



Principles of Complex Systems, Vols. 1 & 2, CSYS/MATH 300 and 303
University of Vermont, Fall 2021
Assignment 17

code name: Raiders of the Lost Art

Due: Wednesday, February 9, by 11:59 pm, 2021.

Relevant clips, episodes, and slides are listed on the assignment's page:

<https://pdodds.w3.uvm.edu//teaching/courses/2021-2022principles-of-complex-systems//assignments/17/>

Some useful reminders:

Deliverator: Prof. Peter Sheridan Dodds (contact through Teams)

Assistant Deliverator: Michael Arnold (contact through Teams)

Office: The Ether

Office hours: Tuesdays, 3:00 to 4:00 pm on Teams

Course website:

<https://pdodds.w3.uvm.edu//teaching/courses/2021-2022principles-of-complex-systems>

All parts are worth 3 points unless marked otherwise. Please show all your workings clearly and list the names of others with whom you collaborated.

For coding, we recommend you improve your skills with Python, R, and/or Julia. The Deliverator uses Matlab.

Graduate students are requested to use \LaTeX (or related \TeX variant). If you are new to \LaTeX , please endeavor to submit at least n questions per assignment in \LaTeX , where n is the assignment number.

Assignment submission:

1. Please send to both the Deliverator and Assistant Deliverator via direct message on Teams.
 2. PDF only! Please name your file as follows (where the number is to be padded by a 0 if less than 10 and names are all lowercase): CSYS300assignment%02d\$firstname-\$lastname.pdf as in CSYS300assignment06michael-palin.pdf
-

1. Using the main text you chose at the start of the semester, plot happiness time series in the following ways using the labMT lexicon.

The labMT word list was published with [1] in 2011, and has been occasionally upgraded. See <https://hedonometer.org> and <https://storywrangling.org> for the current version.

- (a) Process (destroy) your text so that it is a simple text file with one 1-gram per line—a vector of 1-grams.

To the extent possible, keep punctuation in as separate 1-grams. Periods, commas, semicolons, em dashes, ellipses, ldots.

(b) First use the full lexical lens provided by labMT.

Make a single figure containing a stacked set of 7 plots with text windows of size $T = \lceil 10^z \rceil$ for $z = 1, 1.5, 2, 2.5, 3, 3.5,$ and 4.0 .

Stacked here means separated and stacked vertically, as opposed to directly overlaid. See examples for Moby Dick at the end of this assignment.

The notation $\lceil \cdot \rceil$ means round to the nearest integer.

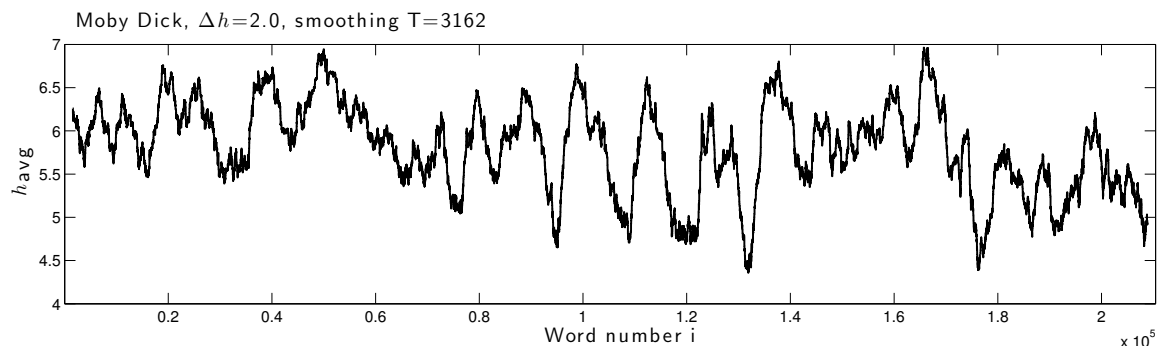
(c) Repeat the above for lenses which exclude the central words around the neutral point.

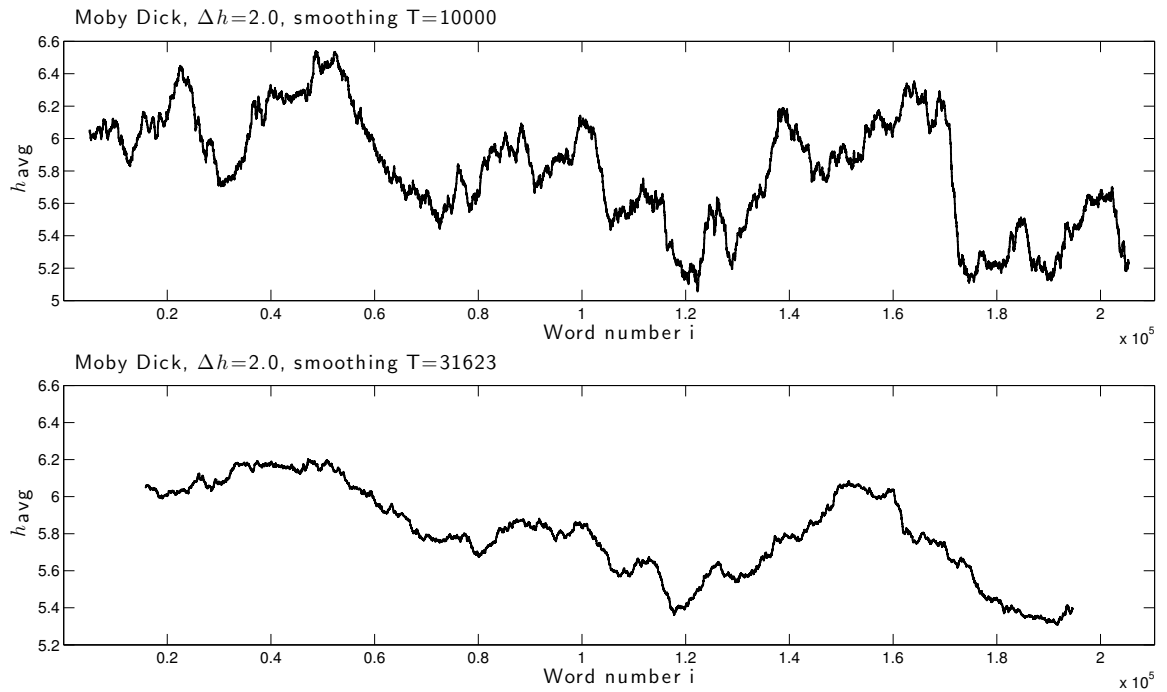
The blocked words are $h_{\text{avg}} \pm \delta h_{\text{avg}}$ where $\delta h_{\text{avg}} = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0,$ and 3.5 .

Notes:

- The horizontal axis is “narrative time” corresponding to 1-grams in the text, running from 1 to N .
- The windows should overlap, sliding one word ahead each time. This is a simple averaging filter.
- Points should be located above the center of each window.
- So the point for the window running from n to $n + T - 1$ (T words) will be located at $n + (T - 1)/2$.
- Do not pre-filter the text for any given lens. Windows will contain variable numbers of words with and without happiness scores.

Three example averaging windows for Moby Dick with $\delta h_{\text{avg}} = 2.0$:





References

- [1] P. S. Dodds, K. D. Harris, I. M. Kloumann, C. A. Bliss, and C. M. Danforth. Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter. *PLoS ONE*, 6:e26752, 2011. [pdf](#)