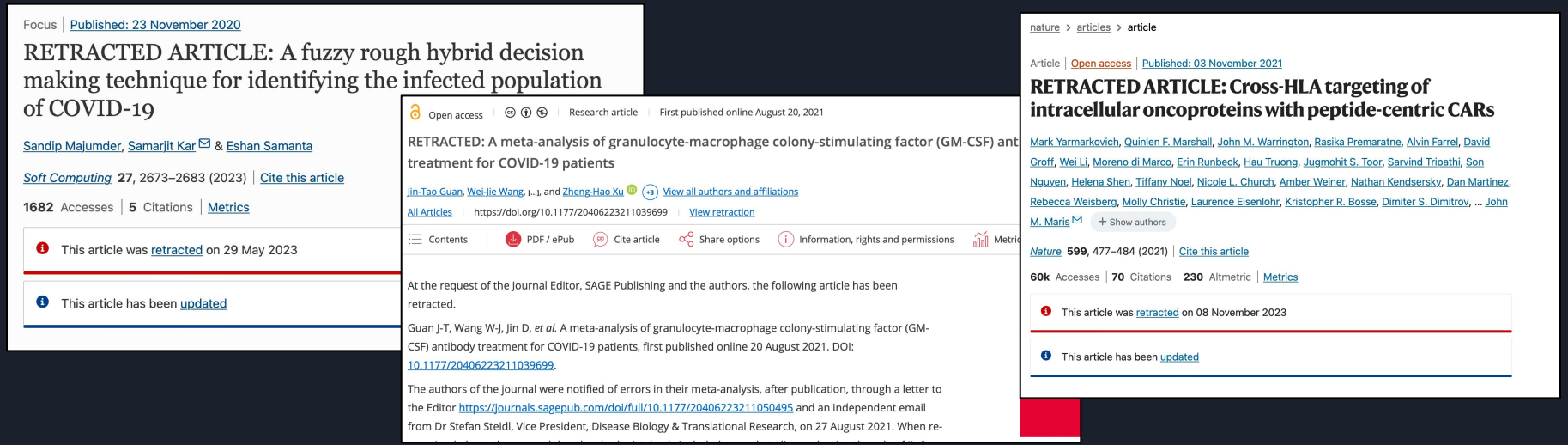
Simulating the Diffusion of Retracted Research

By Joe Menke

Background

- In scientific research, publications are sometimes retracted.
- What does this look like?
  - Usually a flag (or some other indicator) is added to the original article and a notice of a retraction is published



Focus | Published: 23 November 2020

RETRACTED ARTICLE: A fuzzy rough hybrid decision making technique for identifying the infected population of COVID-19

Sandip Majumder, Samarjit Kar & Eshan Samanta

Soft Computing **27**, 2673–2683 (2023) | Cite this article

1682 Accesses | **5** Citations | Metrics

This article was retracted on 29 May 2023

This article has been updated

Open access | Research article | First published online August 20, 2021

RETRACTED: A meta-analysis of granulocyte-macrophage colony-stimulating factor (GM-CSF) ant treatment for COVID-19 patients

Jin-Tao Guan, Wei-Jie Wang, (...), and Zheng-Hao Xu +3 View all authors and affiliations

All Articles | https://doi.org/10.1177/20406223211039699 | View retraction

Contents | PDF / ePub | Cite article | Share options | Information, rights and permissions | Metric

At the request of the Journal Editor, SAGE Publishing and the authors, the following article has been retracted.

Guan J-T, Wang W-J, Jin D, *et al.* A meta-analysis of granulocyte-macrophage colony-stimulating factor (GM-CSF) antibody treatment for COVID-19 patients, first published online 20 August 2021. DOI: 10.1177/20406223211039699.

The authors of the journal were notified of errors in their meta-analysis, after publication, through a letter to the Editor https://journals.sagepub.com/doi/full/10.1177/20406223211050495 and an independent email from Dr Stefan Steidl, Vice President, Disease Biology & Translational Research, on 27 August 2021. When re-

nature > articles > article

Article | Open access | Published: 03 November 2021

RETRACTED ARTICLE: Cross-HLA targeting of intracellular oncoproteins with peptide-centric CARs

Mark Yarmarkovich, Quinlen F. Marshall, John M. Warrington, Rasika Premaratne, Alvin Farrel, David Groff, Wei Li, Moreno di Marco, Erin Runbeck, Hau Truong, Jugmohit S. Toor, Sarvind Tripathi, Son Nguyen, Helena Shen, Tiffany Noel, Nicole L. Church, Amber Weiner, Nathan Kendsersky, Dan Martinez, Rebecca Weisberg, Molly Christie, Laurence Eisenlohr, Kristopher R. Bosse, Dimiter S. Dimitrov, ... John M. Maris + Show authors

Nature **599**, 477–484 (2021) | Cite this article

60k Accesses | **70** Citations | **230** Altmetric | Metrics

This article was retracted on 08 November 2023

This article has been updated

Background

- What does it mean to be retracted?
  - In general, something was found that casts significant doubt over the reported data/findings.
- More specifically, why are articles retracted?
  - Scientific misconduct
  - Errors that impact findings
    - Unintentional - e.g., bug in code or miscalibrated machine
    - Fraudulent - e.g., making up data
    - For small errors - typos, etc. - corrections are issued.
      - Corrections ≠ Retractions
  - Plagiarize previous works
  - Violating ethical guidelines
- Retracted research articles should be avoided unless you are intimately familiar with the reasons for retraction (or your work focuses on retractions).

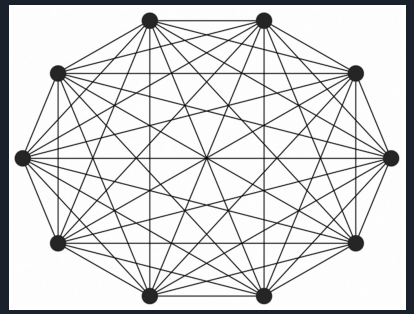
Background

- How often do retractions occur? (Statistics from Brainard and You, 2018)
  - Relatively rare
    - ~ 4 out of every 10,000 papers
  - More retractions today than in the past - more policing, not more fraud
    - 100 per year (<2000)
    - 1000 per year (2014)
- In scientific research, articles cite previously published articles forming citation networks.
  - How often are retractions cited?
    - Non-retracted articles
      - Field dependent
    - Retracted articles
      - No difference in citations before/after retraction (Bordignon, 2020; Hsiao & Schneider, 2021)
        - Only 5.4% of these citations reference the retraction (Hsiao & Schneider, 2021)
- Why do retractions matter?
  - They can spread false information and mislead future researchers.
  - For example, linking MMR vaccine to autism (Wakefield et al., 1998)

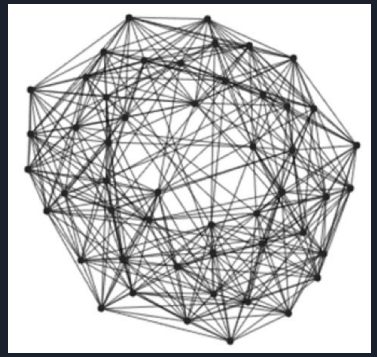
Prior Work - The Dynamics of Retraction in Epistemic Networks

- Models researchers within social networks, spreading ideas amongst each other
- Evaluated a variety of network types:
  1. Connected
  2. Small world (Watts and Strogatz, 1998)
  3. Preferential-attachment networks (Barabási and Albert, 1999)

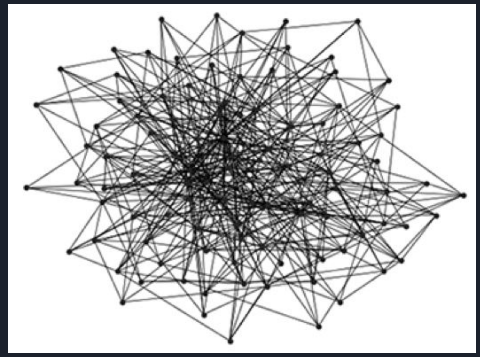
1.



2.



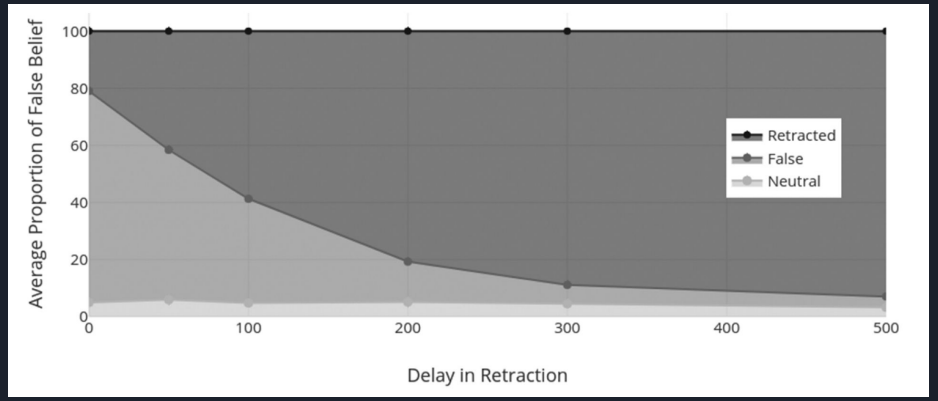
3.

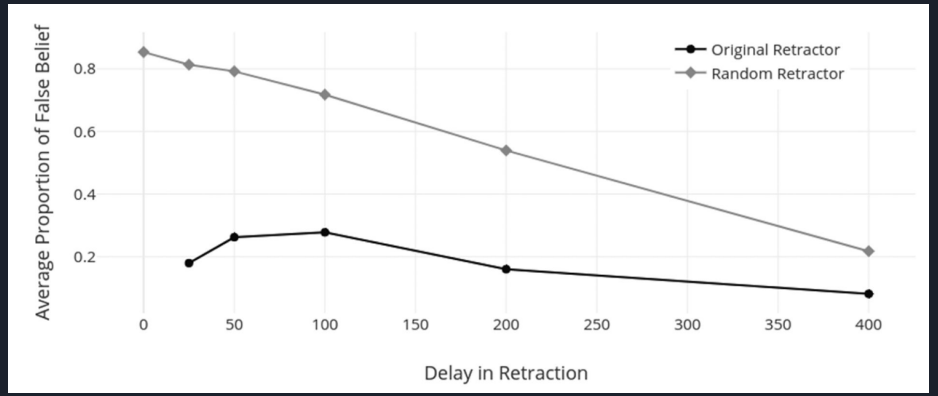


Prior Work - The Dynamics of Retraction in Epistemic Networks

Major Findings:

- False information often perseveres despite retractions.
- Delaying retraction can increase its effectiveness.
- Retractions are most successful when issued by the original source.





Enhancements

- Overall: Adapt model from researchers within social networks to research articles within citation network
- Code Quality
  - Improved documentation - docstrings + type hinting
  - Adding doctests
  - Rewriting main.py
- Modeling Changes
  - More realistic transmission rates
  - In progress - custom network

Enhancements

Improved documentation

```python
class PopAgent(Agent):
    """An Agent with some initial knowledge."""

    def __init__(self, unique_id, model, neighbors,
        super().__init__(unique_id, model)

        # Default params
        self.belief = Belief.Neutral
        self.neighbors = neighbors  # list of agent
        self.clock = 0  # internal timer (absolute
        self.beliefTime = 0  # time current belief
        self.shareTime = sharetime   # time limit
```

```python
if TYPE_CHECKING:
    # To prevent cyclic dependency...
    from model import AcademicLiterature


class Article(Agent):
    """An Agent with some initial information."""
    def __init__(self, unique_id: int, model: AcademicLiterature, neighbors: list, share_time: float):
        """
        Initialize an Article.
        :param unique_id: an integer unique to a single agent within the predefined model
        :param model: the model in which the agent will exist
        :param neighbors: a list of neighboring agents represented as nodes in a NetworkX graph
        :param share_time: a float indicating the time limit within which new beliefs are shared; np.inf if endless
        """
        super().__init__(unique_id, model)

        # Default params
        self.belief = Belief.Neutral
        self.neighbors = neighbors
        self.clock = 0  # internal timer (absolute time)
        self.belief_time = 0  # time current belief has been held
        self.share_time = share_time
        self.delay = model.delay  # time to delay before introducing retraction
```

Added:

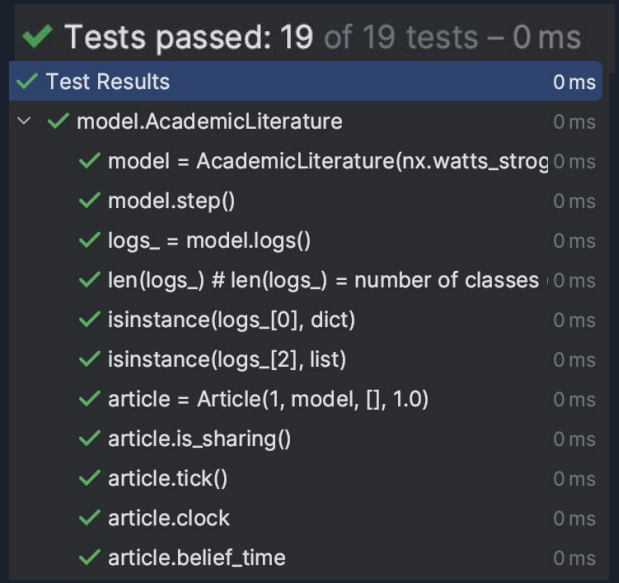
- Improved docstrings w/ argument descriptions
- Type hinting
- Changed names

Enhancements

Adding doctests

- Implemented doctests for various methods and attributes
- Due to cyclic dependencies, all tests were put under the AcademicLiterature model class

```
>>> model = AcademicLiterature(nx.watts_strogatz_graph(n=100, k=8, p=0.1), Mode.Default, 1.0, 0)
>>> model.step()
>>> logs_ = model.logs()
>>> len(logs_)  # len(logs_) = number of classes (neutral, false, retracted)
3
>>> isinstance(logs_[0], dict)
True
>>> isinstance(logs_[2], list)
True
>>> article = Article(1, model, [], 1.0)
>>> article.is_sharing()
True
>>> article.tick()
>>> article.clock
1
>>> article.belief_time
1
>>> article.tick()
>>> article.is_sharing()
False
>>> article_1 = Article(2, model, [], np.inf)
>>> article_1.set_belief(Belief.Fake)
>>> article_2 = Article(2, model, [article_1], np.inf)
>>> article_2.set_belief(Belief.Retracted)
>>> article_1.update(article_2)
>>> article_1.belief
<Belief.Retracted: 2>
```



...

Enhancements

Rewriting main.py

- Added a function to run an experiment
- Utilized if __name__ == '__main__':

Before

```
# Hyperparameters
N = 100      # Number of agents in the ne
T = 1200     # Number of time steps per s
S = 1000      # Number of simulations to

# Agent belief sharing constraints
mode = Mode.Default            # Set age
shareTimeLimit = np.infty      # Time an

# Introducing the retracted belief
delay = 0             # Time delay before
singleSource = False    # Retracted source
samePartition = None    # Retracted source
```

After

```
if __name__ == '__main__':
    # Watts Strogatz Experiment - Small World
    run_experiment(
        num_agents=100,
        num_sims=1000,
        time_steps=1200,
        mode=Mode.Default,
        share_time_limit=np.inf,
        delay=0,
        single_source=False,
        same_partition=None,
        graph=nx.watts_strogatz_graph,
        network_name="SmallWorlds",
        experiment="Watts_Strogatz_Model",
        save=True,
    )
```

Enhancements

Realistic transmission rates

- Articles are still cited even after retraction, but this is not 100% as encoded
- Authors sometimes check to see if an article is retracted before citing
  - 5.4% of citations note the article is retracted
  - 94.6% chance of transmission after retraction is introduced (Hsiao & Schneider, 2021)

```
# Convert self to false belief
if self.belief == Belief.Neutral and other.belief == Belief.Fake and is_sharing_fake:
    # if original info has been retracted, 5.4% chance false belief is not passed on
    # https://direct.mit.edu/qss/article/2/4/1144/107356/Continued-use-of-retracted-papers-Temporal-trends
    if self.clock >= self.delay:
        if random.randint(0, 1000) > 54:
            self.set_belief(Belief.Fake)
    else:
        self.set_belief(Belief.Fake)
```

Future Work - In progress

- Develop a custom network - combine small world with preferential attachment
  - Information should flow one way: older nodes to newer nodes
- Nodes should interact, iteratively, across every link instead of randomly
- Add additional classes
  - Currently, there is only neutral, false, and retracted.
  - This incorporates the incorrect belief that identifying and correcting articles with false information is equally simple across all articles linked to the false information.
  - Realistically, the further away from the retracted article, the harder it is to identify as having potentially used false information.
- Experiment with different number of nodes and edges to be more realistic
  - Average citations varies by field (and arguably only cite other papers within that field), so I plan to model a specific field

References

Brainard, Jeffrey, and Jia You. "What a massive database of retracted papers reveals about science publishing's 'death penalty'." *Science* 25.1 (2018): 1-5.

Bordignon, Frederique. "Self-correction of science: A comparative study of negative citations and post-publication peer review." *Scientometrics* 124.2 (2020): 1225-1239.

Hsiao, Tzu-Kun, and Jodi Schneider. "Continued use of retracted papers: Temporal trends in citations and (lack of) awareness of retractions shown in citation contexts in biomedicine." *Quantitative Science Studies* 2.4 (2021): 1144-1169.

Wakefield, Andrew J., et al. "RETRACTED: Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children." *The lancet* 351.9103 (1998): 637-641.

LaCroix, Travis, Anders Geil, and Cailin O'Connor. "The dynamics of retraction in epistemic networks." *Philosophy of Science* 88.3 (2021): 415-438.

Watts, Duncan J., and Steven H. Strogatz. "Collective dynamics of 'small-world' networks." *nature* 393.6684 (1998): 440-442.

Barabási, Albert-László, and Réka Albert. "Emergence of scaling in random networks." *science* 286.5439 (1999): 509-512.

Thank you!
Any questions?