# Novelty Assurance by a Cognitively Inspired Analogy Approach to Uncover Hidden Similarity

**Diarmuid P. O'Donoghue, Conor Brady, Shane Smullen, Gary Crowe**
Department of Computer Science, Maynooth University, Co. Kildare, Ireland.
diarmuid.odonoghue@mu.ie

### Abstract

The novelty of artefacts is central to creativity but detecting obfuscated versions has become increasingly difficult. Intelligent manipulation of information can render plagiarism detection system virtually useless, allowing nefarious actors to mis-represent modified artefacts as their own creations. We focus on detecting hidden similarities that are likely to elude existing novelty assurance systems, outlining a model inspired by metaphor, analogy and conceptual blending. We focus on text and outline a model that combines parsing, information extraction and graph matching to find hidden similarities between documents knowledge graphs. We present results for a paraphrase corpus, with various degrees of similarity between sentence pairs. Quantitative evaluations are accompanied by evidence detailing different types of similarity between the sentences: 1) identical counterparts 2) alignable counterparts 3) novel elements. The prospects for further development are briefly outlined.

## Introduction

Recent technologies make it easy for nefarious actors to transform creations and present the results as (apparently) novel creations. Recent advances in text processing including translation and paraphrasing tools (many using transformers), are easily misused to falsely present outputs as though they are original creations (Prentice & Kinden, 2018). These and some related challenges are known as The Global Cheating Industry.

Boden (1992) identified *novelty* along with *quality*, as one of only two defining qualities of creativity. Runco and Jaeger (2012) identify *originality* and effectiveness as definitional, while *unusual, unique* and *surprising* are strongly related to creativity. SPECS (Jordanous, 2012) highlights that creativity produces outputs *"that didn't exist before"*, whose "*Originality*" relates to "*innovation / originality / new / novel"*. We believe that novelty's importance can benefit from improved support tools this paper aims to detecting *false* novelty arising from modifications that obfuscate the true origins of creations.

Figure 1 depicts different types of cognitively inspired similarity, to which we have added and obvious and latent similarity (highlighted), which appear to be a somewhat overlooked types of similarity. We focus on comparisons that are less obvious than literal similarity, but stronger than many analogies and metaphors. We aspire to detect cases of fake novelty that might elude existing authentication systems.

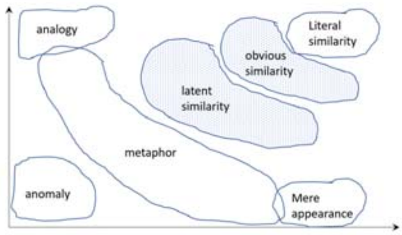


**Figure 1:** Types of similarity (Gentner & Markman, 1997), with highlighted areas added indicating the focus of this paper

We adapt an analogy-based model currently under development, to uncover latent similarities between texts. We shall present results including both quantitative scores and also, itemized details on the latent similarities that have been identified. Ramscar & Yarlett (2003) showed Latent Semantic Analysis is useful in supporting analogy retrieval from texts but not analogical mapping and thus, is unable to identify the detailed comparisons described later in this paper.

## Evidence of Novelty and the Search Report

Patent applications are supported by "Evidence of Novelty" in the form of a search report, serving to inspire our approach. We wish to identify the closest "prior art" for a creations and to detail the obvious and latent similarities to that artefact.

Many plagiarism detection tools are based on identical word sequences, though such services often concede that students *"paraphrase thoroughly"* to avoid detection. A recent review (Vrbanec & Meštrović, 2020) compared systems for text comparison (tf-idf, LSA, Word2vec, GloVe, ELMO, etc), but we believe that these systems can not detect the latent similarities that are the subject of this paper. Weber-Wulff (2019) highlight that plagiarism detectors frequently disagree with one another and their "originality scores" are often relied upon too heavily. Rogerson &

McCarthy (2017) have shown that paraphrasing tools represent a serious problem for some plagiarism detection tools. Foltýnek, et al., (2020) review 15 plagiarism detection tools, concluding they should be improved to detect plagiarism arising from *"...synonym replacement, translation, or paraphrase."* Fakebox (Zhou, 2019) employs fact checking to detect fake news, but doesn't address plagiarism. Some fake reviews are detected using graph structures, but graphs aren't widely used for plagiarism detection.

Publications and patents can be easily copied made seemingly anew new using technologies like paraphrasing and translation tools. Surprisingly, many instances of fake novel publications on www.retractionwatch.com were identified by human readers rather than computational systems. This paper aims to help the detection of such fake novelty.

**Questionable Similarity**

We define *Questionable Similarity* as involving firstly, few if any identical terms that might reveal a documents true origins using standard originality checkers. Secondly, they use terms that are similar to the existing artefact. Thirdly, there is a consistency in the use of terms between the new and the "prior art" that is unlikely to occur by accident. S1 and S2 below bear questionable similarity to one another, and our objective is to detect this latent similarity and identify the itemized correspondences that it contains. The animals in S1 have been replaced by visually similar ones in S2 below.

**S1:** *The leopard chased the rabbit, but he escaped from it.*
**S2:** *A jaguar hunted the hare, but she eluded the jaguar.*

The use of different (if related) terms presents a challenge to detecting latent similarity, with systems using *tf-idf* unlikely to produce useful results, especially when a large list of stopword is used. We also highlight that some comparison system use embeddings but they don't generally itemize the discovered similarities. We believe identified similarities should be supported by direct evidence from prior artefacts.

## Analogies and Blends between Text

As stated previously we take inspiration from cognitve processes like analogy and conceptual blending. We will outline a model for identifying latent similarities between texts. But first we briefly review some related work.

Eppe *et al*, (2018) present a framework for conceptual blending, but not a computational model for mining blends from text. Comparable computational systems focusing on deep semantics and document understanding includes KnIT (Nagarajan *et al*, 2015), Dr Inventor (O'Donoghue, Abgaz, Hurley, Ronzano, & Saggion, 2015) (Abgaz, O'Donoghue, Hurley, Chaudhry, & Zhang, 2017), CrossBee (Lavrač *et al*, 2019), Divago (Martins *et al*, 2019), IBID (Petit-Bois *et al*, 2021). However, none search for concealed similarity that hides the origin of supposedly novel text. Word2Vec (Mikolov *et al*, 2013) can retrieve simple proportional analogies between words, like; *king* is-to *man* as *woman* is-to *?* yielding a vector close to *queen*. However, its ability to accurately predict novel analogies is less certain. Furthermore, the comparisons of interest in his paper involve novel collections of arbitrary named relations between structured collection of named concepts. Knowledge graphs containing temporal information were used to detect fake reviews (Fang, 2020). RoboChair (Pollak, *et al*, 2021) uses text information for reviewing purposes. Blendville (Gonçalves *et al*, 2019) explores existing semantic structures using an optimization approach, but doesn't explore similarities between texts. Aris (Pitu, et al., 2013), (Aiyankovil, Monahan, & O'Donoghue, 2021) uses graphs to improve software reliability by adding formal specifications from similar source code by using analogical inference.

**The Cre8blend System**

Cre8blend is a system to discover latent similarities between semantic structures. Cre8blend extracts a concept map directly from the text and then performs homomorphic Graph Matching to find similarities to another artefact. This approach compliments existing originality systems by shifting the focus from syntactic similarity to identifying certain types of semantic similarity. We point out that while this paper uses text data, it can in principle be adapted to other artefacts. We outline the main components of Cre8blend.

**Text2pred:** The predicate-argument structure is extracted from a tree generated by the Stanford parser, where predicates (triples) generate the document knowledge graph. Alternative information extraction systems include Reverb, TextRunner, ReLink and DeepKE. A survey of open information extraction and identified coreference resolution as an overlooked area in information extraction (Niklaus, Cetto, Freitas & Handschuh, 2018). Our results include details on the coreference chains in our knowledge graphs, as identified by Stanford's deterministic coreference model. The following example shows a coreference chain (node) "*leopard_it*" participating in two (predicates) edges. We note that both nodes and edges contain textual information sourced from the original documents.

***S1:*** *(leopard_it chase rabbit) (rabbit avoid leopard_it)*
***S2:*** *(jaguar hunted hare_she) (hare_she eluded jaguar)*

**Graph Matching - Counterpart Identification**

We take inspiration from Gentner's (1983) Structure Mapping Theory to identify latent similarity between knowledge graphs. A graph matching process identifies comparisons between tiples from the two document graphs. The graph matching algorithms ISMAGS and VF3 impose constraints that inhibit their use in this instance. For example, VF3 is limited to identifying *induced* subgraph to graphs isomorphisms.

Our goal requires identifying subgraph to subgraph matching. For input graphs G1 and G2 we need to identify the largest subgraph of G1 that is isomorphic with the largest subgraph of G2. However, this non-induced subgraph to subgraph matching problem has not yet attracted much attention in graph matching. We developed our own system balancing semantic and topological factors and it's also used by Aris (Aiyankovil, O'Donoghue, & Monahan, 2021) to match graphs containing source code. Semantic similarity between matched words is quantified using Sense2vec (Trask *et al*, 2015), which incorporates part of speech (noun, verb, etc) in the similarity estimate, so *dove#noun =/= dove#verb*.

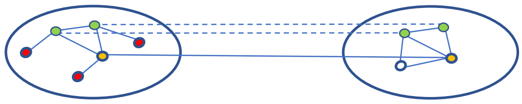

**Figure 2:** Subgraph-subgraph matching. Novelty is influenced by identical pairings (dashed lines), non-identical pairings (solid line), and unmatched items from the inputs.

Overly flexible similarity detection might easily become overwhelmed by false positives. But novel texts should not have highly similar prior art, while longer texts will quickly reduce the problem posed by false positives.

## MRPC - Document Knowledge Graphs

The Microsoft Research Paraphrase Corpus (MRPC) contains pairs of sentences gleaned from news sources, with a judgement representing whether *"the two sentences to be close enough in meaning to be considered close paraphrases"* (Dolan & Brockett, 2005). Our working hypothesis is that sentence pairs may contain differing combinations of identical, similar and dissimilar elements. We treat the first sentence as a target whose novelty we wish to assess, while the second sentence is the closest identified prior art.

The MRPC is challenging as the similarity between sentence pairs is more nuanced than suggested by the binary categorization as either paraphrased (Para) or non-paraphrased (Orig). The paraphrase sentences contain significant amounts of differences while the non-paraphrased pairs also contain various differences. The MRPC includes a training set but this was not used to fine-tune our model.

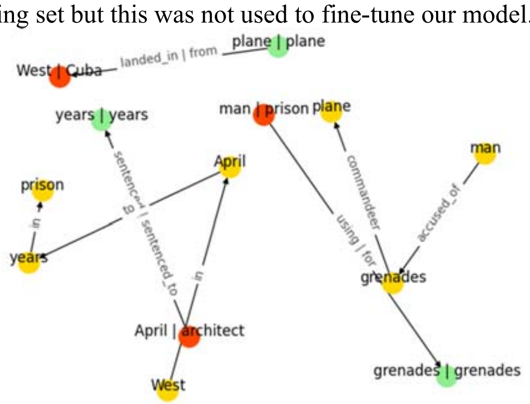


**Figure 3:** Red nodes map non-identical terms between sentences, revealing a possible instance of false novelty.

1458 pairs of graphs were extracted from 1641pairs of text, with failures often attributed to unsuccessful parsing of either sentence in a corpus pairs; eg *"The broader Standard & Poor's 500 Index <.SPX> was 0.46 points lower, or 0.05 percent, at 997.02."* Graphs contained an average of 3.8 edges (SD=3.4) ranging from 1 to 87 edges. There was a moderate difference between the sizes of the original and paraphrased graphs, with average sizes of 3.72 (SD=3.04) and 3.85 (SD=3.76) edges respectively. Paraphrased graphs were slightly larger and more diverse than the originals.

Figure 3 shows similarities between two sentence-graphs. The edge *(plane landed_in West)* was mapped with *(plane from Cuba)*. The items of greatest concern for plagiarism detection are the red nodes depicting paired non-identical concepts and the paired non-identical relations that are separated by "|". Of less concern are orange nodes showing unmapped concepts and green nodes indicate paired identical concepts. Identical paired edges are not repeated.

## Quantitative Results for MRPC Sentence-Pairs

This analysis focuses on quantitative results, but each is accompanied by detailed lists of paired words or paired coreference chains, fostering deep expert or automated investigation of any discovered similarity. Table 1 *Para* indicates the similarity between paraphrased sentence-pairs, while *Orig* assesses Original (or non-paraphrased) sentence pairs. 95 pairs were identified as identical for the Para condition and just 30 for the Orig condition. Only 6 of 1482 sentence were identified by our system as having no detectable similarity. This indicates the prevalence of similarity between sentence pairs in this corpus, highlighting the challenge of distinguishing between them.

The average number of mapped edges for the Para condition was 2.65 and 2.18 for Orig. This moderate difference between the sentence types highlights that even Orig. sentence pairs contain much semantic overlap. The number of identical edges mapped in the paraphrase condition was 1.02 but only 0.58 for the Orig condition. Such overt similarity is not a source of concern for originality assurance.

The Para condition aligned 3.08 concept nodes on average, compared to just 2.80 for Orig. This quantifies the number of overt and latent similarities found. Over 1/3 of the graphs and approximately 50% of comparisons involved at least one node containing a coreference, showing the importance of intra-sentential coreferences.

| **Average Result** | **Para** | **Orig** |
|---|---|---|
| Number of Identical Graphs | 95.00 | 30.00 |
| Avg. num. mapped edges | 2.65 | 2.18 |
| Avg. num. identical edges | 1.02 | 0.58 |
| Total S2v similarity | 4.80 | 2.97 |
| % total S2v similarity | 0.53 | 0.44 |
| Num. mapped concept nodes | 3.08 | 2.80 |
| Coreference Chain in mapping | 0.38 | 0.24 |
| % of target in LCC | 0.58 | 0.66 |

**Table 1:** Comparison of MRPC sentence-pairs

We also estimated the semantic (sense2vec) similarity between mapped edges, each edge including two nouns and one verb. The average similarity for the Para condition was 1.60 but just 1.32 for Orig. from a maximum of 3. The Para condition accounted for 53% of the maximum possible similarity, while this was 44% for Orig. We identified the largest connected component (LCC) of the mapping. Surprisingly, the Orig sentences produced a stronger result, possibly indicating that further improvements are required.

Thus, Cre8blend identified a larger amount of stronger similarity between the paraphrased sentences than the non-paraphrased (Orig). We reiterate, these results are accompanied by detailed comparisons between the two sentences.

### Qualitative Results for MRPC Sentence Pairs

We now illustrate some qualitative results from detecting latent similarity between potentially creative sentences and obfuscated versions that attempt to hide its true origins. Instances of questionable similarity between paraphrased sentences are presented. In the examples in this section, the aligned terms are generally located above one another allowing the non-literal similarity to be interpreted.

**Synonym Replacement:** Synonym replacement is a common strategy to feign novelty and avoid plagiarism detectors, but is detectable by our synta-semantic system.

| |
|---|
| *... by **two miles**, ... a seven-mile section ...* |
| *...by **three kilometres**, ... an 11-kilometre section ....* |

Replacing multiple synonyms can also be detected.

| |
|---|
| *… to topple **Saddam** but to stabilize **Iraq**...* |
| *...to topple **Mr. Hussein** but to stabilize the **country**.* |

**Semantically Distant Term Replacement:** Replacing multiple semantically distant words represents even greater challenge for plagiarism detection. Graph matching identified the following word-pairs: *replacement* ↔ *work; company* ↔ *officials.*

| |
|---|
| *The **company** didn't detail the costs of the **replacement** and repairs.* |
| *But company **officials** expect the costs of the replacement **work** to run into the millions of dollars.* |

**Unknown Term Introduction:** Novel terms (like '5m' below) can also hide a documents' true origin but can be uncovered using context, such as aligning the following edges from 2 sentences: *(5m, over, violations), (million, settle, violations)*. We note also that this novel term was used in a somewhat dissimilar lexical context.

| |
|---|
| *PwC itself paid $**5m** last year ...* |
| *...PWC paid $5 **million** to settle alleged ...* |

### Questionable Similarity

We previously described three hallmarks of questionable similarity. Figure 4 depicts the results of applying one metric for questionable similarity to all sentence-pairs in the MRPC. We observe an exponential style distribution highlighting a small number of MRPC pairs displaying the three hallmarks of questionable similarity. While we cannot conclude these sentences are deliberate fakes, but we believe the authors of one of the following texts may be interested in the latent similarities identified by Cre8blend.

The highest questionable similarity score was for the following sentence pair, aligning 4 predicates and including 4 non-identical terms within that mapping.

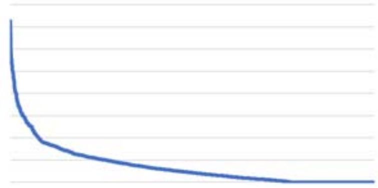

**Figure 4**: Few MSPR pairs have high questionable similarity

| |
|---|
| *Doctors who knowingly **violate** the **ban** could face up to two years in **prison**.* |
| *Under the measure, doctors who **perform** the **procedure** would be subject to two years in prison and unspecified **fines**.* |

This pleasing result identified a large collection of parallels between the two texts, despite the small level of obvious similarity. The next highest result was for the following:

| |
|---|
| *Feith said **people** have **misconstrued** the **purpose** of the small intelligence review **team** he **assembled** in October* |
| *Feith said **critics** have **misrepresented** the **work** of the special intelligence **group** he **set** up in October* |

This paired 4 edges from each graph, aligning 5 non-identical term-pairs between the two graphs. However, the next highest score can be considered a false positive arising from inaccurate identification of the predicate argument structure.

| |
|---|
| *They **were at** Raffles **Hospital** over the weekend **for** further **evaluation**.* |
| *They **underwent** more **tests** over the weekend, and are now warded **at** Raffles **Hospital**.* |

### Conclusions and Future Work

Our model successfully identified some instances of hidden similarity but requires further work with longer texts, as well as comparison to embedding and other approaches. A greater range of lexical information must also be extracted for the graphs. Refining our model may reduce instances of false positives, but its computational expense seems worthwhile only for valuable artefacts like publications, patents *etc* such as may arise from serious creativity. Examining suitable corpora may help identify typical similarity ranges for novelty assurance and for plagiarism detection.

### Author Contributions and Acknowledgements

Diarmuid wrote the paper, Conor contributed broadly to similarity assessment, Shane contributed to the semantic evaluation and Gary contributed to graph matching. Other Maynooth contributors include Michael Griffith, Eoin Manning, Yalemisew Abgaz, Donny Hurley.

## References


Abgaz, Y., O'Donoghue, D., Hurley, D., Chaudhry, E., & Zhang, J. (2017). Characteristics of Pro-c Analogies and Blends between Research Publications. *International Conference on Computational Creativity (ICCC).* Atlanta, GA, USA.

Aiyankovil, K. G., Monahan, R., & O'Donoghue, D. P. (2021). Upcycling Formal Specifications for Similar Implementations with Aris. *International Conference on Case Based Reasoning (ICCBR).*

Aiyankovil, K. G., O'Donoghue, D. P., & Monahan, R. (2021). Creating new Program Proofs by Combining Abductive and Deductive Reasoning. *International Conference on Computational Creativity (ICCC)*, (pp. 394-399). Meixco.

Boden, M. (1992). *The Creative Mind.* Abacus.

Carletti, V., Foggia, P., Saggese, A., & Vento, M. (2017). Introducing VF3: A new algorithm for sub-graph isomorphism. *International Workshop on Graph-Based Representations in Pattern Recognition*, (pp. 128-139).

Cordella, L. P., Foggia, P., Sansone, C., & Vento, M. (2004). A (sub) graph isomorphism algorithm for matching large graphs. *IEEE transactions on pattern analysis and machine intelligence, 26*(10), 1367-1372.

Dolan, B., & Brockett, C. (2005). Automatically constructing a corpus of sentential paraphrases. *Third International Workshop on Paraphrasing (IWP).*

Eppe, M., Maclean, E., Confalonieri, R., Kutz, O., Schorlemmer, M., Plaza, E., & Kühnberger, K.-U. (2018). A computational framework for conceptual blending. *Artificial Intelligence, 256*, 105-129.

Fang, Y., Wang, H., Zhao, L., Yu, F., & Wang, C. (2020). Dynamic knowledge graph based fake-review detection. *Applied Intelligence, 50*(12), 4281-4295.

Foltýnek, T., Dlabolová, D., Anohina-Naumeca, A., Razı, S., Kravjar, J., Kamzola, L. G.-D., . . . Weber-Wulff, D. (2020). Testing of support tools for plagiarism detection. *International Journal of Educational Technology in Higher Education, 17*(1), 1-31.

Gonçalves, J., Bembenek, A., Martins, P., & Cardoso, A. (2019). Going into Greater Depth in the Quest for Hidden Frames. *Intl. Conf. Computational Creativity ICCC*, (pp. 291-295).

Houbraken, M., Demeyer, S., Michoel, T., Audenaert, P., Colle, D., & Pickavet, M. (2014). The Index-based Subgraph Matching Algorithm with General Symmetries (ISMAGS): exploiting symmetry for faster subgraph enumeration. *PloS One, 9*(5).

Jordanous, A. (2012). A standardised procedure for evaluating creative systems. *Cognitive Computation, 4*(3), 246-279.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory & Cognition, 24*(2), 235-249.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *arXiv preprint arXiv:1310.4546.*

Niklaus, C., Cetto, M., Freitas, A., & Handschuh, S. (2018). *A survey on open information extraction.* arXiv preprint arXiv:1806.05599.

O'Donoghue, D., Abgaz, Y., Hurley, D., Ronzano, F., & Saggion, H. (2015). Stimulating and Simulating Creativity with Dr Inventor. *International Conference on Computational Creativity (ICCC)*, (pp. 220-227). Utah, USA.

Pitu, M., Grijincu, D., Li, P., Saleem, A., Monahan, R., & O'Donghue, D. P. (2013). Arís: Analogical Reasoning for reuse of Implementation & Specification. *4th Artificial Intelligence for Formal Methods Workshop (AI4FM).* France.

Pollak, S., Podpecan, V., Kranjc, J., Borut Lesjak, & Lavrac, N. (2021). Scientific question generation: pattern-based and graph-based RoboCHAIR methods.

Prentice, F., & Kinden, C. (2018). Paraphrasing tools, language translation tools and plagiarism: an exploratory study. . *International Journal for Educational Integrity, 14(1), pp.1-16., 14*(1), 1-16.

Ramscar, M., & Yarlett, D. (2003). Semantic grounding in models of analogy: an environmental approach. *Cognitive Science, 27*(1), 41-71.

Rogerson, A., & McCarthy, G. (2017). Using Internet based paraphrasing tools: Original work, patchwriting or facilitated plagiarism? *International Journal for Educational Integrity, 13*(2).

Trask, A., Michalak, P., & Liu, J. (2015). sense2vec-a fast and accurate method for word sense disambiguation in neural word embeddings. *arXiv preprint arXiv:1511.06388.*

Vrbanec, T., & Meštrović, A. (2020). Corpus-based paraphrase detection experiments and review. *Information, 11*(5).

Weber-Wulff, D. (2019). Plagiarism detectors are a crutch, and a problem. *Nature, 567*(no. 7749), 435-436.

Zhou, Z. H. (2019). Fake news detection via NLP is vulnerable to adversarial attacks. arXiv preprint arXiv:1901.09657.