

Novelty Detection Using Ontological Semantics

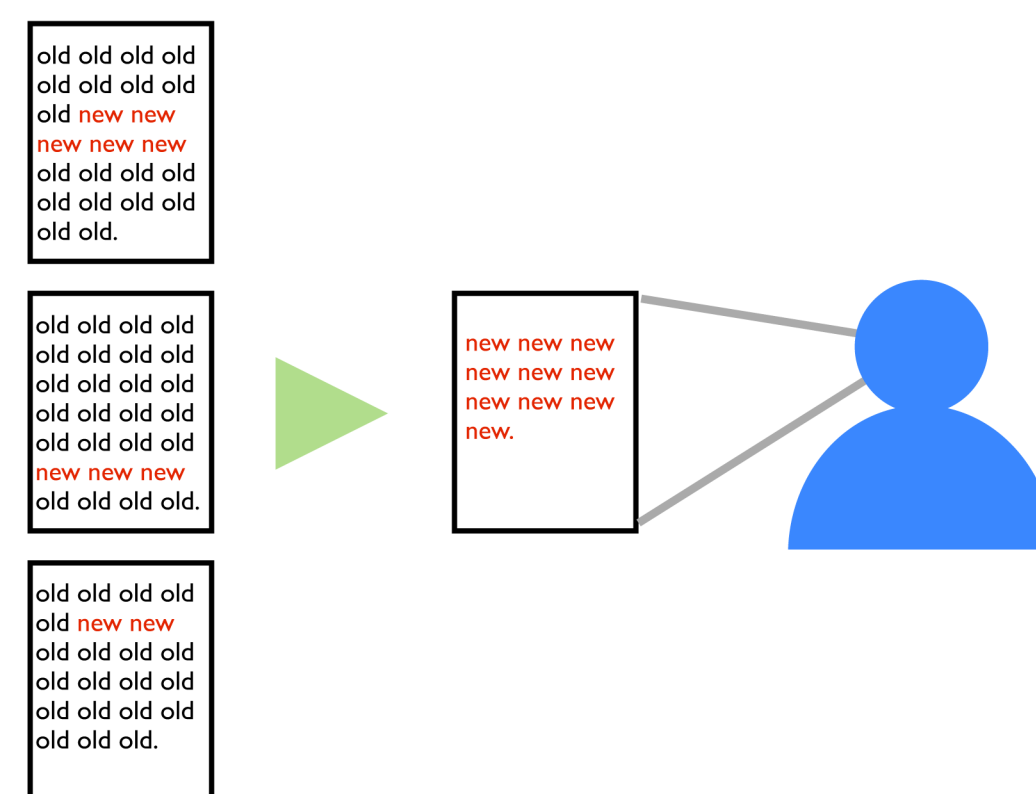
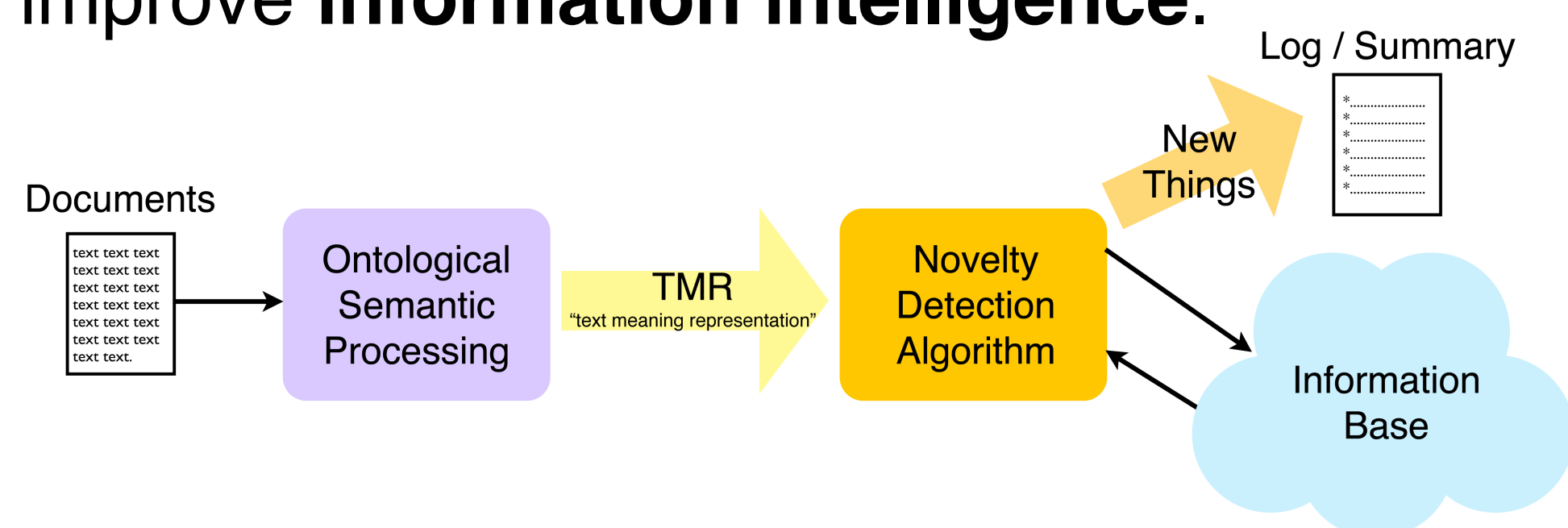
Angela Graci¹, Victor Raskin^{2,3}, Julia Taylor^{2,4}, Lauren Stuart²

¹Department of Computer Science, SUNY Oswego; ²CERIAS, Purdue University;

³English and Linguistics, Purdue University; ⁴Computer and Information Technology, Purdue University

1. Problem Statement

The purpose of this project is to examine the potential for **novelty detection** using **ontological semantics**, with a view to improve **information intelligence**.

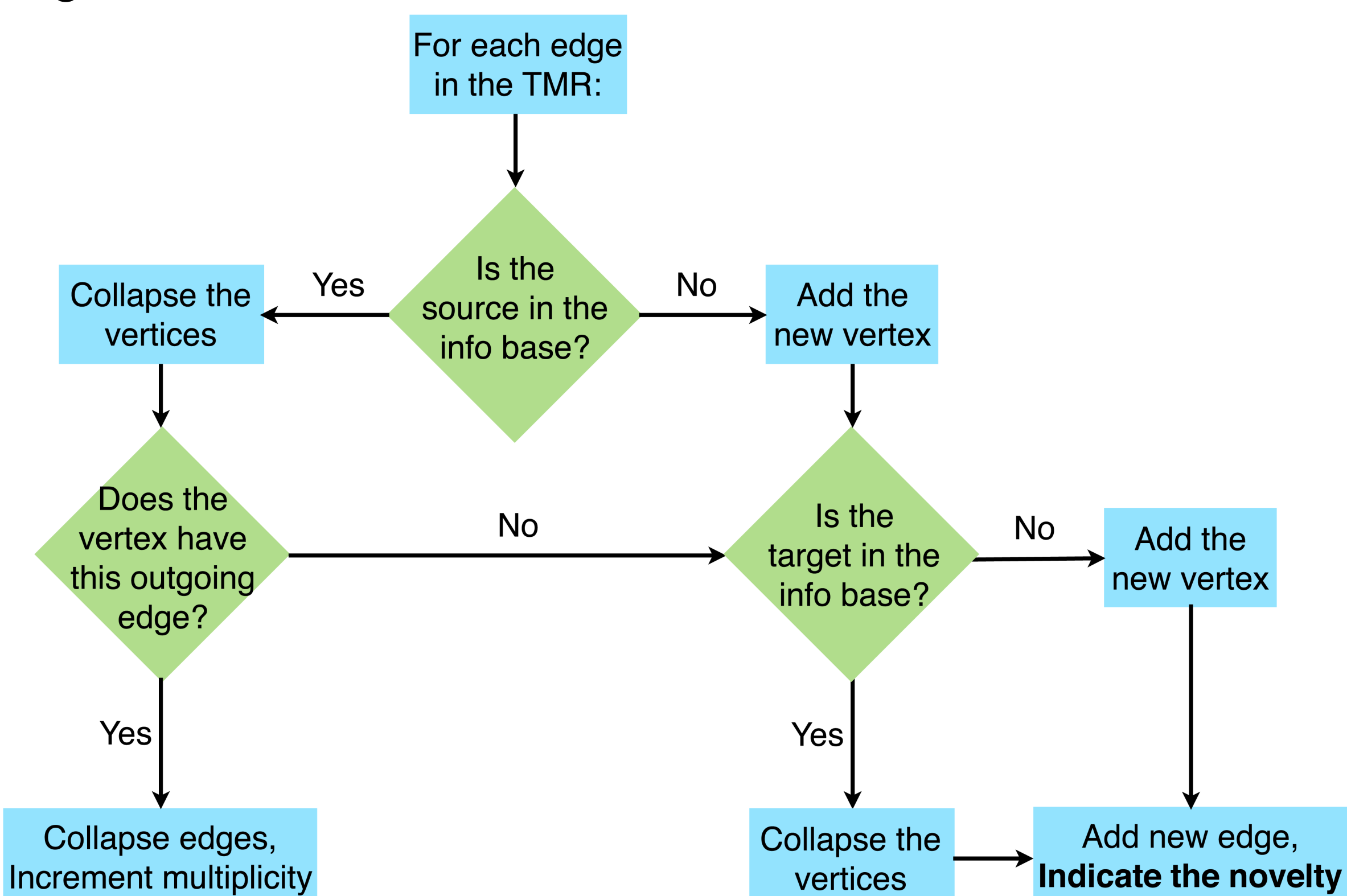


Reliable novelty detection could:

- computationally process large volumes of intelligence documents.
- increase efficiency.
- reduce instances of duplication and human error.
- help to mitigate the effects of information overload.

2. A Simple Algorithm

The TMRs that the novelty algorithm takes as input have the form of directed graphs, where each edge has a source and a target vertex:

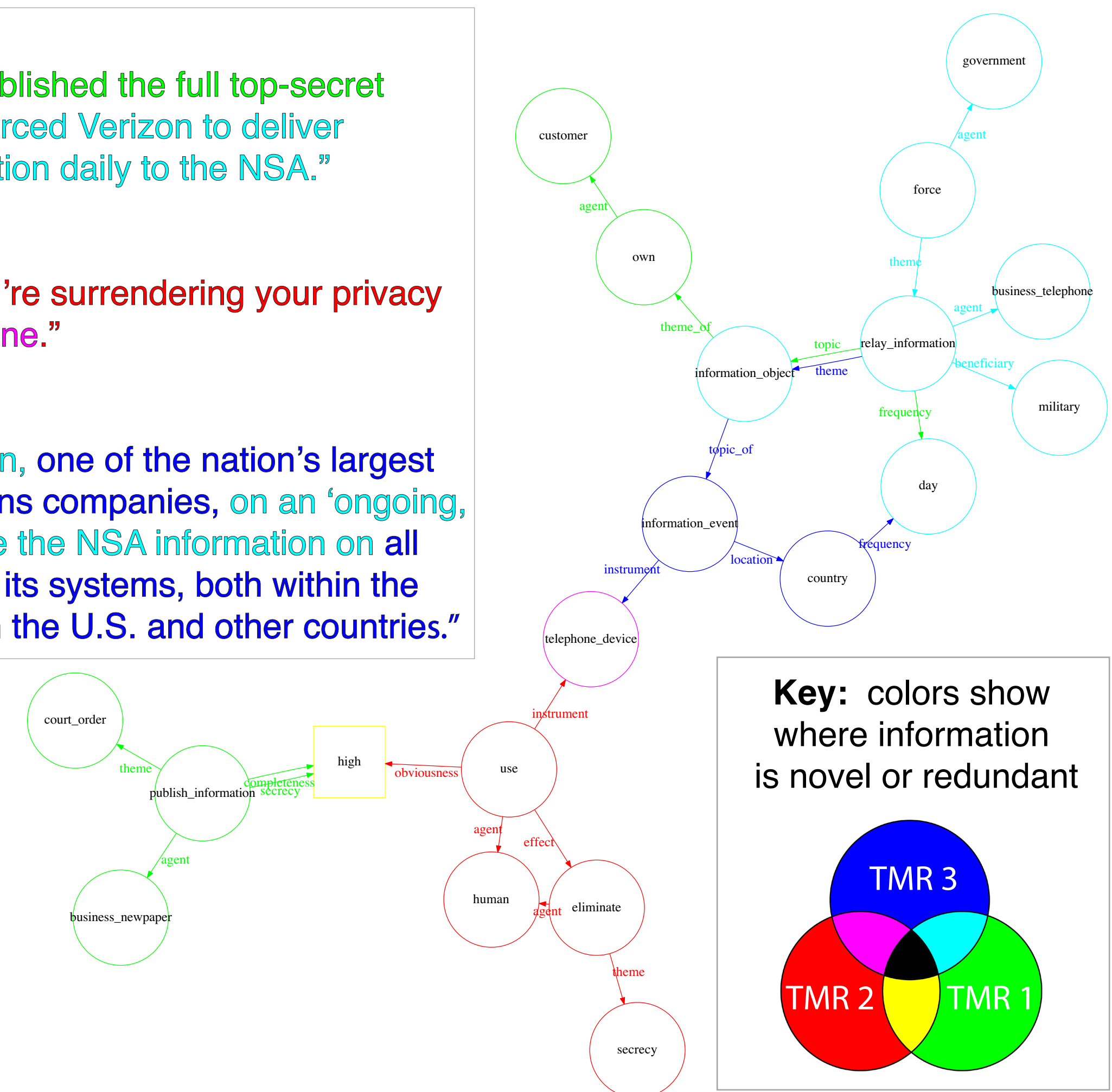


3. Example Sentences (taken from various newspapers)

Sentence 1:
"The Guardian published the full top-secret court order that forced Verizon to deliver customer information daily to the NSA."

Sentence 2:
"By definition, you're surrendering your privacy by using your phone."

Sentence 3:
"It requires Verizon, one of the nation's largest telecommunications companies, on an 'ongoing, daily basis' to give the NSA information on all telephone calls in its systems, both within the U.S. and between the U.S. and other countries."



4. Adding Complexity

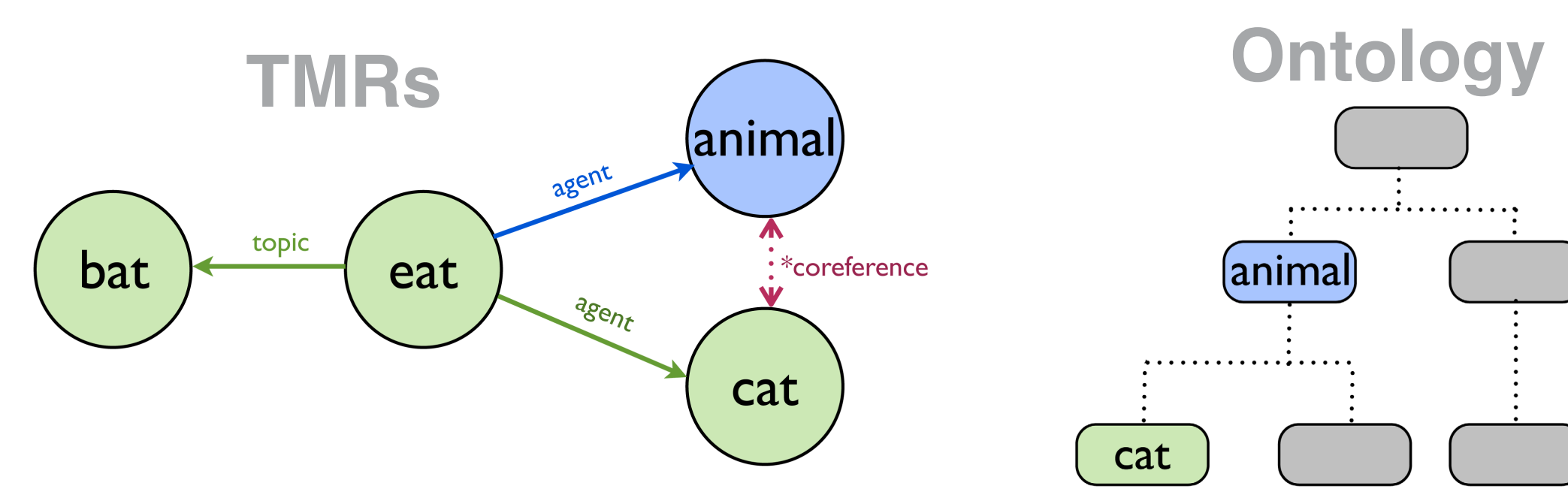
Problem: The simple algorithm outputs **any** previously unknown information, which results in too much **trivial novelty**.

Solution: Identify **types** of novelty and find ways to measure the **degree** of each kind of novelty.

New information could then be filtered by "**how new**" it is.

5. Grain Size Novelty

One measure of novelty that we can use is the **grain size** relative to known vertices:



Smaller grain size = more specific / detailed
Larger grain size = more general information

6. Next Steps

- **Test** the revised (grain size) algorithm with more texts.
- Combine the algorithm's output with a text generator to make a human-readable **summary**.
- Identify more **types** of novelty and **ways to measure** them.
- Find ways to **score the novelty** of graphs by the weights of their subgraphs.



This material is based upon work supported by the National Science Foundation under grant #1062970. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.