# Text Analysis of Ethical Influence in Bioinformatics and its Related Disciplines

**Oliver Bonham-Carter, Ph.D.**
Dept of Computer and Information Science
Meadville, PA 16335
https://cis.allegheny.edu/

FICC 2024
$4^{th}$-$5^{th}$ April 2024
Berlin, Germany

# Ethics and Responsibility

## Terms

Two terms to differentiate ...

## What is **ethically**-influenced research?

- Research that has been completed with governance of moral principles which applies a standard of accepted methodology

## What is **responsible** research?

- Accountable research that where the liability of wrong-doing or blame has been defined

In this work ...

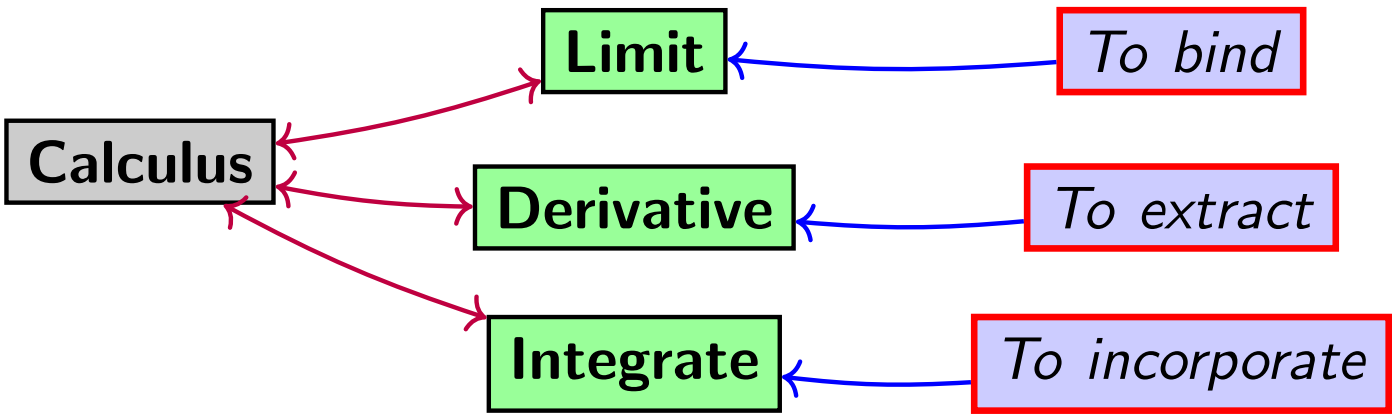
## Main Motivations

- Determine the landscape of ethical and responsible thinking in bioinformatics
- To describe visually a landscape of shared ethical ideas across *Bioinformatics*, in addition to some of its related subjects: *Biology*, *Computer Science* and *Mathematics*

# Specifications

- **Corpus**
  - Published scientific articles
- **Supervised text analysis**
  - Pre-selected keywords for a bag-of-words approach
- **Abstract-centric**
  - Carefully crafted text by authors: abstract contains about 250 relevant words to study
  - Prominent articles contain terms of ethical conduct
- **Discipline-centric**
  - Articles contain terms of subject areas: *Bioinformatics*, *Biology*, and others

# Keywords




- By inspection, we determined the ethical and responsible language (keywords) according to studied disciplines (i.e., *Biology*, *Computer Science*, and *Mathematics*)
- Choose **generic** keywords as meanings vary across diverse disciplines

# Selected Generic Keywords

## Selected Keywords

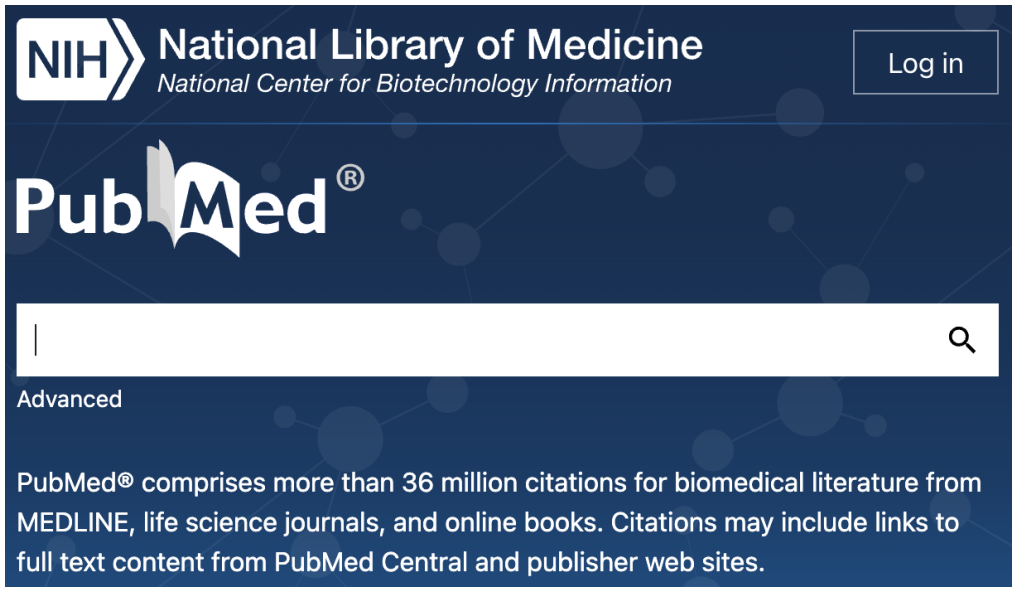
### Chosen Keywords

| | |
|---|---|
| analysis | informatics |
| analytical | informed consent |
| bioinformatics | liability |
| biology | mathematics |
| code of ethics | responsibility |
| computer science | responsible |
| ethic | stem |
| ethical | trust |
| ethics | whistle-blowing |
| general science | |

- Generic Keywords: *discipline-neutral* definitions

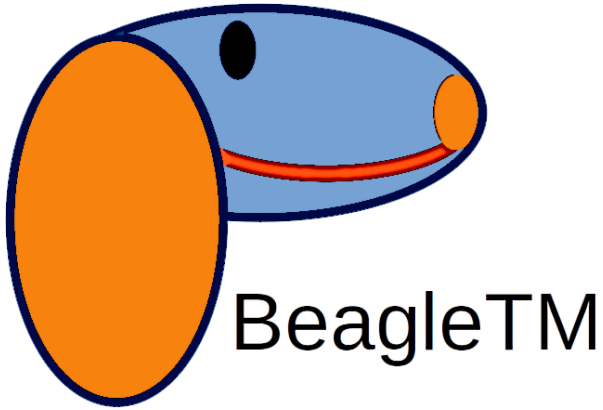
# Corpus

## Non-commercial article data



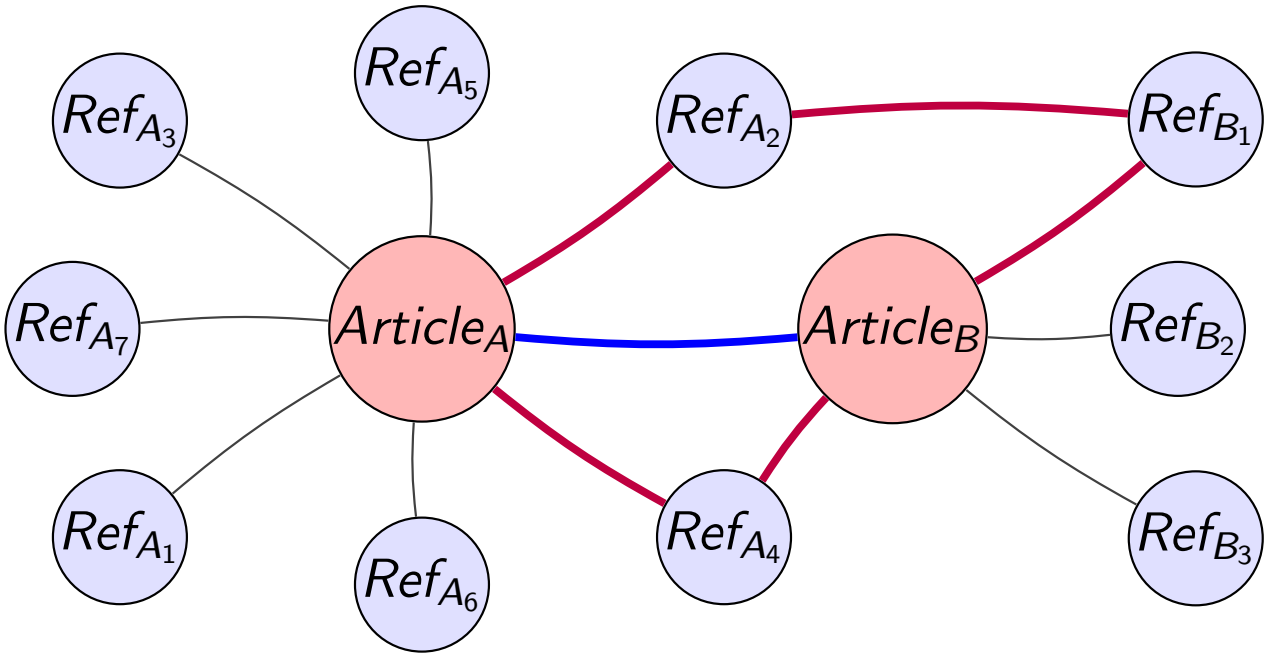

- PubMed: Millions of articles
- Our study: Over 1.3 Million articles used as related to Bioinformatics, or its related subjects
- `https://ftp.ncbi.nlm.nih.gov/pub/pmc/oa_bulk/oa_noncomm/`

# BeagleTM2
## Text Analysis




- Parsing corpus for specific keywords pertaining to ethical identities
- Create Relationship Networks
- Network connectivity describes magnitude of ethical idea sharing
- https://github.com/developmentAC/BeagleTM2

# Relationship Networks




- Relationship Networks to determine connectivity
- Articles $A$ and $B$ exhibit shared information, directly and indirectly
- Connected articles have some common language between them

# Articles Containing SINGLE Keywords

Overview: Keywords parsed across 1,380,596 articles

**Keywords, Articles and Proportions**

| | | | | | |
|---|---|---|---|---|---|
| informatics | 6931 | 0.005 | informed consent | 2128 | 0.002 |
| analytical | 5532 | 0.004 | ethical | 5281 | 0.004 |
| ethic | 10810 | 0.008 | mathematics | 296 | 0 |
| stem | 21023 | 0.015 | responsibility | 1498 | 0.001 |
| liability | 303 | 0.0002 | whistle-blowing | 1 | 0 |
| analysis | 168036 | 0.122 | general science | 6 | 0 |
| bioinformatics | 6305 | 0.005 | trust | 1868 | 0.001 |
| ethics | 6366 | 0.005 | computer science | 74 | 0 |
| responsible | 16093 | 0.012 | biology | 6777 | 0.005 |
| code of ethics | 22 | 0 | | | |

- Number of articles in which the keywords pairs were found
- Word usage directed types of Relationship Networks to study

# PAIRED Simultaneously Occurring Keywords (I)

## Pairs of Keywords parsed across 1,380,596 articles

| Keywords, Articles | | | |
|---|---|---|---|
| ethic, ethics | 6366 | analysis, responsible | 2732 |
| bioinformatics, informatics | 6305 | analysis, ethic | 2328 |
| ethic, ethical | 5281 | analysis, analytical | 1722 |
| analysis, informatics | 3384 | analysis, ethics | 1430 |
| analysis, stem | 3270 | analysis, biology | 1220 |
| analysis, bioinformatics | 3252 | analysis, ethical | 1084 |

- Number of articles in which the keywords pairs were found simultaneously
- A non-exhaustive list ranking pairs of keywords acoss corpus.

# PAIRED Simultaneously Occurring Keywords (II)

Non-exhaustive pairs across 1,380,596 articles

| Keywords and Articles | | | |
|---|---|---|---|
| analytical, ethic | 98 | informed_consent, responsible | 39 |
| responsibility, responsible | 96 | responsible, trust | 39 |
| ethic, trust | 95 | **ethical, informatics** | **32** |
| mathematics, stem | 87 | **biology, ethics** | **27** |
| analytical, informatics | 83 | analytical, responsibility | 24 |
| analytical, responsible | 78 | analytical, trust | 23 |
| analytical, bioinformatics | **73** | code_of_ethics, ethics | 22 |
| ethical, trust | 63 | code_of_ethics, ethic | 22 |
| **biology, ethic** | **63** | informed_consent, stem | 20 |
| analytical, biology | 62 | informatics, trust | 20 |
| biology, mathematics | 61 | informed_consent, trust | 19 |

# Relationship Networks: PAIRS

*Responsible* and *Bioinformatics*



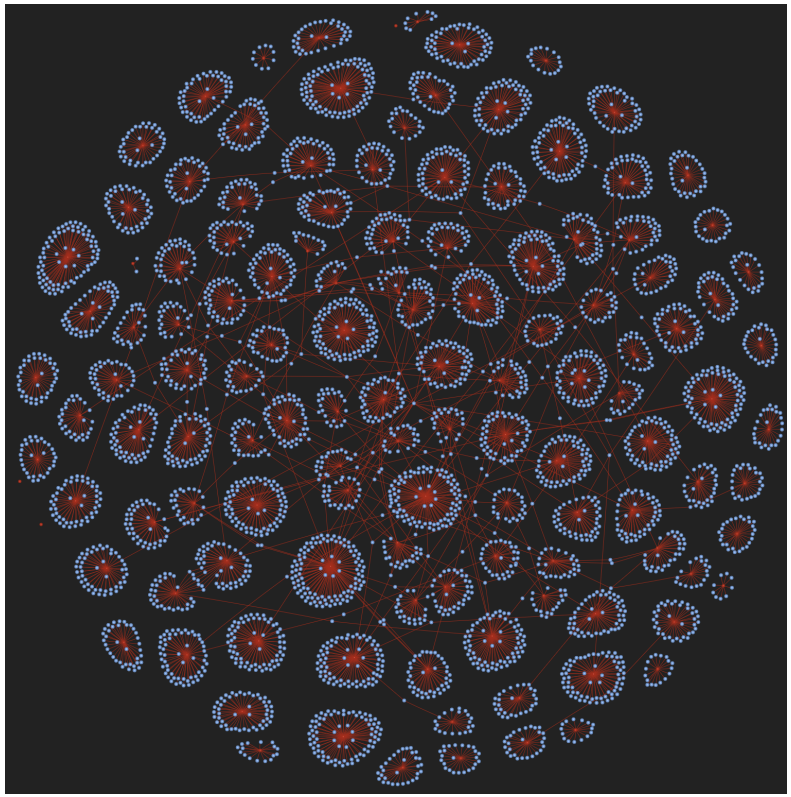


Figure: A *Relationship Network* showing common language of articles (red nodes) and their supporting references (blue nodes). High connectivity – there are many supporting articles sharing these terms.

# Relationship Networks: PAIRS

*Responsible* and *Biology*



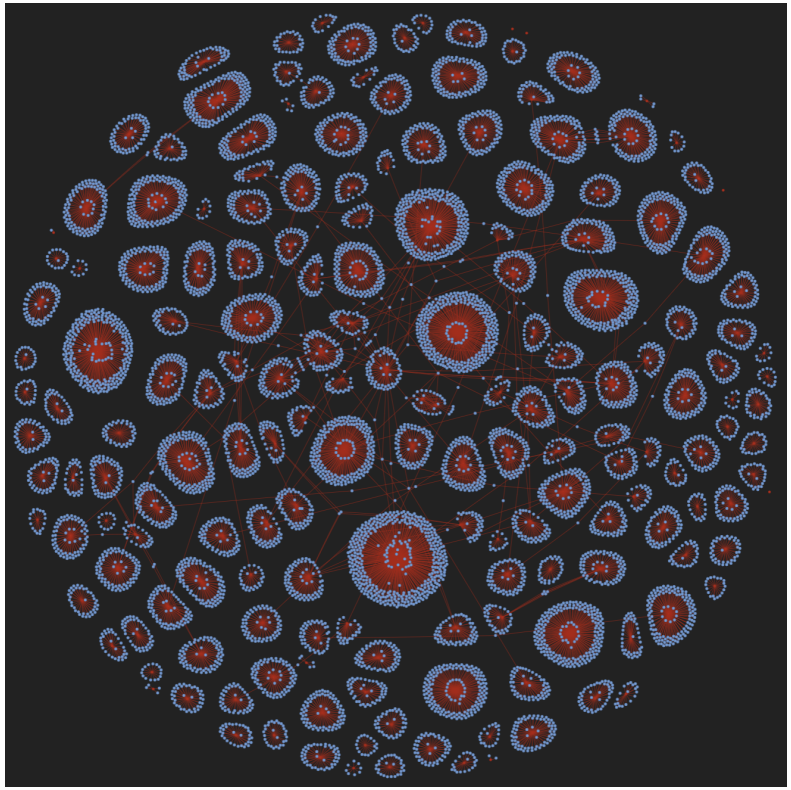


Figure: A *Relationship Network* showing common language of articles (red nodes) and their supporting references (blue nodes). Low to Medium connectivity – implying articles are not often connected by references containing the same types of terms.

# Summary

- Ideas in research may be signalled by the use of specific keywords (ex: relating to *ethics* and *responsibility*)
- Relationship Networks are created by keywords from the corpus
- Connections indicate idea sharing: increasing connections may imply increasing interest

## Final Thoughts

- Not all disciplines we checked had the same amount of connections between nodes in Relationship Networks.
- Idea sharing appears inconsistent across research areas

# Reach-Out with Questions, Comments or *Whatever*!


## Questions and comments are welcome!


- obonhamcarter@allegheny.edu
- https://www.oliverbonhamcarter.com/

- https://github.com/developmentAC/BeagleTM2