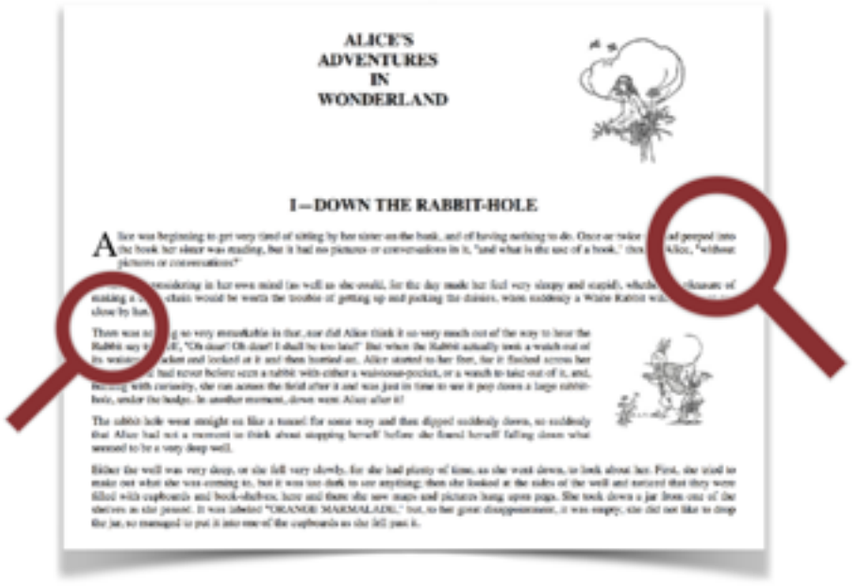
Embedding narrative of books and lyrics

Filipi Nascimento Silva

IUNI

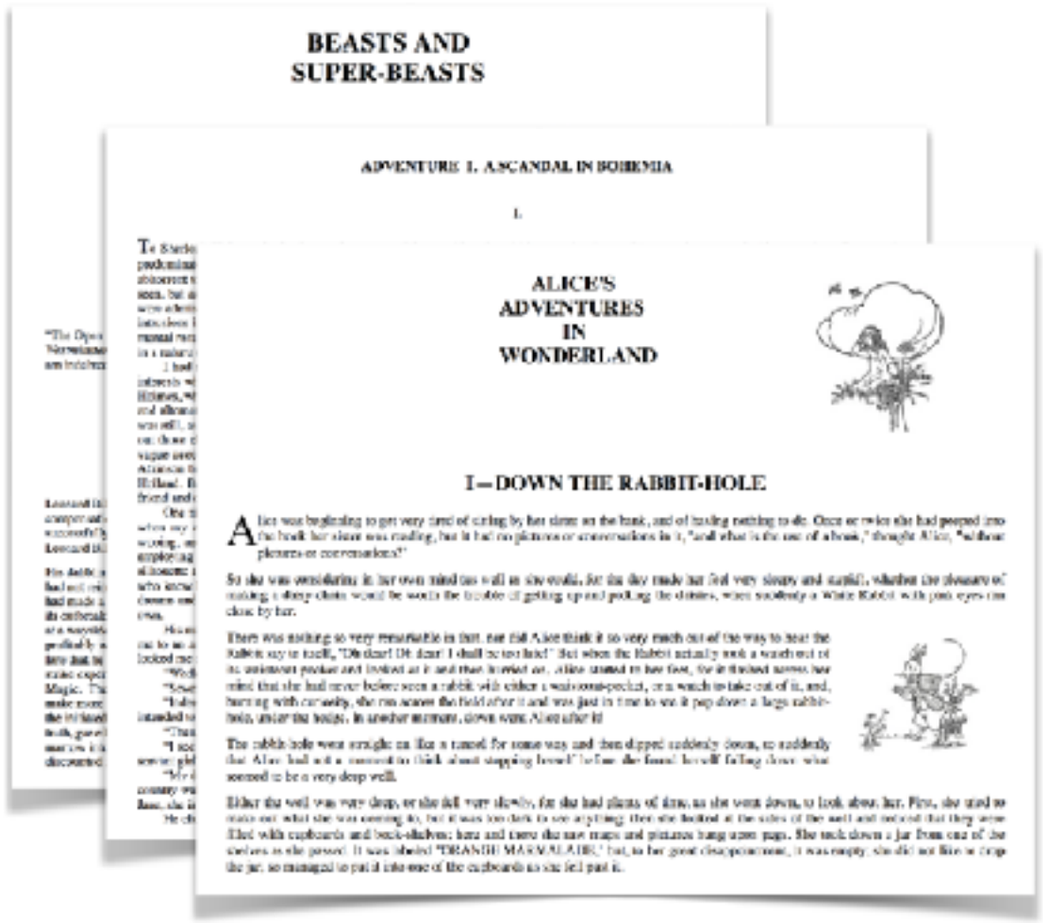
Motivation

- Extracting knowledge from information - understanding unstructured data
- Fast processing of huge amounts of data



Network science applied to text analysis

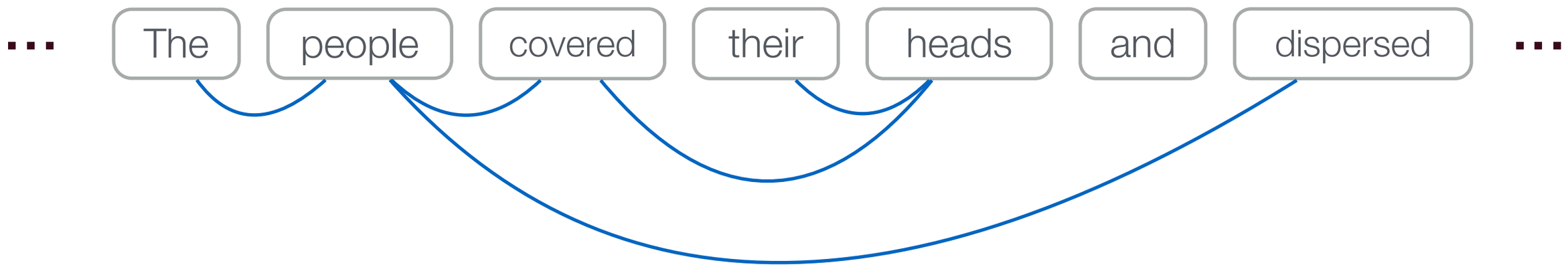
Traditional textual analysis



… The people covered their heads and dispersed …

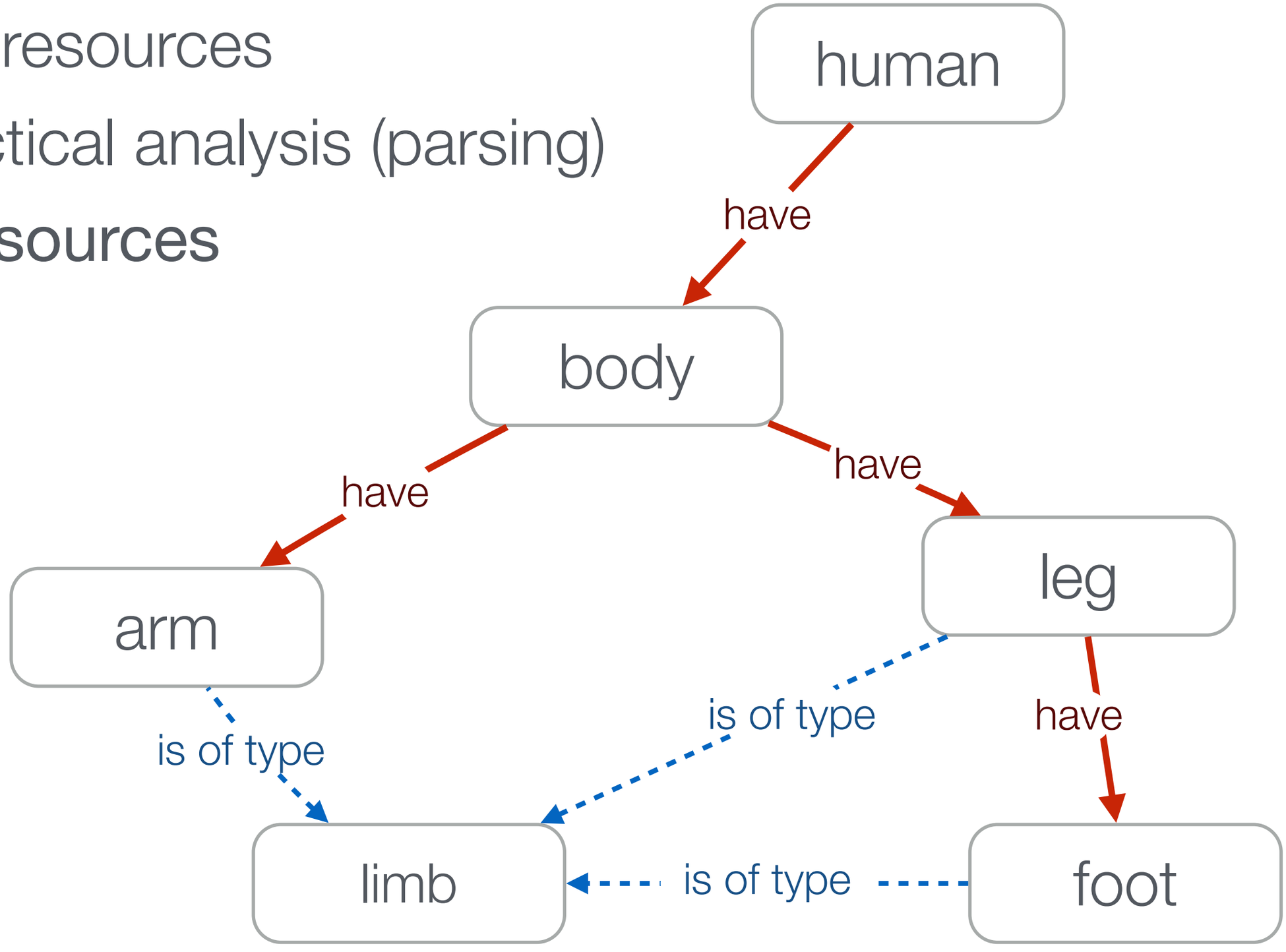
Traditional textual analysis

- Shallow resources
  - Syntactical analysis (parsing)



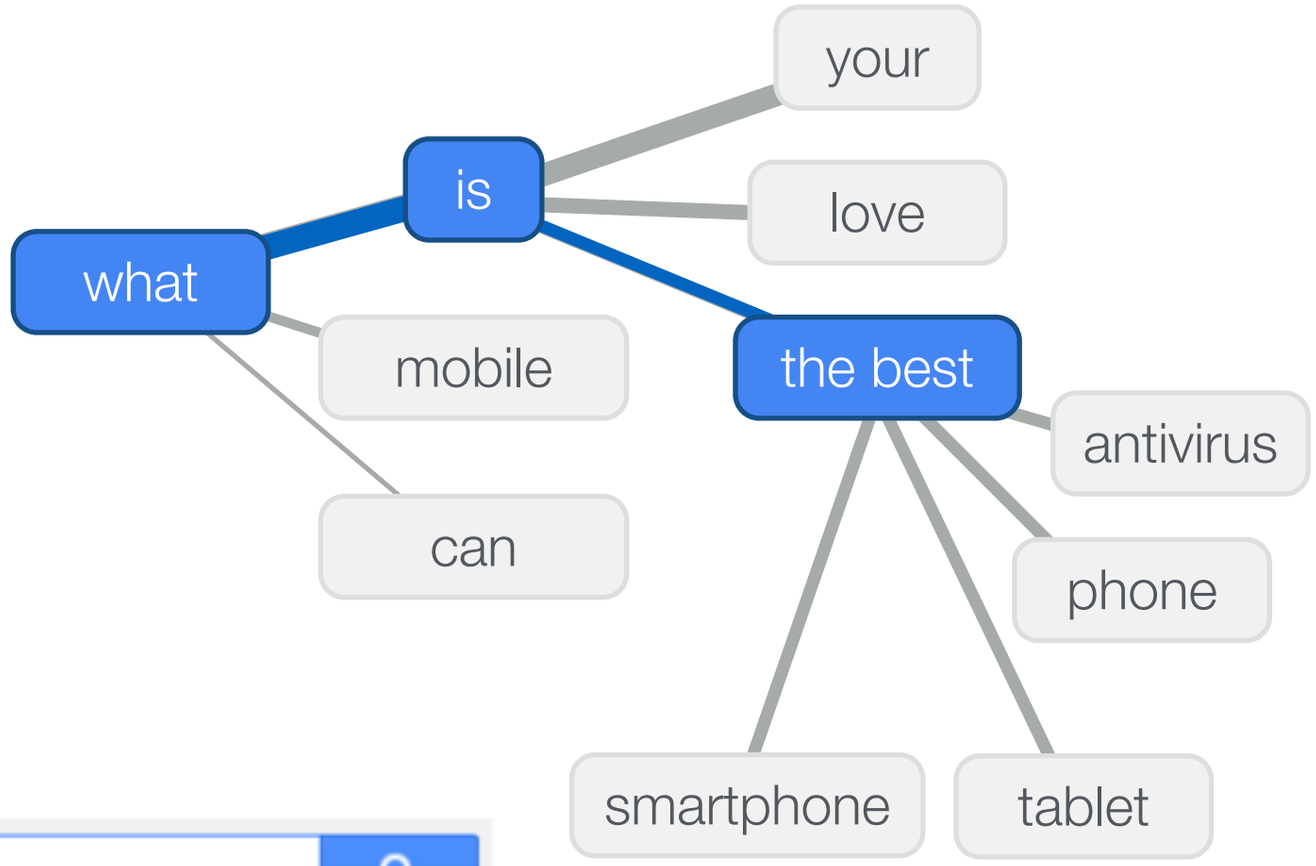
Traditional textual analysis

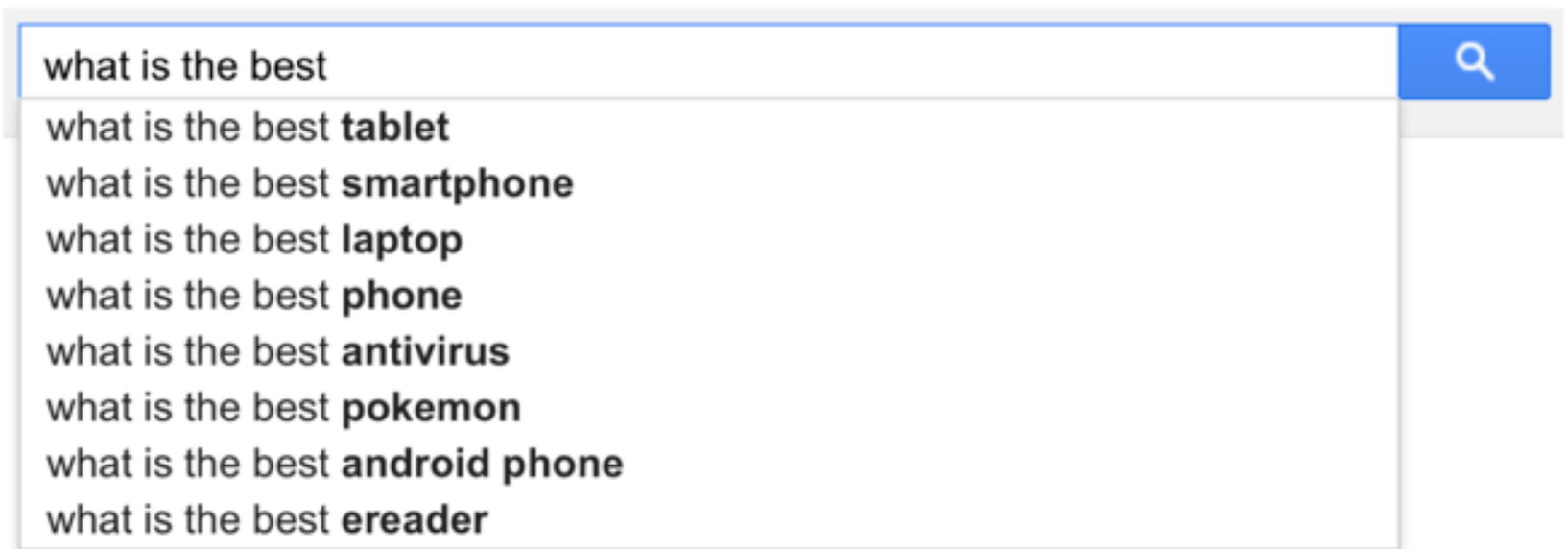
- Shallow resources
  - Syntactical analysis (parsing)
- Deep resources



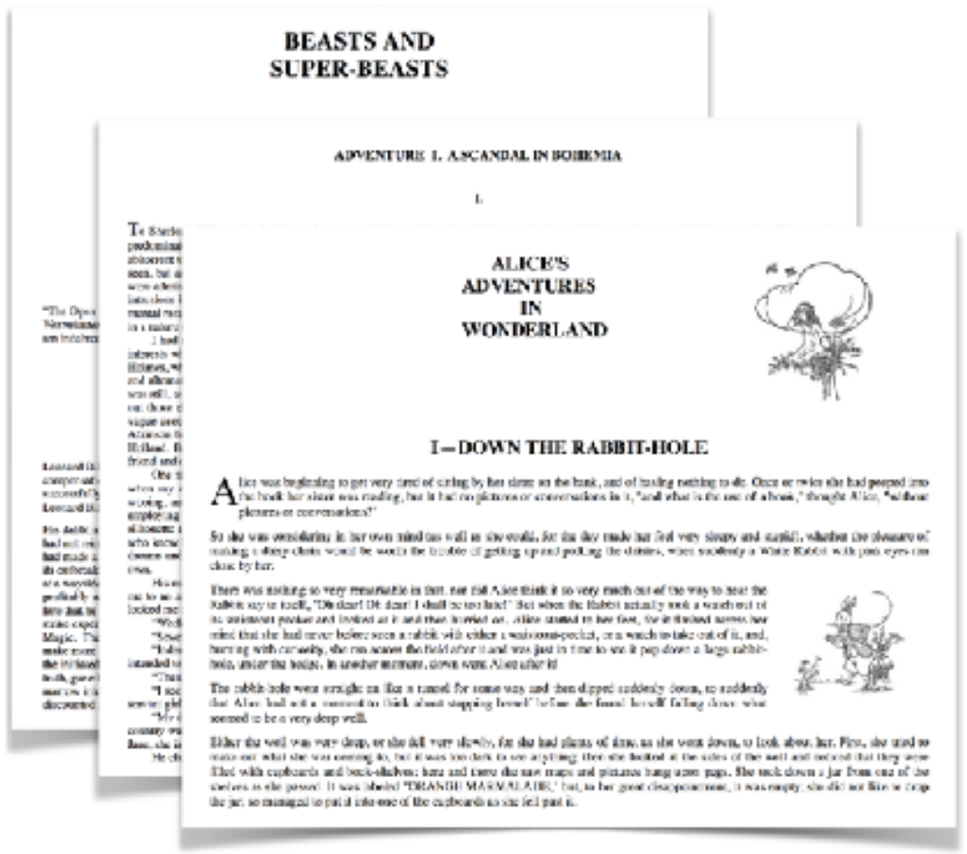
Traditional textual analysis

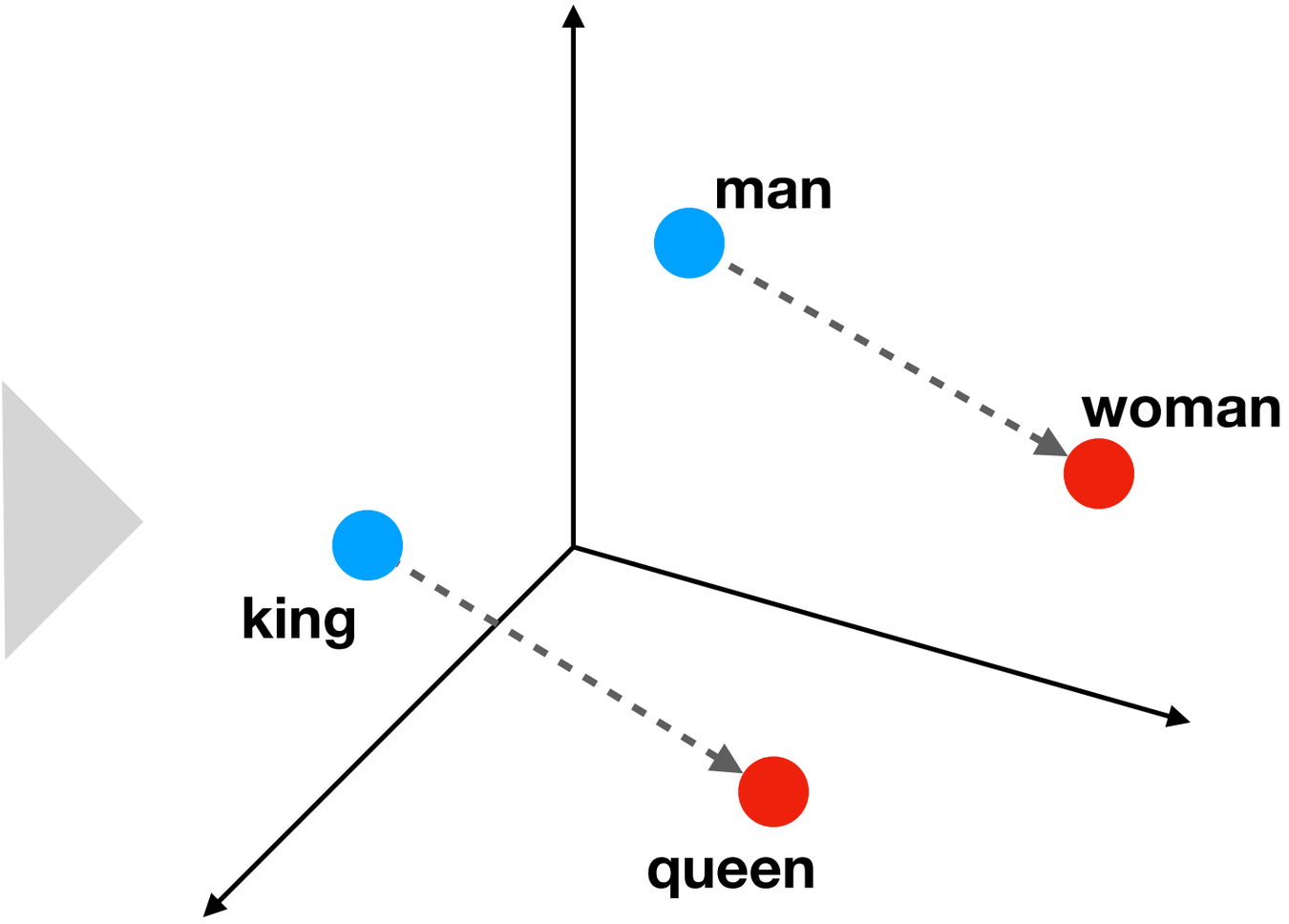
- Shallow resources
  - Syntactical analysis
- Deep resources
- **Statistical methods**





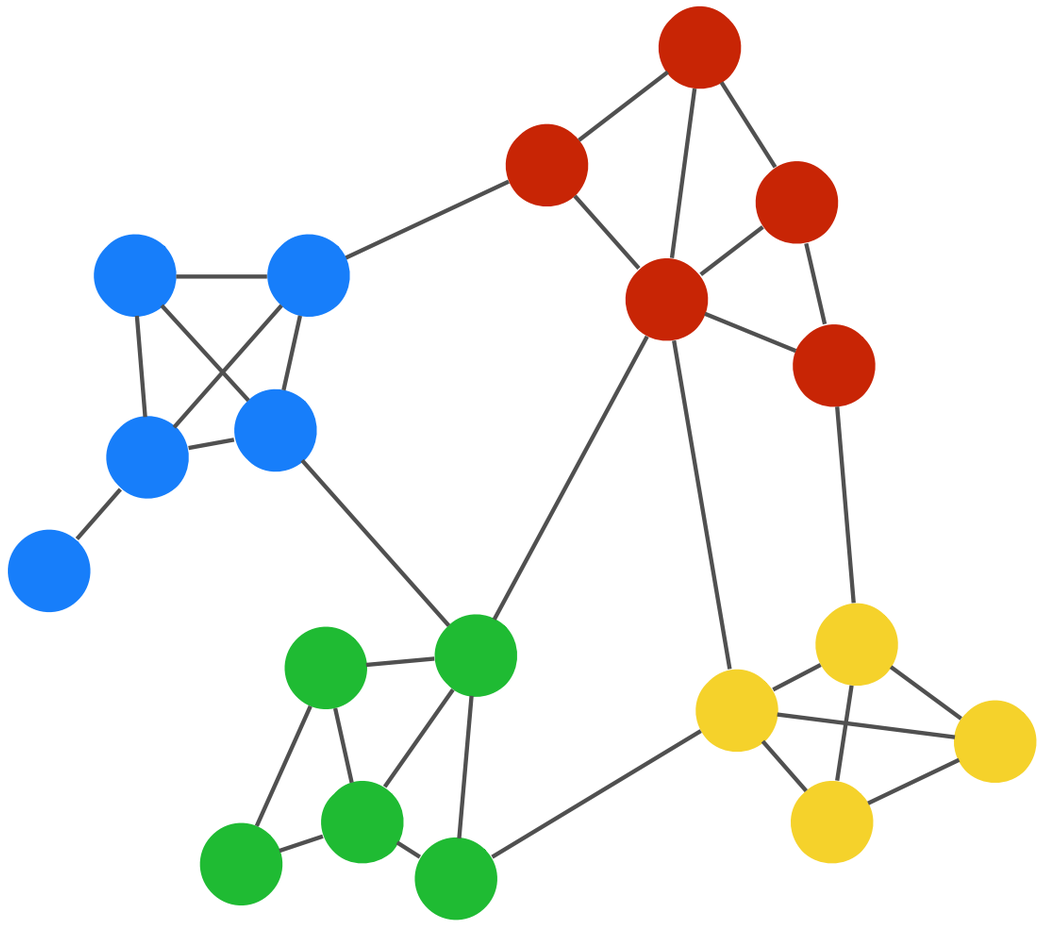
Word Embeddings





Network-based methods

- Shallow resources
  - Syntactical analysis
- Deep resources
- Statistical methods
- Deep learning
- Networks



“... wondering in his darker moments what Sally would say about that project when she heard of it, and he had hoped that she would not hear of it until all the preparations were so complete that interference would be impossible. He was extremely fond of Sally, but there was, he knew, a lamentable ...”

P. G. Wodehouse
– The Adventures of Sally

Removing functional words

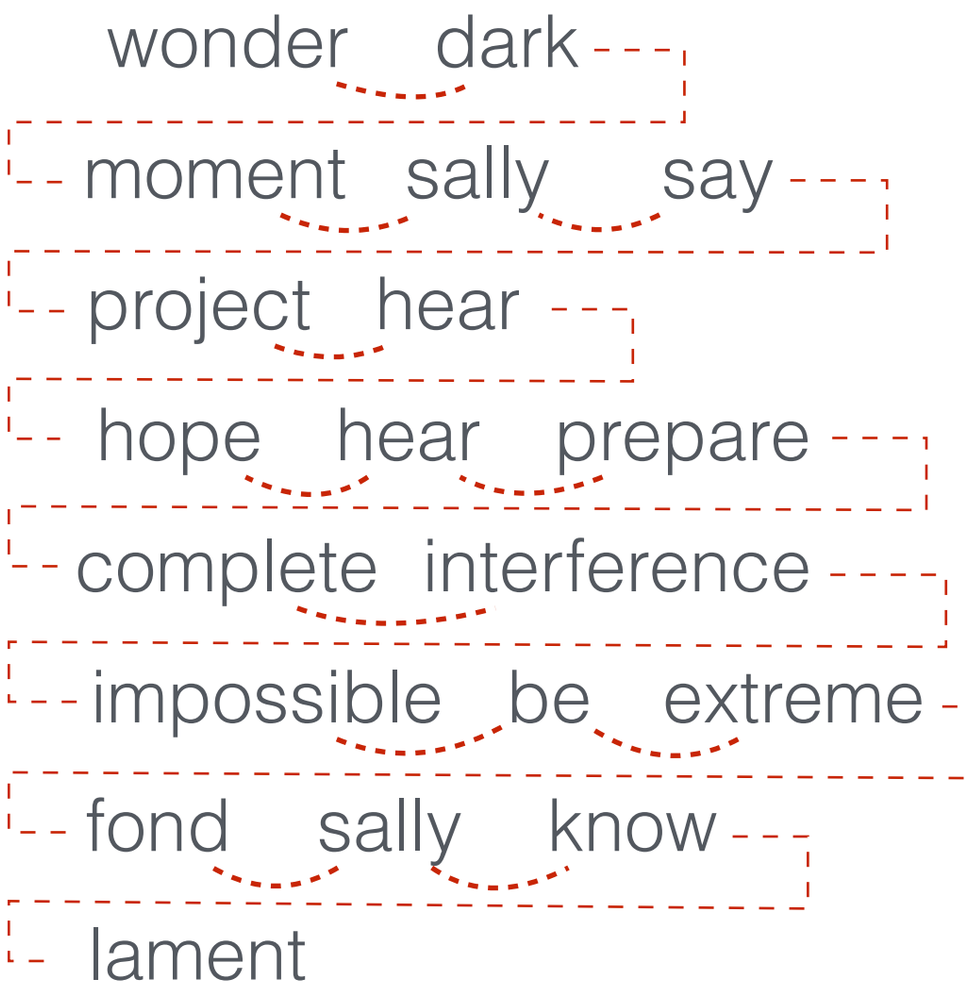
"... wondering in his darker moments what Sally would say about that project when she heard of it, and he had hoped that she would not hear of it until all the preparations were so complete that interference would be impossible. He was extremely fond of Sally, but there was, he knew, a lamentable ..."

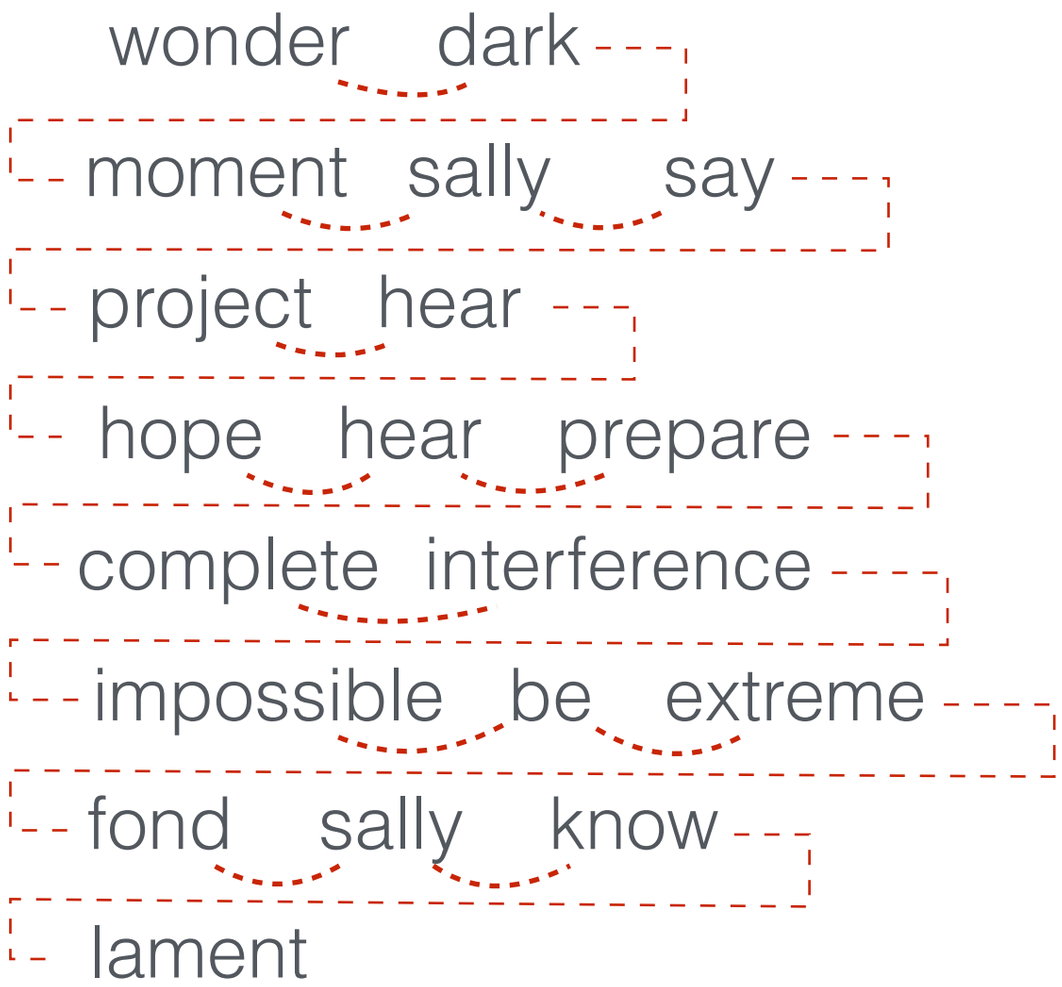
P. G. Wodehouse
– The Adventures of Sally

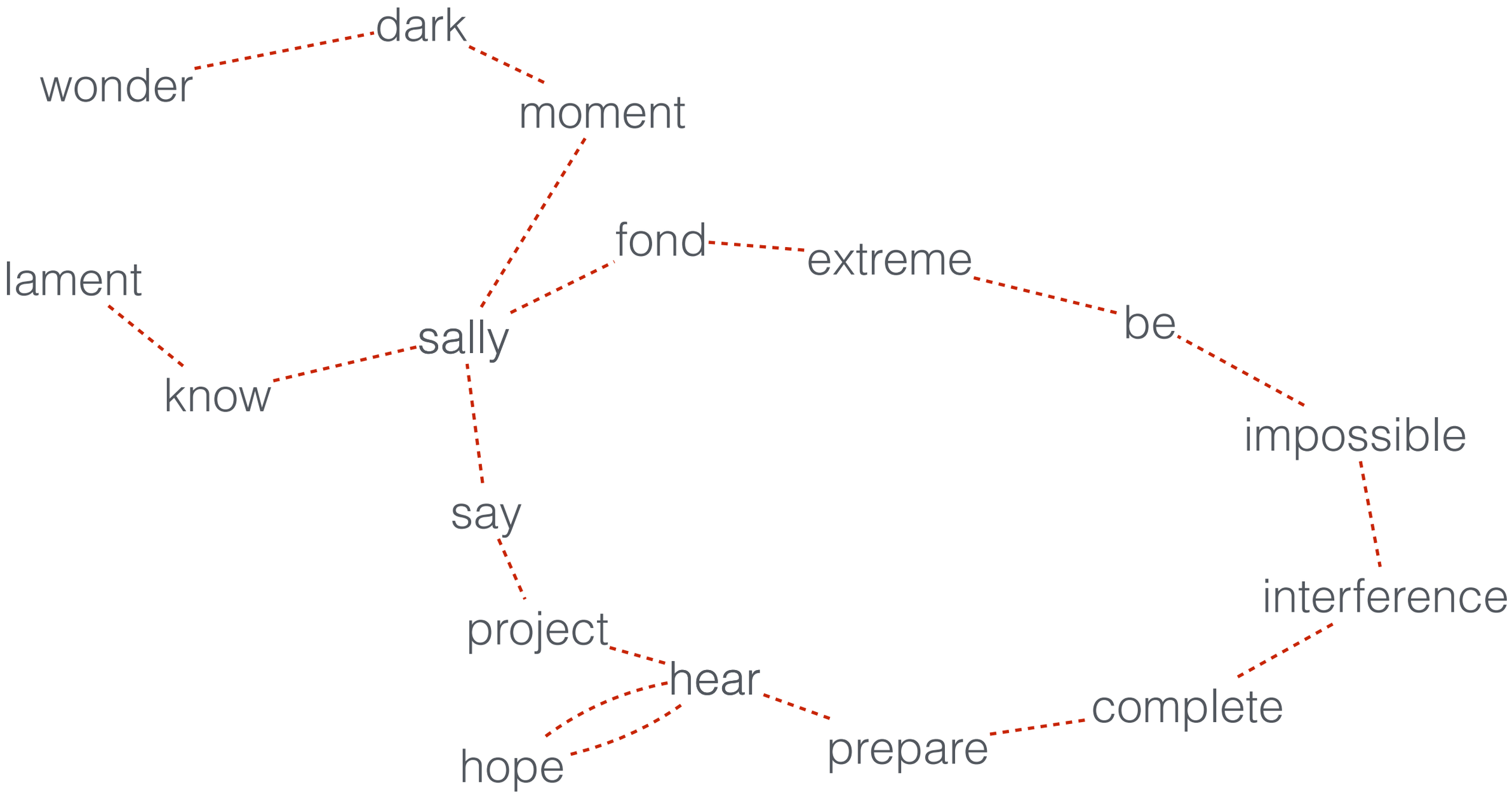
Lemmatization

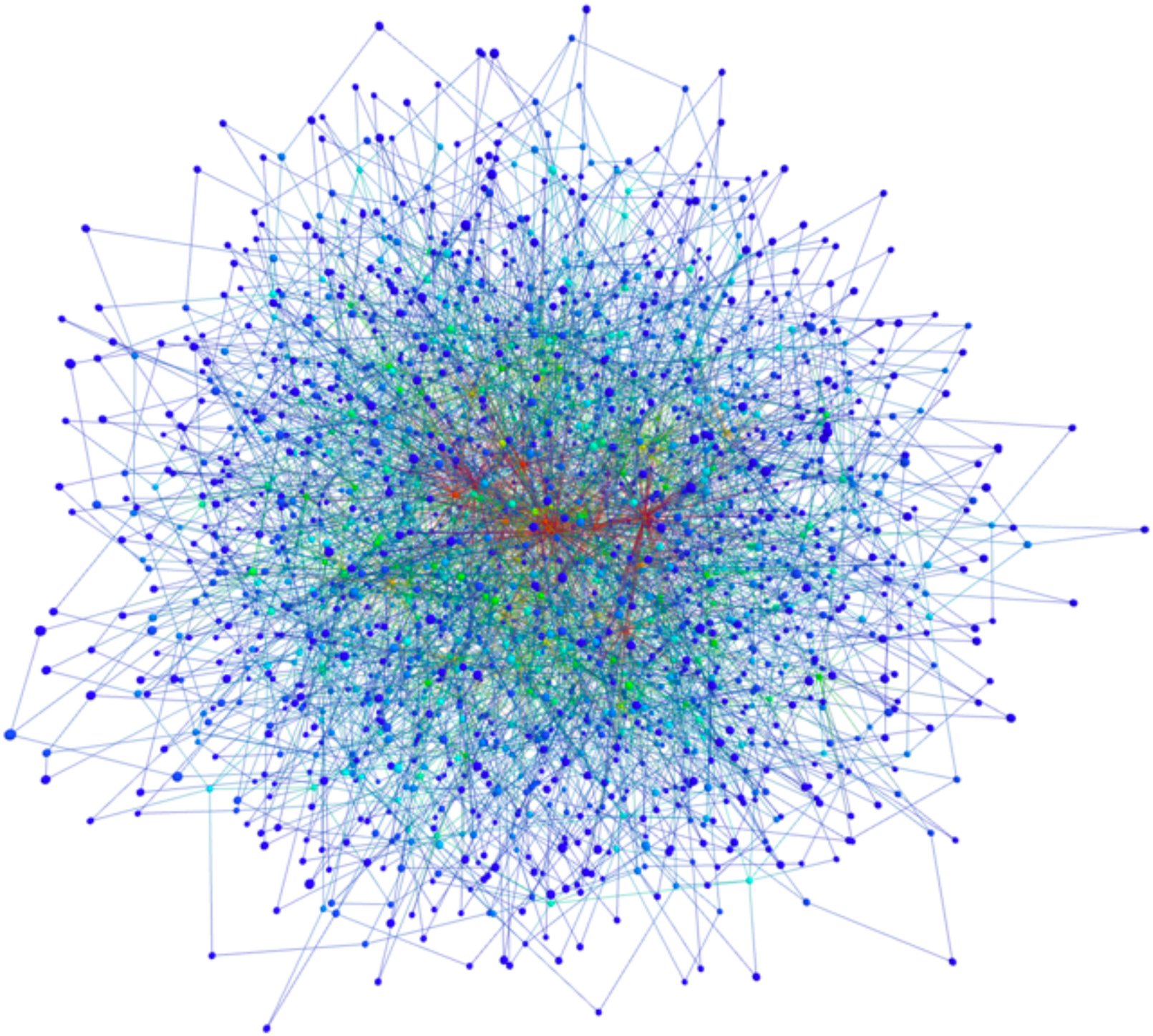
“… wondering in his darker moments what Sally would say about that project when she heard of it, and he had hoped that she would not hear of it until all the preparations were so complete that interference would be impossible. He was extremely fond of Sally, but there was, he knew, a lamentable …”

P. G. Wodehouse
– The Adventures of Sally



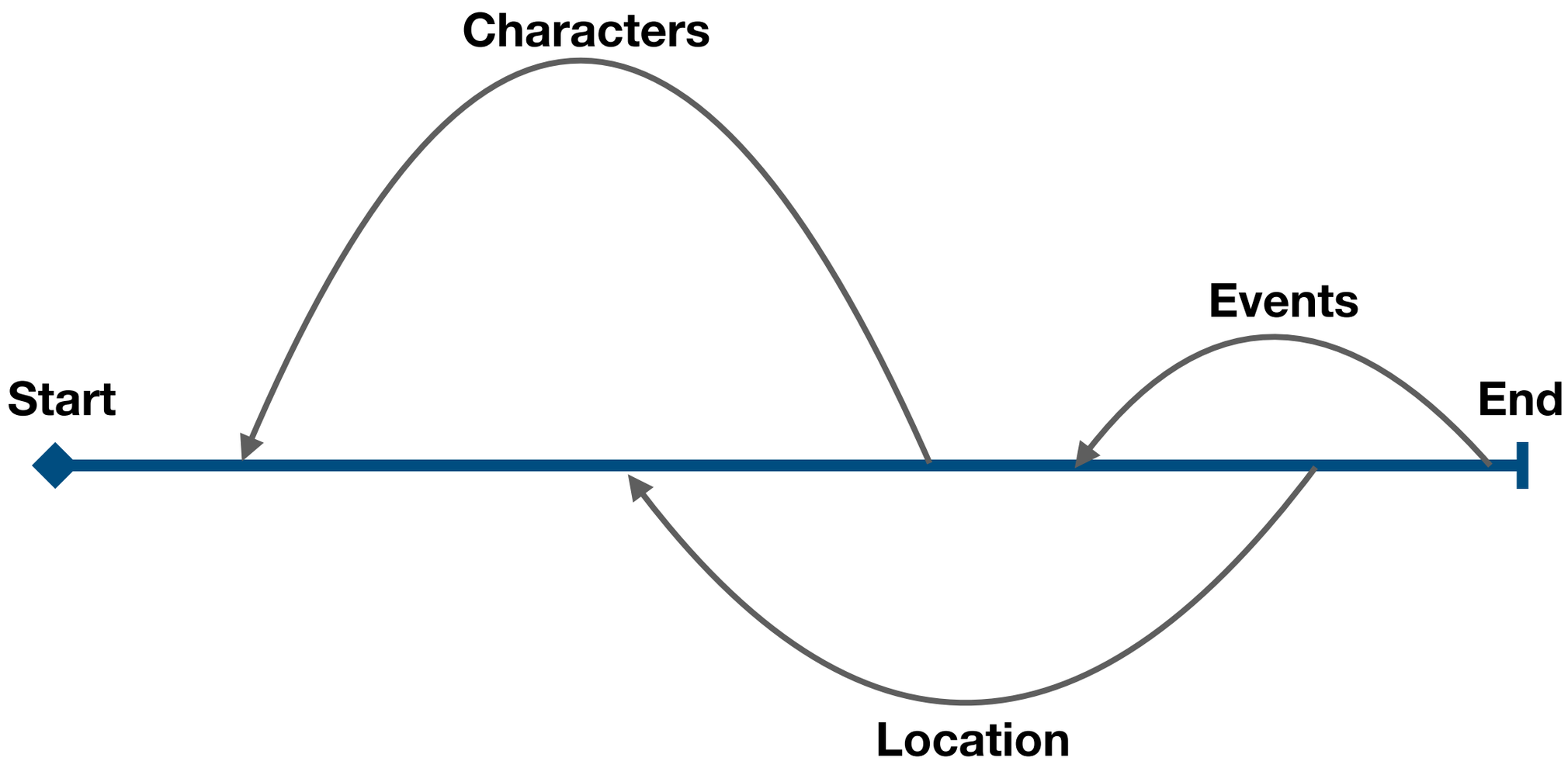

wonder dark
moment sally say
project hear
hope hear prepare
complete interference
impossible be extreme
fond sally know
lament


dark
wonder
moment
fond
extreme
lament
be
sally
know
impossible
say
interference
project
hear
complete
hope
prepare

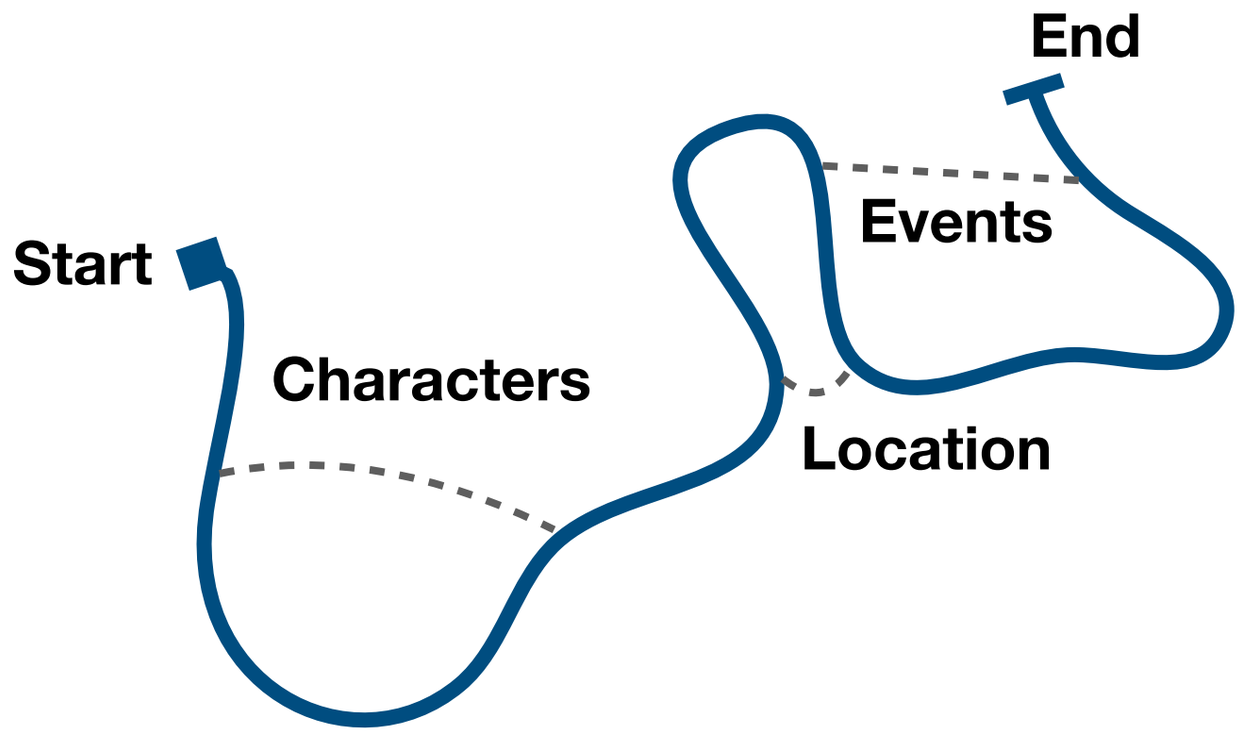


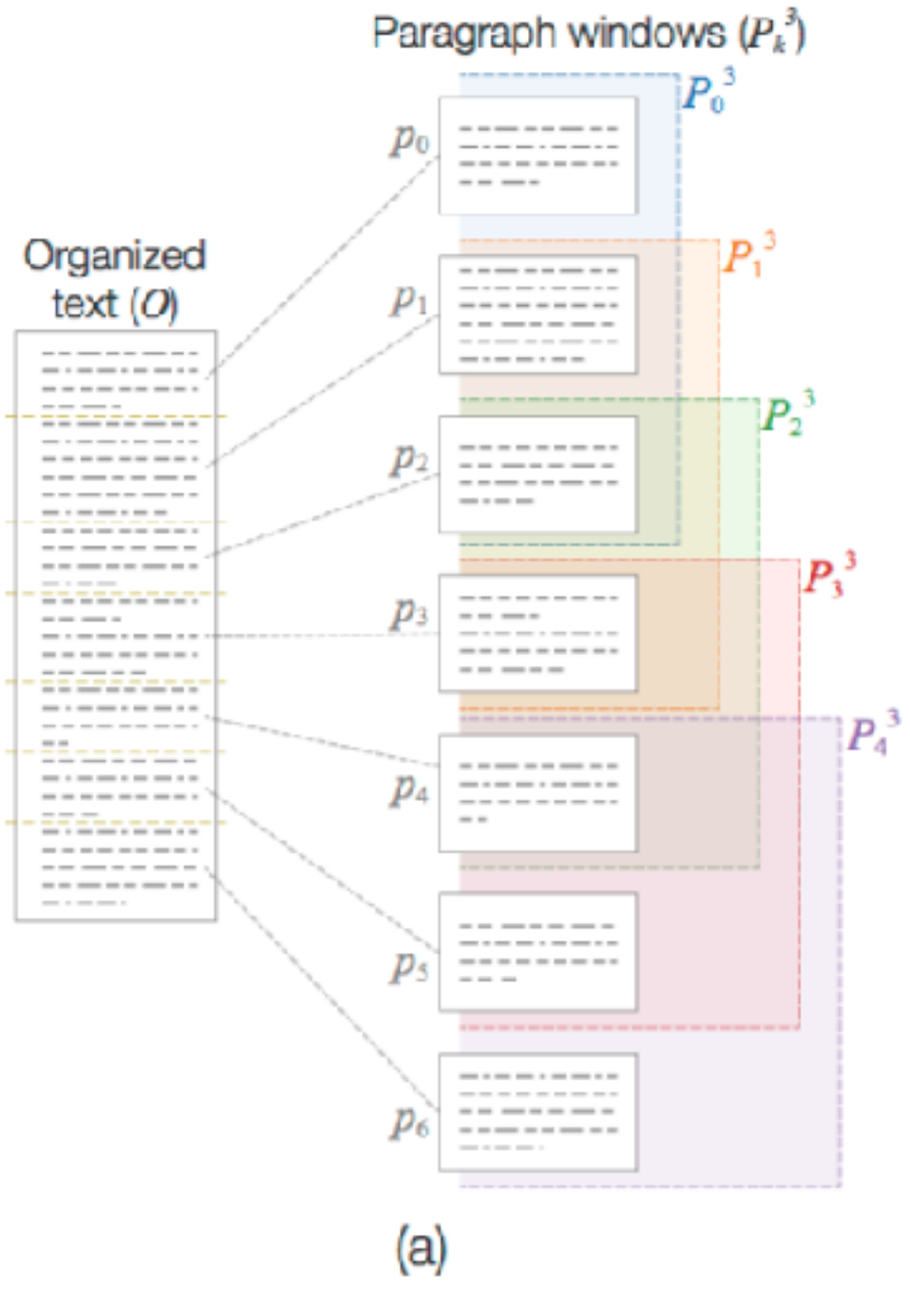
Narrative

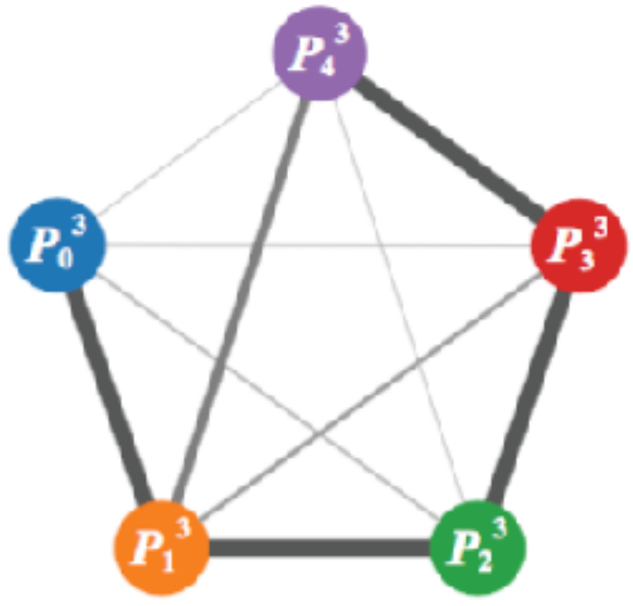
Narrative

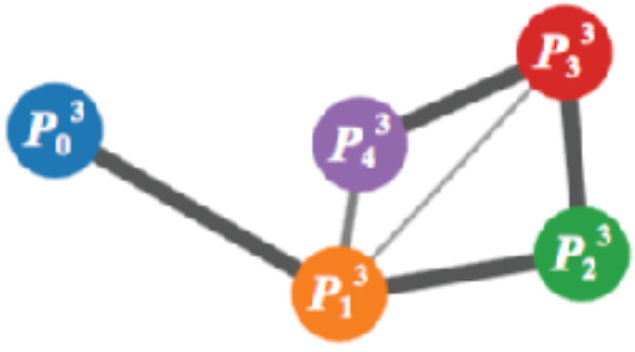


Narrative



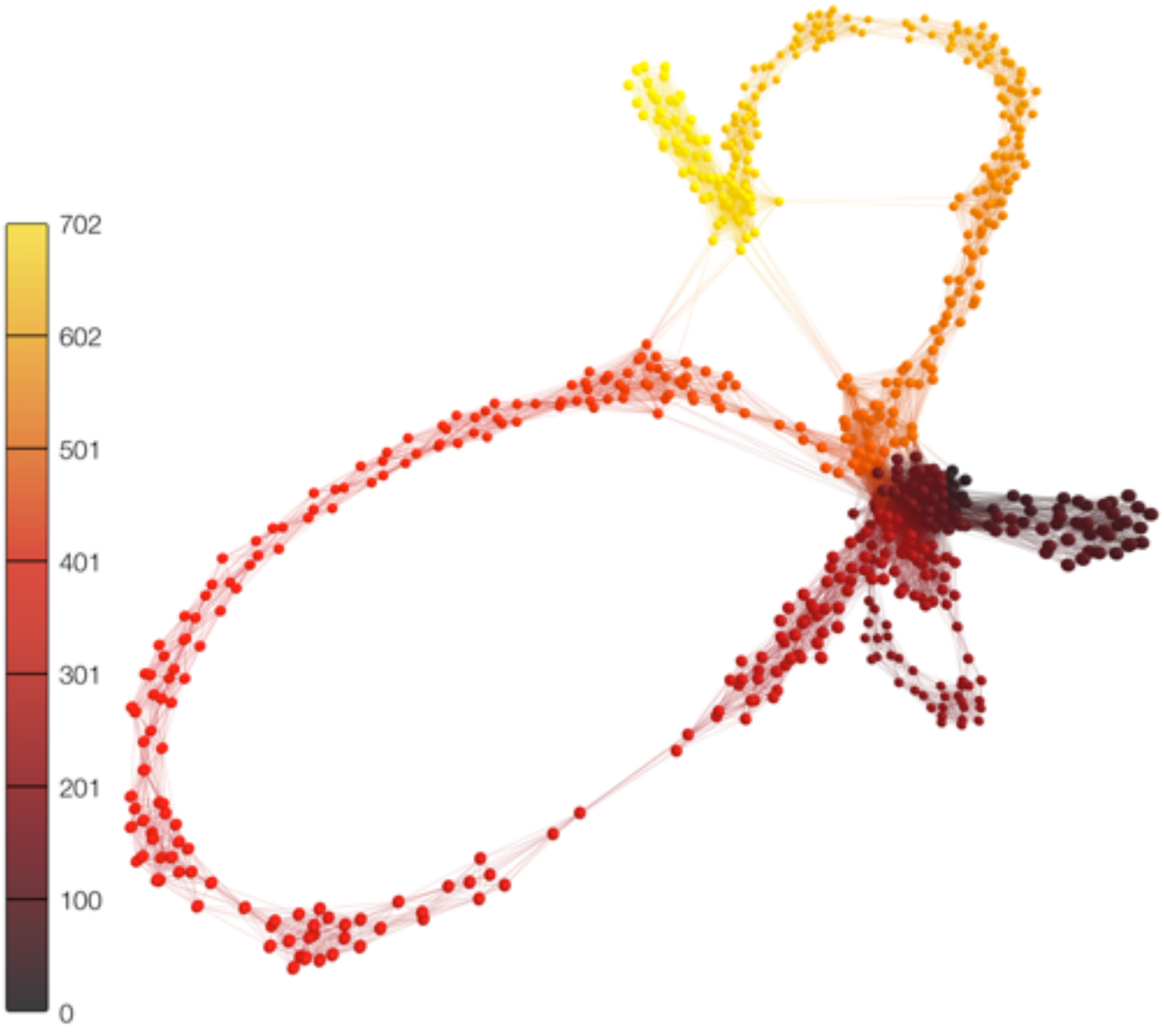


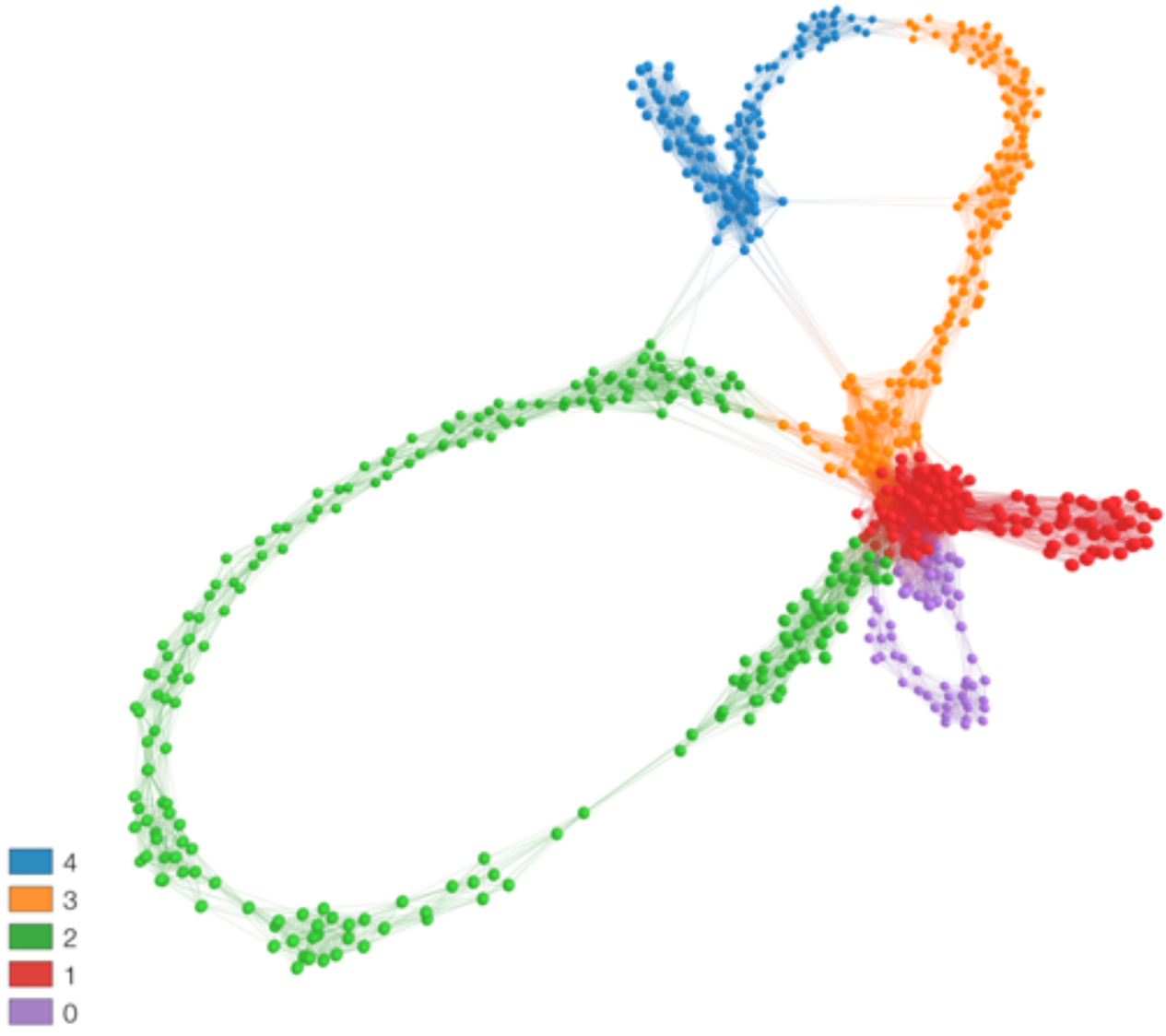




Mesoscopic representation of texts as complex networks
H. F. de Arruda, F. N. Silva, V. Q. Marinho, D. R. Amancio, L. da F. Costa
https://arxiv.org/abs/1606.09636

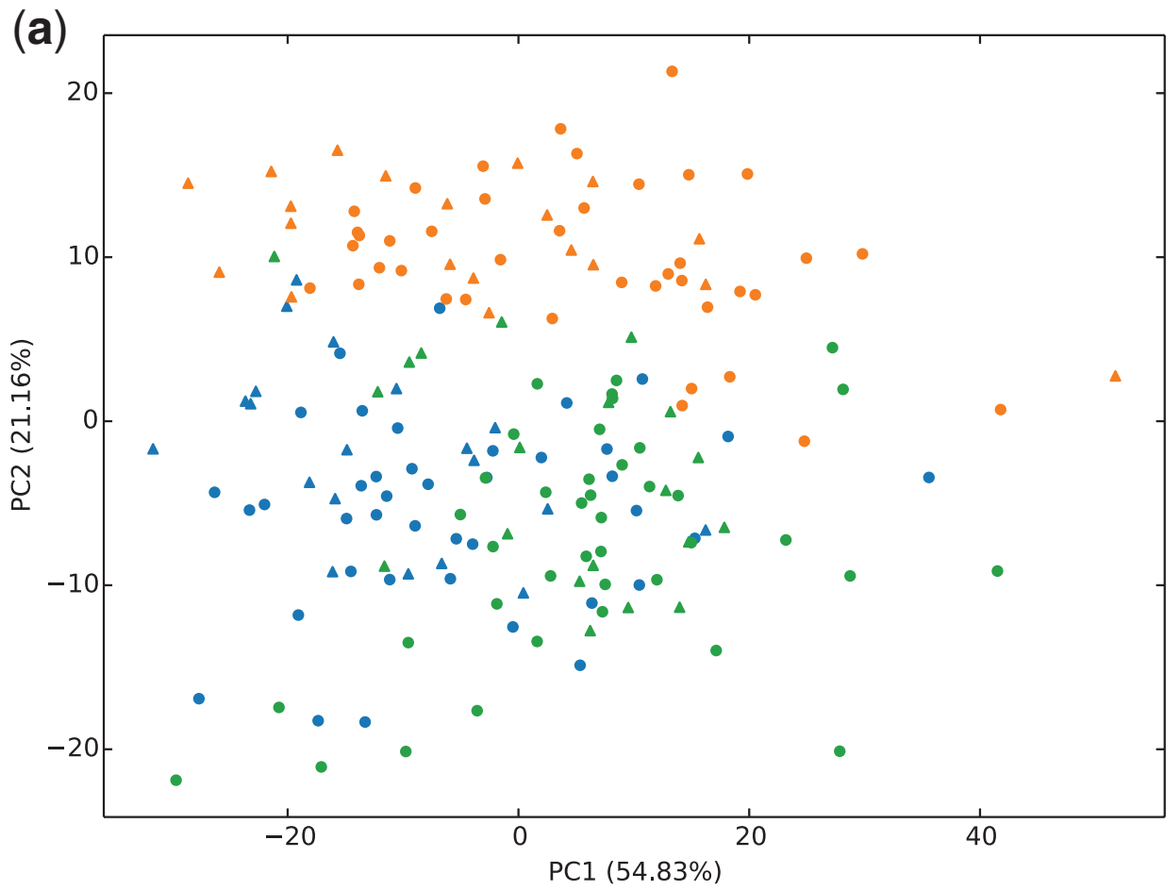
Alice in Wonderland



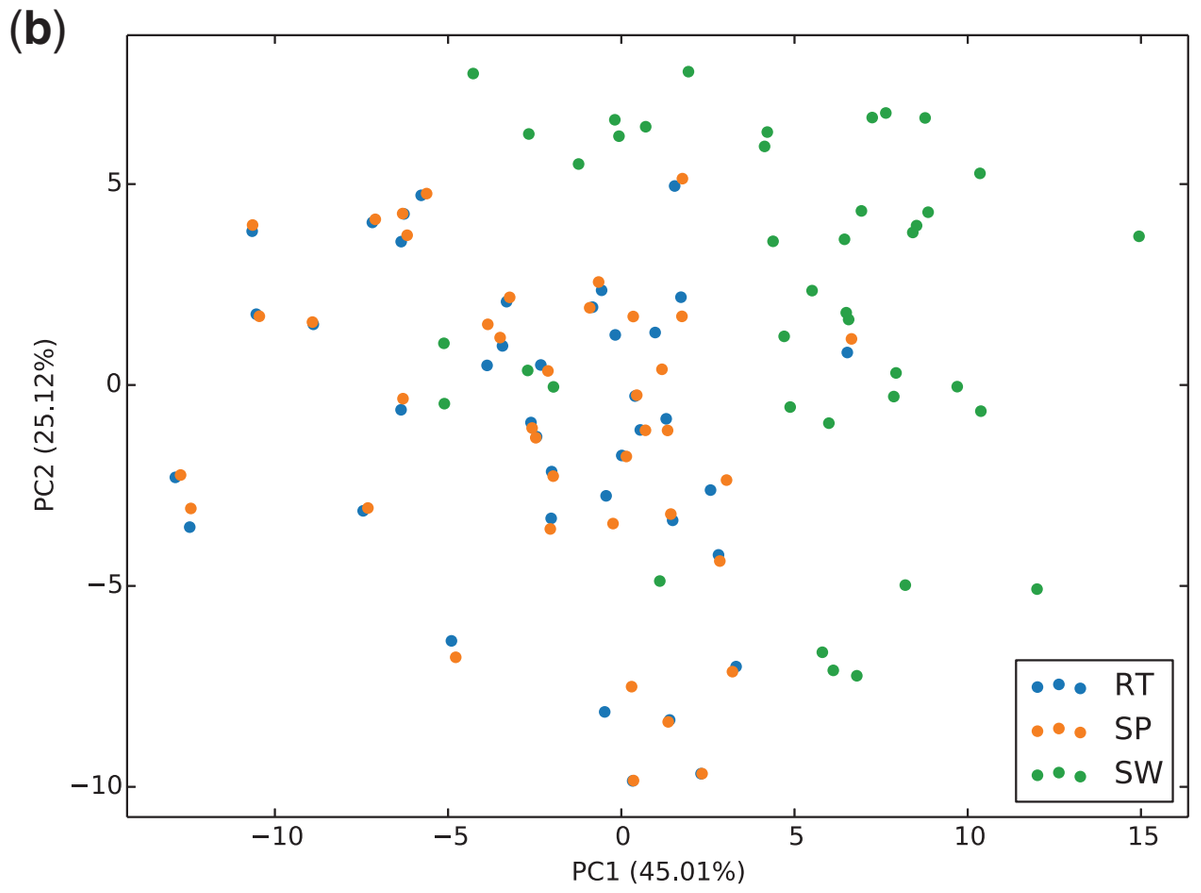


Mesoscopic representation of texts as complex networks
H. F. de Arruda, F. N. Silva, V. Q. Marinho, D. R. Amancio, L. da F. Costa
https://arxiv.org/abs/1606.09636

Classification using mesoscopic representation



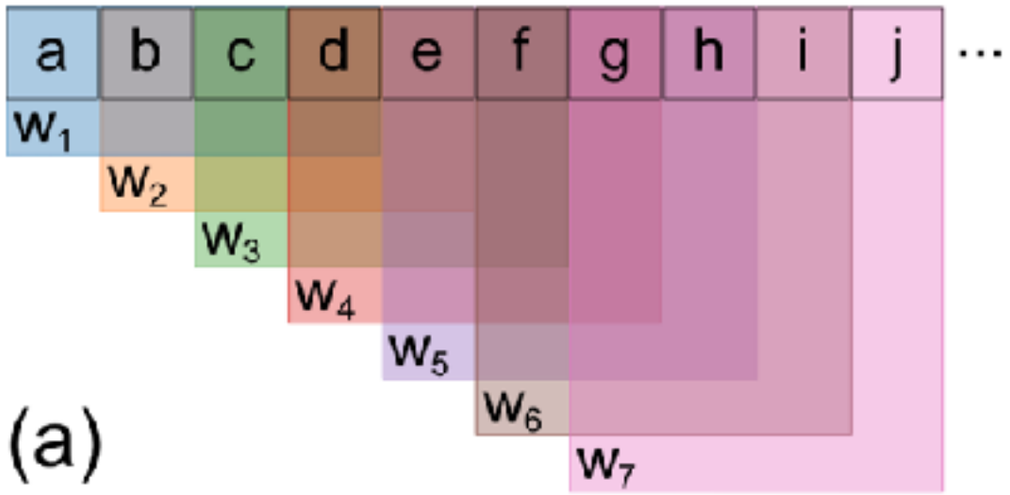
Mesoscopic networks.



Co-occurrence networks.

Mesoscopic representation of texts as complex networks
H. F. de Arruda, F. N. Silva, V. Q. Marinho, D. R. Amancio, L. da F. Costa
https://arxiv.org/abs/1606.09636

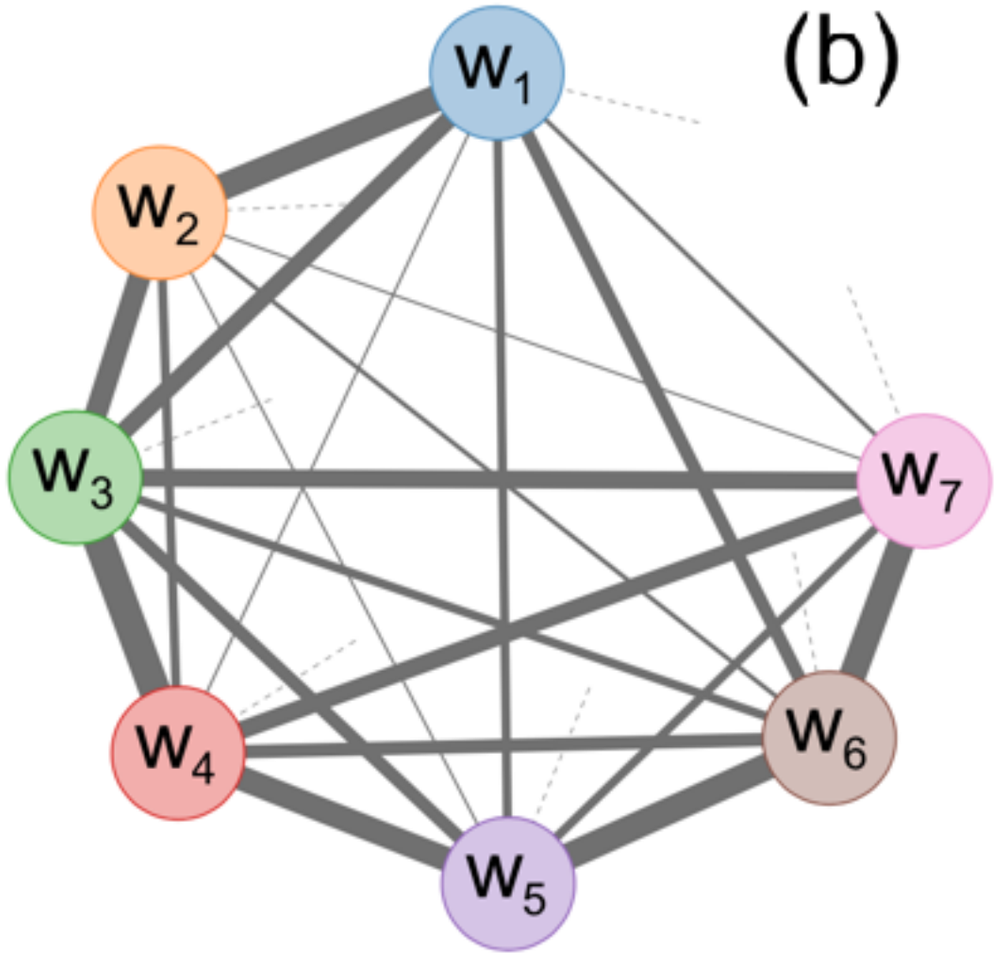
Lyrics



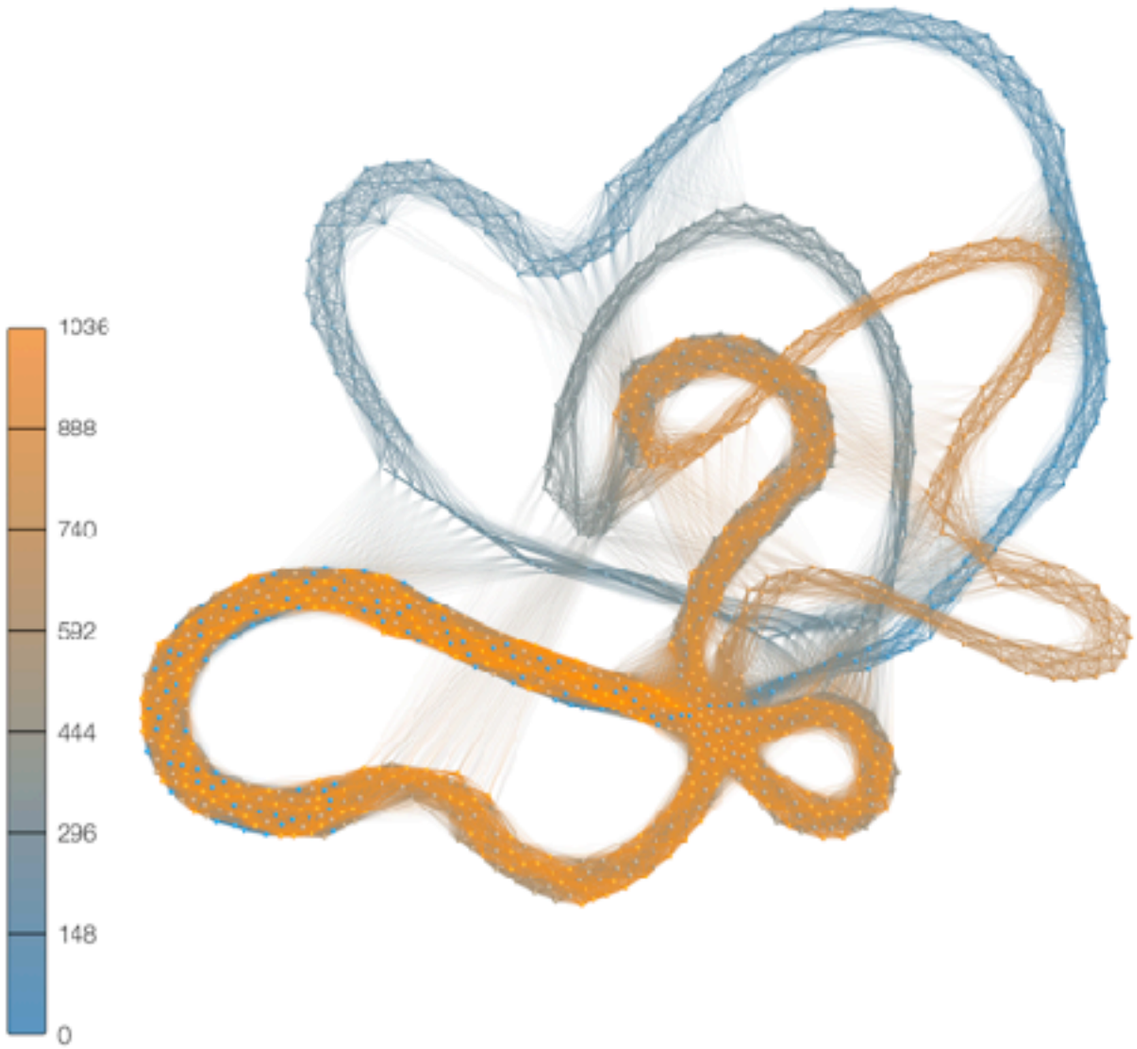
(a)

w_1: "ab", "bc", and "cd"
w_2: "bc", "cd", and "de"
w_3: "cd", "de", and "ef"
w_4: "de", "ef", and "fg"
w_5: "ef", "fg", and "gh"
w_6: "fg", "gh", and "hi"
w_7: "gh", "hi", and "ij"

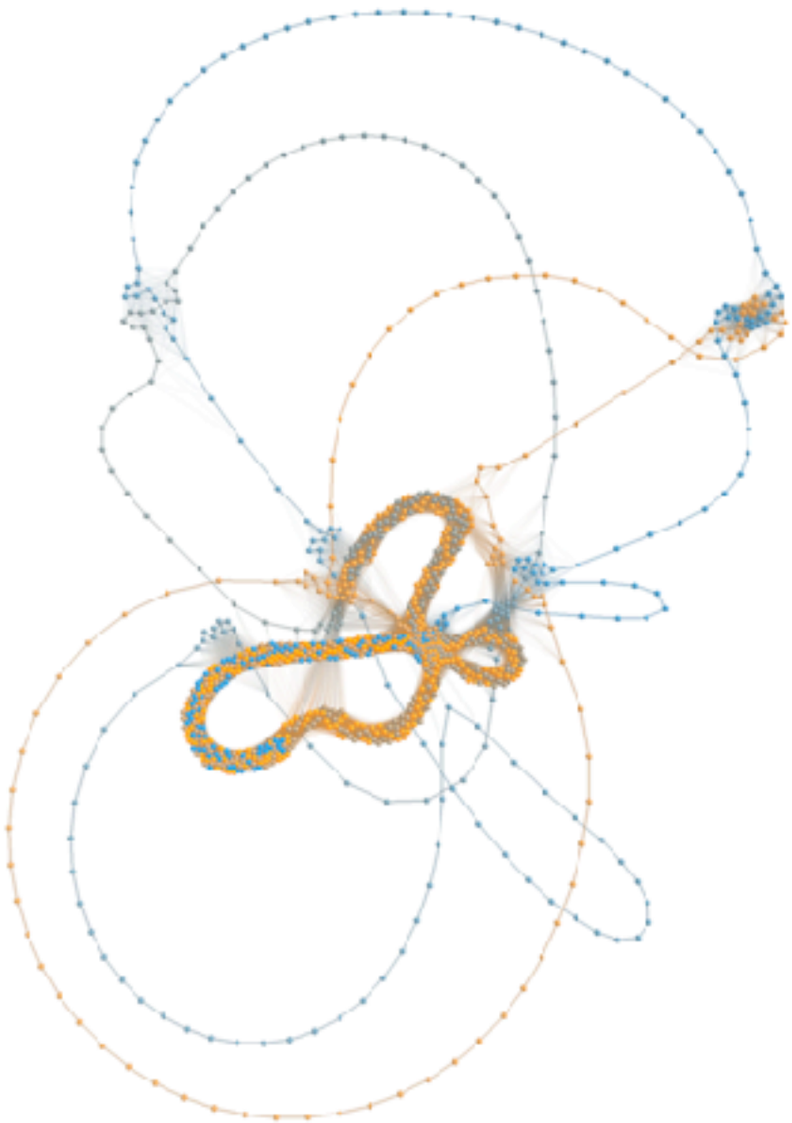
(b)



Lyrics

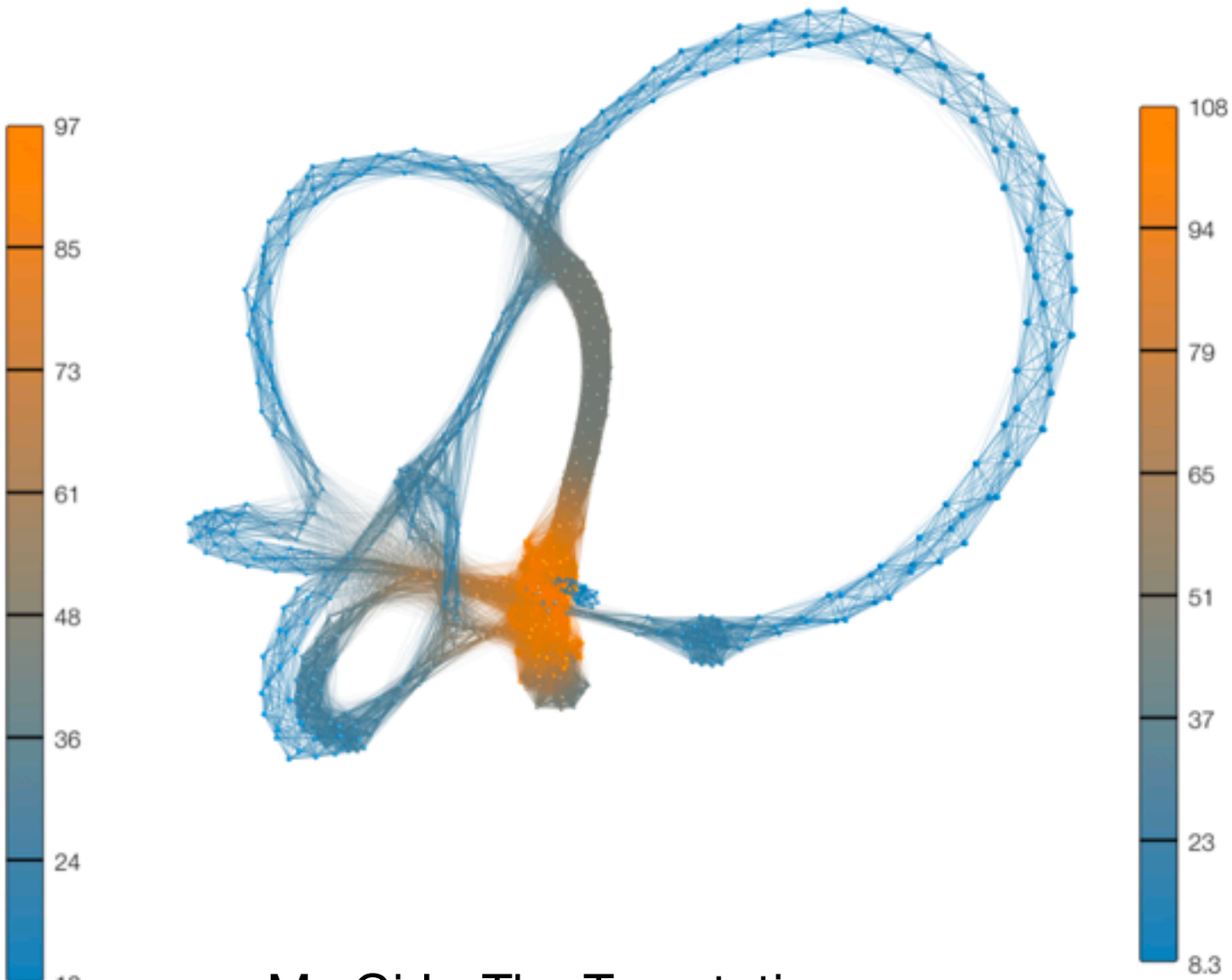


Dancing Queen - ABBA

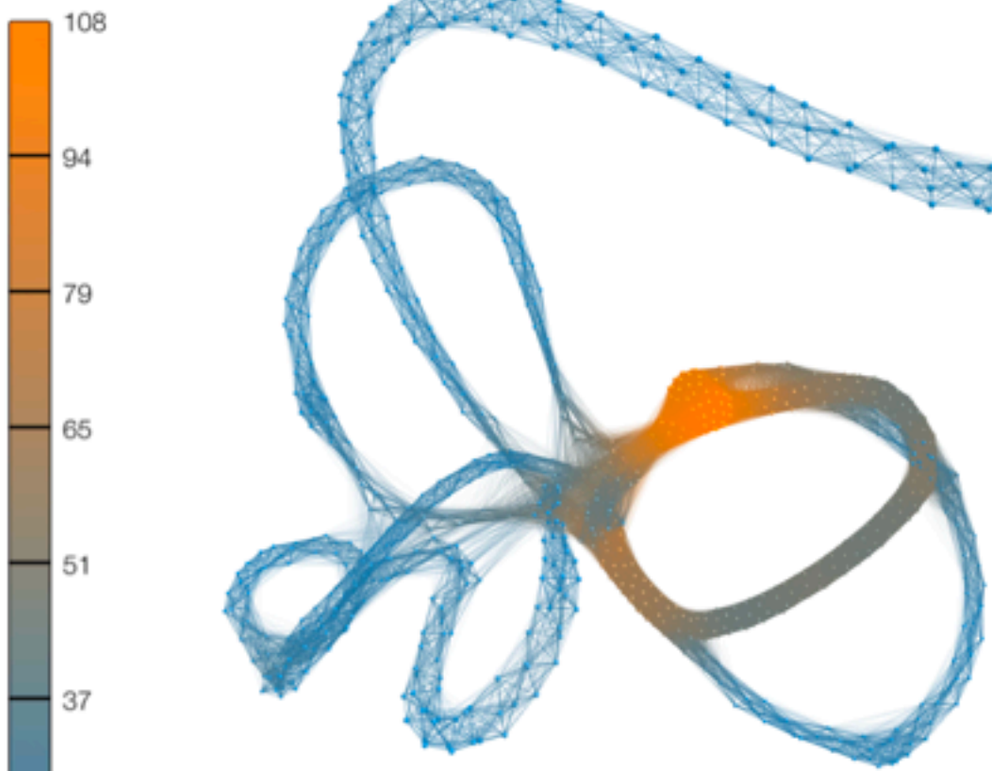


Pruned version

Lyrics



My Girl - The Temptations



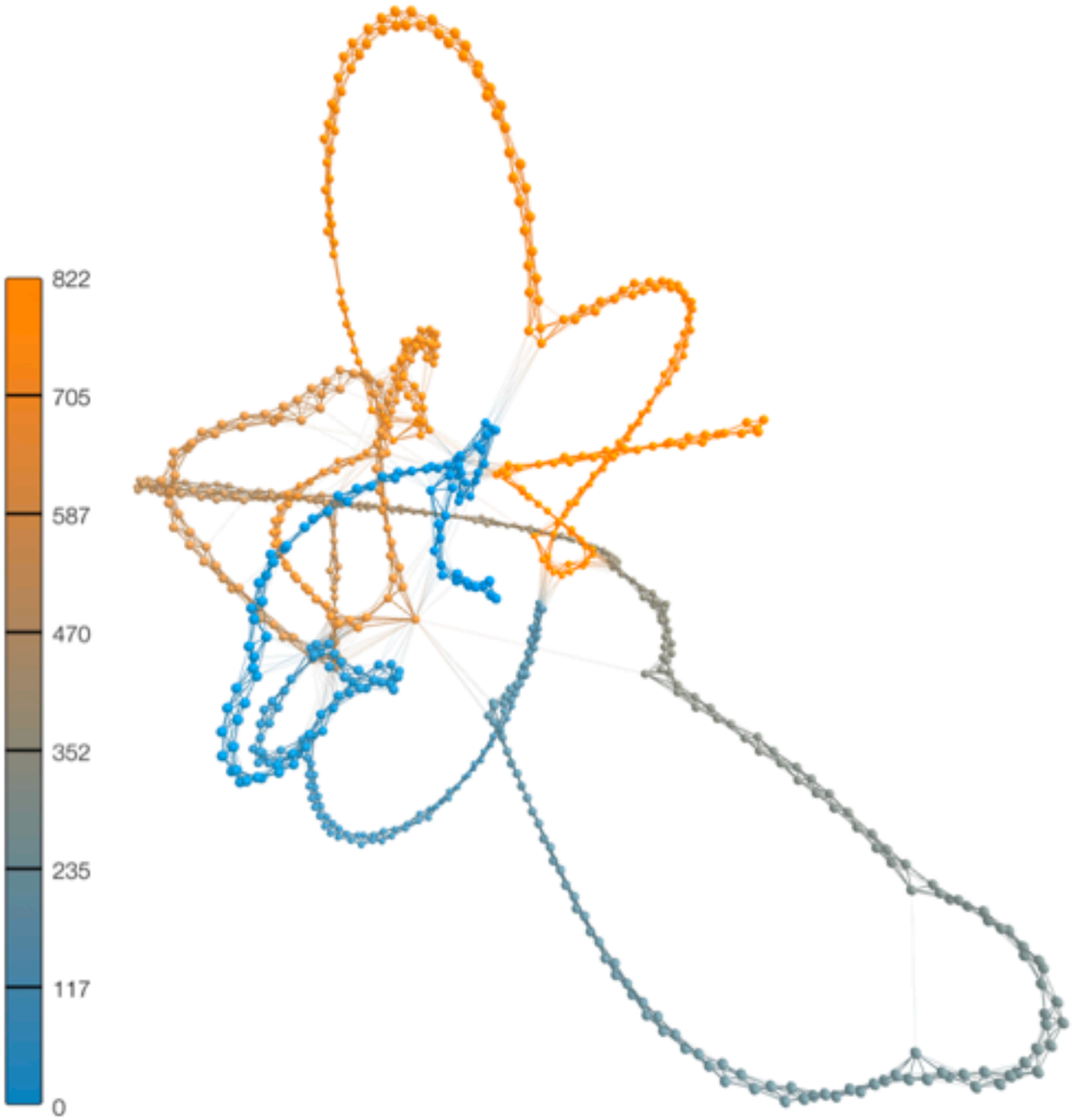
Footloose - Kenny Loggins

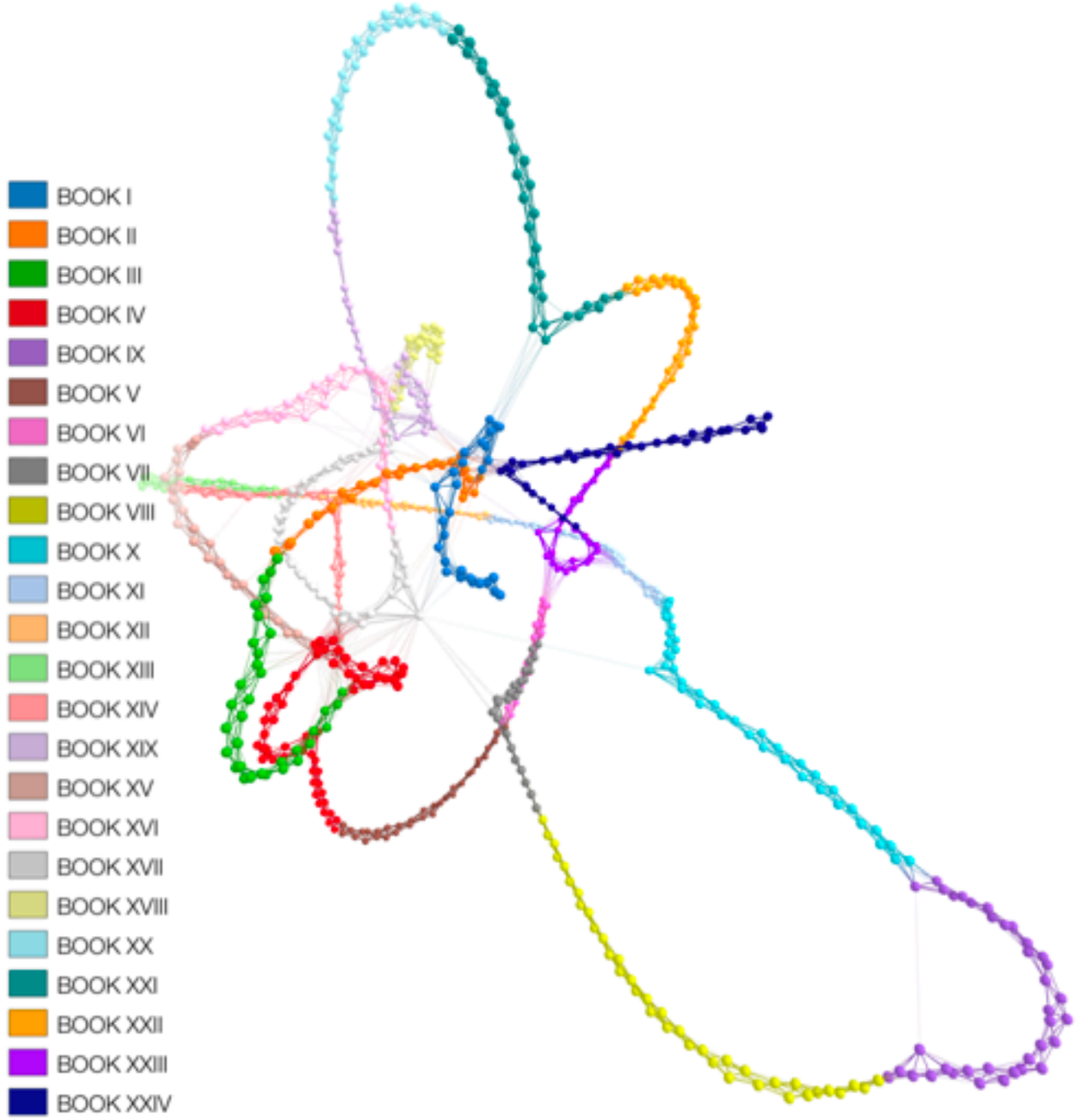
Lyrics

Roar - Kate Perry

Baby - Justin Bieber

Odyssey


822
705
587
470
352
235
117
0


BOOK I
BOOK II
BOOK III
BOOK IV
BOOK IX
BOOK V
BOOK VI
BOOK VII
BOOK VIII
BOOK X
BOOK XI
BOOK XII
BOOK XIII
BOOK XIV
BOOK XIX
BOOK XV
BOOK XVI
BOOK XVII
BOOK XVIII
BOOK XX
BOOK XXI
BOOK XXII
BOOK XXIII
BOOK XXIV

Demo

Is it fake? Is it a narrative?