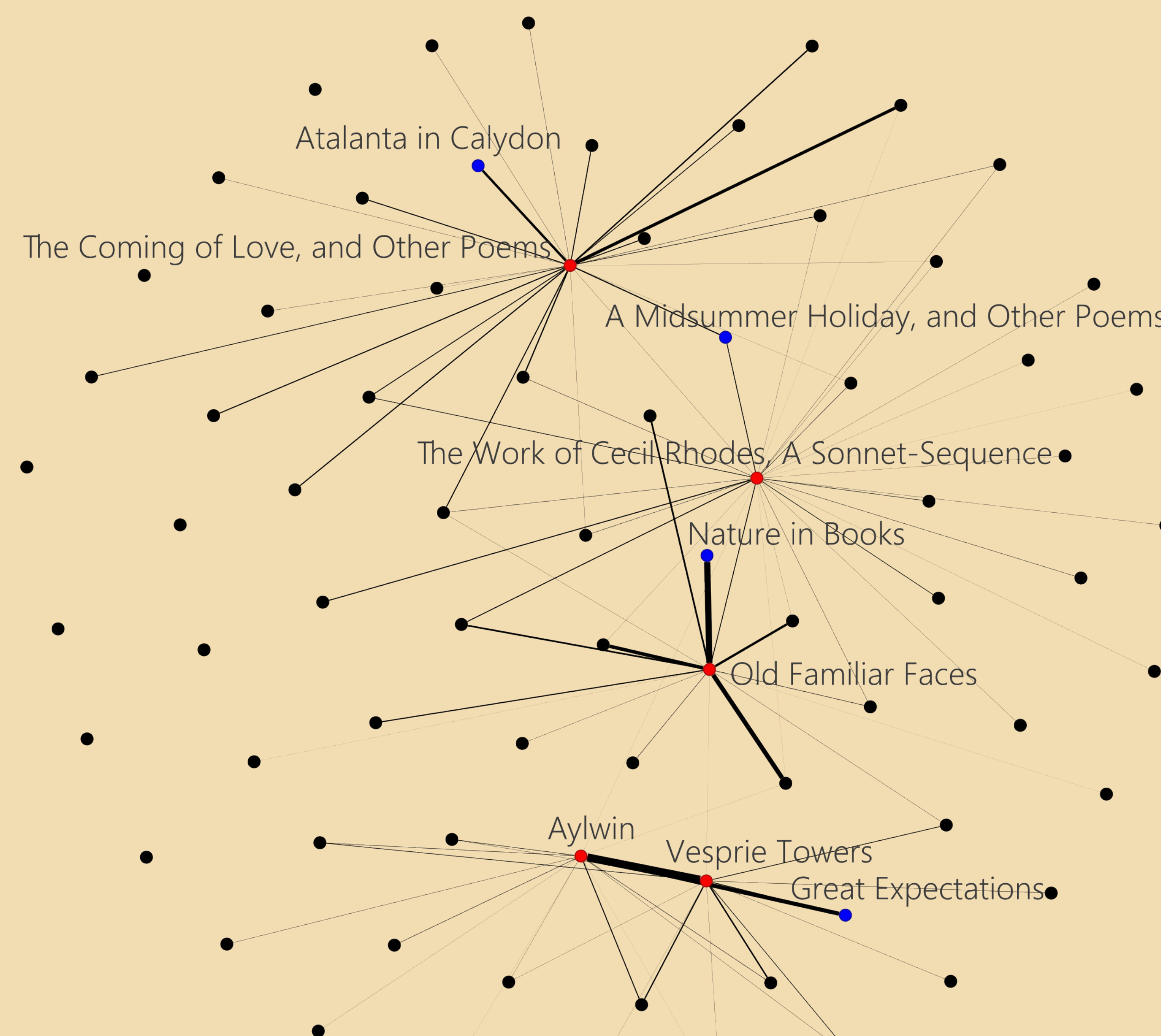


Old Familiar Faces

The Library & Works of Walter Theodore Watts-Dunton



A network graph showing each of **Watts-Dunton's** works, along with significant connections to each other and **gifted volumes**.

I. BACKGROUND

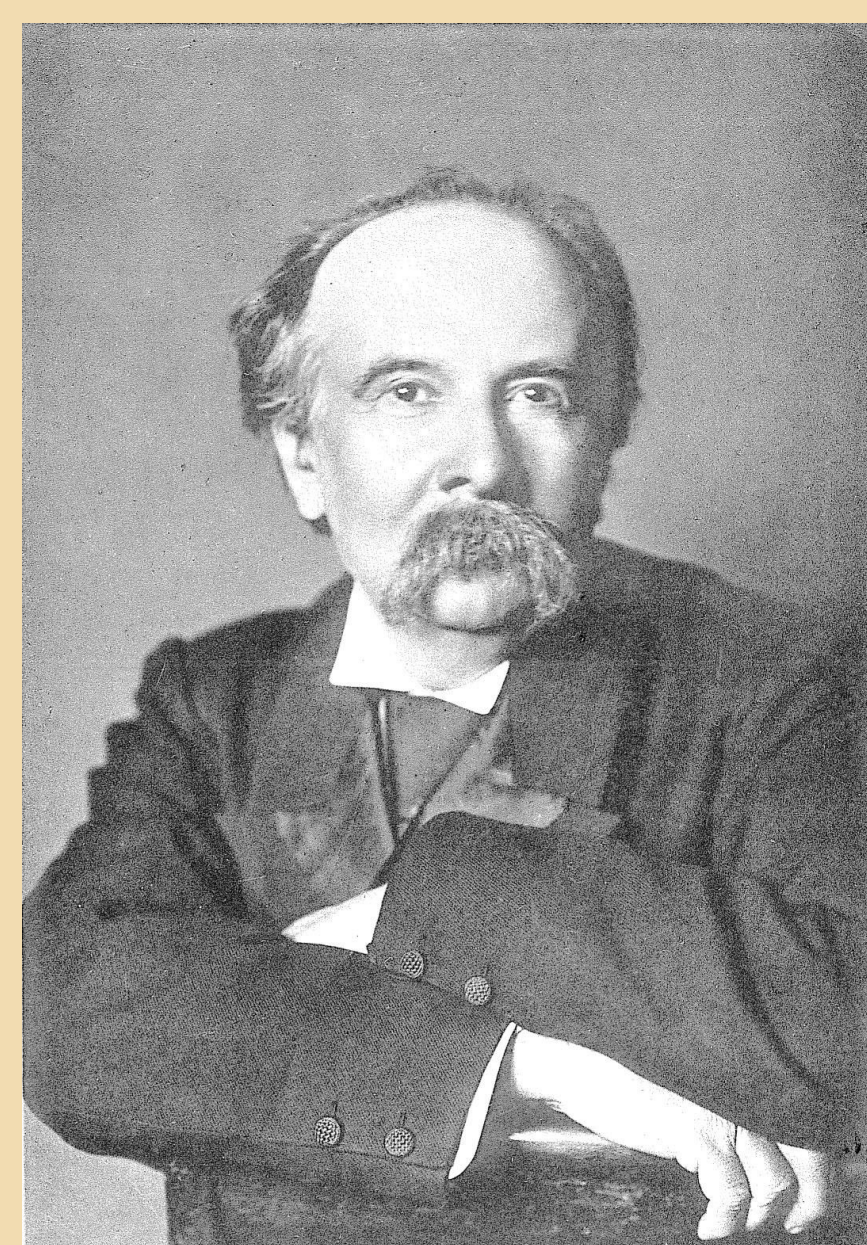
stylometry (sty-lom-ē-try ; stī-lām-ē-trē)
The study of... an author's work based especially on the **recurrence of particular turns of expression or trends of thought**.

Victorian writer Walter Theodore Watts-Dunton received nearly one hundred volumes from his confidant and literary compatriot Algernon Charles Swinburne over their decades-long friendship.

This project uses stylometric analysis to examine Watts-Dunton's body of work and compare it against these gifted volumes.



Swinburne



Watts-Dunton

II. RESEARCH QUESTIONS

RQ1: Which gifted volumes in Watts-Dunton's library have stylistic characteristics that match up with Watts-Dunton's own stylistic characteristics, and what specifically are the commonalities between matching pairs?

RQ2: Based on the inscriptions and dates in gifted volumes, can we determine that some gifted volumes may have influenced written works?

III. TOOLS & METHODS

- **stylo** - this R package is used to perform stylometric analysis of volumes and compare volumes to each other.
 - It has two dimensionality reduction methods and ten distance measures, for 20 modes of analysis.
 - All of these were performed on the full dataset with the top 5000 unigrams, bigrams, and trigrams.
- **Voyant** - this web-based suite of small-scale text analysis tools was used to compare pairs of volumes with high similarity, as determined by the stylo analysis.
 - The **TermsBerry** tool (pictured right) was the most helpful in visualizing which words were present in both documents.
 - The **Phrases** tool was used to manually find common trigrams between the two works.



Voyant's TermsBerry tool with the top 100 terms in 'Atalanta in Calydon' and 'The Coming of Love' shown (excluding stopwords). Grey circles indicate terms in both volumes, while white circles are terms unique to just one.

IV. FINDINGS

Four significant pairs of volumes were identified, but three were eliminated due to irrelevance to the research questions.

- *Nature in Books* and *Old Familiar Faces* are stylistically similar because they discuss the same historical subject matter.
- *Aylwin* and *Great Expectations* were marked as similar, but only due to the presence of common English stopwords.
- *A Midsummer Holiday* and *The Work of Cecil Rhodes* were similar only because of the terms used by catalogs in the back pages.

However, one promising pair was identified: **Swinburne's *Atalanta in Calydon* and Watts-Dunton's *The Coming of Love, and Other Poems***. 400 of the top 500 terms across these two works were present in both of them, as well as 90 of the top 100, pictured lower-middle.

Trigrams found in both volumes include the following:

- all the world
- as thou art
- in the air / dark / world
- mine own heart
- of the gods / wind
- than the sea

In addition to the stylometric similarities, consider the following:

- *Atalanta in Calydon* was written by Swinburne.
- It was gifted to Watts-Dunton on August 8th, 1894, and *The Coming of Love* was published in 1897.
- *The Coming of Love* was also Watts-Dunton's first published work, though some of the poems had appeared in periodicals prior to that year.

V. CONCLUSIONS & FUTURE RESEARCH

While it is difficult to definitively say that *Atalanta in Calydon* influenced some of the poems in *The Coming of Love*, there is significant stylometric similarity and other outside factors which suggest a strong connection between the two volumes, independent of the connective qualities of the false positive pairs that the project also identified.

Some limitations faced by this project:

- Watts-Dunton's library consisted of thousands of volumes. While the volumes in this project may have been emotionally significant to him, it's still a small fraction of the works he owned and presumably read.
- The stylometric analysis is an aggregation of many different measures, taking a shotgun approach despite the fact that some measures may not be appropriate to the data.

Some remedies and ideas for future research:

- I'm currently working on a structured XML file of Watts-Dunton's entire library, which could be used to collect and analyze the full collection rather than a subset.
- A new stylometric approach may be taken, where the best possible distance measure is determined, and then the parameters of that individual measure are tweaked.