In addition to be being a highly current issue with broad socio-cultural implications, it would be interesting to see how a deliberately introduced concept spreads in comparison to the more organic concepts studied in this paper.

More generally, one could study how concepts spread on other social media platforms such as Facebook or Instagram to verify the generalizability of the findings from Twitter. Although this possibility is limited by restrictions in the APIs of these platforms. It would also be interesting to expand the model by studying entire discourses and narratives surrounding e.g. immigration, instead of focusing on just a concept.

Another interesting direction could be to simulate the spread of concepts using agent-based models such as EpiSimdemics described in Barrett, Bisset, Eubank, Xizhou Feng, & Marathe (2008). In that way you could investigate not only how the overall diffusion patterns develop but also which type of individuals play a key role in spreading the new concepts.

4. Conclusions

In this paper, I have investigated how epidemiological models can be used to understand the spread of concepts on Twitter. Using methods and assumptions borrowed from epidemiological sociology, I created a model accurately describing the spread of the concepts *klimatosse* and *Paludan.* The results showed that the epidemiological models accurately described the spread as seen in the data. Furthermore, I have argued, that the SEIZ-model clearly outperforms the SI-model, thus confirming both my hypotheses. I also found that the SEIZ-model specifically designed for the spread of social phenomena gives a more accurate and nuanced description of the phenomena.

References

Barrett, C. L., Bisset, K. R., Eubank, S. G., Xizhou Feng, & Marathe, M. V. (2008). EpiSimdemics: An efficient algorithm for simulating the spread of infectious disease over large realistic social networks. *SC '08: Proceedings of the 2008 ACM/IEEE Conference on Supercomputing*, 1–12. https://doi.org/10.1109/SC.2008.5214892

Bettencourt, L. M. A., Cintrón-Arias, A., Kaiser, D. I., & Castillo-Chávez, C. (2006). The power of a good idea: Quantitative modeling of the spread of ideas from epidemiological models. *Physica A: Statistical Mechanics and Its Applications*, *364*, 513–536. https://doi.org/10.1016/j.physa.2005.08.083

Casasanto, D., & Boroditsky, L. (2008). Time in the mind: Using space to think about time. *Cognition*, *106*(2), 579–593. https://doi.org/10.1016/j.cognition.2007.03.004

Casasanto, D., Boroditsky, L., Phillips, W., Greene, J., Goswami, S., Bocanegra-Thiel, S., Santiago-Diaz, I., Fotokopoulu, O., Pita, R., & Gil, D. (2004). How deep are effects of language on thought? Time estimation in speakers of English, Indonesian, Greek, and Spanish. *Proceedings of the Annual Meeting of the Cognitive Science Society*, *26*.

Centola, D. (2010). The spread of behavior in an online social network experiment. *Science*, *329*(5996), 1194–1197.

Centola, D. (2011). An experimental study of homophily in the adoption of health behavior. *Science*, *334*(6060), 1269–1272.

Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American Journal of Sociology*, *113*(3), 702–734.

Christoffersen, P. (2019, May 27). *Pia Kjærsgaard: Høj stemmeprocent skyldes klimatosser - TV 2*. nyheder.tv2.dk. https://nyheder.tv2.dk/2019-05-26-pia-kjaersgaard-hoej-stemmeprocent-skyldes-klimatosser

Chung, C.-F., Agapie, E., Schroeder, J., Mishra, S., Fogarty, J., & Munson, S. A. (2017). When personal tracking becomes social: Examining the use of Instagram for healthy eating. *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, 1674–1687.

Clark, H. H. (1996). *Using language*. Cambridge university press.

Dayhoff, S. A. (1983). Sexist language and person perception: Evaluation of candidates from newspaper articles. *Sex Roles*, *9*(4), 527–539. https://doi.org/10.1007/BF00289792

Derczynski, L., Albert-Lindqvist, T. O., Bendsen, M. V., Inie, N., Pedersen, J. E., & Pedersen, V. D. (2019). Misinformation on Twitter During the Danish National Election: A Case Study. *Proceedings of the Conference for Truth and Trust Online*.

Fusaroli, R., & Tylén, K. (2016). Investigating Conversational Dynamics: Interactive Alignment, Interpersonal Synergy, and Collective Task Performance. *Cognitive Science*, *40*(1), 145–171. https://doi.org/10.1111/cogs.12251

Garrod, S., & Pickering, M. J. (2009). Joint Action, Interactive Alignment, and Dialog. *Topics in Cognitive Science*, *1*(2), 292–304. https://doi.org/10.1111/j.1756-8765.2009.01020.x

Goldfarb, A., & Tucker, C. (2011). Online Display Advertising: Targeting and Obtrusiveness. *Marketing Science*, *30*(3), 389–404. https://doi.org/10.1287/mksc.1100.0583

Gong, W., Lim, E.-P., & Zhu, F. (2015, April 21). Characterizing Silent Users in Social Media Communities. *Ninth International AAAI Conference on Web and Social Media*. Ninth International AAAI Conference on Web and Social Media. https://www.aaai.org/ocs/index.php/ICWSM/ICWSM15/paper/view/10462

Gorodnichenko, Y., Pham, T., & Talavera, O. (2018). *Social Media, Sentiment and Public Opinions: Evidence from #Brexit and #USElection* (Working Paper No. 24631). National Bureau of Economic Research. https://doi.org/10.3386/w24631

Gramsci, A., & Hoare, Q. (1971). *Selections from the prison notebooks*. Lawrence and Wishart London.

Granovetter, M. S. (1977). The strength of weak ties. In *Social networks* (pp. 347–367). Elsevier.

Howard, P. N., Duffy, A., Freelon, D., Hussain, M. M., Mari, W., & Maziad, M. (2011). *Opening Closed Regimes: What Was the Role of Social Media During the Arab Spring?* (SSRN Scholarly Paper ID 2595096). Social Science Research Network. https://papers.ssrn.com/abstract=2595096

Howard, P. N., & Kollanyi, B. (2016). *Bots, #Strongerin, and #Brexit: Computational Propaganda During the UK-EU Referendum* (SSRN Scholarly Paper ID 2798311). Social Science Research Network. https://papers.ssrn.com/abstract=2798311

Ipsos. (2019, May). *Ny undersøgelse: Danskerne og klimaforandringerne*. Ipsos. https://www.ipsos.com/da-dk/ny-undersogelse-danskerne-og-klimaforandringerne

Jensen, K. B. (2009). Three-step flow. *Journalism*, *10*(3), 335–337. https://doi.org/10.1177/1464884909102594

Jin, F., Dougherty, E., Saraf, P., Cao, Y., & Ramakrishnan, N. (2013). Epidemiological Modeling of News and Rumors on Twitter. *Proceedings of the 7th Workshop on Social Network Mining and Analysis*, 8:1–8:9. https://doi.org/10.1145/2501025.2501027

Katz, E. (1957). The Two-Step Flow of Communication: An Up-To-Date Report on an Hypothesis. *Public Opinion Quarterly*, *21*(1), 61–78. https://doi.org/10.1086/266687

Larsen, M. C., & Kofoed, J. (2016). A snap of intimacy: Investigating photo sharing practices on Snapchat and Instagram. *AoIR Selected Papers of Internet Research*, *6*.

Mair, P., & Wilcox, R. (2019). Robust statistical methods in R using the WRS2 package. *Behavior Research Methods*. https://doi.org/10.3758/s13428-019-01246-w

Majid, A., Bowerman, M., Kita, S., Haun, D. B. M., & Levinson, S. C. (2004). Can language restructure cognition? The case for space. *Trends in Cognitive Sciences*, *8*(3), 108–114. https://doi.org/10.1016/j.tics.2004.01.003

Michel, M., & Smith, B. (2017). Measuring lexical alignment during L2 chat interaction:An eye-tracking study. In S. M. Gass, P. Spinner, & J. Behney (Eds.), *Salience in Second Language Acquisition* (pp. 244–267). Taylor and Francis. https://doi.org/10.4324/9781315399027

Olof Larsson, A., & Moe, H. (2013). Representation or participation? Twitter use during the 2011 Danish election campaign. *Javnost-The Public*, *20*(1), 71–88.

Parker, I. (1990). Discourse: Definitions and contradictions. *Philosophical Psychology*, *3*(2–3), 187–204. https://doi.org/10.1080/09515089008572998

Pinker, S. (1994). *The Language Instinct. How the Mind Creates Language*.

R Core Team. (2019). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Romero, D. M., Meeder, B., & Kleinberg, J. (2011). *Differences in the mechanics of information diffusion across topics: Idioms, political hashtags, and complex contagion on Twitter*.

Sapir, E. (1929). The Status of Linguistics as a Science. *Language*, *5*(4), 207. https://doi.org/10.2307/409588

Soetaert, K., & Petzoldt, T. (2010). Inverse Modelling, Sensitivity and Monte Carlo Analysis in *R* Using Package **FME**. *Journal of Statistical Software*, *33*(3). https://doi.org/10.18637/jss.v033.i03

The Chromium Authors, McCandless, M., Sanford, M., & FIRAT, A. (2013). *cldr: Language Identifier based on CLD library*. https://CRAN.R-project.org/package=cldr

Tomasello, M. (1995). Joint attention as social cognition. *Joint Attention: Its Origins and Role in Development*, *103130*.

Tomasello, M. (2001). First steps toward a usage-based theory of language acquisition. *Cognitive Linguistics*, *11*(1–2). https://doi.org/10.1515/cogl.2001.012

Trecca, F., Bleses, D., Madsen, T. O., & Christiansen, M. H. (2018). Does sound structure affect word learning? An eye-tracking study of Danish learning toddlers. *Journal of Experimental Child Psychology*, *167*, 180–203. https://doi.org/10.1016/j.jecp.2017.10.011

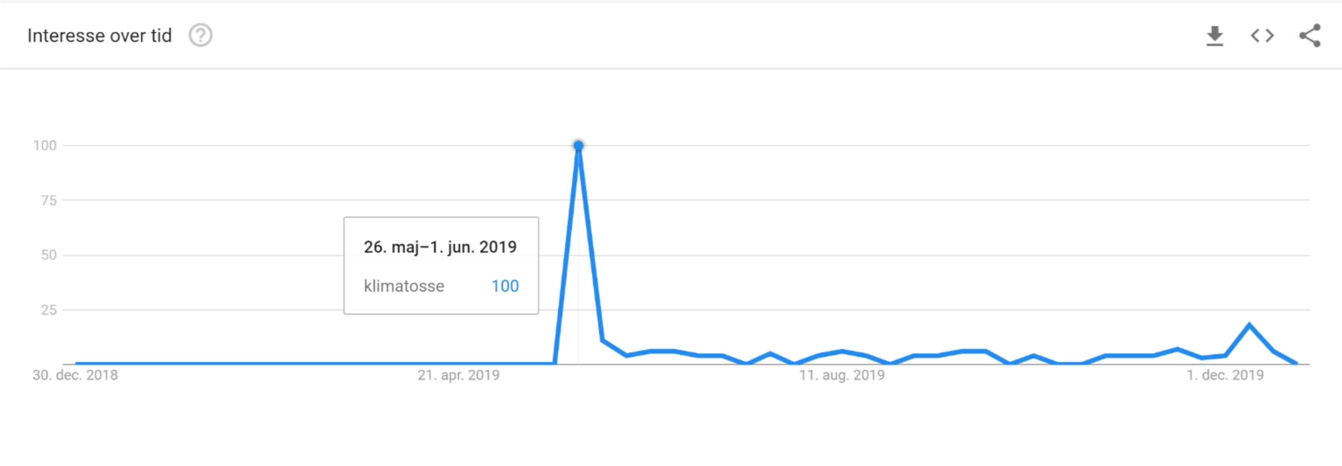
Vestergaard, N. (2019, December 13). *Klimatosse er årets ord 2019*. DR. https://www.dr.dk/nyheder/indland/klimatosse-er-aarets-ord-2019

Whorf, B. L. (1940). *Science and linguistics*. Bobbs-Merrill Indianapolis, IN.

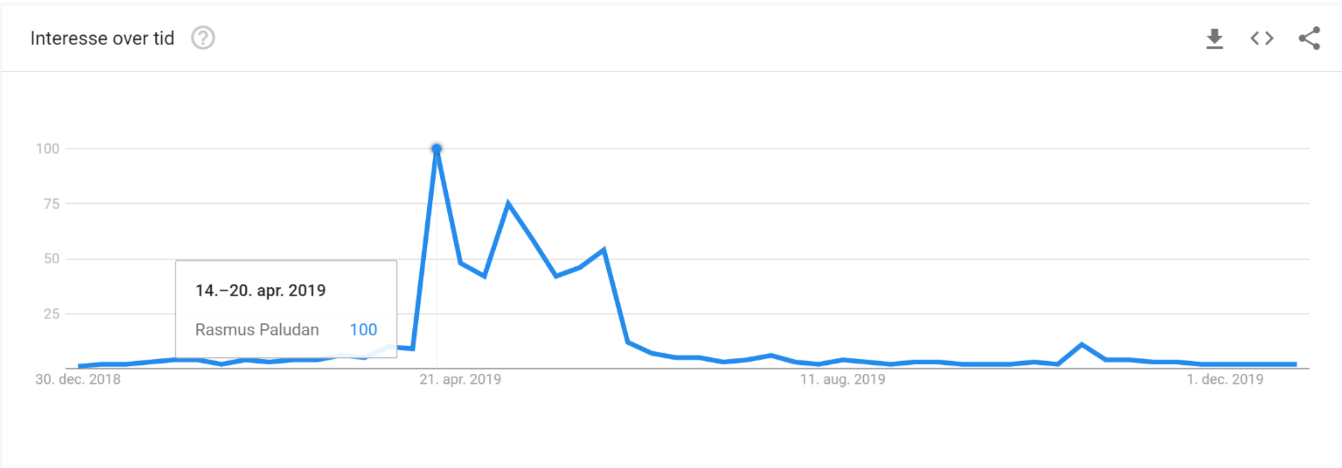
Winawer, J., Witthoft, N., Frank, M. C., Wu, L., Wade, A. R., & Boroditsky, L. (2007). Russian blues reveal effects of language on color discrimination. *Proceedings of the National Academy of Sciences*, *104*(19), 7780–7785. https://doi.org/10.1073/pnas.0701644104

Zeller, F., & Hermida, A. (2015). When tradition meets immediacy and interaction. The integration of social media in journalists' everyday practices. *Sur Le Journalisme, About Journalism, Sobre Jornalismo*, *4*(1), 106–119.

Appendix



Interest in klimatosse, Source: Google Trends (https://trends.google.com/trends/explore?geo=DK&q=Klimatosse)



Interest in Paludan, Source: Google Trends (https://trends.google.com/trends/explore?geo=DK&q=%2Fg%2F11g697wr2f)