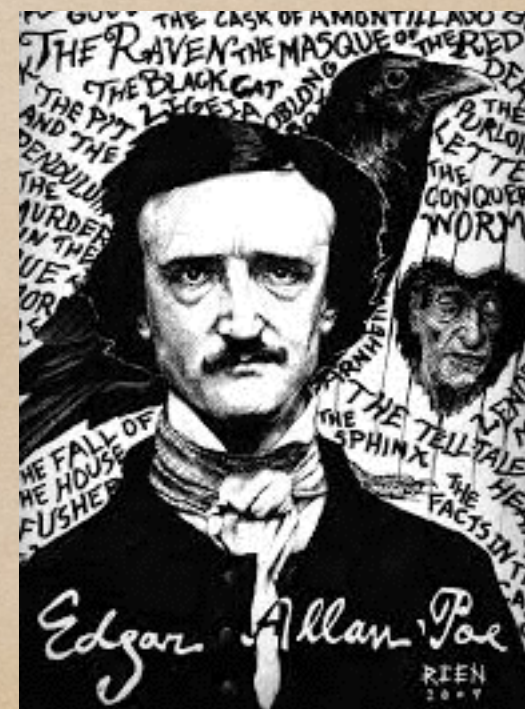


Topic Modeling on Poe-larity

Michael Laster





Motivation

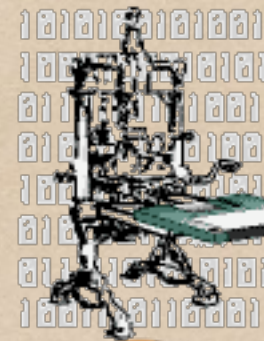


(From Wikipedia): "[Poe's] most recurring themes deal with questions of death... the reanimation of the dead, and mourning.... Beyond horror, Poe also wrote satires, humor tales, and hoaxes."

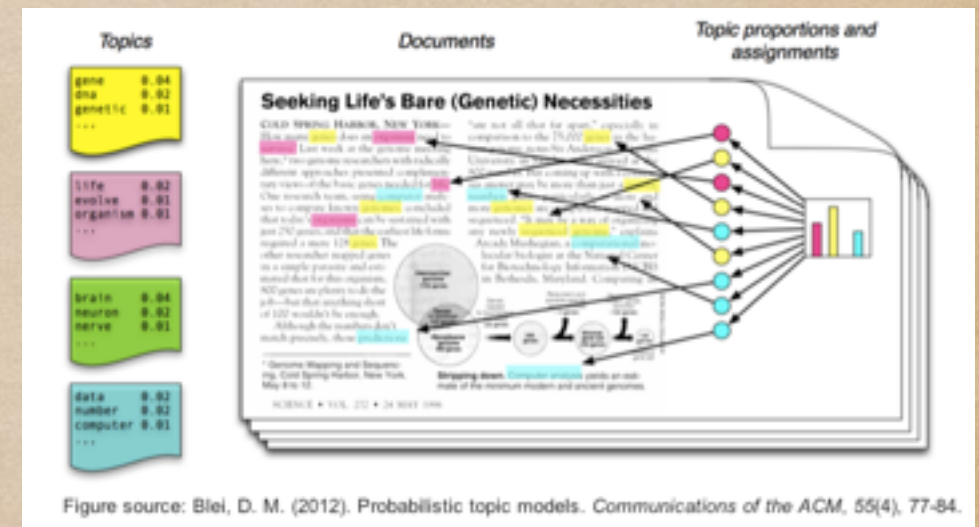
Can we tease out meaningful distinct topics?

Resources Used

- ◆ Culled 66 short stories from Project Gutenberg
- ◆ Used SKlearn's count-vectorizer
- ◆ LDA topic-modeling
- ◆ TextBlob sentiment tool
- ◆ Loaded results into MongoDB



PROJECT
GUTENBERG
"Fine Literature Digitally Re-Published" ☺



Process

- ◆ Grouped 66 short stories by Poe and treated each as its own document
- ◆ But...how many topics to model on?
 - ◆ Ran 24 topic models ranging from 1-24 topics
 - ◆ Measured the cumulative polarity of the 25 most common words of each topic in each model

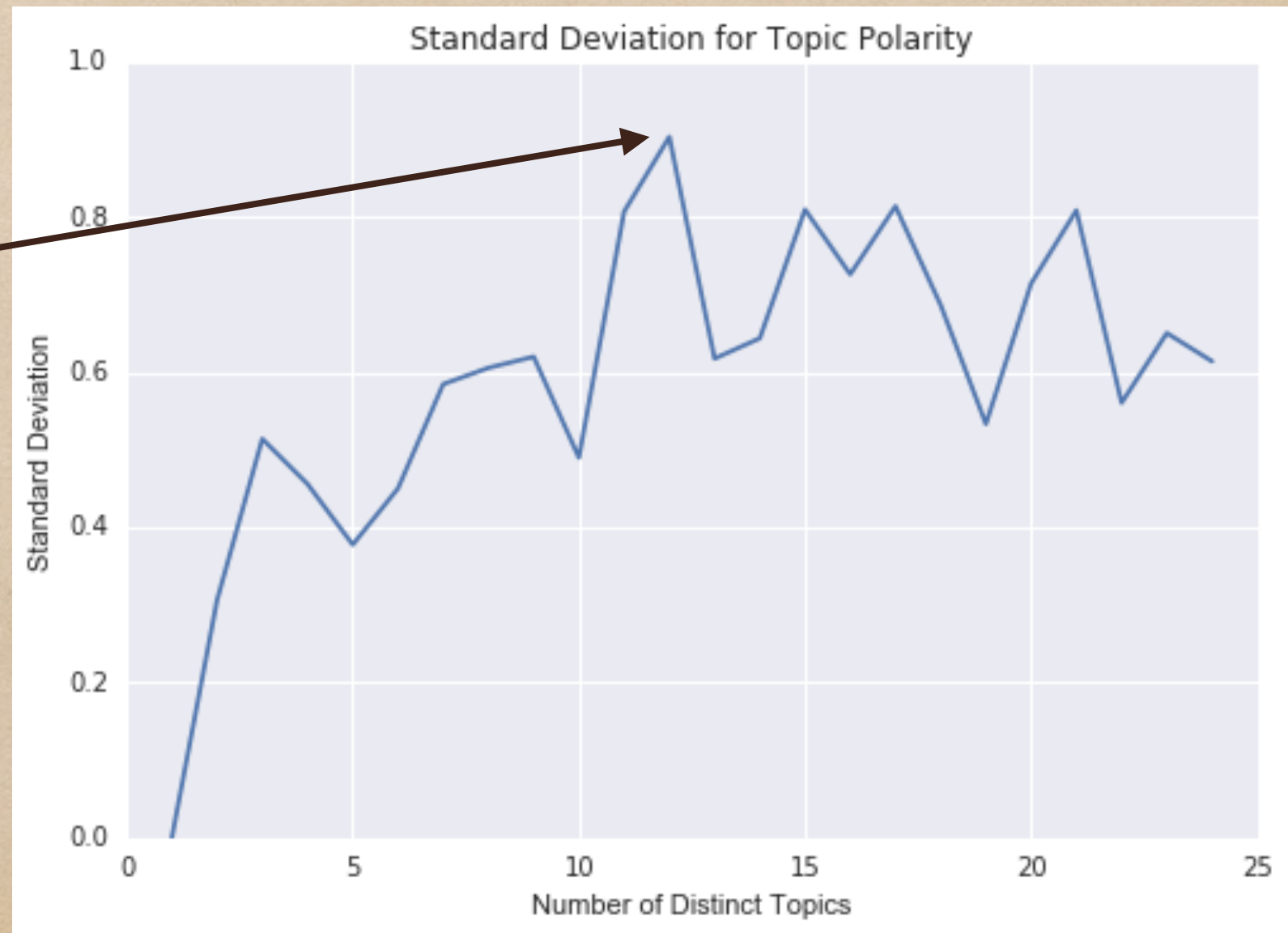
Intuition

- ◆ Fewer topics will capture fewer of a range in sentiment
- ◆ Too many topics will spread it out too thinly
- ◆ There must be some optimized amount in the middle



Looked for the model with the widest
variability in polarity from topic to
topic

n=12



Some Topics

Mortality

- ♦ long death yet spirit life soul
within character condition shadow
spoke thus days things feeling
never lips sense memory fancy died
earth god body evil

Art & Nature

- ♦ one beauty yet art seen valley
light scene young hundred character
green ellison whose love sense
flowers magnificent form became
trees eye less sentiment color

More Topics

Abstract Reflection

- ♦ may would upon matter us thus idea
nature even **general** every say let
reason said possible case man
thought person **merely mere** present
men **certain**

Mechanical things

- ♦ door arm open box back machine
automaton within body table person
left interior **game** **right** baron man
hand company never book maelzel
drawer **main** portion

Limitations

- ◆ TextBlob seems to be a crude indicator of sentiment
 - ◆ Many words register as neutral
- ◆ Language is dynamic, hard to use tools trained on modern language usage for 150 year old literature
- ◆ Domain knowledge can really contribute to interpreting meaningfulness of topics

Quoth the raven...

Thank you!

