

Monitoring Term Drift Based on Semantic Consistency in an Evolving Vector Field

Peter Wittek^{*†}, Sándor Darányi[†], Efstratios Kontopoulos[‡], Theodoros Moysiadis[§] and Ioannis Kompatsiaris[‡]

^{*}ICFO-The Institute of Photonic Sciences, 08860 Castelldefels, Barcelona, Spain

[†]University of Borås, 50190 Borås, Sweden

[‡]Information Technologies Institute, Centre for Research & Technology - Hellas, 57001 Thessaloniki, Greece

[§]Institute of Applied Biosciences, Centre for Research & Technology - Hellas, 57001 Thessaloniki, Greece

Abstract—Based on the Aristotelian concept of potentiality vs. actuality allowing for the study of energy and dynamics in language, we propose a field approach to lexical analysis. Falling back on the distributional hypothesis to statistically model word meaning, we used evolving fields as a metaphor to express time-dependent changes in a vector space model by a combination of random indexing and evolving self-organizing maps (ESOM). To monitor semantic drifts within the observation period, an experiment was carried out on the term space of a collection of 12.8 million Amazon book reviews. For evaluation, the semantic consistency of ESOM term clusters was compared with their respective neighbourhoods in WordNet, and contrasted with distances among term vectors by random indexing. We found that at 0.05 level of significance, the terms in the clusters showed a high level of semantic consistency. Tracking the drift of distributional patterns in the term space across time periods, we found that consistency decreased, but not at a statistically significant level. Our method is highly scalable, with interpretations in philosophy.

I. INTRODUCTION

The modeling of semantic content for statistical analysis, prominently by means of computational and theoretical linguistics, has been quietly inspired by physics and chemistry over the past two decades. Strictly on a metaphoric basis, the idea was to compare language as a rule-based system to domains of natural science as like systems for innovative model design. Such endeavours typically go back to two kinds of physical phenomena, i.e. attraction acting on its own like in gravity, a force with non-polar roots, vs. a system of attraction and repulsion based in polarity like electromagnetism. Both word meaning and sentence meaning show statistical behaviour compliant with the idea of non-polar [15], [14] vs. polar [5], [80], [57] binding forces, allowing for latent analytic thinking for category building in many areas, including natural language processing [65], bioinformatics networks [45], [79], quantum theory [6], or chemometry [10]. Clearly, similarity of meaning as an attractor vs. difference of meaning as a repellent are organizing principles of conceptual processing one cannot ignore, and an interesting way ahead is to explore the implications of this observation.

Below we will consider word semantics as the “behaviour” of linguistic signs of a dual nature, i.e. intertwined form and content, leading to the emergence of conceptual categories over objects, and ultimately to the applicability of artificial neural networks [38] for machine learning. Mathematical

“energy” [8], [72] and machine learning are related, the latter often being based on minimizing a constrained multivariate function such as a loss function. Concepts in feature space “sit” at energy minima, representing the cost of a classification decision as an energy minimizing process. This suggests that machine learning must identify concepts with such minima, and since potential energy in physics is carried by a field or a respective topological mapping, concepts naturally have something to do with energy as work capacity. As this general process is practically isomorphic with the theory of reaction paths over a potential hypersurface leading to the probabilistic composition of chemical compounds in computational chemistry [50], [43], we believe that evolving fields as a metaphor to simulate category formation by semantic content is a legitimate approach. Furthermore, attractor networks [1], [14] establishing a quasi-continuous field [75] and capable of processing both word and sentence meaning link the above considerations with the study of neural networks.

Key to our current line of thought is the *semantic continuity hypothesis*, i.e. the assumption that any vocabulary modelled by term space consists of both actual and potential word content, the former mapped to observable locations, the latter filling in the so-called “lexical gaps” between them. Linguistics offers innumerable examples for the existence of such gaps where a language lacks spelled out, i.e. actualized content in contrast to another one. This continuity is best modelled as an evolving field, with both actual and potential word content constantly dislocated over time. Due to such dislocations, both the actual positions and their embedding potential contexts may change, offering a rich texture of semantic substance quasi charging actual term locations vs. discharged potential ones. The same line of thought applies to the vector space model of sentence content [13]. Given timestamped data, one can measure such dislocations, called the *semantic drift*, an important indicator of ongoing language change [3], [19], prominently affecting the monitoring of novelties in document indexing terminology.

We aim to demonstrate the following objectives:

- 1) We are interested in evaluating semantic consistency within single time periods of an evolving data set.
- 2) We would like to see if semantic drift can be detected by analysing the change in semantic consistency.

II. BACKGROUND

In what follows we introduce four considerations leading to our methodology underlying the experiment design.

A. Semantic similarity

As object or feature categorization by neural networks depends on the concept of similarity as a fundamental “binding force”, we briefly review *measures of semantic relatedness* (MSR) to express thematic coherence [68]. In linguistics relevant for text processing, there are two prominent theories of word meaning, the distributional hypothesis [33], and the referential theory of word semantics [25]. According to the first, meaning depends on word use, i.e. is contextual, whereas for the second one, it is referential, i.e. goes back to convention expressed e.g. by definitions in ontology entries. Because habitual word use as context clearly implies agreements about the sense in which certain word forms are being used in certain contexts, there is a dependency between the two approaches.

Automated systems assign a score of semantic relatedness to a given pair of terms calculated from a relatedness measure. The absolute score itself is typically irrelevant on its own; what is important is that the measure assigns a higher score to term pairs which humans think are more related and comparatively lower scores to term pairs that are less related [52].

Distributional similarity and its predecessors go back a long way, building on the notion of term dependency and structures derived therefrom [53]. The underlying distributional hypothesis is often cited for explaining how word meaning enters information processing [37]. Before attempts to utilize lexical resources for the same purpose, this used to be the sole source of word semantics in information retrieval, inherent in the exploitation of term occurrences – most notably, in the term frequency-inverse document frequency (TFIDF) measure – and co-occurrences [26], [56], [63], including multiple-level term co-occurrences [39]. On the other hand, the referential approach relies these days on lexical resources. A lexical resource in computer science is a structure that captures semantic relations among terms, quasi “charging” word occurrences in context with external information.

The reason for combining the two approaches is that statistical techniques typically suffer from the sparse data problem: they perform poorly when the terms are relatively rare. Hybrid methods attempt to address this problem by supplementing sparse data with information from a lexical database [59], [35]. In a semantic network, to differentiate between the weights of edges connecting a node and all its child nodes, one needs to consider the link strength of each specific child link. This is a situation in which corpus statistics can contribute. The following types of resources are commonly used in measuring semantic similarity between terms: dictionaries [42], semantic networks, such as WordNet [51], thesauri modelled on Roget’s Thesaurus [54], and ontologies.

B. Semantic fields

We find the tradition of using a combination of two planes to describe a phenomenon in several disciplines. E.g. the

general practice of evaluating the effectiveness of information retrieval and text categorization models by measures like recall, precision, accuracy, and many more [68]. There is ongoing work to build semantic spaces from distributional vs. compositional semantics [55], [23], representing both word and sentence meaning as locations in high-dimensional space where for phrase or sentence component binding, recursive matrix-vector spaces [64], the tensor product [4], [6], [30], or circular holographic reduced representation are routinely used [13]. In these models, the representation of semantic content in documents is compared to an ideal state of language use, provided by the human standards of interpretation inherent in the evaluation method [22]. Using geometry or probability as a vehicle of meaning, i.e. building a new medium of language, aims at maximizing similarity between the human standard and its statistical reconstruction. This hypothetical original, a correlate of spoken language called a mental state or internal state in neuroscience [21], recalls the “language of thought hypothesis” in philosophy [24], also called *mentalese*. A joint element in the above is that whereas language as a mental phenomenon is assumed to be continuous, its uttered or mathematically modelled representations are discrete.

The same duplicity returns as “hidden metaphysics” in traditional mentalist and more recent generalist-universalist theories about language: language is but a tool operated by something deeper – thought, reason, logic, cognition – which functions in line with biological-neurological mechanisms common to all human beings [34]. Moreover, a linguistic school of thought orthogonal to the above theories, called Neo-Humboldtian field theories of word meaning, goes back to the same dual model where discrete distributions of related content called lexical or semantic fields, based on language use, are underpinned by the assumption of conceptual fields in the mind. Then, the lexical field of related words is only an outward manifestation of the underlying conceptual field so that the sum total of conceptual fields describes one’s world view [67]. In yet another unrelated school of thought, Saussure’s structural linguistics, language (*langue*) is a mental grammar with a rule set specifying ideal content pronunciation, whereas speech (*parole*) stands for the exemplification of those rules [16].

An important symptom of lexical fields is that regions of related content are separated by *lexical gaps*. These are nonexistent names for things where one could exist by rules of a particular language, and indicate possible conceptual distinctions not mapped to actual language use, such as mother’s father (Swedish *morfar*) vs. father’s father (Swedish *farfar*), both called *grandfather* in English, or father’s brother (Swedish *farbror*) not distinguished from mother’s brother (Swedish *morbror*), both called *uncle*. Such language-specific discontinuities of semantic content play a prominent role in our methodology.

The assumption that products of the mind are continuous while their mapping to spoken language is discrete goes back ultimately to Aristotle’s *Metaphysics*. In this, existence or reality is described as the sum total of two components,

conceivable potentiality (*dynamis*) plus observable-measurable actuality (*energeia*). These are names for the latent vs. manifest capacity of existents to induce change. Therefore in our current thinking, existence consists of two layers, potentiality (a continuum) and actuality (a discrete distribution sampling the former). Importantly, one ascribes a field nature to mental experience because of the potentiality layer which we indirectly perceive by the actualized values of events.

C. Measuring semantic consistency

For any model departing from the idea of similarity between instances in a semantic field, a logical next question is, how coherent are the groups of instances in that field? Relating term similarity and semantic consistency, the *domain restriction hypothesis* answers that question [28]. Based on the filtering away of extracted but false sense relations, semantically related terms extracted from a corpus tend to be semantically coherent. To this end, *semantic domains* are used as filters by integrating pattern-based and distributional approaches to capture two characteristic properties of semantic relations:

- Syntagmatic properties: if two terms X and Y are in a given relation, they tend to co-occur in texts, and are mostly connected by specific lexical-syntactic patterns (e.g., the pattern “X is a Y” connects terms in *is-a* relations). This aspect is captured using a pattern-based approach;
- Domain properties: if a semantic relation among two terms X and Y holds, both X and Y should belong to the same semantic domain (i.e. they are semantically coherent), where semantic domains are sets of terms characterized by very similar distributional properties in a (possibly domain specific) corpus.

This approach is detailed in [27]. On the other hand, in a recent reincarnation, semantic consistency is a new distant supervision method which can identify reliable instances from noisy instances by inspecting whether an instance is located in a semantically consistent region. One way to find out is to first model the local subspace around an instance as a sparse linear combination of training instances, then estimate the semantic consistency by exploiting the characteristics of the local subspace [32].

D. Semantic drifts

In the context of Semantic Web dynamics [2], there is a growing body of literature about the semantic drift [40], [31], the language-related version of abrupt parameter value changes in data mining called concept drifts [18], [73], [61], [29]. By semantic drift we mean how the features of ontology concepts gradually change as their knowledge domain evolves, or, alternatively, how different user communities reinterpret the same concept in a different context so that the risk is having these concepts lose their rhetorical, descriptive and applicative power [11]. In a more general sense, the topic is important beyond its linguistic implications, especially for managing semantic interoperability for federations; respective research to date has focused on the generation of semantic

mappings and has tended to ignore the problem of dealing with the dynamism of both the data and the schemata that is characteristic of real-world integration problems [7].

III. METHODOLOGY

We build a vector space model by random indexing of an evolving text collection that is able to closely track the changes. The corpus is split into time periods, and a separate vector space is built for each period. We use self-organizing maps to project the vector space to a two-dimensional surface where clusters and shifts are more apparent. This projection preserves the local topology of the high-dimensional space and allows us to model dynamic semantic fields. We use WordNet-based referential similarity measures to evaluate semantic consistency and also to detect semantic drifts over time. Refer to Figure 1 for an outline.

A. Distributional similarity and random indexing

We build a TFIDF vector space model of the corpus, which provides the foundation for most distributional semantic distance measures. The basic TFIDF space is known to be extremely sparse, having 1–5 % nonzero elements. Latent semantic analysis, or latent semantic indexing [17] measures semantic information through co-occurrence analysis in the corpus, but it reduces the dimensionality and solves the problem of sparsity. The dimension of the vector space is reduced by singular value decomposition.

Random indexing is a similar idea which does not rely on the use of computationally intensive matrix decomposition. This makes random indexing a much more scalable technique in practice. Instead of first constructing a huge co-occurrence matrix and then using a separate dimension reduction phase, random projection builds an incremental word space model [36]. The random projection technique can be described as a three-step operation:

- First, each document in the corpus is assigned a unique and randomly generated representation called an index vector. These index vectors are sparse, high-dimensional, and ternary, which means that their dimensionality (d) is on the order of hundreds, and that they consist of a small number of randomly distributed values, with the rest of the elements of the vectors set to 0.
- Then, context vectors are produced by scanning through the text, and each time a word occurs in a context (e.g. in a document, or within a sliding context window), that context’s d -dimensional index vector is added to the context vector for the word in question. Words are thus represented by d -dimensional context vectors that are effectively the sum of the words’ contexts.

Comparing the term vectors in the random indexed space by a similarity measure such as the Euclidean distance or cosine dissimilarity enables a quantitative framework for semantic analysis.

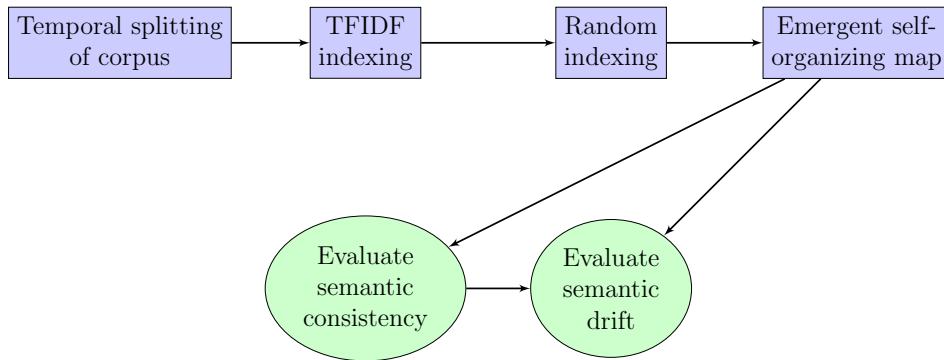


Fig. 1. Outline of methodology. We split the corpus to time periods, each containing approximately the same number of documents. We build a vector space on each period and reduce the dimensions to 200 by random indexing. Next we train an ESOM on the first period, and update it using the random indices for the subsequent periods. We use the sequence of trained maps to evaluate consistency within time periods and the drift between time periods.

B. Semantic fields in emergent self-organizing maps

The vectors of the term space are disjoint locations in a high-dimensional space. A field, on the other hand, is defined at all points in a space. To bridge this problem, we embed the vector space on a two-dimensional surface using emergent self-organizing maps [76].

A self-organizing map is a two-dimensional grid of artificial neurons. Each neuron is associated with a weight vector that matches the dimension of the training data. We take an instance of the training data, find the closest weight vector, and pull it closer to the data instance. We also pull the weight vectors of nearby neurons closer to the data instance, with decreasing weight as we get further from the best matching unit. We repeat this procedure with every training instance. This constitutes one epoch. We repeat the same process in the second epoch, but with a smaller neighbourhood radius, and a lower learning rate when adjusting the weight vectors. Eventually, the neighbourhood function decreases to an extent that training might stop. The time needed to train an SOM grows linearly with the data set size, and it grows linearly with the number of neurons in the SOM. The resulting network reflects the local topology of the high-dimensional space [38]. Emergent self-organizing maps are a type of SOM that contain a much larger number of target nodes for embedding, and thus capture the topology of the original space more accurately [70]. Using a toroid map avoids edge effects.

Some nodes of the emergent self-organizing map correspond to one or more terms; these nodes or best matching units have a special role as they identify semantic content with one or more terms. The rest of the nodes act as an interpolation of the semantic field. Since the field is continuous in nature, we use toroid maps – a planar map would introduce an artificial discrete cut-off at the edges.

We train a map on the first period, and continue training the same map as new instances enter from the next period. Hence the basic layout and the number of nodes do not change across time periods.

C. WordNet-based similarity metrics and validating consistency

WordNet (WN) is a large lexical database of English, created and maintained by Princeton University [51]. It is publicly available to research and commercial users free of charge and its latest version is 3.1 (released Nov’12). WordNet’s popularity arguably lies on the fact that, besides merely offering short definitions and usage examples of the contained nouns, verbs, adverbs and adjectives, it also introduces certain types of semantic relations between terms. Examples of such relations include synonymy, hyper/hyponymy (*is-a* relationship), meronymy (part-of relationship) etc. WN can, thus, be viewed as a combination of a dictionary and a thesaurus.

Due to its structure described above, WN has been extensively deployed in tasks related to determining the semantic similarity between terms, primarily in automatic text analysis and artificial intelligence applications. Towards this direction, many semantic similarity metrics have been proposed, which can be grouped into four different categories [49]: path-based, content-based, feature-based and hybrid metrics.

- In *path-based metrics* the similarity between two terms depends on their relative position in the WN taxonomy as well as on the length of the path linking the concepts. Representative examples deploying path-based measures include [9], [41], [44] and [78].
- *Content-based metrics* are based on the information content available for each concept in WN. The more common information two concepts share, the more similar they are. Examples belonging to this category include [35], [46], [47] and [59].
- *Feature-based metrics* are based on the properties of the WN ontology for obtaining a similarity value. The more common (and the fewer non-common) characteristics two concepts have (e.g. their definitions or “glosses” in WN), the more similar they are. Relationships to other similar terms in the taxonomy are also taken into consideration. Some related approaches are the classical model proposed by Tversky [69] or the more recent approach presented in [62].

- Finally, the *hybrid metrics* combine the ideas presented above. The following constitute paradigms of applying hybrid similarity metrics: [20], [60] and [81].

Apart from the above distinction, Varelas et al. also distinguish semantic similarity measures in *single ontology* and *cross ontology* methods [71]; the former assume that the terms compared all come from the same reference ontology, while the latter compare terms from two different ontologies. Since it is not easy to directly compare the structure and information content of different ontologies, the case of cross ontology similarity typically employs hybrid or feature-based methods (e.g. see [60] and [44]).

D. Quantifying drifts

The foundation for measuring the drift is based on random indexing of subsequent TFIDF spaces with a fixed random seed. By following this method, we are able to derive subsequent low-dimensional spaces which can be compared against one another. We train an emergent self-organizing map on each of these term vector spaces: after the first period, we continue training the map with a lower learning rate to arrive at smoothly changing dynamics.

As a potential source of confusion, we point out that the time-like variable of the iterations and the epochs in the training of the ESOM are unrelated to the temporal aspects of the corpus. By time we always refer to the time related to the corpus, and by a period we refer to documents belonging to a certain time interval of the corpus. Epoch, on the other hand, refers to the training rounds of the ESOM.

IV. EXPERIMENT DESIGN

The experiments were based on a large text corpus on which the TFIDF spaces and random indices were created. The random indices were used to generate the sequence of self-organizing maps. We analysed the topology of these maps for consistency and drifts.

A. Corpus

Book reviews are the literary genre equivalent of abstracts to scientific articles, cross-pollinated by the idea of crowd-sourcing underlying recommender systems [58], i.e. one summary per article produced per one professional abstracting and indexing service vs. potentially many summaries of the same item, written by users as part professional, part lay contributors. In a sense both approaches represent user feedback. From a methodological perspective, due to the nature of condensed semantic content in them, book reviews processing falls in the category of e.g. text summarization [12] on the one hand, combined with sentiment analysis [66] on the other hand. Due to this blend, they represent an interesting and scalable resource of complex semantic content for “neuromorphic” studies.

The experiments described here were based on Stanford’s Amazon book reviews data set [48], which is publicly available as part of the University’s SNAP project¹. The data set spanned

a period of 18 years and included approximately 12.8M book reviews up to March 2013. Every item in the data set included product and user information, ratings, as well as a plain text content description. The collection contained 51 degenerate time stamps, the corresponding instances were discarded.

Period 1	Period 2	Period 3
Until 30 Jan 2003 45162 terms	Until 03 Aug 2008 49400 terms	Until 04 Mar 2013 50672 terms

TABLE I
KEY STATISTICS OF THE TEMPORAL SPLIT OF THE CORPUS.

We split the corpus in three periods, each containing close to 4.3M reviews. The key characteristics are summarised in Table I.

B. Computational background

We used Lucene², which is an information retrieval software library that builds an inverted index, which can be interpreted as a row-major sparse representation of a term-document matrix. We used the SemanticVectors package [74] for reducing the dimensionality of the space. For training the emergent self-organizing maps, we used Somoclu [77].

The implementation of the semantic similarity metrics was based on WS4J³, a Java API for several of the semantic relatedness/similarity algorithms presented in section III-C. WS4J works over WordNet 3.0 and constitutes an improved version of the Perl-based WordNet-Similarity-2.05⁴. In the experiments described subsequently, we deployed a representative path-based semantic similarity method (Wu and Palmer’s [78]), while in the imminent future we plan to investigate the behaviour of more methods (path-based and content-based). Note that all experiments are open source⁵.

C. Consistency in a single time period

This subsection is aimed at assessing whether proximity of terms in the toroid plane indicates a strong underlying semantic similarity. Our focus lied on neurons that had more than 1 term assigned to them. In order to evaluate the consistency of our approach in the case of a single time period, we needed to check whether the terms within a randomly selected neuron were assigned at random to this neuron. In case this was true, obviously our approach failed to implement its purpose. In case the assignment (of terms to each neuron) was not random, we needed to conclude whether it was better than random assignment. To evaluate the assignment, we utilized the average similarity between the terms of a neuron. The main question was how to assess its value. The key idea was to approximate the empirical probability distribution of the average semantic similarities. Towards this direction, we randomly divided the terms of the period into clusters of 5 terms each (5 is the average count of terms assigned to non-empty neurons in the toroid plane). For each of the groups, we

²<http://lucene.apache.org/>, last accessed: Jan’15

³<https://code.google.com/p/ws4j/>, last accessed: Jan’15.

⁴<http://wn-similarity.sourceforge.net/>, last accessed: Feb’15.

⁵https://github.com/peterwittek/concept_drifts

¹<http://snap.stanford.edu/index.html>, last accessed: Jan’15.

Term count	Period 1			Period 2			Period 3		
	$a = 0.05$	$a = 0.1$	$a = 0.2$	$a = 0.05$	$a = 0.1$	$a = 0.2$	$a = 0.05$	$a = 0.1$	$a = 0.2$
$N \geq 3$	0.287	0.334	0.404	0.269	0.320	0.406	0.259	0.316	0.400
$N \geq 5$	0.339	0.380	0.429	0.327	0.366	0.421	0.304	0.353	0.410
$N \geq 10$	0.383	0.416	0.453	0.372	0.401	0.448	0.332	0.380	0.431

TABLE II

PERCENTAGES OF REJECTION OF THE NULL HYPOTHESIS FOR THE t -TEST FOR EACH PERIOD AND FOR DIFFERENT NEURON POPULATIONS.

Terms-Periods	$a = 0.05$	$a = 0.1$	$a = 0.2$
3 - $p1 - p2$	0.051	0.130	0.860
3 - $p2 - p3$	0.211	0.614	0.505
5 - $p1 - p2$	0.330	0.251	0.515
5 - $p2 - p3$	0.051	0.301	0.351
10 - $p1 - p2$	0.318	0.494	0.833
10 - $p2 - p3$	0.060	0.344	0.451

TABLE III

P-VALUES OF THE COMPARISON OF PERCENTAGES DISPLAYED IN TABLE II BETWEEN CONSECUTIVE PERIODS AND FOR DIFFERENT NEURON POPULATIONS.

computed the average semantic similarity; then, based on all the derived average similarities, we determined their empirical probability distribution. The latter was a good approximation of a normal distribution with mean μ_0 and variance σ_0^2 . This motivated us to make use of statistical tests to evaluate whether the average similarity in the neurons (based on our approach) was significant in value, comparing with the case of random term assignment.

We considered three generic cases: (a) neurons containing 3 or more terms ($N \geq 3$), (b) neurons containing 5 or more terms ($N \geq 5$), and, finally, (c) neurons containing 10 or more terms ($N \geq 10$). For each of the three cases we followed the following procedure.

Assuming that the terms within each neuron constituted a random group from the total term population, the respective average similarity should have followed the normal distribution with mean equal to the empirical mean μ_0 and variance equal to the empirical variance σ_0^2 . We performed a one-sided t -test at a predefined level of significance (a) for each neuron, in order to assess whether the average similarity within each neuron was statistically significantly greater than μ_0 ($H_0 : \mu = \mu_0$, $H_1 : \mu > \mu_0$). We calculated the percentage (\hat{p}) of the cases where we rejected the null hypothesis. A percentage \hat{p} greater than a (the level of significance of the t -test) would indicate that the number of samples with high average similarity was greater than expected from the assumption that the average similarities followed the normal distribution $N(\mu_0, \sigma_0^2)$.

We performed a one-sided binomial test ($H_0 : p = a$, $H_1 : p > a$) at a level of significance $a' = 0.05$ to assess whether \hat{p} was statistically significantly greater than a . The rejection of H_0 would mean that both \hat{p} and the overall level of similarity of terms *within the neurons* were statistically significantly greater than expected, based on the initial assumption of random term assignment. Table II displays the percentages \hat{p} for each of the three cases ($N \geq 3, 5, 10$) and for different values of a for each of the three periods (see next subsection). The p-values derived from the application of the binomial test

are not displayed, since they are all negligible (< 0.001). The above procedure was repeated independently for Periods 1, 2, 3. Therefore, we deduced that the clustering in neurons was meaningful, since overall, the terms within a neuron demonstrated significantly greater similarity in comparison to a randomly selected group of terms.

As an example, consider the following neurons: $ex_1 = \{\text{thymine, cytosine, uracil}\}$ and $ex_2 = \{\text{pair, harvard, scorpio, monsignor, misrepresentation}\}$. Intuitively, ex_1 displays apparent semantic similarity (all terms are nucleobases), in contrast to the terms included in ex_2 that do not. We would expect that $H_0 : \mu = \mu_0$ would be rejected for ex_1 and not rejected for ex_2 , which is indeed the case. However, there are some cases when the semantic similarity of terms within a neuron is not that apparent, but the null hypothesis is again rejected. This is due to the chosen path-based semantic similarity metric [78]; thus, we leave as future work the application of further path-based and information content-based methods. All in all, the robustness of this statistical approach is based on our initial assumption that the population of averages followed a normal distribution which we tested empirically. We believe this is the groundwork for interesting future investigations.

D. Dynamics of semantic consistency

We were interested to find out if over several periods the overall similarity within each neuron converges. Our experiment in this paper involved three periods and for each of them we repeated the process described in the previous subsection. We calculated the percentages of the cases where we rejected the null hypothesis of the t -test for all three periods, as displayed in Table II. As observed in the table, there is a slight decrease of the percentages from period to period. To assess whether this decrease is statistically important, we performed a test for proportional comparison at $a = 0.05$. The resulting p-values are shown in Table III and indicated that in every case this decrease was not statistically significant. This means that there are indeed some slight differences that do

not demonstrate divergence. More periods would be needed in order to investigate whether there is convergence, which we will study in the imminent future.

V. CONCLUSION AND FUTURE WORK

Recently increased attention has been paid to models of evolving semantic content, something we represented as a dynamic vector field. This approach was based on the hypothesis of semantic continuity in the vocabulary, allowing both for manifest (actual) and latent (potential) word content mapped to lexical forms, with word meaning in term space behaving like “energy” while constructing conceptual categories. In spite of the simple design to analyse book reviews over 18 years in just three periods, the t -test confirmed that the overall level of similarity of terms within the ESOM grid neurons was significantly higher than expected, i.e. neuron content was statistically consistent.

We plan to continue this line of research in different directions, including the following:

- Increase the number of periods to compute more detailed visual maps over shorter time spans;
- Increase grid granularity to reduce term overlap on neurons;
- Upgrade this model by introducing reflexive random indexing to smoothen transition between periods;
- Interpret the vector field model in terms of e.g. process philosophy.

VI. ACKNOWLEDGEMENT

This project has received funding from the European Union’s Seventh Framework Programme for research, technological development and demonstration under grant agreement no FP7-601138 PERICLES. S. Darányi and P. Wittek are grateful to Dominic Widdows (Microsoft Bing) and Trevor Cohen (University of Houston) for assistance with the Semantic Vectors package. The authors would also like to thank Julian McAuley (Stanford University) for granting access to the Amazon data set.

REFERENCES

[1] D. J. Amit, *Modeling Brain Function: The World of Attractor Neural Networks*. Cambridge University Press, 1992.

[2] G. Antoniou, M. d’Aquino, and J. Z. Pan, “Semantic web dynamics,” *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 9, no. 3, pp. 245–246, 2011.

[3] A. Baker, “Computational approaches to the study of language change,” *Language and Linguistics Compass*, vol. 2, no. 3, pp. 289–307, 2008.

[4] M. Baroni and A. Lenci, “Distributional memory: A general framework for corpus-based semantics,” *Computational Linguistics*, vol. 36, no. 4, pp. 673–721, 2010.

[5] D. Beeferman, A. Berger, and J. Lafferty, “A model of lexical attraction and repulsion,” in *Proceedings of ACL-97, 35th Annual Meeting of the Association for Computational Linguistics*, July 1997, pp. 373–380.

[6] W. Blacoe, E. Kashefi, and M. Lapata, “A quantum-theoretic approach to distributional semantics,” in *Proceedings of NAACL-HLT-13, Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, June 2013, pp. 847–857.

[7] R. Brennan, B. Walshe, and D. O’Sullivan, “Managed semantic interoperability for federations,” *Journal of Network and Systems Management*, vol. 22, no. 3, pp. 302–330, 2014.

[8] E. Bruce, *Biomedical signal processing and signal modeling*. Wiley-Interscience, 2001.

[9] H. Bulskov, R. Knappe, and T. Andreasen, “On measuring similarity for conceptual querying,” in *Flexible Query Answering Systems*, ser. Lecture Notes in Computer Science, T. Andreasen, A. Motro, H. Christiansen, and H. L. Larsen, Eds. Springer, 2002, vol. 2522, pp. 100–111.

[10] M. Bylesjö, M. Rantalainen, O. Cloarec, J. K. Nicholson, E. Holmes, and J. Trygg, “OPLS discriminant analysis: Combining the strengths of PLS-DA and SIMCA classification,” *Journal of Chemometrics*, vol. 20, no. 8–10, pp. 341–351, 2006.

[11] F. Cabitza, “At the boundary of communities and roles: Boundary objects and knowledge artifacts as resources for IS design,” in *From Information to Smart Society*. Springer, 2015, pp. 149–160.

[12] N. Chatterjee and P. K. Sahoo, “Random indexing and modified random indexing based approach for extractive text summarization,” *Computer Speech & Language*, vol. 29, no. 1, pp. 32–44, 2015.

[13] T. Cohen, D. Widdows, R. W. Schvaneveldt, and T. C. Rindflesch, “Discovery at a distance: Farther journeys in predication space,” in *Proceedings of BIBMW-12, IEEE International Conference on Bioinformatics and Biomedicine Workshops*, 2012, pp. 218–225.

[14] D. L. Cooper, *Linguistic attractors: The cognitive dynamics of language acquisition and change*. John Benjamins Publishing, 1999, vol. 2.

[15] S. Darányi and P. Wittek, “Connecting the dots: Mass, energy, word meaning, and particle-wave duality,” in *Proceedings of QI-12, 6th International Quantum Interaction Symposium*, June 2012, pp. 207–217.

[16] F. De Saussure, *Course in General Linguistics*, 1916.

[17] S. Deerwester, S. Dumais, G. Furnas, T. Landauer, and R. Harshman, “Indexing by latent semantic analysis,” *Journal of the American Society for Information Science*, vol. 41, no. 6, pp. 391–407, 1990.

[18] S. J. Delany, P. Cunningham, A. Tsymbal, and L. Coyle, “A case-based technique for tracking concept drift in spam filtering,” *Knowledge-Based Systems*, vol. 18, no. 4, pp. 187–195, 2005.

[19] A. Deo, “Diachronic semantics,” *Annual Review of Linguistics*, vol. 1, no. 1, pp. 179–197, 2015.

[20] H. Dong, F. K. Hussain, and E. Chang, “A hybrid concept similarity measure model for ontology environment,” in *On the Move to Meaningful Internet Systems: OTM 2009 Workshops*, ser. Lecture Notes in Computer Science, R. Meersman, P. Herrero, and T. S. Dillon, Eds., vol. 5872. Springer, 2009, pp. 848–857.

[21] J. L. Elman, “An alternative view of the mental lexicon,” *Trends in Cognitive Sciences*, vol. 8, no. 7, pp. 301–306, 2004.

[22] K. Erk, D. McCarthy, and N. Gaylord, “Measuring word meaning in context,” *Computational Linguistics*, vol. 39, no. 3, pp. 511–554, 2013.

[23] K. Erk and S. Padó, “A structured vector space model for word meaning in context,” in *Proceedings of EMNLP-08, 13th Conference on Empirical Methods in Natural Language Processing*, October 2008, pp. 897–906.

[24] J. A. Fodor, *The Language of Thought*. Harvard University Press, 1975, vol. 5.

[25] G. Frege, “Sense and reference,” *The Philosophical Review*, vol. 57, no. 3, pp. 209–230, 1948.

[26] S. I. Gallant, “A practical approach for representing context and for performing word sense disambiguation using neural networks,” *Neural Computation*, vol. 3, pp. 293–309, 1991.

[27] A. Gliozzo and C. Strapparava, *Semantic domains in computational linguistics*. Springer Science & Business Media, 2009.

[28] A. M. Gliozzo, M. Pennacchiotti, and P. Pantel, “The domain restriction hypothesis: Relating term similarity and semantic consistency,” in *Proceedings of NAACL-HLT-07, Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, April 2007, pp. 131–138.

[29] P. M. Gonçalves Jr and R. S. M. d. Barros, “Rcd: A recurring concept drift framework,” *Pattern Recognition Letters*, vol. 34, no. 9, pp. 1018–1025, 2013.

[30] E. Grefenstette, G. Dinu, Y.-Z. Zhang, M. Sadrzadeh, and M. Baroni, “Multi-step regression learning for compositional distributional semantics,” *arXiv:1301.6939*, 2013.

[31] J. A. Gulla, G. Solskinnsbakk, P. Myrseth, V. Haderlein, and O. Cerrato, “Concept signatures and semantic drift,” in *Web Information Systems and Technologies*. Springer, 2011, pp. 101–113.

[32] X. Han and L. Sun, “Semantic consistency: A local subspace based method for distant supervised relation extraction,” in *Proceedings of ACL-14, 52nd Annual Meeting of the Association for Computational Linguistics*, June 2014, pp. 718–724.

- [33] Z. Harris, "Distributional structure," in *Papers in structural and transformational Linguistics*, ser. Formal Linguistics, Z. Harris, Ed. New York, NY, USA: Humanities Press, 1970, pp. 775–794.
- [34] J. House, "Linguistic relativity and translation," *Amsterdam Studies in the Theory and History of Linguistic Science*, vol. 4, pp. 69–88, 2000.
- [35] J. Jiang and D. Conrath, "Semantic similarity based on corpus statistics and lexical taxonomy," in *Proceedings of ROCLING-97, International Conference on Research in Computational Linguistics*, 1997, pp. 19–33.
- [36] P. Kanerva, J. Kristofersson, and A. Holst, "Random indexing of text samples for latent semantic analysis," in *Proceedings of CogSci-00, 22nd Annual Conference of the Cognitive Science Society*, vol. 1036, 2000.
- [37] J. Karlgren and M. Sahlgrén, "From words to understanding," in *Foundations of Real-World Intelligence*, Y. Uesaka, P. Kanerva, and H. Asoh, Eds. CSLI Publications, 2001, pp. 294–308.
- [38] T. Kohonen, *Self-Organizing Maps*. Springer, 2001.
- [39] A. Kontostathis and W. Pottenger, "A framework for understanding latent semantic indexing (LSI) performance," *Information Processing and Management*, vol. 42, no. 1, pp. 56–73, 2006.
- [40] A. Lauriston, "Criteria for measuring term recognition," in *Proceedings of EACL-95, 7th Conference of the European Chapter of the Association for Computational Linguistics*, March 1995, pp. 17–22.
- [41] C. Leacock and M. Chodorow, "Combining local context and WordNet similarity for word sense identification," in *Wordnet: An Electronic Lexical Database*, C. Fellbaum, Ed. Cambridge, Massachusetts: MIT Press, 1998, pp. 265–283.
- [42] M. Lesk, "Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone?" in *Proceedings of SIGDOC-86, 5th Annual International Conference on Systems Documentation*, 1986, pp. 24–26.
- [43] I. N. Levine and P. Learning, *Quantum Chemistry*. Pearson Prentice Hall, 2009, vol. 6.
- [44] Y. Li, Z. Bandar, and S. McLean, "An approach for measuring semantic similarity between words using multiple information sources," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 4, pp. 871–882, July 2003.
- [45] C.-S. Liao, K. Lu, M. Baym, R. Singh, and B. Berger, "IsoRankN: Spectral methods for global alignment of multiple protein networks," *Bioinformatics*, vol. 25, no. 12, pp. i253–i258, 2009.
- [46] D. Lin, "An information-theoretic definition of similarity," in *Proceedings of ICML-98, 15th International Conference on Machine Learning*, June 1998, pp. 296–304.
- [47] P. W. Lord, R. D. Stevens, A. Brass, and C. A. Goble, "Investigating semantic similarity measures across the Gene Ontology: the relationship between sequence and annotation," *Bioinformatics*, vol. 19, no. 10, pp. 1275–1283, 2003.
- [48] J. McAuley and J. Leskovec, "Hidden factors and hidden topics: Understanding rating dimensions with review text," in *Proceedings of RecSys-13, 7th ACM Conference on Recommender Systems*. ACM, 2013, pp. 165–172.
- [49] L. Meng, R. Huang, and J. Gu, "A review of semantic similarity measures in WordNet," *International Journal of Hybrid Information Technology*, vol. 6, no. 1, pp. 1–12, January 2013.
- [50] P. G. Mezey, *Potential Energy Hypersurfaces*. Elsevier Amsterdam, 1987.
- [51] G. A. Miller, "WordNet: A lexical database for english," *Communications of the ACM*, vol. 38, no. 11, pp. 39–41, November 1995.
- [52] S. Mohammad and G. Hirst, "Distributional measures as proxies for semantic relatedness," 2005, submitted for publication.
- [53] J. Morris, C. Beghtol, and G. Hirst, "Term relationships and their contribution to text semantics and information literacy through lexical cohesion," in *Proceedings of CAIS-03, 31st Annual Conference of the Canadian Association for Information Science*, May 2003.
- [54] J. Morris and G. Hirst, "Lexical cohesion computed by thesaural relations as an indicator of the structure of text," *Computational Linguistics*, vol. 17, no. 1, pp. 21–48, 1991.
- [55] S. Padó and M. Lapata, "Dependency-based construction of semantic space models," *Computational Linguistics*, vol. 33, no. 2, pp. 161–199, 2007.
- [56] H. Peat and P. Willett, "The limitations of term co-occurrence data for query expansion in document retrieval systems," *Journal of the American Society for Information Science*, vol. 42, no. 5, pp. 378–383, 1991.
- [57] A. Renouf and J. Banerjee, "The search for repulsion: A new corpus analytical approach," *Towards Multimedia in Corpus Studies*, vol. 2, 2007.
- [58] P. Resnick and H. R. Varian, "Recommender systems," *Communications of the ACM*, vol. 40, no. 3, pp. 56–58, 1997.
- [59] P. Resnik, "Using information content to evaluate semantic similarity in a taxonomy," in *Proceedings of IJCAI-95, 14th International Joint Conference on Artificial Intelligence*, vol. 1, August 1995, pp. 448–453.
- [60] M. A. Rodríguez and M. J. Egenhofer, "Determining semantic similarity among entity classes from different ontologies," *IEEE Transactions on Knowledge and Data Engineering*, vol. 15, no. 2, pp. 442–456, 2003.
- [61] G. J. Ross, N. M. Adams, D. K. Tasoulis, and D. J. Hand, "Exponentially weighted moving average charts for detecting concept drift," *Pattern Recognition Letters*, vol. 33, no. 2, pp. 191–198, 2012.
- [62] D. Sánchez, M. Batet, D. Isern, and A. Valls, "Ontology-based semantic similarity: A new feature-based approach," *Expert Systems with Applications*, vol. 39, no. 9, pp. 7718–7728, 2012.
- [63] H. Schütze and T. Pedersen, "A co-occurrence-based thesaurus and two applications to information retrieval," *Information Processing and Management*, vol. 3, no. 33, pp. 307–318, 1997.
- [64] R. Socher, B. Huval, C. D. Manning, and A. Y. Ng, "Semantic compositionality through recursive matrix-vector spaces," in *Proceedings of EMNLP-CoNLL-12, Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning*. Association for Computational Linguistics, 2012, pp. 1201–1211.
- [65] R. Taira, V. Bashyam, and H. Kangarloo, "A field theoretical approach to medical natural language processing," *IEEE Transactions on Information Technology in Biomedicine*, vol. 11, no. 4, p. 364, 2007.
- [66] D. Tang, "Sentiment-specific representation learning for document-level sentiment analysis," in *Proceedings of WSDM-15, 8th ACM International Conference on Web Search and Data Mining*. ACM, 2015, pp. 447–452.
- [67] J. Trier, "Das sprachliche feld," *Neue Jahrbucher fur Wissenschaft und Jugendbildung*, vol. 10, pp. 428–449, 1934.
- [68] P. D. Turney and P. Pantel, "From frequency to meaning: Vector space models of semantics," *Journal of Artificial Intelligence Research*, vol. 37, no. 1, pp. 141–188, 2010.
- [69] A. Tversky, "Features of similarity," *Psychological Review*, vol. 84, no. 4, pp. 327–352, July 1977.
- [70] A. Ultsch and F. Mörchen, "ESOM-Maps: Tools for clustering, visualization, and classification with emergent SOM," Data Bionics Research Group, University of Marburg, Tech. Rep., 2005.
- [71] G. Varelas, E. Voutsakis, P. Raftopoulou, E. G. Petrakis, and E. E. Milios, "Semantic similarity methods in WordNet and their application to information retrieval on the web," in *Proceedings of WIDM-05, 7th Annual ACM International Workshop on Web Information and Data Management*. ACM, 2005, pp. 10–16.
- [72] C. Wang and X. Wang, "Indexing very high-dimensional sparse and quasi-sparse vectors for similarity searches," *The VLDB Journal*, vol. 9, no. 4, pp. 344–361, 2001.
- [73] S. Wang, S. Schlobach, and M. Klein, "Concept drift and how to identify it," *Web Semantics: Science, Services and Agents on the World Wide Web*, vol. 9, no. 3, pp. 247–265, 2011.
- [74] D. Widdows and K. Ferraro, "Semantic vectors: a scalable open source package and online technology management application," in *LREC-08, 6th International Conference on Language Resources and Evaluation*, May 2008.
- [75] D. Widdows and T. Cohen, "Reasoning with vectors: A continuous model for fast robust inference," *Logic Journal of IGPL*, p. jzu028, 2014.
- [76] P. Wittek, S. Darányi, and Y. Lin, "A vector field approach to lexical semantics," in *Proceedings of QI-14, 8th International Conference on Quantum Interaction*, June 2014, pp. 78–92.
- [77] P. Wittek, S. C. Gao, I. S. Lim, and L. Zhao, "Somoclu: An efficient parallel library for self-organizing maps," *arXiv:1305.1422*, 2015.
- [78] Z. Wu and M. Palmer, "Verb semantics and lexical selection," in *Proceedings of ACL-94, 32nd Annual Meeting on Association for Computational Linguistics*. Association for Computational Linguistics, 1994, pp. 133–138.
- [79] L. Yang, L. Xu, and L. He, "A CitationRank algorithm inheriting google technology designed to highlight genes responsible for serious adverse drug reaction," *Bioinformatics*, vol. 25, no. 17, pp. 2244–2250, 2009.
- [80] D. Yuret, "Discovery of linguistic relations using lexical attraction," *arXiv:cmp-1g/9805009*, 1998.
- [81] Z. Zhou, Y. Wang, and J. Gu, "New model of semantic similarity measuring in Wordnet," in *Proceedings of ISKE-08, 3rd International Conference on Intelligent System and Knowledge Engineering*, vol. 1, Nov 2008, pp. 256–261.