TrustResearch

Check in #2

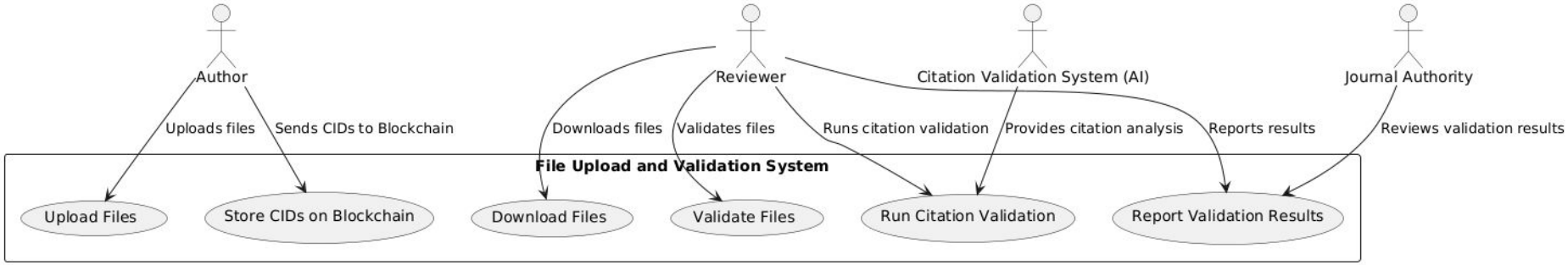
How to Validate Claims

- Receive claim (accuracy, precision, recall, and f1-score), model, dataset, hyperparameters and validation code from the author
- Apply spaCy on the claim data to parse the values via Named Entity Recognition
- Then run the validation code on the dataset with hyperparameters
- Finally check the values with the claims automatically

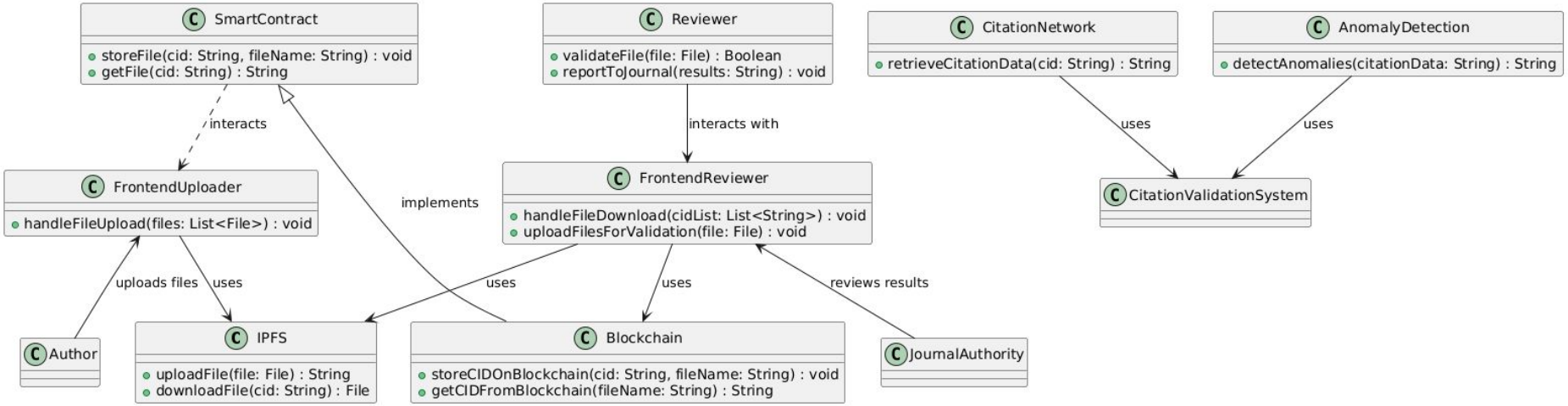
Citation Analysis

- Knowledge Graph General Structure:
  - Entities:
    - Papers
    - Authors
  - Relations:
    - Authored by (paper -> author)
    - Cites (paper -> paper)

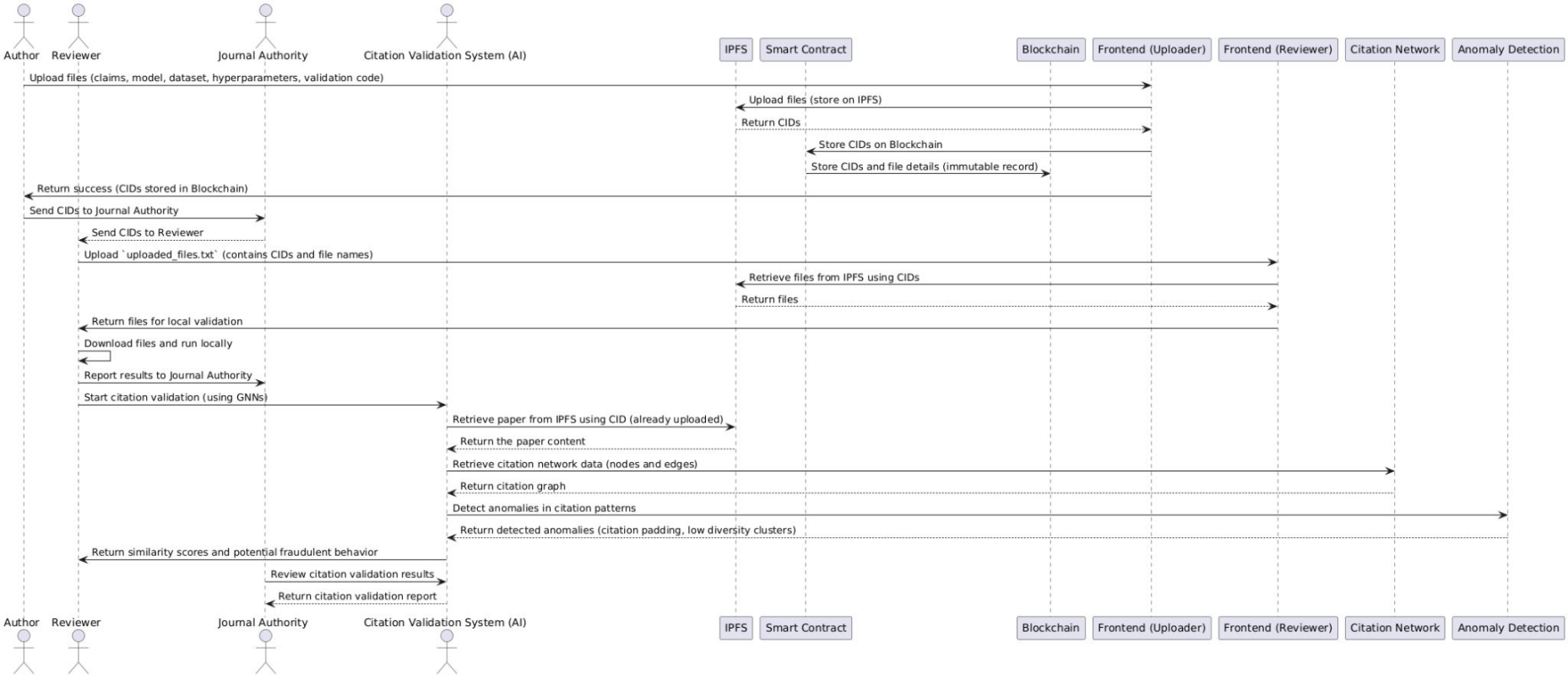
Architecture of the Project



Class Diagram of the Project



Sequence Diagram of the Project



High Level Overview of AI

- Graph Neural Networks (GNNs):
  - We use GNNs to model citation networks as graphs, where nodes represent papers and edges represent citation relationships.
- Citation Clustering:
  - Instead of parsing full-text data across millions of documents, we cluster papers based on their citation patterns.
- Anomaly Detection:
  - The system detects unusual citation behavior, such as papers citing an identical group of sources (citation padding) or forming tight clusters with low diversity.
- Similarity Scoring:
  - Identifies papers with highly overlapping references, suggesting potential redundancy or fraudulent replication.
- Scalability:
  - Graph-based processing allows us to analyze large academic datasets efficiently without heavy reliance on full document parsing.

High Level Overview of Blockchain

- All the claims, model, dataset, hyperparameters, validation code are stored in the IPFS
- A smart contract is deployed on the Blockchain to store the CIDs and manage the validation process
  - Used Ganache and Truffle for local Blockchain deployment
- The CIDs are sent to the journal authority for validation by the reviewers
- Reviewer can download the files in just one click and run the validation code to validate claims
  - This will automatically load all the necessary files to reproduce the result