Principles of Complex Systems, Vols. 1, 2, & 3D
CSYS/MATH 6701, 6713, & a pretend number
University of Vermont, Fall 2024
"... in the shadow of a unicorn's dream ..."
Assignment 14

P
o What's
C The
S Story?

☑, Community, Conspiracy Theories and Interior Design, .
Episode links: Wikipedia ☑, IMDB ☑, Fandom ☑, TV Tropes ☑.

Due: Monday, February 24, by 11:59 pm

https://pdodds.w3.uvm.edu/teaching/courses/2024-2025pocsverse/assignments/14/

Some useful reminders:

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

Office: The Ether and/or Innovation, fourth floor

Office hours: See Teams calendar

Course website: https://pdodds.w3.uvm.edu/teaching/courses/2024-2025pocsverse

Overleaf: LaTeX templates and settings for all assignments are available at
https://www.overleaf.com/read/tsxfwwmwdgxj.

Some guidelines:

1. Each student should submit their own assignment.

2. All parts are worth 3 points unless marked otherwise.

3. Please show all your work/workings/workingses clearly and list the names of others with whom you ~~conspired~~ collaborated.

4. We recommend that you write up your assignments in LaTeX (using the Overleaf template). However, if you are new to LaTeX or it is all proving too much, you may submit handwritten versions. Whatever you do, please only submit single PDFs.

5. For coding, we recommend you improve your skills with Python, R, and/or Julia. **Please do not use any kind of AI thing unless directed.** The (evil) Deliverator uses (evil) Matlab.

6. There is no need to include your code but you can if you are feeling especially proud.

Assignment submission:

Via Brightspace (which is not to be confused with the death vortex of the same name).

Again: One PDF document per assignment only.

1. Continue to explore the emotional content of Pride and Prejudice, Frankenstein, and Moby Dick.

   Task: Compute emotional arcs by sliding a window of width T 1-grams through each book as described below.

   Reminder: The labMT word list was published with Ref. [1] in 2011, and has been occasionally upgraded when necessary to accommodate changes in language use.

   See https://hedonometer.org ⌇ and https://storywrangling.org ⌇ for the current version.

2. (3×3 = 9) First use the full lexical lens provided by labMT ($\delta h_{\mathrm{avg}} = 0$).

   Make a single figure containing a stacked set of 7 plots with text windows of size $T = [10^{\mu}]$ for $\mu = 1, 1.5, 2, 2.5, 3, 3.5$, 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 $[\cdot]$ means round to the nearest integer.

3. (3×3 = 9)

   Choose a 'good' text window from above, and repeat the analysis with lenses which exclude the central words around the neutral point.
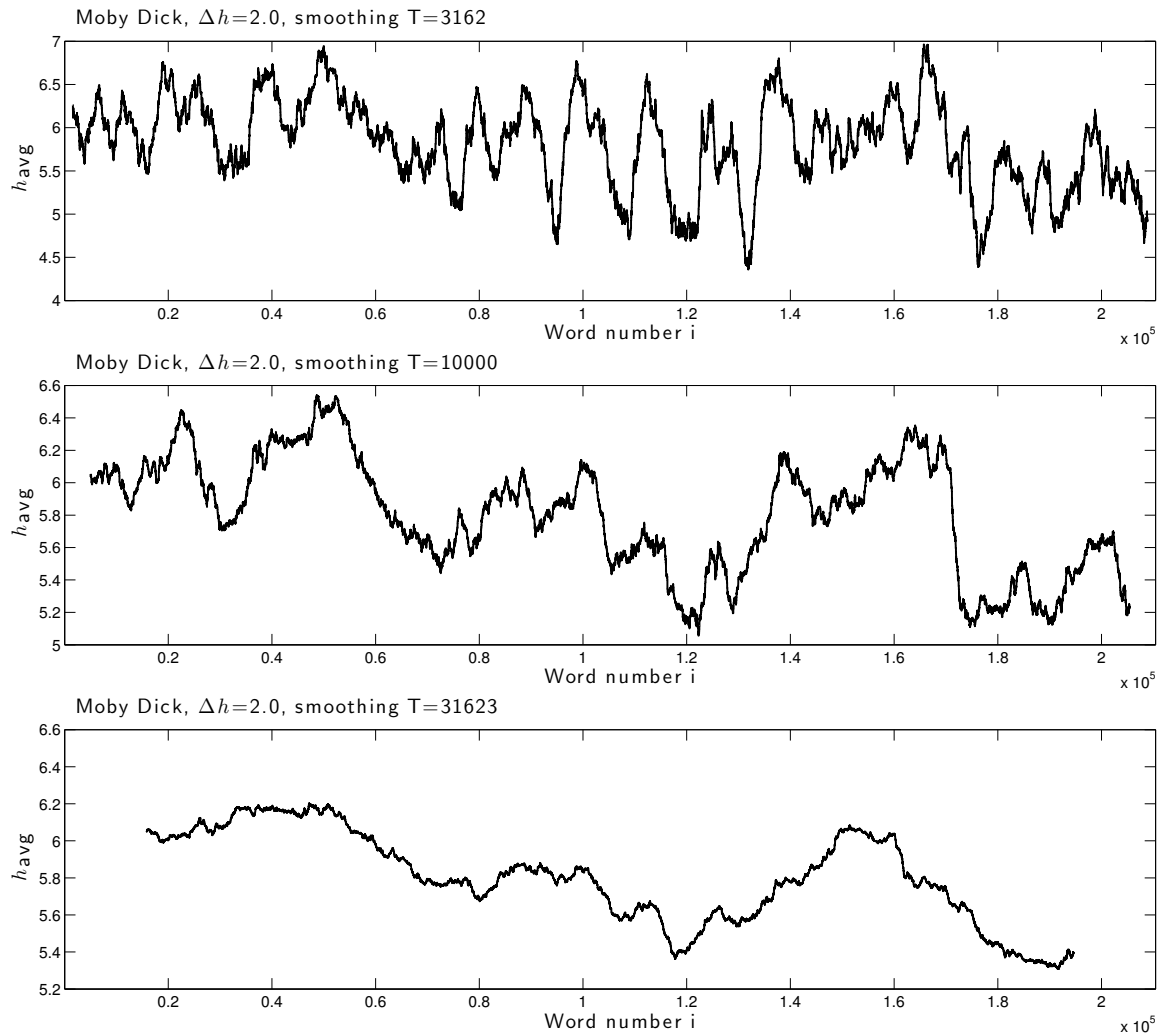
   The blocked words are $5 \pm \delta h_{\mathrm{avg}}$ where $\delta h_{\mathrm{avg}} = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0$, and 3.5.

   By 'good', we mean one that seems to you to produce a reasonable smoothing. Not too choppy, not too washed out.
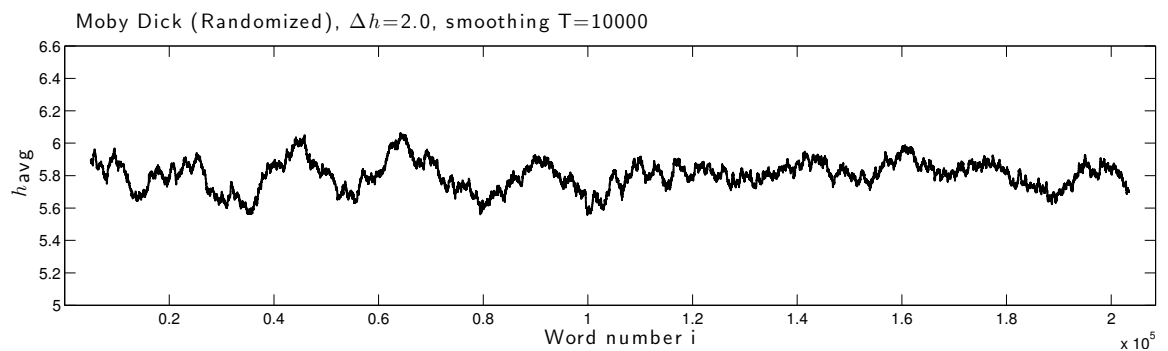
   Notes:

   - The horizontal axis is "reading-experience 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.
   - Only average 1-grams that are in the lexical lens—do not count all possible T words for the normalization.
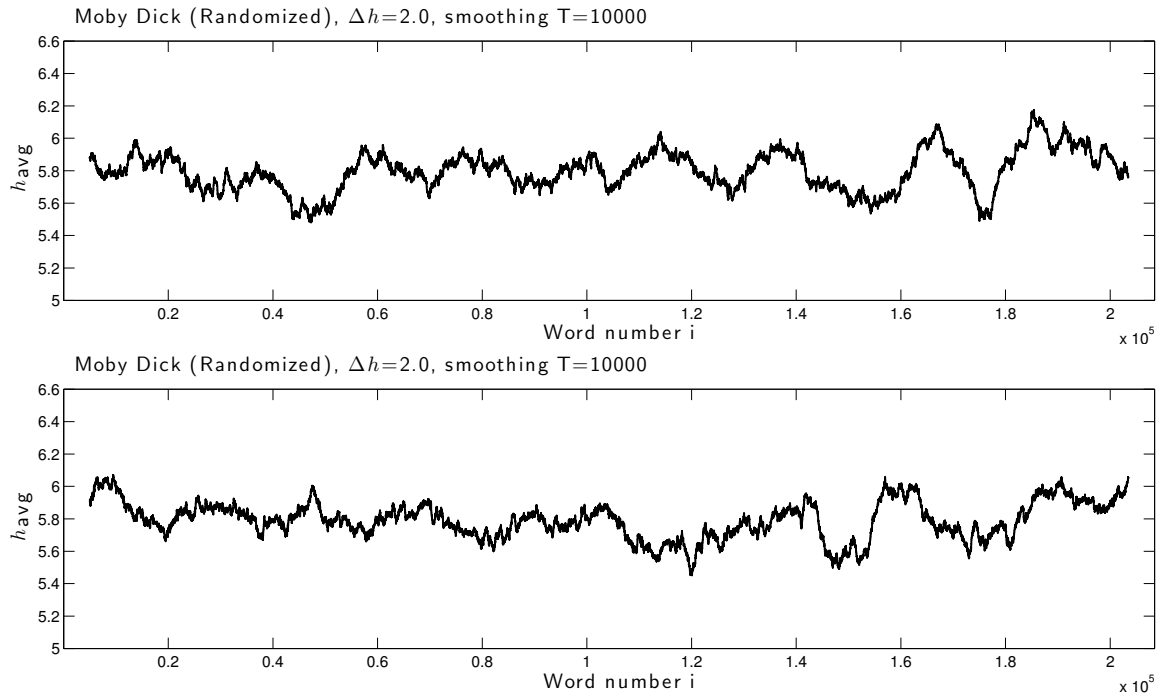
Three example averaging windows for Moby Dick with $\delta h_{\mathrm{avg}} = 2.0$ (from Ref. [2]):

Moby Dick, $\Delta h = 2.0$, smoothing T=3162

Moby Dick, $\Delta h = 2.0$, smoothing T=10000

Moby Dick, $\Delta h = 2.0$, smoothing T=31623

4. Using a smoothing of $T = 10^4$, plot time series for four randomized time series of Moby Dick.

Some examples (from Ref. [2]):

Moby Dick (Randomized), $\Delta h = 2.0$, smoothing T=10000

Moby Dick (Randomized), $\Delta h = 2.0$, smoothing T=10000

Moby Dick (Randomized), $\Delta h = 2.0$, smoothing T=10000

5. For some fun/madness/science, see what you can get out of your favorite/most despised LLM or chatbot for the emotional arcs of Moby Dick, Pride and Prejudice, and Frankenstein.

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 ☑

[2] P. S. Dodds, E. M. Clark, S. Desu, M. R. Frank, A. J. Reagan, J. R. Williams, L. Mitchell, K. D. Harris, I. M. Kloumann, J. P. Bagrow, K. Megerdoomian, M. T. McMahon, B. F. Tivnan, and C. M. Danforth. Human language reveals a universal positivity bias. Proc. Natl. Acad. Sci., 112(8):2389–2394, 2015. Available online at http://www.pnas.org/content/112/8/2389. pdf ☑