

# Computational Social Science & Humanities

A cold take by a self-taught dabbler in computational methods and techniques

---

Kristian G. Kjellmann

December 8th 2022

Department of Sociology & Social Work

What is CSSH seen from your  
perspective?

---

# What it is not

- Anything involving large amounts of data
- Anything "beyond" econometrics
- Anything involving programming

# What it is

*Techniques for condensing, making sense of or deriving meaningful measurements from high dimensional and (mainly) unstructured data*

- Exploratory

# What it is

*Techniques for condensing, making sense of or deriving meaningful measurements from high dimensional and (mainly) unstructured data*

- Exploratory
- Experimental

*Techniques for condensing, making sense of or deriving meaningful measurements from high dimensional and (mainly) unstructured data*

- Exploratory
- Experimental
- Complimentary

# What it is

*Techniques for condensing, making sense of or deriving meaningful measurements from high dimensional and (mainly) unstructured data*

- Exploratory
- Experimental
- Complimentary
- Integrated (in research domain)

# My areas of interest

Generally: *Changes and variations in societal discourses*

- The legitimization and de-legitimization of views and opinions
- Emergence and manifestation of hegemonic discourses
- In- and outgroup dynamics as expressed in speech
- Interplay of societal developments and discourses on societal developments

Specifically (currently): *Co-variation in discourses and ethnic integration outcomes at local community levels*



How could you and your group  
benefit from a CSSH Hub?

---

# Discussing methods and techniques

Me



# Discussing methods and techniques

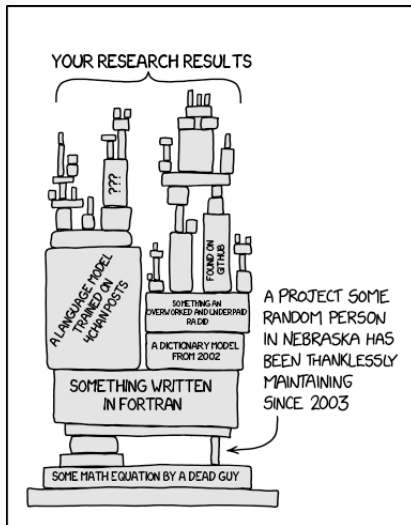
Me



My colleagues



# Unpacking the black boxes



Adapted from: <https://xkcd.com/2347/>.

# Navigating the tools

What I know



What is available



Where do you picture that CSSH  
will be at AAU in three years?

---

# From novel approaches...

Suppose that  $x$  is a vector of token counts for the particular text of interest, and that  $\hat{\theta} = \theta_0 + C\hat{\beta}$  is the affinity vector estimate gotten from  $\hat{\beta}$ , the maximizer of the corresponding log likelihood  $l(\beta)$  defined in (3). Making the dependence on  $x$  explicit, the score and observed information functions are

$$u(\beta; x) = C^T Q x, \quad I(\beta; x) = C^T Q X X^T C,$$

where  $X \in \mathbb{R}^{\mathcal{V} \times \mathcal{V}}$  is a diagonal matrix with  $X_{vv} = x_v$  for  $v \in \mathcal{V}$  and  $Q = Q(\beta)$  is as defined in Section 5.

For an arbitrary word type  $v \in \mathcal{V}$ , consider the effect of setting  $x_v = 0$ . This defines a new vector of token counts  $x_{(v)}$  defined by  $x_{(v)v} = 0$  and  $x_{(v)w} = x_w$  for all  $w \neq v$ . Let  $e_v$  denote the  $v$ th standard basis vector in  $\mathbb{R}^{\mathcal{V}}$  and define  $h_v = C^T \hat{Q} e_v$ , where  $\hat{Q} = Q(\hat{\beta})$ . Note that  $x = x_{(v)} + x_v e_v$ , so that

$$u(\hat{\beta}; x) = u(\hat{\beta}; x_{(v)}) + x_v h_v, \quad I(\hat{\beta}; x) = I(\hat{\beta}; x_{(v)}) + x_v h_v h_v^T.$$

Since  $u(\hat{\beta}; x) = 0$ , this implies that evaluating the score function with the new data at the old estimate gives

$$(9) \quad u(\hat{\beta}; x_{(v)}) = -x_v h_v.$$

The maximizer  $\hat{\beta}_{(v)}$  of the new log likelihood is roughly equal to the first Newton scoring step from  $\hat{\beta}$ . We can compute this step explicitly by first computing the inverse of the observed information matrix:

$$(10) \quad \begin{aligned} \{I(\hat{\beta}; x_{(v)})\}^{-1} &= \{I(\hat{\beta}; x) - x_v h_v h_v^T\}^{-1} \\ &= \{I(\hat{\beta}; x)\}^{-1} + (x_v^{-1} - \tilde{h}_v^T h_v)^{-1} \tilde{h}_v \tilde{h}_v^T \end{aligned}$$

where  $\tilde{h}_v = \{I(\hat{\beta}; x)\}^{-1} h_v$ .

Approximating the maximizer by the first Newton step from  $\hat{\beta}$  gives

$$\begin{aligned} \hat{\beta}_{(v)} &\approx \hat{\beta} + \{I(\hat{\beta}; x_{(v)})\}^{-1} u(\hat{\beta}; x_{(v)}) \\ &= \hat{\beta} - (x_v^{-1} - \tilde{h}_v^T h_v)^{-1} \tilde{h}_v, \end{aligned}$$

In the multivariate analyses, we estimate ISEI scores using ordinary least squares (OLS) models.

In our models, we effectively control for composition, with variables including level of education, numeracy skills, readiness to learn, host language usage, years since migration, and relevant socio-demographics.<sup>5</sup>



## From single to team authorship



## (Co-)Developed tools

