



What's
The
Story?

Principles of Complex Systems, Vols. 1 and 2
CSYS/MATH 6701, 6713
University of Vermont, Fall 2025
“Bird law in this country is not governed by reason”
Assignment 14

[It's Always Sunny in Philadelphia](#) [The Gang Exploits the Mortgage Crisis, S5E01](#)
Episode links: [IMDB](#), [Fandom](#), [TV Tropes](#).

Due: Friday, March 6, by 11:59 pm

<https://pdodds.w3.uvm.edu/teaching/courses/2025-2026pocsverse/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/2025-2026pocsverse>

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. And it's going to be a no for the catachrestic Excel. 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, just a weird coincidence). Again: One PDF document per assignment only.

Continue to explore the emotional content of *Pride and Prejudice*, *Frankenstein*, *Moby Dick*, and *Les Misérables*.

Task: Compute emotional arcs by sliding a window of width T 1-grams through each book as described below. The main paper on emotional arcs is by Reagan et al. [1], but was first introduced in [2].

Reminder: The labMT word list was published with Ref. [3] 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.

1. ($4 \times 3 = 12$) First use the full lexical lens provided by labMT ($\delta h_{\text{avg}} = 0$).

For each text, make a single figure containing a stacked set of 7 plots with text windows of size $T = \lceil 10^\mu \rceil$ 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* below.

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

2. ($4 \times 3 = 12$)

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 $5 \pm \delta h_{\text{avg}}$ where $\delta h_{\text{avg}} = 0.5, 1.0, 1.5, 2.0, 2.5, 3.0,$ and 3.5 .

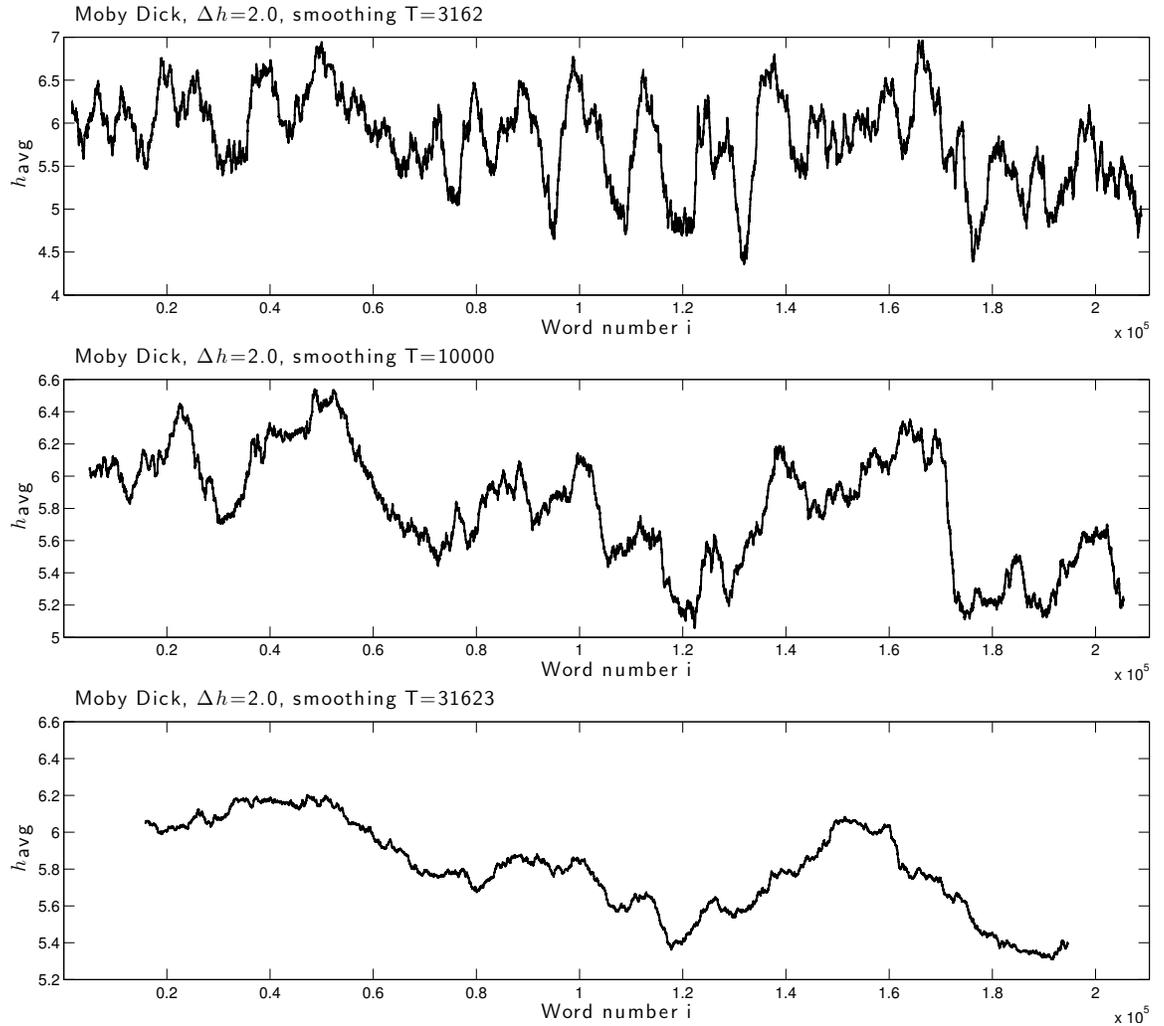
By 'good', we mean one that seems to you to produce a reasonable smoothing. Not too choppy, not too washed out.

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 1-gram 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 $n + T - 1$ will be located at $n + (T - 1)/2$.
- Do not pre-filter the text for any given lens. Windows will contain variable numbers of 1-grams with and without happiness scores.

- Only average 1-grams that are in the lexical lens—do not count all possible T 1-grams for the normalization.

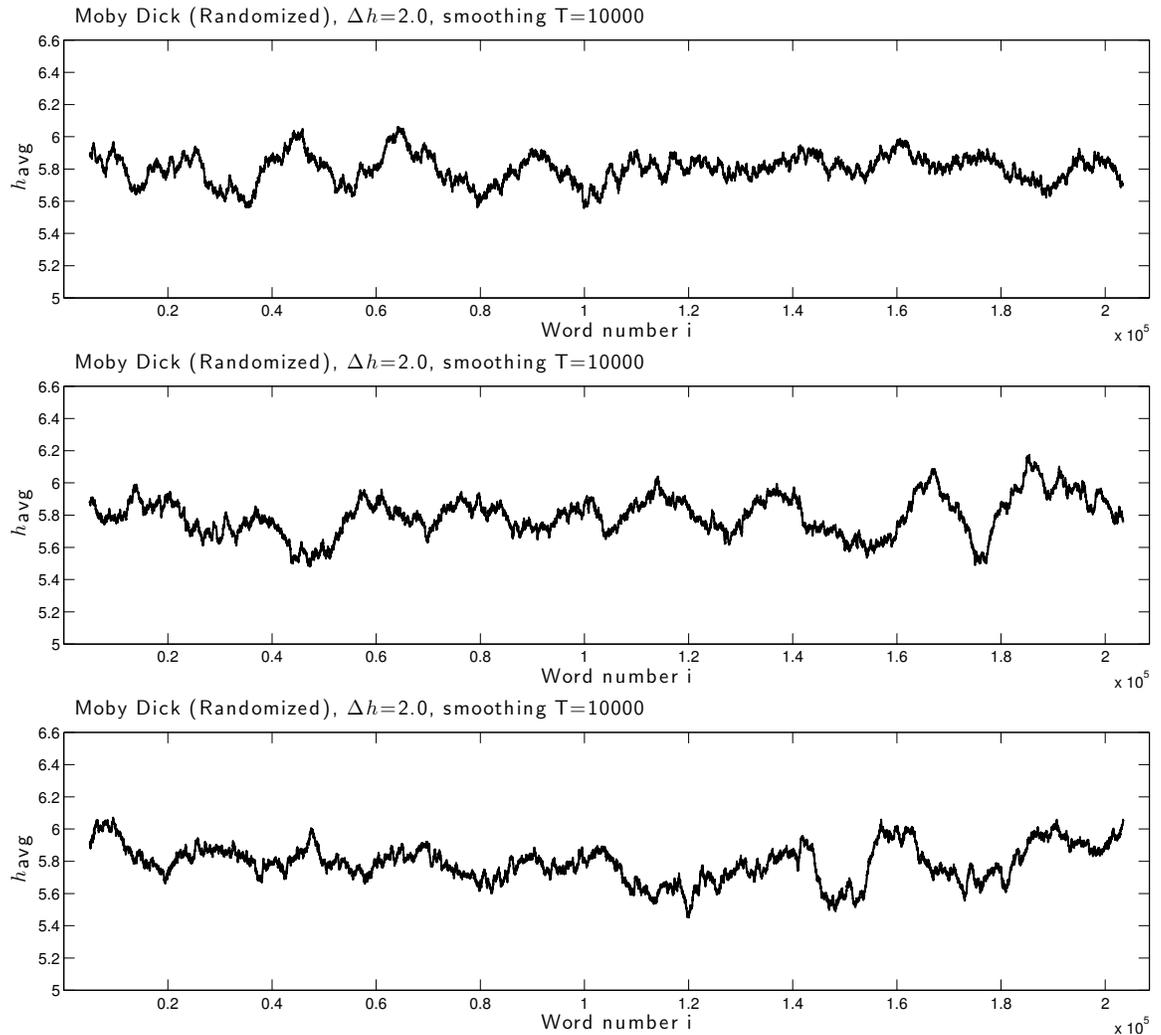
Three example averaging windows for Moby Dick with $\delta h_{\text{avg}} = 2.0$ (from Ref. [2]):



3. ($4 \times 3 = 12$)

Using a smoothing of $T = 10^4$, plot time series for four randomized time series for each of the four texts.

Some examples for Moby Dick (from Ref. [2]) with $\delta h_{\text{avg}} = 2.0$:



References

- [1] A. J. Reagan, B. F. Tivnan, J. R. Williams, C. M. Danforth, and P. S. Dodds. Sentiment analysis methods for understanding large-scale texts: A case for using continuum-scored words and word shift graphs. *EPJ Data Science*, 6, 2017. [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. Megerdooian, 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](#) 
- [3] 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](#) 