

Topic Mapping

So new, google is barely aware of it.

The proof of awesome:

A lengthy and thorough academic paper:

<https://journals.aps.org/prx/pdf/10.1103/PhysRevX.5.011007>

A python implementation:

<http://amaral-lab.org/resources/software/topic-mapping>

The process

Step 1: Consider a body of text comprising six documents, three in the topic biology and three in the topic math. We exclude meaningless words from those documents and stem words (convert them to their simplest form) in order to denoise the data. We then build a network connecting words by similarity.

Process Part 2

Steps 2 and 3: We filter out insignificant connections, using a p value of 5%, and we run Infomap [44](A community detection algorithm) to obtain the community structure of the network. In this case, we find two clusters and two isolated words, “study” and “research”. Step 4: We refine the word clusters using a topic model: The two isolated words are now assigned to both topics

Lets try it!

1. Words from wikipedia:
 - a. In Haitian folklore, a zombie (Haitian French: zombi, Haitian Creole: zonbi) is an animated corpse raised by magical means, such as witchcraft.

Denoising the data.

1. Remove Stop-words (meaningless words)
 - a. Haitian folklore zombie Haitian French zombi, Haitian Creole zonbi animated corpse raised magical witchcraft.

Denoising part II

Stemming:

Haiti folklore zombie Haiti French zombi Haiti
Creole zonbi animate corpse raise magic
witchcraft.

The word Haitian becomes just Haiti, Magical-
>Magic

Denoising part III

Removing duplicates:

haiti folklore zombie haiti zombi french creole
zonbi animate corpse raise magic witchcraft.

Community Detection:

1. Pruning

- a. Construct a randomized model of this text based on the word frequency and see if the pairing of words happen more often than can be explained by random chance
- b. "haiti folklore zombie haiti zombi french creole zonbi animate corpse raise magic witchcraft."

c. <http://icbin.com/lobuaguna/1/edit?html.is.console>

Community detection

(Switch to white board)