categorical

Introduction to Recurrence Quantification Analysis (RQA)

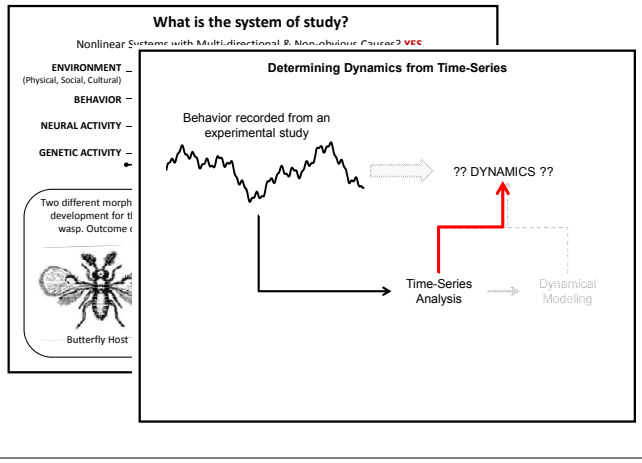
Rick Dale
Department of Communication
University of California, Los Angeles
co-mind.org/ati

Nonlinear Methods for Psychological Science
APA Advanced Training Institute
University of Cincinnati

co-mind.org/ati

additional materials related to dynamical methods for categorical data

from presentation Monday

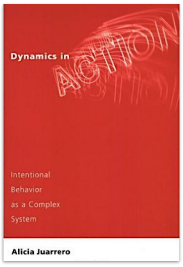


from presentation Monday

Dynamics & Behavior

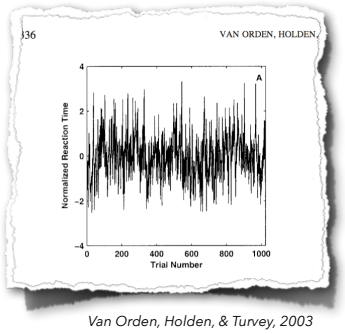
- Behavioral dynamics can be measured & quantified
- Behavioral dynamics can be a rich source of information to help understand behavior
- But, we often find ourselves dealing with snapshots of behavior
  - Single measurements, or averages of measures that were collected over time
  - Can be misleading or tell an incomplete story

Prof. Alicia Juarrero

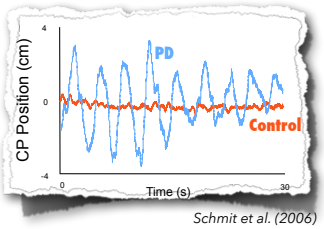


MIT Press, 1999

…we leverage ***sequential structure of behavior***



Van Orden, Holden, & Turvey, 2003



Schmit et al. (2006)

Linguistic Levels...

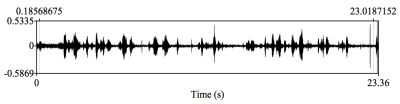
conversational topic, pragmatics, ...

phrase structure, syntactic frames, ...

syllables, words, ...

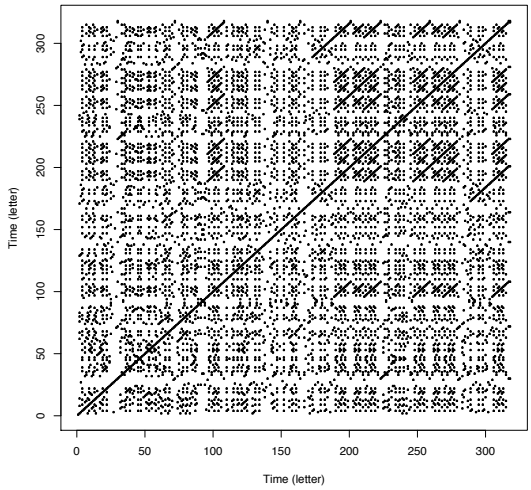
speech rate, pause duration, ...

amplitude, frequenc(ies), pitch, ...



"Higher-Order" States

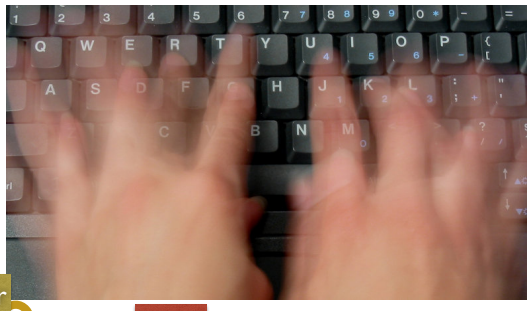
- Dynamics of the cognitive system produce **higher-order properties** that can be subjected to dynamic methods.
- These measurements are on a **nominal** scale, such as behavioral categories over time (e.g., emotions)
- **Language** is often studied in this manner: sounds, words, sentence structures, topics of conversation, etc.



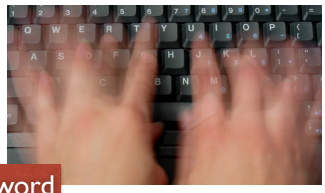
Outline

1. Time series of higher-order states
   Analysis of series of behavioral *categories*
2. The recurrence plot (RP) and "textures"
3. Quantifying the plot (RQA)
4. Examples and exercises

Example…



Example…



letter

word

nonlinear, nonlinear, let's hear it for today,
the content is delicious, fractal in every way.
nonlinear, nonlinear, rich in data too.
let's consult about your research - i dig, do you?

sentence

Example

Spontaneous

```
i am attending the nonlinear ati
this year and i enjoy showcasing
nonlinear and dynamic methods. i
also enjoy the practical
sessions in which we discuss the
research of attendees too.
```

spontaneous.txt

Poetic

```
nonlinear, nonlinear, let's hear it for today
the content is delicious, fractal in every wa
nonlinear, nonlinear, rich in data too.
let's consult about your research - i dig, do
```

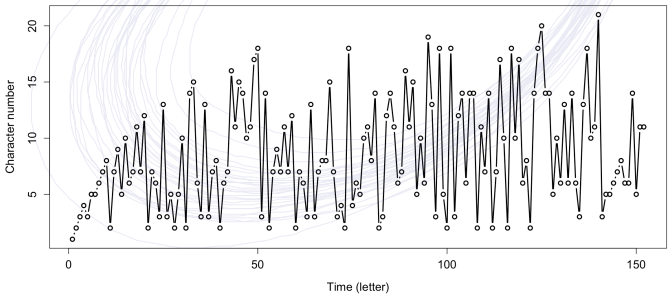
poetic.txt

Conversion…

- How do we get **text** into a format that permits use of nonlinear software packages?
- Let's **convert** transcripts to sequence of numeric identifiers:
  - *"i am attending the nonlinear …"*
    i = 1, <space> = 2, a = 3 …
  - *1 2 3 4 2 3 5 5 …*
- Level of analysis here: **letters**

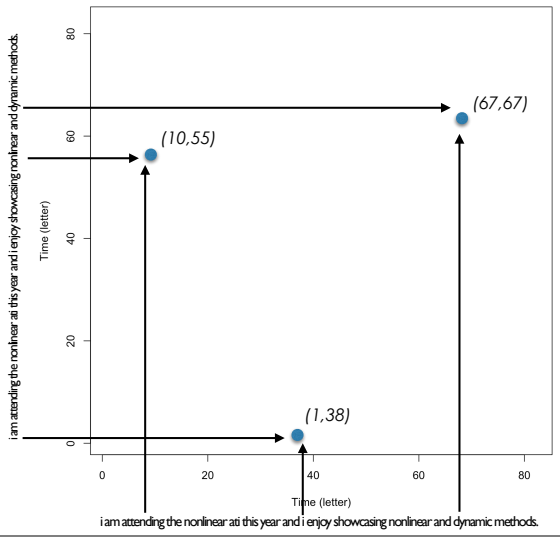
Trajectories

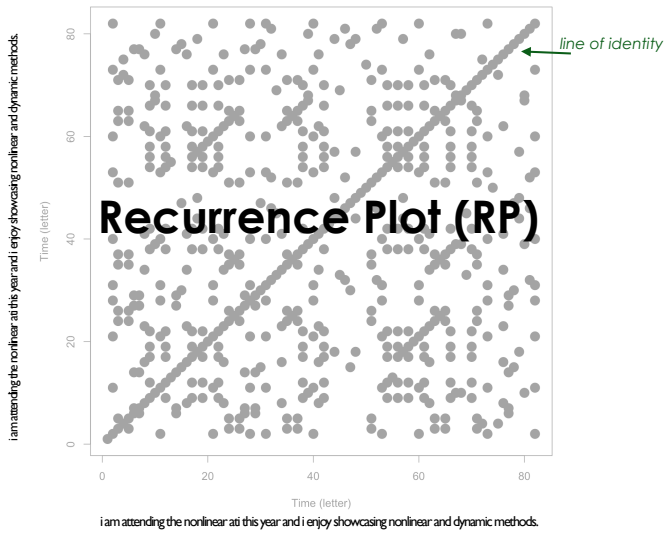
- Think of the behavior sequences in geometric terms: **trajectories in "category space."**



Recurrence Plot (RP)

- Bird's-eye view of the system's trajectories through its behavior space.
- With recurrence plots, we visualize various features of this path.
  - How repetitive is a sequence? If repetitive, how *long* are the repetitions? Lots of short bursts, or fewer but lengthy repeated trajectories?





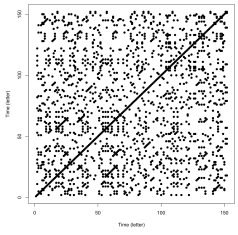
Under the Hood

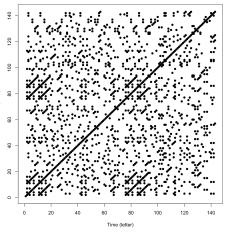
- If time series is:
  x = *1, 3, 2, 3, …,* x_N
- RP = set of points (*i, j*) such that:
  $x_i - x_j = 0$
- In other words, set of points such that the numeric identifiers *have a distance of zero from each other*.

Textures (Eckmann et al., 1987)

Spontaneous

Poetic





Exercises 1 & 2:

1. Convert Some Text
2. Build Some RPs

Quantification

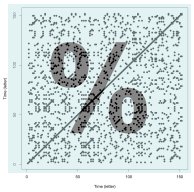
- You can eyeball a plot, but in real contexts we want some **quantification** so that plots (or, e.g., conditions) can be compared.
- Enter: *recurrence quantification analysis* (RQA).
  - These are measures that describe the **extent and distribution** of points on the plot.

RQA Measures

%REC (or, RR)

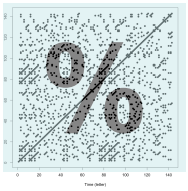
Recurrence rate (*%REC*): Total percentage of the plot occupied by points.

Spontaneous



7.6%

Poetic



7.8%

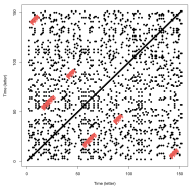
http://www.recurrence-plot.tk $RR = \frac{1}{N^2} \sum_{i,j=1}^{N} \mathbf{R}_{i,j}$

RQA Measures

%DET (or, DET)

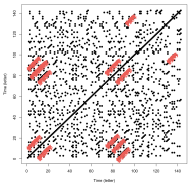
Percent determinism (*%DET*): Percentage of the points on the plot that fall on diagonal lines (length > 1).

Spontaneous



28.7%

Poetic



31.7%

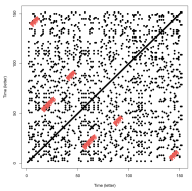
http://www.recurrence-plot.tk $DET = \frac{\sum_{l=l_{min}}^{N} l P(l)}{\sum_{i,j}^{N} \mathbf{R}_{i,j}}$

RQA Measures

MEANLINE

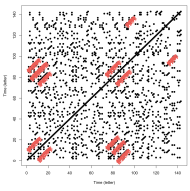
Average diagonal line length (*MEANLINE*): Average length of diagonal lines on the plot excluding the line of incidence (length > 1).

Spontaneous



3.4

Poetic



4.4

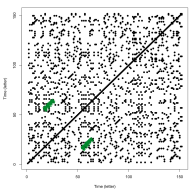
http://www.recurrence-plot.tk $L = \frac{\sum_{l=l_{min}}^{N} l P(l)}{\sum_{l=l_{min}}^{N} P(l)}$

RQA Measures

MAXLINE

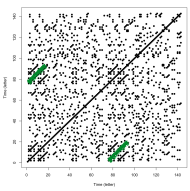
Maximum line length (*MAXLINE*): The longest diagonal line on the plot (excluding the line of incidence).

Spontaneous



10

Poetic



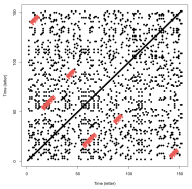
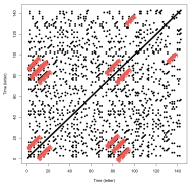
18

http://www.recurrence-plot.tk $L_{max} = \max(\{l_i;\ i = 1 \ldots N_l\})$

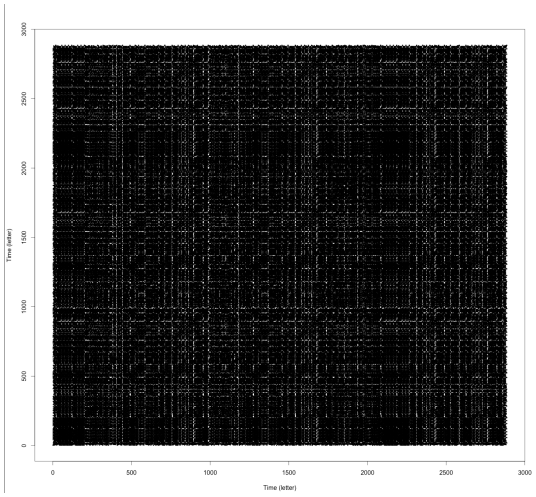
RQA Measures

ENTROPY

Entropy (*ENTROPY*): The entropy of the distribution of diagonal lines on the plot (how much "disorder" is there in the sequences).

Spontaneous	Poetic
	
0.7	1.1

http://www.recurrence-plot.tk $ENTR = -\sum_{l=l_{min}}^{N} p(l) \ln p(l)$

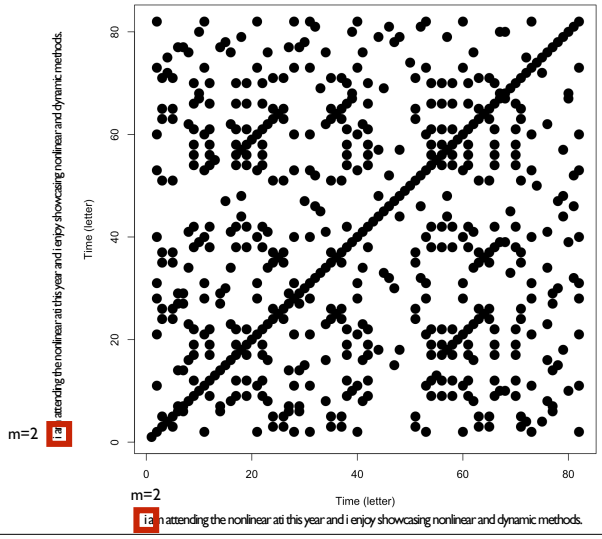


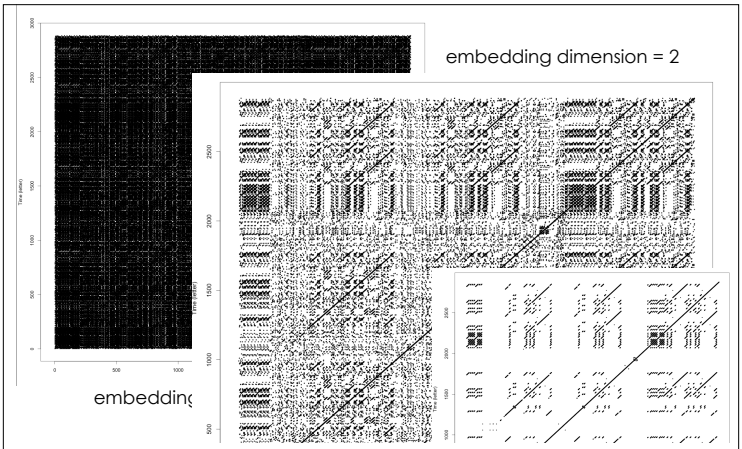
Three Important Concepts

- Parameters important for later…
  - embedding dimension,
  - delay,
  - radius
- These parameters will be *crucial* when we apply these RQA methods to *continuous* data (e.g., posture).

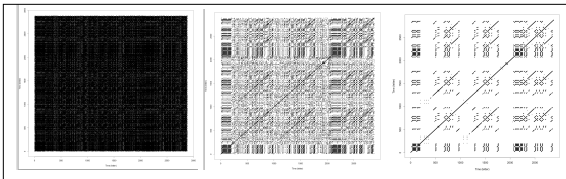
Embedding Dimension

- Interpretation in categorical data: how many states must match in order to count it as a recurrence.
- In previous analysis, dimension = 1
- What about 2? 3?
- *Window, vector, sequence, etc.*

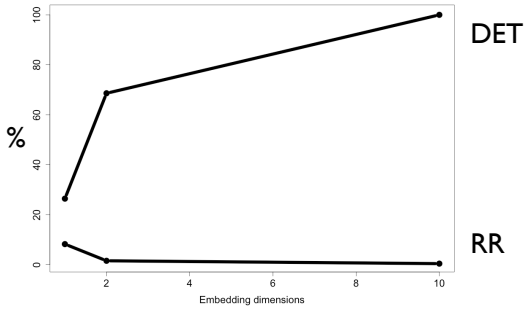




Dimension is useful with categorical data, and the window size can be determined by theoretical or practical concerns (e.g., avoiding single-letter recurrence, as above).



1 2 … 10



Delay

- With most categorical data (behavior sequences, linguistic sequences, etc.) temporal ordering should probably be preserved (delay = 1).
- NB: Situation more complex with continuous data (tomorrow).

Radius

- The distance between units (or windows if dimension > 1) required in order to count (*i, j*) as a *recurrent point*.
- With category data: radius = 0.
- NB: Situation more complex with continuous data.

Radius

But, with categories...

- With most data of nominal codes (e.g., letters, words, etc.), the following parameters often suffice:
  - Embedding dimension: 1
  - Delay: 1
  - Radius: .0001

Exercise 3:

Quantify Some Plots

Outline

- Time series of higher-order states
  - Analysis of series of behavioral *categories*
- The recurrence plot (RP) and "textures"
- Quantifying the plot (RQA)
- Examples and exercises

Outline

- Time series of higher-order states
  - Analysis of series of behavioral *categories*
- The recurrence plot (RP) and "textures"
- Quantifying the plot (RQA)
- Examples and exercises
- **Extensions and applications**

Extensions and Applications

- Windowed recurrence analysis
- Fresh work: dynamics of texts
- RQA as "dynamical NLP"

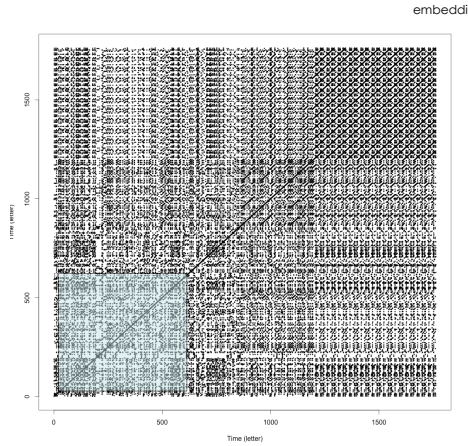
Extensions and Applications

- **Windowed recurrence analysis**
- Fresh work: dynamics of texts
- RQA as "dynamical NLP"

Nonuniformity in
behavioral dynamics

Behavioral "modes"
Stable, but temporary,
functional structures

Windowed Recurrence



Obtain several new time series of recurrence measures as they change across windows

Coco & Dale, 2014

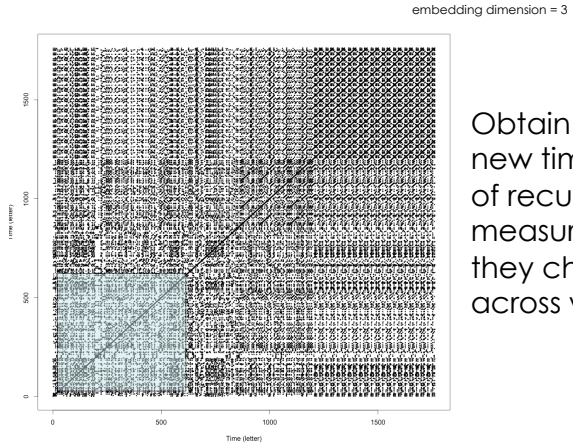
- With Dr. Moreno Coco
  University of Edinburgh
- **R** package for categorical recurrence (adaptable for continuous recurrence)
  - Basic (C)RQA measures
  - Diagonalwise recurrence (tomorrow afternoon)
  - Windowed recurrence measures

Cattail Down

by MeWithoutYou

headed east out of st. paul,
we stopped for water.
rested in the cemetery,
watched the mississippi.
running out of food stamps,
found a bag along the footpath
off highway 61 filled with
what looked like marijuana.
(don't worry mom, we left it there)
hopped a grainrail out of pig's eye
toward milwaukee, a deer
between the tower and the tracks,
saw right through us.
said, "you don't know where you came
you don't know where you're going,
you think you're you-

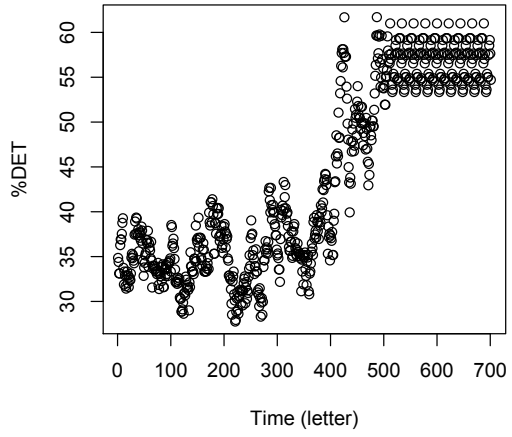
Windowed Recurrence

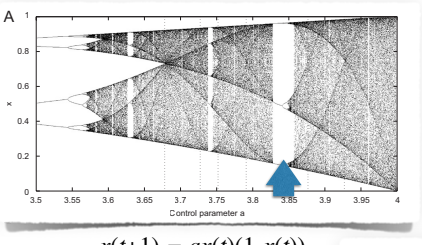


Obtain several new time series of recurrence measures as they change across windows

runcrqa in R

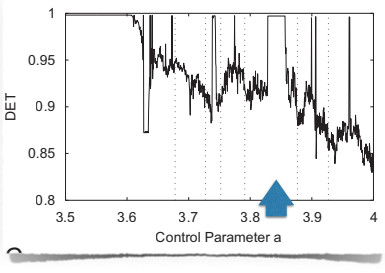
```
runcrqa(  ts1,
          ts2,
          par   )

      # par = list of parameters
      par = list(
                  type = 2, step =  1,
                  windowsize =  50,
                  lagwidth = 40,
                  method = "window",
                  datatype = "categorical",
                  thrshd = 8   )
```





$$x(t+1) = ax(t)(1-x(t))$$

each value computed with 800 data points!



Trulla et al., 1996, Physics Letters A; Marwan et al., 2007, Physics Reports

Extensions and Applications

- **Windowed recurrence analysis**
- Fresh work: dynamics of texts
- RQA as "dynamical NLP"

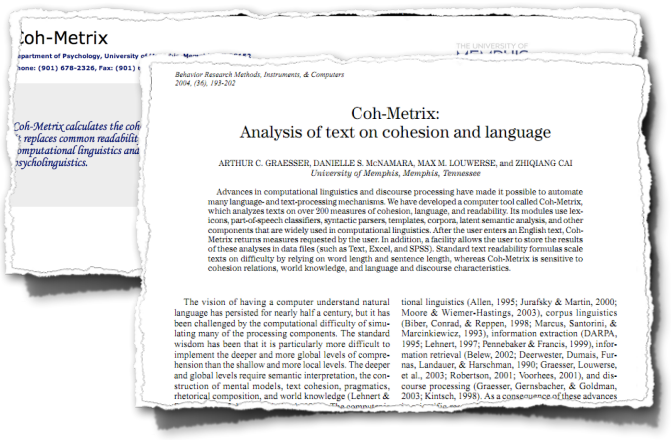
Extensions and Applications

- Windowed recurrence analysis
- **Fresh work: dynamics of texts**
- RQA as "dynamical NLP"

RQA for Text Analysis

- **Genre** identification in educational data mining contexts.
- Do "**textual dynamics**" differ across history, science, etc. texts?
- Do these dynamic patterns correlate with accessibility, learning gains, etc.?

“Cohesion”


Coh-Metrix

Behavior Research Methods, Instruments, & Computers
2004, (36), 193-202

Coh-Metrix:
Analysis of text on cohesion and language

ARTHUR C. GRAESSER, DANIELLE S. McNAMARA, MAX M. LOUWERSE, and ZHIQIANG CAI
University of Memphis, Memphis, Tennessee

Advances in computational linguistics and discourse processing have made it possible to automate many language- and text-processing mechanisms. We have developed a computer tool called Coh-Metrix, which analyzes texts on over 200 measures of cohesion, language, and readability. Its modules use lexicons, part-of-speech classifiers, syntactic parsers, templates, corpora, latent semantic analysis, and other components that are widely used in computational linguistics. After the user enters an English text, Coh-Metrix returns measures requested by the user. In addition, a facility allows the user to store the results of these analyses in data files (such as Text, Excel, and SPSS). Standard text readability formulas scale texts on difficulty by relying on word length and sentence length, whereas Coh-Metrix is sensitive to cohesion relations, world knowledge, and language and discourse characteristics.

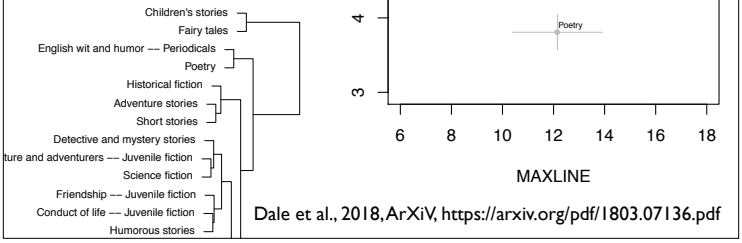
The vision of having a computer understand natural language has persisted for nearly half a century, but it has been challenged by the computational difficulty of simulating many of the processing components. The standard wisdom has been that it is particularly more difficult to implement the deeper and more global levels of comprehension than the shallow and more local levels. The deeper and global levels require semantic interpretation, the construction of mental models, text cohesion, pragmatics, rhetorical composition, and world knowledge (Lehnert & tional linguistics (Allen, 1995; Jurafsky & Martin, 2000; Moore & Wiemer-Hastings, 2003), corpus linguistics (Biber, Conrad, & Reppen, 1998; Marcus, Santorini, & Marcinkiewicz, 1993), information extraction (DARPA, 1995; Lehnert, 1997; Pennebaker & Francis, 1999), information retrieval (Belew, 2002; Deerwester, Dumais, Furnas, Landauer, & Harshman, 1990; Graesser, Louwerse, et al., 2003; Robertson, 2001; Voorhees, 2001), and discourse processing (Graesser, Gernsbacher, & Goldman, 2003; Kintsch, 1998). As a consequence of these advances

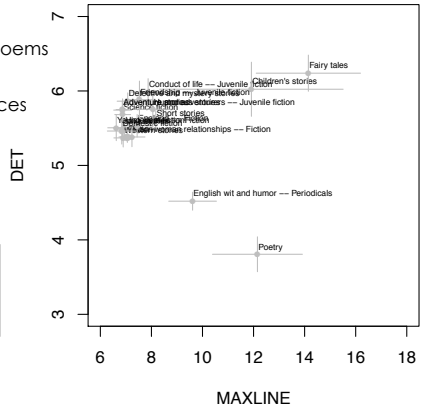
Gutenberg corpus

6,408 sections of books and poems

Extracted 5,000 word sequences

Analyzed an array of genres: poetry, fiction, etc.





Dale et al., 2018, ArXiV, https://arxiv.org/pdf/1803.07136.pdf

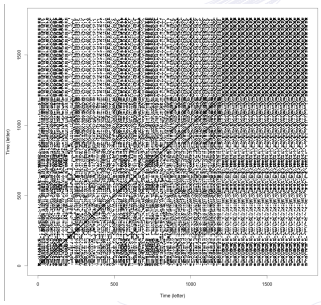
Extensions and Applications

- Windowed recurrence analysis
- **Fresh work: dynamics of texts**
- RQA as “dynamical NLP”

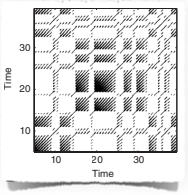
Extensions and Applications

- Windowed recurrence analysis
- Fresh work: dynamics of texts
- **RQA as "dynamical NLP"**

RQA as "Dynamical NLP"



- Can be used inside an interpretive framework anchored to "**complex dynamical systems**."
- When interpreted simply, can be used as a **technique** for obtaining **descriptive quantities** for behavioral streams

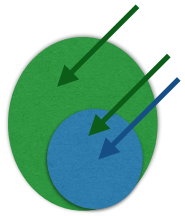


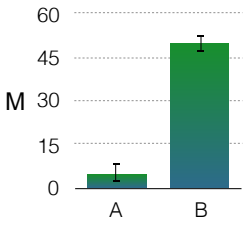
RQA	NLP Connection
DET	compressibility ratio
RR	co-occurrence or frequency
MAXLINE	longest common subsequence (LCS)
ENTROPY	n-gram variability

Dale et al., 2018, ArXiV, https://arxiv.org/pdf/1803.07136.pdf

Nonlinear?

*nonlinear **system***
*nonlinear **statistics***
*linear **statistics***





RQA as "generalized [auto-/]cross-correlation…"

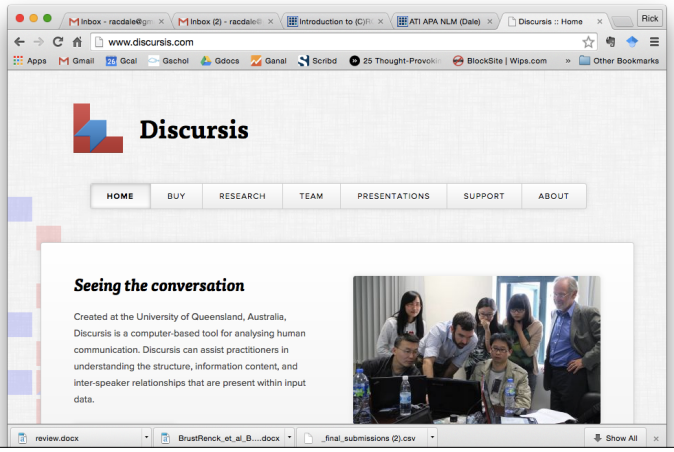
Recurrence plots for the analysis of complex systems

Norbert Marwan*, M. Carmen Romano, Marco Thiel, Jürgen Kurths

Nonlinear Dynamics Group, Institute of Physics, University of Potsdam, Potsdam 14415, Germany

Accepted 3 November 2006
Available online 12 January 2007
editor: I. Procaccia

Discursis (Angus et al.)



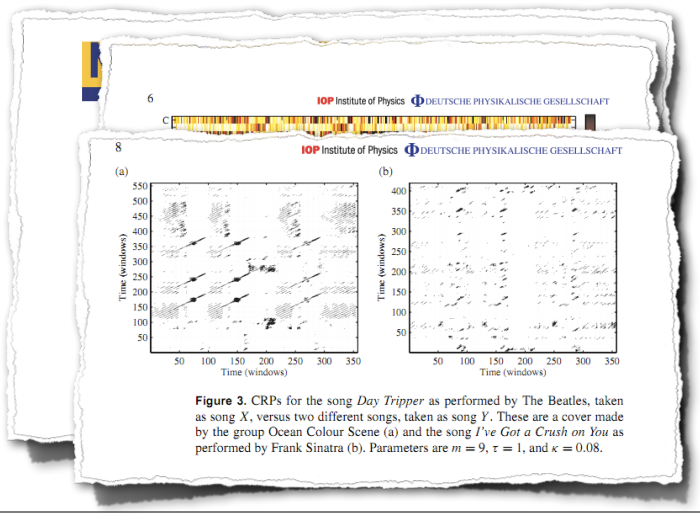


Figure 3. CRPs for the song *Day Tripper* as performed by The Beatles, taken as song X, versus two different songs, taken as song Y. These are a cover made by the group Ocean Colour Scene (a) and the song *I've Got a Crush on You* as performed by Frank Sinatra (b). Parameters are $m = 9$, $\tau = 1$, and $\kappa = 0.08$.

Exercise 4 (wrapping up):

Extended Exercises on RQA