Memetic or Epidemic: Modelling Internet-Based Neologisms in China 2008-2016

Menghan Jiang, Xiangying Shen, Chu-ren Huang, Kathleen Ahrens

Contagious words/Epidemic Behaviours: Language Big Data Approaches to COVID-19

17 Oct 2020

The Hong Kong Polytechnic University
香港理工大學

Opening Minds • Shaping the Future
啟迪思維 • 成就未來

Introduction

> Neologisms, or newly coined words,
  - in traditional communication formats
  - in internet-based format.

> Nowadays internet-based neologisms have become part of our daily life as people switched to communicate with others, seek information, or provide opinions online.

Introduction

> However, linguistic research on these “internet neologisms” has not been extensively conducted and a comprehensive theoretical model that can clearly describe a neologism's life cycle has yet to be proposed.

Introduction

>Previous empirical studies on neologisms were driven by **lexicographic concerns** and focused on research issues:

- **Life cycle of neologisms;**
- **Prediction of which new words will be successfully incorporated into the lexicon of a language**

(e.g., Schmid, 2008; Renouf, 2013; Kerremans, 2015; Baayen, 1996).

>e.g., Word frequency (Altmann et al., 2011, 2013)

Introduction

- Internet neologisms evolve much faster and have the potential to spread widely than ever before, but at the same time may fade and disappear quickly.

e.g., 洪荒之力 *hong2huang1zhi1li4* 'with all one's might'

Google Trends

> Google Trends： an online tool which provides access to a largely sample of actual search requests made to Google.

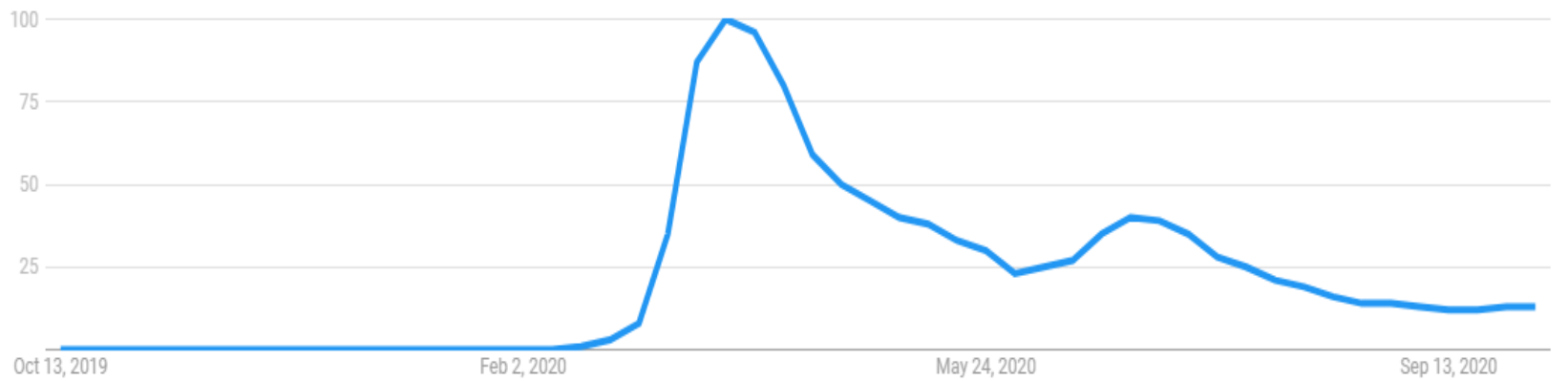
- It provides the popularity of a selected word within a flexible range of timescales.
- The granularity of the reported data is accurate to each day.
- The reported popularity is normalized on a scale of zero to one hundred.

COVID 19
Search term

+ Compare

United States | Past 12 months | All categories | Web Search

Interest over time

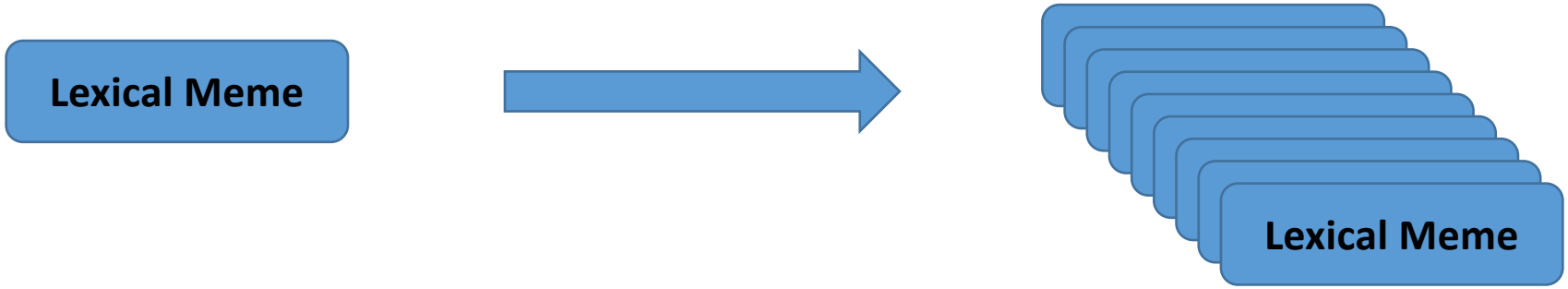


Memetic Model

- This approach models the growth and spread of new words, which are sometimes called **lexical memes**, similar to cultural memes.

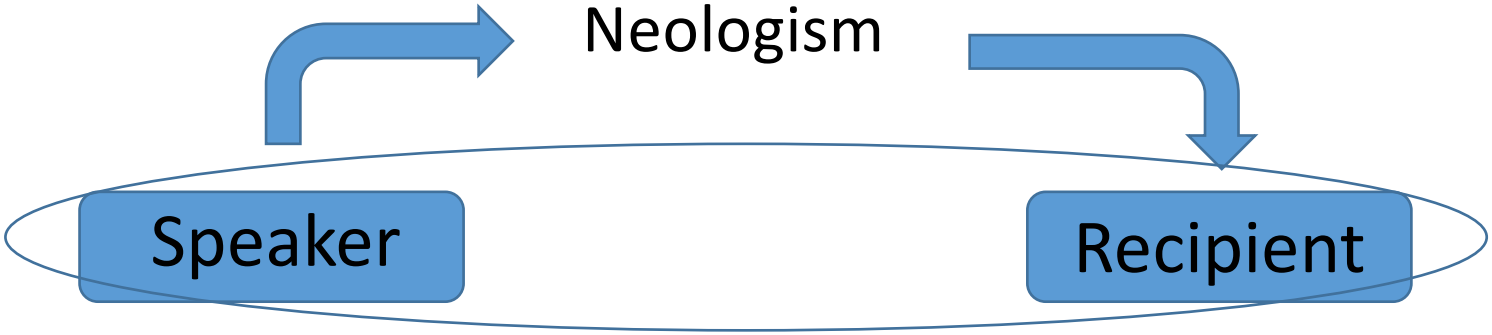
- The dominant memetic models are genetic replications (Hsieh, 2019, Dawkins 1976; Hull 1982, 2000; Dennett 1991; Blackmore 2001, 2007).

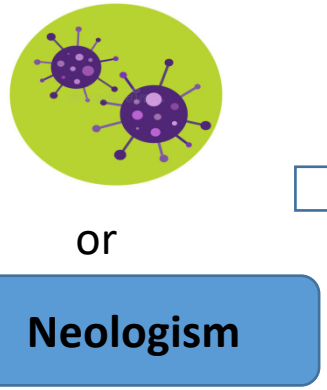
Memetic Model

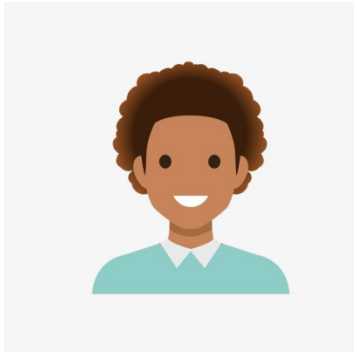


- Model the development of memes in terms of their ability to propagate just like the ability of genes to self-replicate in biology

Epidemic Model







Research Questions

> 1.What is the typical life cycle and most important features of a successful neologism?

> 2.Does the memetic model or epidemic model better predict the development of neologisms? In particular, is the propagation of neologisms better predicted by the object of propagation (i.e., the new words) in the memetic model or by the host (i.e., speaker) in the epidemic model?

Data Collection

> The words selected are from the journal 咬文嚼字 *Yao3 Wen1 Jiao2 Zi4*

> The criteria of selecting the neologisms for 咬文嚼字 *Yao3Wen1 Jiao2 Zi4*

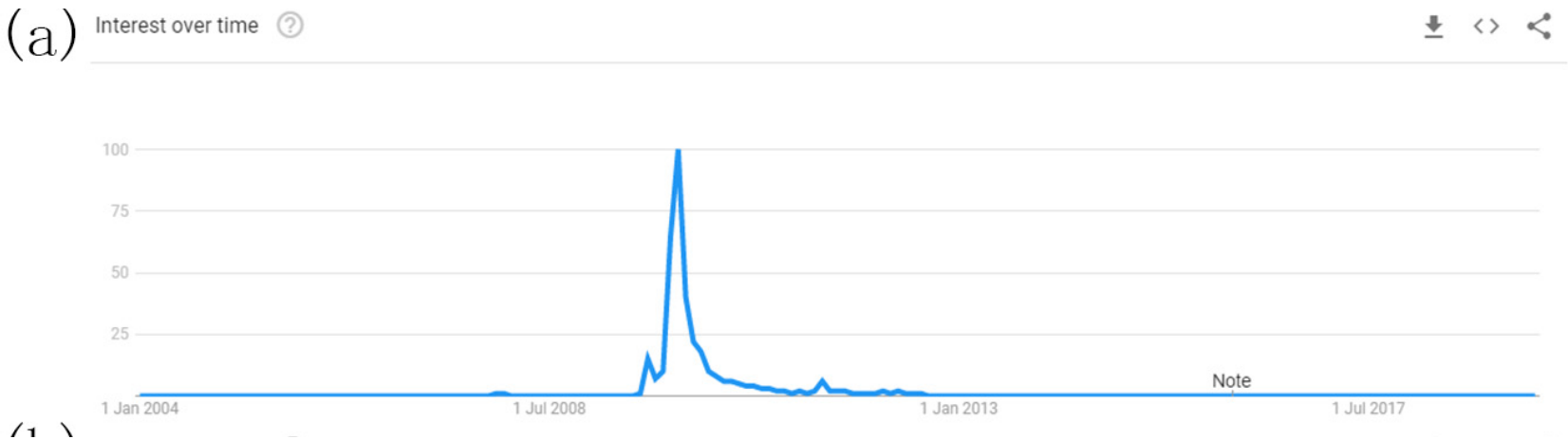
- 1) Fashionable.
- 2) Popular.
- 3) Expressive.

> **We take 10 new words of the year each for the 9 year duration of 2008-2016, for a total of 90 new words.**

RQ1. Pattern of the Internet Neologism

The data from Google Trends shows that the popularity *P (t)* over time *t* of 62 of the 90 Internet neologisms have similar sharp rise and decay patterns.

RQ1. Pattern of the Internet Neologism



The example of 蜗居 ''living in a tight space'

Model Fitting

> **Epidemic Model**

S (Susceptible) **I** (Infectious) **R** (Recovered)

$$\begin{cases} \dfrac{dS(t)}{dt} = -bI(t)S(t) \\ \dfrac{dI(t)}{dt} = bI(t)S(t) - aI(t) \\ \dfrac{dR(t)}{dt} = aI(t) \end{cases}$$

Model Fitting

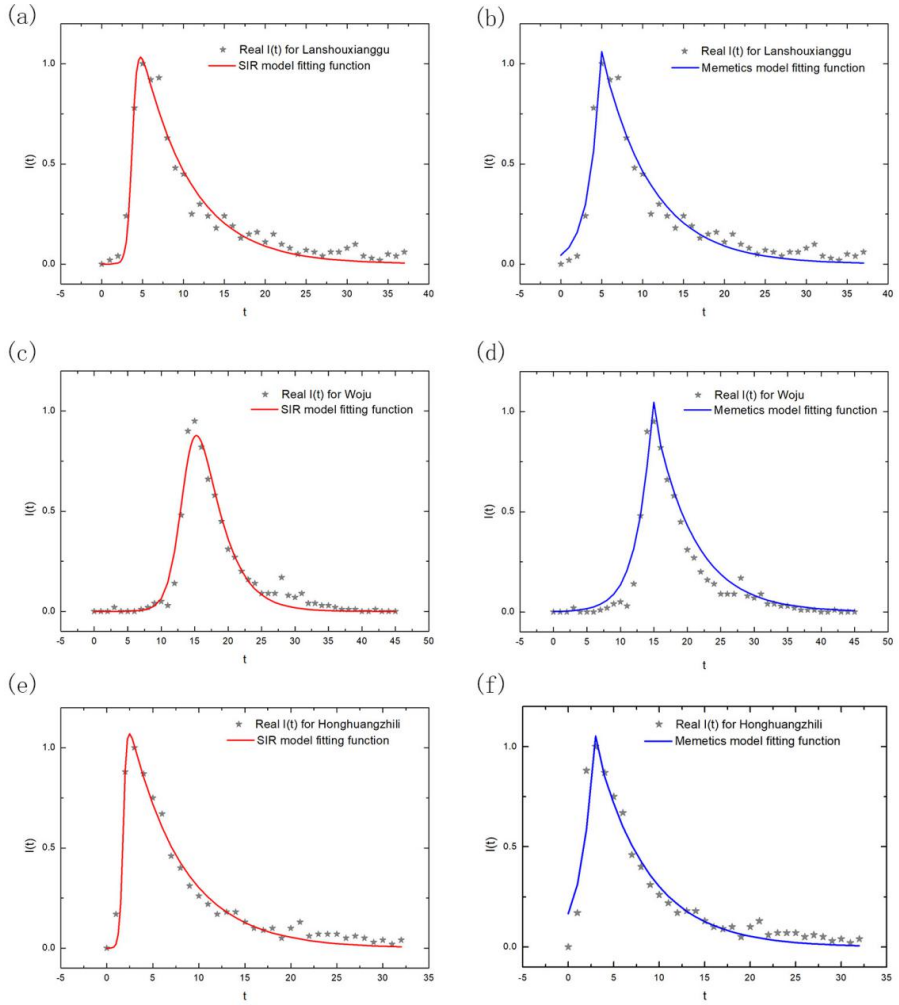
>Memetic Model

$$F(m) \equiv \frac{N(m,t+1)}{N(m,t)} = A(m)R(m)E(m)T(m)$$

This is the product of the survival rates for each of the four stages:

Assimilation A;
Retention R;
Expression E;
Transmission T

Model Fitting



$$R^2 \equiv 1 - \frac{\sum (f_i - x_i)^2}{\sum (\bar{x} - x_i)^2}$$

Epidemic:R2=0.8975
Memetic: R2=0.8960

Results

> **An epidemic model is better in describing the life cycle of neologisms**

Memetic Model vs. Epidemic Model

> 1) monotonic vs. non-monotonic

> 2) *a posterior* vs. *a priori*

Summary

- We adopted internet-based data from Google Trends for the temporally marked popularity of neologisms in Chinese in order to model their propagation and life cycle.
- We found the SIR model, a host-driven dynamic computational model for epidemiology, provides a better descriptive and predictive ability over a spreader-driven self-adaptive memetic model.
- Our study provides a new method to obtain statistical data for neologisms from an Internet search engine, and proposes a classical mathematical model used for epidemics to describe the life cycle of neologisms.

Questions?