Concepts-Bridges: Uncovering Conceptual Bridges Based On Biomedical Concept Evolution

Kishlay Jha, Guangxu Xun, Yaqing Wang, Vishrawas Gopalakrishnan, Aidong Zhang
State University of New York at Buffalo, Buffalo, NY
{kishlayj,guangxux,yaqingwa,vishrawa,azhang}@buffalo.edu

ABSTRACT

Given two topics of interest (A and C) that are otherwise disconnected - for instance two concepts: a disease ("Migraine") and a therapeutic substance ("Magnesium") - this paper attempts to find the *conceptual bridges* (e.g., serotonin (B)) that connects them in a meaningful way. This problem of mining implicit linkage is known as hypotheses generation and its potential to accelerate scientific progress is widely recognized. Almost all of the prior studies to tackle this problem ignore the temporal dynamics of concepts. This is limiting because it is known that the semantic meaning of a concept evolves over time. To overcome this issue, in this study, we define this problem as mining time-aware Top-k conceptual bridges, and in doing so provide a systematic approach to formalize the problem. Specifically, the proposed model first extracts relevant entities from the corpus, represents them in time-specific latent spaces, and then further reasons upon it to generate novel and experimentally testable hypotheses. The key challenge in this approach is to learn a mapping function that encodes the temporal characteristics of concepts and aligns the across-time latent spaces. To solve this, we propose an effective algorithm that learns precise mapping sensitive to both global and local semantics of the input query. Both qualitative and quantitative evaluations performed on the largest available biomedical corpus substantiate the importance of leveraging temporal dynamics and suggests that the generated hypotheses are novel and worthy of clinical trials.

CCS CONCEPTS

• **Information systems applications** → **Data mining**; • **Artificial intelligence** → *Knowledge representation and reasoning*;

KEYWORDS

Hypotheses generation, temporal dynamics, word embeddings

ACM Reference Format:
Kishlay Jha, Guangxu Xun, Yaqing Wang, Vishrawas Gopalakrishnan, Aidong Zhang. 2018. Concepts-Bridges: Uncovering Conceptual Bridges Based On Biomedical Concept Evolution. In *KDD '18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, August 19–23, 2018, London, United Kingdom.* ACM, New York, NY, USA, 9 pages. https://doi.org/10.1145/3219819.3220071

KDD '18, August 19–23, 2018, London, United Kingdom

ACM ISBN 978-1-4503-5552-0/18/08…$15.00
https://doi.org/10.1145/3219819.3220071

1 INTRODUCTION

Scientific knowledge is growing at an unprecedented rate as evident from the growing body of research publications, grants, clinical trials and other scientific endeavors. A large body of this knowledge available in the free-form text has provided practitioners access to a staggering amount of information; however, at the same time, it has also made it increasingly difficult for them to keep up with the latest information, trends and findings in their field of interest in a reasonable amount of time. Imagine a researcher attempting to formulate a new hypothesis in the research area of *autism* (a serious developmental disorder). To do so, first, one has to thoroughly study and understand the existing body of literature already available. At present, a simple search in MEDLINE (a popular bibliographical database) for autism yields more than 50,000 results. While technologies based on text summarization would help users get a high level idea of the papers, it fails to stitch together disparate and seemingly uncorrelated facts together to present novel and "actionable" insights that can drive new research frontiers. Motivated by this, hypotheses generation, a sub-branch of biomedical text mining, aims at identifying non-trivial implicit assertions within a large body of documents. Simply put, the task of hypotheses generation is to answer questions like: Is there an implicit linkage between two seemingly related but explicitly disjoint topics of interest (A and C)? Consider the example shown in Figure 1. It can be observed that a direct relationship between two topics A and C might not be known/studied but there might exist an implicit linkage between them via bridging terms (B). Finding these *conceptual bridges* might reveal hitherto unknown but potentially interesting relationships. This is the crux of the problem that this paper attempts to address.

Prior studies tackle this problem through a range of solutions based on approaches such as distributional statistics [5, 23], graph theoretic measures [1, 21] and supervised machine learning techniques [12, 20]. However, in a broad sense, they are afflicted with three major drawbacks:

(1) **Rigid schema**: Almost all of the previous approaches rely on a "hard-wired" schema (e.g. graph) that results in finding only those linkages that are en route. Consequently, it risks missing the connections that are surprising or radical. More often, these radical linkages have the potential of shedding novel insights into pathways that would remain otherwise hidden.
(2) **Strict query reliance**: Existing approaches find implicit connections by strictly relying on the given input pairs; thereby ignoring the subtle cues from concepts present in their local neighbourhood.
(3) **Static domain**: The prior studies mainly assume the prevailing domain to be static; nevertheless, it is known that the domains in general (and in particular bio-medicine) are usually dynamic with new facts being added every single day [3].

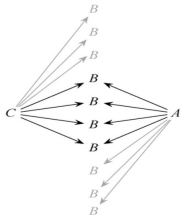

Figure 1: An Overview Schematic of Hypotheses generation

To tackle the problem of **rigid schema**, we model the problem of finding key conceptual bridges in the latent continuous space which allows us to include even those terms in our search-space that have not yet been rigorously investigated; thereby nudging the system to perform novel and radical discovery. We use the concept of word embedding techniques [4, 10] in conjunction with temporal information to identify bridge terms that have the highest likelihood of creating a meaningful connection.

The use of word-embeddings also allows us to circumvent the second issue of **strict query reliance**. Because word-embeddings project semantically similar terms closer in the vector space [10], we can leverage the terms that are deemed 'close enough' to the query to augment our search-space. This differs markedly from the classical approaches [1, 15] that find conceptual bridges by relying solely on the user provided input query terms. Their idea being that those concepts that have high semantic relatedness to both the start and end concepts (A and C) are promising candidates for bridge concepts. In this study, we extend this intuition and argue that good bridge concepts are those that, apart from being connected to A and C, are also connected to their semantically similar neighbours.

The infusion of temporal information into our word-embedding generation process enables the proposed model to be sensitive to the dynamic nature of the domain; thus alleviating the limitations of modeling it under **static domain**. While some prior studies [4, 6] have attempted to generate temporally sensitive word-embeddings, they cannot handle the current problem setting, wherein it is important for the temporal embeddings to factor in the fundamental relationship between input query and its informative local context in order to find promising conceptual bridges. To this end, we propose a new approach that allows us to first train the distributed representation of words in temporally distant time scopes and then learn a mapping function/transformation matrix being sensitive to both the global and query-specific semantics; thereby enabling the system to learn precise transformation.

Thus, our contributions can be summarized as:

(1) We propose a novel model for hypotheses generation, namely *Concepts-Bridges*, that infers implicit relations by capturing the latent evidence manifested in the temporal drift.
(2) The proposed technique for capturing temporal dynamics is sensitive to both local and global correspondence of input query, thereby capturing the semantics at a granular level.
(3) The experimental results corroborate the efficacy of the proposed model - we obtain a 20% improvement over baselines in terms of Mean Average Precision @ top-K. Qualitative evaluation of the bridge terms also validate that the hypotheses generated are plausible and worthy of further investigations.

2 RELATED WORK

Hypothesis generation from unstructured text has long been an important problem of text mining [14, 15]. This area of study in particular started gaining attention after the seminal work of Don R. Swanson in 1986 [16]. In this study, the researchers demonstrated the potential of combining facts from multiple documents to discover new knowledge. However, their approach required significant manual labor. To overcome this issue, the subsequent studies focused on automating it.

Distributional approaches: Some of the previous studies in this area of research relied on statistical analysis of concept co-occurrence (term frequency, inverse document frequency, record frequency and so on) [5, 15, 23]. Their notion being, new associations are likely to be found if the conceptual bridges are highly or rarely connected to the disparate topics of interest. However, a drawback of these approaches lie in the fact that term frequencies indicate strong but not necessarily semantically meaningful associations. Another disadvantage is their neglect of temporal dimension. This is troublesome because it is known that the semantic meaning of a concept evolves over time. Furthermore, it promptly affects domain such as bio-medicine where some new facts emerge and some are rendered obsolete every now and then.

Graph theoretic approaches: Another line of research tends to model the problem of hypotheses generation using graph based approaches [1, 2, 21]. In [21], the authors proposed a graph-based approach utilizing semantic predicates present in the form of subject-verb-object. However, their performance was tied to the accuracy and coverage of such predicate extracting tools. More recently, [1] proposed a context-driven approach wherein the sub-graphs are automatically generated for the user provided input. The essence of this study was to utilize the idea of shared context to find relevant bridge concepts. While these graph based approaches have been shown more successful than distributional approaches, they still suffer from scalability issues. Moreover, as these models rely on a rigid schema, they risk missing surprising association that are not in their route. This may be limiting because one of the main objectives of hypotheses generation is to provide users with radical (but meaningful) associations.

Machine learning based techniques: Recently, several studies [12, 20] proposed supervised machine learning based approaches to generate novel hypotheses. In [20], the authors proposed a logistic regression based model to learn the characteristic path patterns of biomedical relations to infer new linkages. The machine learning based techniques have shown the promise to find novel associations; however, a potential drawback lies in the monetary cost associated with the process of gathering training data.

Some of the motivation for this study stems from the research area of automatic language translation and temporal information retrieval [4, 22, 25]. While close in spirit, we differ from these studies in two aspects. Firstly, the goals are different. Our study focuses to capture temporal dynamics of concepts to find conceptual bridges. Secondly, our problem is more difficult in a sense that the given input is a pair of terms (instead of a single concept), and to learn accurate temporal change one has to factor in the nature of relationships between the given input pair too.

3 OVERVIEW OF PROPOSED MODEL

In this section, we outline our proposed methodology at a high level by providing the necessary intuition behind various components in our proposed model.

Recall that the input to our system is a pair of topics of interest, which we interchangeably call as query terms. Our goal is to find temporally charged top-k bridge concepts that are most likely to connect them in future. To find these concepts in a large-scale setting, we first need a text corpus collected across time. This corpus is then split into distant time scopes to obtain the collection of articles occurring within overlapping time windows. Based on this time-specific set of articles, we extract relevant entities, represent them into the latent embedding space and then reason upon it to find novel conceptual bridges. Since the focus of this study is to capture the temporal dynamics, it is important to track the semantic evolution of concepts over time. However, due to the prevailing stochastic nature of initialization for word embedding models, a direct comparison of vector spaces to quantify bio-medical concept evolution cannot be performed. To tackle this problem, we propose to learn a transformation matrix that aligns vectors spaces across time slices and thus correspondingly encapsulates the dynamics of medical concepts. Once this alignment is performed, we can capture and rank the bridge terms by their evolving proximity to the query terms in the latent embedding space.

To learn the aforementioned transformation, we propose two ways: a) Global and b) Query-biased. While the global transformation captures the more "general" information from the corpus, the query-specific transformation captures the information particular to the semantics of input query. To achieve the latter, we need a way to identify concepts similar to the input query so as to learn the transformation matrix utilizing them. This is where we leverage the principles of collaborative filtering. The importance of combining information from these two sources and the speculation that they complement each other is experimentally validated. Having learned the transformation matrices, we use them to calculate the likelihood of a concept to be potential conceptual bridge between the input query terms. Figure 2 provides a high level intuition of the proposed framework.

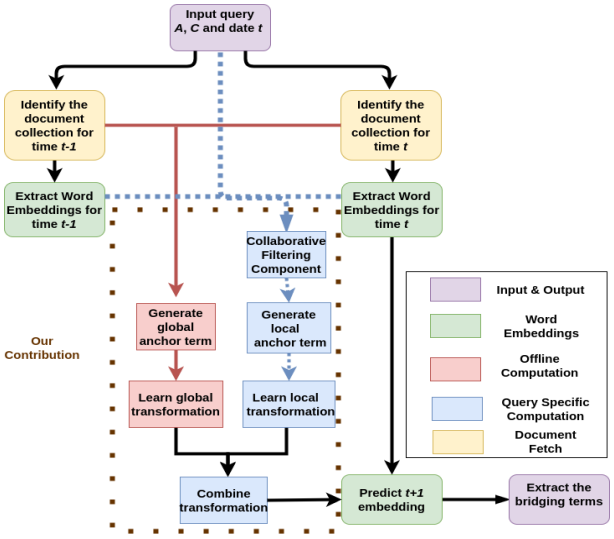

Figure 2: An Overview of the Proposed Framework

4 METHODOLOGY

This section describes our methodology in detail. It is primarily divided into three sections. Section 4.2 provides details on how the transformation matrix at a global level is learned from word-embeddings corresponding to the individual time-slices. Section 4.3 extends this idea to include the information from the local context of individual query and describes the technique to find transformation matrix in a query sensitive fashion. Having calculated both the transformation matrices, Section 4.4 calculates the ranked list of bridge terms.

4.1 Preliminaries

In this sub-section, we introduce some definitions and background information on word-embeddings.

Definition 1. Those concepts that do not change their semantic meaning over time are referred to as semantically stable concepts. An example of semantically stable concepts is "Animals". The meaning of concept "Animals" in (1850) is equivalent to its meaning in (2018).

Definition 2. Those concepts that change their meaning over time are referred to as semantically unstable concepts. An example of semantically unstable concept is "Cell". The meaning of concept cell during 1850's used to be associated with "cave", "dungeon" and "prision", however, at present (2018) it is associated with "cytoplasm", "tumor" and "epithelial cells".

Word Embeddings: To learn the distributed representation of concepts in each snapshot, we utilize a popular word embedding model, namely Continuous Bag-of-Words Model (CBOW) [10]. Given a target word w_v and its u neighboring words, the model aims at maximizing the log-likelihood of each word given its context. The objective function is shown below:

$$J = \frac{1}{V}\sum_{v=1}^{V} \log p(w_v | w_{v-u}^{v+u}) \tag{1}$$

where V refers to the overall size of Vocabulary. The probability $p(w_v|w_{v+u}^{v-u})$ is calculated as:

$$\frac{exp(e'^{\top}_{w_v} \cdot \sum_{-u \le j \le u, j \neq 0} e_{w_{v+j}})}{\sum_w exp(e'^{\top}_{w} \cdot \sum_{-u \le j \le u, j \neq 0} e_{w_{v+j}})}$$

where e_w and e'_w denote the input and output embeddings respectively. In this study, to generate word embeddings for different time slots, we first collect all the concepts occurring in the corpus and prepare an overall vocabulary. Based on this vocabulary, we train CBOW model for each consecutive time unit. The time unit is aggregated to the granularity of ten years (e.g., 1981-1990, 1982-1991 and so on) to handle the data sparsity issue. In this setting, every concept present in the vocabulary (from the beginning of time unit) has a certain position in the vector space. Then, for each consecutive time unit, we iterate over epochs and train the word vectors until convergence. As suggested by the previous studies [10, 25], the number of embedding dimension is set to 300.

4.2 Global transformation

The focus of this section is to discuss the methodology of learning a global transformation matrix. This matrix is expected to capture the global temporal dynamics of concepts present in the corpus. Another objective is to align the two different vector spaces.

Having learned the distributed representation of words on distinct snapshots through Equation 1, the next step is to learn the transformation matrix that captures temporal change and also aligns them. In this direction, the main idea in learning a global mapping is to utilize the semantically stable terms across time as anchors to bridge the two distinct vector spaces. Once the mapping is found using anchors, other semantically unstable concepts within the two spaces can be aligned by the similarity of their positions relative to the anchor terms in their own spaces. However, this gives rise to a new challenge of selecting the candidate anchor terms. To circumvent this issue, we rely on an approximate method and choose the anchor pairs based on two criteria: a) they should have same syntactic/literal form and b) they are sufficiently frequent in both the time periods. A few examples of such terms in medical domain include "humans", "animals", "male" and so on. The rationale behind choosing frequent terms as anchors is their tendency to have high degree centrality/connectedness; this causes their position in the vector space to be semantically stable [4].

For the ease of explanation, we present the technique to learn global transformation matrix using two time stamps (t_0, t_1). Formally, given P pairs of global anchor terms (w_1^0, w_1^1),...,(w_p^0, w_p^1), where w_i^0 denotes the anchor term at time t_0 and w_i^1 denotes the anchor term at time t_1 respectively. The transformation matrix M_1 is then found by minimizing the differences between $M_1 \cdot \vec{w}_i^0$ and $\vec{w}_i^1$ (See Equation 2). To prevent over-fitting, a regularization component is added to Equation 2 with γ as its corresponding weight.

$$M_1 = \arg\min_{M_1} \sum_{i=1}^{P} ||M_1.\vec{w}_i^0 - \vec{w}_i^1||_2^2 + \gamma||M_1||_2^2 \quad (2)$$

where $\vec{w}_i^0$ and $\vec{w}_i^1$ refers to the vector position of w_i^0 and w_i^1 at t_0 and t_1 respectively. In our implementation, the top 5% frequent terms in the corpus is chosen as the size of P. Both the threshold for P and $\gamma = 0.02$ is empirically set as suggested by some of the previous studies [25].

4.3 Query biased transformation

The global transformation explained in Section 4.2 is query independent; therefore, the mappings generated are not sensitive to the specific semantics of input query. This is problematic because it generates transformation matrix that neglects the fundamental relation between query and its local context. Furthermore, this also leads to insufficient characterization of temporal dynamics that in particular affects the current problem of interest, wherein the quality of conceptual bridges is highly dependent on the given input query. To overcome this issue, we propose an approach to train the transformation matrix in a query-biased way by leveraging upon the principles of collaborative filtering. The collaborative filtering provides a systematic approach to identify terms similar to the input query terms - refer Section 4.3.1. These terms act as a "seed" to the process of generating local anchors - refer Section 4.3.2.

Algorithm 1 Generate similar concepts

Given: Set of clusters $\mathbf{C}_1$, $\mathbf{C}_2$, $\mathbf{C}_3$, ..., Semantic Dictionary of medical concepts *Dict*
Input: Input concept (A) and cutoff-date (t')
Output: A set (Max Heap of N) of similar concepts which are similar to a - *setSimilarConcepts*
$setSimilarConcepts \leftarrow \phi$
$\{\mathbf{C}_A\} \leftarrow clusterLookUpOn(A)$
for $x \in \mathbf{C}_{A_i}$ **do**
    $\{cand_A\} \leftarrow$ Extract all terms that have $\mathbf{C}_x$ as the cluster with highest membership probability
end for
$\{Sem_A\} \leftarrow Dict(A)$//Get the semantic type of A
for $x \in cand_A$ **do**
    $\{Sem_x\} \leftarrow Dict(x)$//Get the semantic type of x
    $CommonSem \leftarrow$Get all types of relationships existing between $\{Sem_A \times Sem_x\}$
    if $CommonSem \neq \phi$ **then**
        $setSimilarConcepts \cup \{x\}$
    end if
end for
Return *setSimilarConcepts*

4.3.1 Generating similar concepts. Given an input concept of interest A (or C) and a date (t'), the goal is to find top-N concepts similar to the input for downstream processing. A straightforward way is to find the similar concepts by comparing the distance (in latent space) of input with each of the concepts present in the vocabulary and choosing the top-N closest neighbours. However, this becomes inefficient if the size of vocabulary scales to millions or billions. To do this in an efficient manner, we perform a soft-clustering of concepts present in the dictionary based on their word-vectors. Gaussian Mixture Model is used to perform the soft-clustering with number of clusters set to 300 as suggested in previous studies [2, 19].

Simply put, for a given input concept, we first find their respective cluster IDs and then all the concepts belonging to those clusters are added to the candidate similar set. However, this resultant set consists of concepts that are both semantically similar and semantically related to the input concept. Note that similarity calculated based on word-vectors captures both the notions of semantic similarity and relatedness [10]. This becomes problematic because in the current problem of interest we are particularly interested in finding only similar concepts. To mitigate this issue, we leverage the categorical information (known as semantic type in medical domain) of concepts. Every concept present in the vocabulary is assigned a semantic type. For example, a disease such as "Migraine" is assigned to a semantic type "Disease or syndrome"[1]. We leverage this semantic information and retain only those concepts whose explicit semantic type is same as the given input. This step allows us to distill only similar concepts. Overall, this technique allows us to efficiently identify similar concepts for any given input. Algorithm 1 provides the pseudo-code for generating similar concepts.

4.3.2 Generating Local Anchors. Having identified a set of concepts similar to input A and C (i.e., S_a, S_c), our objective is to find

[1]The explicit semantic types of medical concepts can be obtained from Unified Medical Language System.

set of anchor pairs, {<a_1, c_1>. . . <a_q, c_q>}, such that $a_i \in S_a$ and $c_i \in S_c$. To do this, a Cartesian product between terms in S_a and S_c has to be performed. However, this leads to $N \times N$ (N refers to the size of similar concept set for both A and C) comparison that is computationally expensive. Therefore, in order to find quality anchors, we define its goodness on the hypothesis that, "a good anchor pair should align well with many other good anchor pairs". This idea is inspired by the theory of PageRank. To implement this, a graph based scenario is considered where a pair is referred to as vertex (V_i') and the degree of alignment between them defines their weight. The formula to calculate the alignment between pairs is shown in Equation 3.

$$\psi_{ij} = cos((\vec{a}_i - \vec{a}_j), (\vec{c}_i - \vec{c}_j)) \quad (3)$$

where ψ_{ij} denotes the two pairs (a_i, c_i) and (a_j, c_j). Here, (a_i, a_j) $\in S_a$ (i.e., concepts similar to A) and (c_i, c_j) $\in S_c$ (i.e., concepts similar to C). The intuition behind this is that the difference in vector points of concepts captures the relational/functional alignment between concepts and it is important to preserve this geometric arrangement to precisely capture the query specific semantics.

Equation 4 is used to calculate the final weight of each pair. Specifically, the importance (λ) of each pair in the candidate set is computed in a way similar to TextRank algorithm [9] by interactively computing Equation 4 until convergence. One crucial advantage of using the idea PageRank is that it promotes pairs with higher authority; as a result, those pairs that have higher connectivity are assigned higher weights. Commonly, generic pairs tend to have higher connectivity than specific pairs. This ensures the pairs that are generic (correspondingly having relatively stable semantic meaning) and simultaneously cognizant to the semantics of input query have a higher impact on the transformation matrix being learned. Algorithm 2 provides the pseudo-code for generating anchor pairs.

$$\lambda(V_i') = (1-d) + d \sum_{V_j' \in Neigh(V_i')} \frac{\psi_{ji}}{\sum_{V_k' \in Neigh(V_j')} \psi_{jk}} * \lambda(V_j') \quad (4)$$

where $Neigh(V_j')$ denotes the neighbours of V_j' and d is the damping factor set to 0.85 by default.

4.3.3 Query biased transformation. Based on Section 4.3.1 and Section 4.3.2, we have identified a set (Q) of quality anchor pairs. Now given that, this section enumerates the process to learn the transformation that is sensitive to the relationship between input query terms. Towards this end, the model builds upon some of the special features provided by word embedding spaces such as linear analogical reasoning *vec("ibuprofen") - vec("pain") ≈ vec("treats")*. In particular, to capture the relation between anchor pair (a, c), where a is a term similar to 'A' and c is a term similar to 'C', we take the difference of their vector representations. Such linear operations are expected to capture the relational/functional aspect of input query. Our intuition behind this is to preserve the geometric arrangements of pairs in vector space that in turn is expected to encapsulate the precise temporal dynamics particular to a given query. The optimization function for learning local transformation M_2 is given in Equation 5.

$$M_2 = \arg\min_{M_2} \left(\sum_{i=1}^{Q} ||M_2.\lambda_i^0(\vec{a}_i^0 - \vec{c}_i^0) - \lambda_i^1(\vec{a}_i^1 - \vec{c}_i^1)||_2^2 + \gamma||M_2||_2^2 \right) \quad (5)$$

Algorithm 2 Generate candidate anchor pairs

Input: Set of concepts similar to A - $setSimilarConcepts(A)$ and Set of concepts similar to C - $setSimilarConcepts(C)$ (From Algorithm 1)
Output: A ranked set (Max Heap of Q) of pair of terms {a,c} which are similar to A and C - $candidateAnchors$
$candidateAnchors \leftarrow \phi$
$\{S_a\} \leftarrow setSimilarConcepts(A)$
$\{S_c\} \leftarrow setSimilarConcepts(C)$
$filteredCandidateAnchors \leftarrow \phi$
for $a \in S_a$ **do**
  for $c \in S_c$ **do**
    if ($cosine(a,c) \approx cosine(A,C)$) **then**
      $filteredCandidateAnchors \cup \{a, c\}$
    end if
  end for
end for
$tempCandidateAnchors \leftarrow \phi$
for $pair1 \in filteredCandidateAnchors$ **do**
  for $pair2 \in filteredCandidateAnchors$ **do**
    $align=calculateAlign(pair1, pair2)$ //According to equation 3
    $tempCandidateAnchors \cup \{pair1\}$
  end for
end for
$candidateAnchors=pageRankScore(tempCandidateAnchors)$ // According to equation 4
Return $candidateAnchors$

where $\vec{a}_i$ and $\vec{c}_i$ refers to the vector position of a_i and c_i at their respective time-slots. λ_i^0 and λ_i^1 are the weights associated with anchor pairs at t_0 and t_1 respectively. The λ_i in Equation 5 is the weight associated to each anchor pair based on its "goodness" as compared to other pairs (Using Equation 4). Similar to global transformation, the value of regularizer component (γ) is set to 0.02. By default, all the anchor pairs generated are chosen as the size of Q.

4.3.4 Combining with global transformation. Our contention is that the temporal change captured by both global and local transformation has valuable information and their amalgamation is necessary to find important bridge concepts. While the global transformation effectively captures the general information present in the corpus, it misses the subtle cues from the local context. On the other hand, relying solely on query specific transformation risks awarding undue importance to overly specific terms. Thus, it is important to leverage the benefits provided by two distinct but complementary transformations. Against this backdrop, we propose to combine Equation 2 and Equation 5 and jointly minimizes the following objective function. This allows us to preserve both the global and local proximity of input query simultaneously.

$$M = \alpha M_1 + (1-\alpha) M_2 \quad (6)$$

It can be observed that the final expression still results in regularized least square form. Thus, similar to solving Equation 2, we find its closed form updates and obtain the unified transformation matrix. Despite its simplicity, this concatenated approach of linear transformation method worked well in our experiments. The value of α is set to 0.5 by default.

4.4 Scoring Conceptual Bridges

Given two previously disconnected terms A and C along with a cut-off time-stamp t' (a meta-constraint to restrict the search space), the goal is to identify plausible bridge concepts k that will connect them in future (t'+1). The candidate for B terms are all the concepts present in vocabulary besides - A, $setSimilarConcepts(A)$, C and $setSimilarConcepts(C)$. Recall that our objective to find bridges concepts that are not only connected to the query pairs but also to their semantically similar local neighbours. To compute the semantic relatedness of bridges, we first learn the transformation matrix (M) particular to this input query between an initial time stamp t'_0 (by default set to t'-10) and t'. This matrix is learned by the methods described in Section 4.3.4 and is expected to encode the temporal dynamics. Note that as the goal is to predict which conceptual bridge has the highest likelihood at $t' + 1$, the corresponding embeddings ($\vec{b}^{t'+1}$, $\vec{a}^{t'+1}$ and $\vec{c}^{t'+1}$) are not available. The following formula is used to compute the likelihood score for each candidate bridge concept (b_k).

$$Score(b_k^{t'+1}) = \frac{1}{2}\{\sum_{i=1}^{N_1} cos(M.\vec{a}_i^{t'}, M.\vec{b}_k^{t'}) + \sum_{j=1}^{N_2} cos(M.\vec{c}_j^{t'}, M.\vec{b}_k^{t'})\} \tag{7}$$

where N_1 and N_2 refers to the number of neighbours of A and C, $a_i \in setSimilarConcepts(A)$ and $c_j \in setSimilarConcepts(C)$. Based on the obtained likelihood score, the candidate bridge concepts are ranked and presented to the user.

5 EXPERIMENTS

The focus of this section is to demonstrate the efficacy of the proposed model through a variety of experiments performed under different settings. In our experiments, we use MEDLINE[2] as our main corpora because it provides access to more than 100 years of time-stamped scientific articles, primarily, from life sciences and bio-medicine. The latest dump (2017) contains more than 24 million articles. Every article contains a unique identifier (PMID), title, abstract, publication date and Medical Subject Headings (MeSH) terms. As a unit of representation for articles, we choose MeSH terms. MeSH terms are the special keywords assigned by subject matter experts to each article in MEDLINE. Since these terms are selected by subject matter experts based on the full text of articles, it is safe to assume that they represent the conceptual meaning of an article without adding noise [15, 23]

DataSets: To evaluate the performance of proposed model and compare them with existing hypotheses generation algorithms, the following test cases are chosen. These test case are widely regarded as the "golden dataset" in this area of study [1, 5, 15, 21, 23]. The test cases are enumerated below:

(1) Fish-oil (FO) and Raynaud's Disease (RD) (1985)
(2) Magnesium (MG) and Migraine Disorder (MIG) (1988)
(3) Somatomedin C (IGF1) and Arginine (ARG) (1994)
(4) Alzheimer Disease (AD) Indomethacin (INN) (1989)
(5) Schizophrenia (SZ) and Calcium - Independent Phospholipase A2 (PA2) (1997)

[2]The source code of Concepts-Bridges is available at https://github.com/kishlayjha/Concepts-Bridges.

For the consideration of being self-contained, we briefly provide a background about these test cases. The pioneers in this area of study [16, 17] applied their hypotheses generation technique and postulated above enumerated hypotheses. Later, these hypotheses were clinically verified in the real world laboratories. Since then the re-discovery of these test cases is widely adopted as a way of demonstrating the effectiveness of proposed approach. Note that the dates given for above test cases acts as a threshold to base our analyses. These are the dates when the association between query terms were known and published in the literature. We run the proposed model and baseline algorithms to generate possible connections using all the articles before threshold (pre-cutoff) and then check their validity in the articles present in post-cutoff period.

Evaluation scheme: We provide both qualitative as well as quantitative validation of our approach. In qualitative evaluation, we present the top-k bridge terms and inspect their correctness. In quantitative evaluation, we compare our approach against a variety of baselines and show the superiority of our approach.

Evaluation baselines for quantitative evaluation: To evaluate effectiveness of the proposed model, the following five previous hypotheses generation algorithms are implemented. The initial four algorithms are based on raw term co-occurrence frequency and fifth is a word embedding based approach.

(1) *Apriori algorithm*: This algorithm [5] uses two important measure of association rule: a) support and b) confidence to rank the bridge concepts. The threshold for support and confidence are chosen as suggested in [5].
(2) *Chi Square ($\tilde{\chi}^2$)*: This study [8] uses Chi-square test to quantify and rank the bridge terms. The threshold for $\tilde{\chi}^2$ is used as suggested in [8].
(3) *Term-frequency and Inverse-document frequency (TF–IDF)*: TF-IDF is a popular metric that measures the importance of a concept present in an article. [15] adopts this measure to identify the bridge concepts.
(4) *Literature Cohesiveness (coh)*: Literature Cohesiveness is a metric proposed by [18], to identify bridge concepts based on the cohesion of literature.
(5) *Static embeddings (Static)*: This algorithm [7] generates cumulative year-wise co-occurence matrix and applies SVD to generate word embeddings. Based on these embeddings, the bridge terms are then ranked using cosine measure.

It should be noted that a direct comparison with the results of above enumerated baselines cannot be performed. This is because of the difference in choice of input, threshold used to select linking terms and the use of domain expertise to prepare gold standard. Nevertheless, to facilitate a fair comparison, their methods have been adjusted to fit the current problem setting.

Evaluation metrics for quantitative evaluation: Two evaluation metrics are used to quantify our results: 1) Precision@k and 2) Mean Average Precision (MAP). Precision@k allows us to measure the coverage of ground truth terms in top-k target set; thus allowing analysis at a granular level. To quantify the system's performance across queries, we report MAP.

Table 1: Top 10 conceptual bridges for all the five test cases

FO-RD	epoprostenol (PMID: 1322453)	arteriosclerosis (PMID: 1285700)	platelet aggregation (PMID: 1285700)	blood viscosity (PMID: 1313369)	alprostadil (PMID: 1322453)	prostaglandins (PMID: 1428263)	fatty acids, nonesterified (No evidence)	thrombosis (PMID:1414987)	beta-thromboglobulin (PMID: 11569481)	lipid metabolism (No evidence found)
MG-MI	basilar artery (PMID: 1559255)	cerebrovascular disorders (PMID: 1379707)	cerebral arteries (PMID: 1322512)	prostaglandins (PMID: 7816789)	carotid arteries (PMID: 1358648)	calcium antagonist (PMID: 1495818)	epilepsy (PMID: 1291896)	ergotamine (No evidence found)	hemiplegia (PMID: 1495342)	cations, monovalent (No evidence found)
AD-INN	prostaglandins (PMID: 8536873)	membrane fluidity (PMID: 8591886)	acetylcholine (PMID: 8605031)	arachidonic acid (PMID: 8567655)	propionates (PMID: 8819185)	receptors, prostaglandin (PMID: 8937434)	prostaglandin d2 (PMID: 9020023)	atrophy (PMID: 11793864)	phenylacetates (PMID: 9217884)	chorea (No evidence)
IGFI-ARG	somatomedins (PMID: 2406696)	growth inhibitors (PMID: 1381713)	somatostatin (PMID:1346379)	growth substances (PMID: 1284245)	thyrotropin (PMID: 1309347)	tyrosine transaminase (PMID: 16981136)	aminoisobutyric acids (No evidence found)	peptides (PMID: 1316907)	cycloheximide (PMID: 1309347)	hypophysectomy (1282673)
SZ-PA2	chlorpromazine (PMID: 2516320)	schizophrenic psychology (PMID: 29105546)	oxidative stress (PMID: 29037473)	lysophosphatidyl cholines (PMID: 11704897)	phosphatidic acids (PMID: 10509868)	psychotropic drugs (PMID: 1282673)	phospholipid ethers (PMID: 28152600)	diglycerides (PMID: 24565079)	phosphatidylserines (PMID: 15219471)	lysophospholipids (PMID: 25849980)

5.1 Qualitative evaluation

In this section, we evaluate our proposed model based on its ability to rediscover the existing knowledge.

Fish-Oil - Raynaud's Disease: In this test case, the pioneers identified that fish oils might prevent raynaud disease by a) inhibiting platelet aggregation, b) reducing blood viscosity and c) preventing vasoconstriction (epoprostenol) [16] and reported them in an article in 1986. These *conceptual bridges* were later experimentally validated. In our experiments, we seed our algorithm with input pairs (A,C) as ("fish oils", "Raynaud disease") and a date (t') as 1985. The results (Top-10 terms) for this and all other test cases are reported in Table 1. As it can be observed, all the significant conceptual bridges for this test case are found in top five.

Migraine - Magnesium: The objective of this test case was to examine the effect of magnesium in treating migraine disorder. Similar to the previous case, several intermediate terms such as *epilepsy, serotonin, prostaglandins, platelet aggregation, calcium antagonist, type A personality, vascular tone and reactivity, calcium channel blockers, spreading cortical depression and substance P* were reported. Unlike the previous case, we are unable to achieve high recall. Nevertheless, we obtain important conceptual bridges such as *epilepsy*, *calcium antagonist*, *prostaglandins*, etc. Note that the previous studies indicate this to be a difficult test case [15].

Indomethacin - Alzheimer Disease: The most significant pathways reported for this case are *Acetylcholine* and *Membrane fluidity*. Both of these pathways were found in top five.

Somatomedin C - Arginine: For this test case, *Somatotropin* and *somatostain* are the most important pathways [18]. In our results, we were able to obtain both of them in top five.

Schizophrenia - CI Phospholipase A2: The initial studies reported oxidative stress to be the key connecting term for this test case. In our results, we found Dopamine Receptors (a derivative of oxidative stress at rank 3).

Overall, the proposed model was able to identify a majority of true connections at top ranks, however, a related questions arises: How novel are the other top terms reported? To this end, the following paragraph presents a discovery example based on terms reported in our Top-10.

Discovery Example: For the first test- case (FO-RD), one of the term reported in Top-10 was *beta-thromboglobulin*. Beta-thromboglobulin is a platelet-specific protein that is released when platelets aggregate. Manually inspecting the literature, we found that an article [13] in 2001 reported the potential role of beta-thromboglobulin in preventing endothelial cell damage that is known to cause Raynaud's disease. Although prior to 1986 there was no reported connection, the proposed model could identify it by analyzing existing connections in the medical literature. Similarly, for another test-case of INN-AD, one of the top ranked connecting term was *Phenylacetates*. More recently, [11] reported the potential role of Phenylacetates in treatment for Alzheimer Disease. While these connections are being reported recently in the literature, the model was able to identify them much in advance. We believe one reason for this lies in the choice of modelling in latent space that enables the algorithm to find connections that might be surprising at the time of being postulated. To further aid the biomedical scientists in conducting extensive study, we provide evidence for our top 10 terms (refer Table 1) in the form of PMID.

Based on the rediscovery of existing knowledge and aforementioned discovery scenario, it can be deduced that the model is able to replicate already known knowledge and possibly originate new knowledge. However, this form of evidence based evaluation does not inform us about the overall quality of result set. To this end, a quantitative evaluation has to be performed.

5.2 Quantitative evaluation

The purpose of this section is to probe the overall quality of output generated. However, to perform a quantitative analysis certain ground truth is required. Unfortunately, there is no standard ground truth available and creating one remain an open problem [24]. One reason behind this is the fact that it is near-impossible to build a comprehensive ground truth set that will presumably have all the future discoveries. Therefore, a "supposedly" ground truth has to be constructed. To accomplish this goal, a split corpus approach is adopted. Specifically, the dataset is divided into two sets: 1) Pre-cut-off segment: this includes articles published before the cut-off date and 2) Post-cut-off segment: this includes articles published after the cut-off date. The proposed model and baseline algorithms are run on the pre-cut-off segment. Then, the generated connections are checked in the post-cut-off segment. The legitimacy of a connection is defined as its presence (co-occurence) in post-cut-off segment and absence in pre-cut-off. Equation 8 presents the formula to rank ground truth bridge term k for a given pair (A, C).

$$gt(k) = \frac{\#(k, A) + \#(k, C)}{\#(k)}, \tag{8}$$

where $\#(i, j)$ is the number of times terms i and j co-occur and $\#(i) = \sum_j \#(i, j)$. In this way, a ranked set of ground truth is constructed. As a post-processing step, all the stop-words (also referred to as check-tags in medical domain) are removed from the resultant set.

Table 2: Precision@k for FO-RD

Algorithm	k=10	k=20	k=30	k=40	k=50
apriori	0.5	0.6	0.6	0.55	0.54
TF-IDF	0.4	0.6	0.567	0.55	0.54
$\tilde{\chi}^2$	0.4	0.4	0.567	0.475	0.54
coh	0.6	0.5	0.467	0.5	0.5
static	0.2	0.35	0.467	0.45	0.46
Concepts-Bridges	**0.8**	**0.7**	**0.667**	**0.625**	**0.62**

Table 3: Precision@k for MG-MIG

Algorithm	k=10	k=20	k=30	k=40	k=50
apriori	0.7	0.65	0.7	0.675	0.64
TF-IDF	0.8	0.65	0.7	0.66	0.66
$\tilde{\chi}^2$	0.4	0.55	0.667	0.6	0.62
coh	0.5	0.45	0.5	0.525	0.54
static	0.5	0.55	0.633	0.675	0.66
Concepts-Bridges	**0.8**	**0.8**	**0.733**	**0.725**	**0.7**

Table 4: Precision@k for AD-INN

Algorithm	k=10	k=20	k=30	k=40	k=50
apriori	0.6	0.7	0.8	0.75	0.66
TF-IDF	0.5	0.55	0.7	0.75	0.7
$\tilde{\chi}^2$	0.6	0.65	0.667	0.675	0.64
coh	0.6	0.7	0.7	0.7	0.66
static	0.7	0.65	0.7	0.675	0.7
Concepts-Bridges	**0.9**	**0.85**	**0.833**	**0.825**	**0.8**

Table 5: Precision@k for IGF1-ARG

Algorithm	k=10	k=20	k=30	k=40	k=50
apriori	0.8	0.85	0.833	0.725	0.7
TF-IDF	0.5	0.45	0.467	0.575	0.64
$\tilde{\chi}^2$	0.6	0.7	0.7	0.7	0.7
coh	0.8	0.85	0.833	0.825	0.7
static	0.6	0.45	0.433	0.525	0.58
Concepts-Bridges	**0.9**	**0.9**	**0.867**	**0.85**	**0.84**

Results: Table 2, 3, 5, 4, 6 reports the Precision@k for each of the five golden datasets. The value of K is gradually increased from 10 to 50 (in the interval of 10) and results are reported. Table 7 reports the Mean Average Precision @k by consolidating numbers across different datasets.

Discussion: It can be observed that the proposed model outperforms all the existing baselines. Across all the datasets, a common pattern noticed for the proposed model is the decrease in precision with the increase in value of K. In contrast, for baseline algorithms the precision increases (in general) with increase in value of K. This trend elucidates the advantage of proposed model to rank relevant connections at higher positions. Analyzing the results further, we observe that Literature Cohesivenes (COH) performs the best among all the baselines. Perhaps, the reason for this lies in the ability of COH to leverage the cohesion of literature effectively.

Another important point to note is that pure frequency based approaches (Top 4 baselines) boosts contextually generic terms at higher positions. Contextually generic terms are those terms that are generic to the given input query. For instance, in the Fish Oils and Raynaud's disease test case, some of the top terms found for

Table 6: Precision@k for SZ-PA2

Algorithm	k=10	k=20	k=30	k=40	k=50
apriori	0.6	0.75	0.767	0.825	0.82
TF-IDF	0.4	0.6	0.7	0.75	0.78
$\tilde{\chi}^2$	0.5	0.7	0.767	0.825	0.86
coh	**1.0**	0.95	**0.967**	0.85	0.82
static	0.4	0.6	0.7	0.775	0.78
Concepts-Bridges	**1.0**	**1.0**	**0.967**	**0.95**	**0.92**

Table 7: Mean Average Precision@k for all test cases

Algorithm	k=10	k=20	k=30	k=40	k=50
apriori	0.616	0.667	0.702	0.728	0.744
TF-IDF	0.538	0.571	0.594	0.615	0.632
$\tilde{\chi}^2$	0.477	0.548	0.587	0.605	0.618
coh	0.731	0.723	0.725	0.723	0.725
static	0.442	0.487	0.528	0.552	0.570
Concepts-Bridges	**0.907**	**0.907**	**0.860**	**0.847**	**0.836**

COH (and other baselines) are "double-blind method", "skin ulcer", "leg ulcer" and so on. Although these terms have relatively lower overall frequency they tend to frequently co-occur with the input query (i.e., Raynaud's disease). Selecting these terms prove counterproductive as they are ranked lower in the ground truth. The reason being, these contextually generic terms have no functional relationship with the input concept. Note that more often the true conceptual bridges have important functional relationship with input query. For example: Fish oils $\xrightarrow{\text{disrupts}}$ platelet aggregation $\xrightarrow{\text{cause}}$ Raynaud's disease. Furthermore, as the fifth baseline (Static embeddings) too does not factor in the "functional" aspect, it suffers from this issue. To mitigate these issue, the proposed model (in particular query-specific transformation component) takes advantage of the analogical relationships provided by word embedding spaces to capture the functional component of medical concepts.

Another reason for the lower performance of baselines lies in the fact that they strictly rely on the given input query (ignoring the cues from local neighbourhood). This is limiting because more often when a potential conceptual bridge (e.g. "platelet aggregation") is being studied/reported in the literature with a particular concept (e.g. "fish oils"), it is highly likely that it is also being reported with the chemical substances/genes associated with them. In this case, the chemical substance being "eicosapentaenoic acid". Models based on strict query reliance attempt to find bridge concepts ("platelet aggregation") by only considering the semantic association with particular input concept ("fish oils"). Ignoring such semantically similar neighbours ("eicosapentaenoic acid") may limit the capability of model to find potential bridge concepts. Note that some of the existing approach [1] manually augment their input query to enrich their relevant document set. However, this requires the user to possess some form of domain knowledge. In the proposed approach, the use of word embeddings automatically enables to find semantically similar concepts that enriches the user provided input queries. Lastly, the fifth baseline chosen for comparison is Static embeddings (Static). This baseline ranks the bridge concepts based on the static embeddings generated from cumulative co-occurrence

Table 8: Effect of global and local transformation. MAP@K

Algorithm	k=10	k=20	k=30	k=40	k=50
global	0.728	0.715	0.697	0.688	0.657
local	0.816	0.797	0.782	0.781	0.764
Concepts-Bridges	**0.907**	**0.907**	**0.860**	**0.847**	**0.836**

matrix. Our intend behind this is to test the necessity of leveraging temporal dynamics itself. Static essentially assumes a static world in which each term is supposed to retain its semantics across different domains. As reported in the results, we can see that the proposed approach outperforms it. This result suggests that it is crucial to consider the temporal change of concepts in order to generate semantically sensible hypotheses.

5.3 Effect of global and local transformation

The only parameter in the proposed approach is the α in Equation 6. The α parameter controls the contribution of global and local transformation. Table 8 compares the influence of each transformation in the form of MAP@k calculated for all the five test cases. As can be seen, the local transformation outperform global transformation. We believe the reason for this lies in the ability of local transformation to encode query-specifics semantics in an effective manner. Furthermore, the best result comes from combination of both global and local, thus validating the need for Equation 6.

6 CONCLUSIONS

In this study, we proposed a new model to discover conceptual bridges between two disparate but complementary topics of inquiry. Specifically, the model leverages upon the temporal information present in the corpus and captures the semantic change of medical concepts at a coarse-grained level. The proposed query-biased transformation technique, in particular, leverages the fundamental relationship between input query and its informative neighbours to encapsulate precise semantics. This enables the model to promote those conceptual bridges that have higher semantic meaning. Empirically, we evaluate the model in a variety of experimental settings. The experimental results demonstrate that the proposed model has the potential of generating practical new knowledge. In future research, we intend to add more semantic expressiveness to our generated hypotheses. Towards this end, we are looking at more specialized biomedical resources such as SEMMEDDB - a repository of semantic predications in the form of 'subject-predicate-object'.

ACKNOWLEDGMENT

This work was supported in part by the US National Science Foundation under grants NSF IIS-1218393 and IIS-1514204. 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.

REFERENCES

[1] D. Cameron, R. Kavuluru, T. C. Rindflesch, A. P. Sheth, K. Thirunarayan, and O. Bodenreider. 2015. Context-driven automatic subgraph creation for literature-based discovery. *J Biomed Inform* 54 (Apr 2015), 141–57.

[2] Vishrawas Gopalakrishnan, Kishlay Jha, Guangxu Xun, Hung Q Ngo, and Aidong Zhang. 2017. Towards Self-Learning Based Hypotheses Generation in Biomedical Text Domain. *Bioinformatics* (2017).

[3] Anika Groß, Cédric Pruski, and Erhard Rahm. 2016. Evolution of biomedical ontologies and mappings: Overview of recent approaches. *Computational and structural biotechnology journal* 14 (2016), 333–340.

[4] William L. Hamilton, Jure Leskovec, and Dan Jurafsky. 2016. Diachronic Word Embeddings Reveal Statistical Laws of Semantic Change. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, ACL 2016, August 7-12, 2016, Berlin, Germany, Volume 1: Long Papers*.

[5] Xiaohua Hu, Xiaodan Zhang, Illhoi Yoo, Xiaofeng Wang, and Jiali Feng. 2010. Mining hidden connections among biomedical concepts from disjoint biomedical literature sets through semantic-based association rule. *International Journal of Intelligent Systems* 25, 2 (2010), 207–23.

[6] Vivek Kulkarni, Rami Al-Rfou, Bryan Perozzi, and Steven Skiena. 2015. Statistically significant detection of linguistic change. In *Proceedings of the 24th International Conference on World Wide Web*. ACM, 625–635.

[7] Jake Lever, Sitanshu Gakkhar, Michael Gottlieb, Tahereh Rashnavadi, Santina Lin, Celia Siu, Maia Smith, Martin Jones, Martin Krzywinski, and Steven J Jones. 2017. A collaborative filtering based approach to biomedical knowledge discovery. *Bioinformatics* (2017).

[8] Guangrong Li and Xiaodan Zhang. 2011. Mining Biomedical Knowledge Using Mutual information ABC. In *Granular Computing (GrC), 2011 IEEE International Conference on*. 848–50.

[9] Rada Mihalcea and Paul Tarau. 2004. Textrank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*.

[10] Tomas Mikolov, Kai Chen, Greg Corrado, and Jeffrey Dean. 2013. Efficient Estimation of Word Representations in Vector Space. *CoRR* abs/1301.3781 (2013).

[11] N Yi Mok, James Chadwick, Katherine AB Kellett, Eva Casas-Arce, Nigel M Hooper, A Peter Johnson, and Colin WG Fishwick. 2013. Discovery of biphenylacetamide-derived inhibitors of BACE1 using de novo structure-based molecular design. *Journal of medicinal chemistry* 56, 5 (2013), 1843–1852.

[12] Shengtian Sang, Zhihao Yang, Zongyao Li, and Hongfei Lin. 2015. Supervised learning based hypothesis generation from biomedical literature. *BioMed research international* 2015 (2015).

[13] Andreina Poggi Stefania Muti Giuseppe Bonapace Franco Argentati Claudio Cervini Ferdinando Silveri, Rossella De Angelis. 2001. Relative roles of endothelial cell damage and platelet activation in primary Raynaud's phenomenon (RP) and RP secondary to systemic sclerosis. *Scandinavian journal of rheumatology* 30, 5 (2001), 290–296.

[14] Scott Spangler, Angela D Wilkins, Benjamin J Bachman, Meena Nagarajan, Tajhal Dayaram, Peter Haas, Sam Regenbogen, Curtis R Pickering, Austin Comer, Jeffrey N Myers, et al. 2014. Automated hypothesis generation based on mining scientific literature. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1877–1886.

[15] P. Srinivasan and B. Libbus. 2004. Mining MEDLINE for implicit links between dietary substances and diseases. *Bioinformatics* 20 Suppl 1 (Aug 2004), i290–96.

[16] Don R Swanson. 1986. Fish oil, Raynaud's syndrome, and undiscovered public knowledge. *Perspectives in biology and medicine* 30, 1 (1986), 7–18.

[17] Don R Swanson. 1988. Migraine and magnesium: eleven neglected connections. *Perspectives in biology and medicine* 31, 4 (1988), 526–557.

[18] Don R Swanson, Neil R Smalheiser, and Vetle I Torvik. 2006. Ranking indirect connections in literature-based discovery: The role of medical subject headings. *Journal of the Association for Information Science and Technology* 57, 11 (2006), 1427–1439.

[19] Xing Wei and W Bruce Croft. 2006. LDA-based document models for ad-hoc retrieval. In *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*. ACM, 178–185.

[20] D. Weissenborn, M. Schroeder, and G. Tsatsaronis. 2015. Discovering relations between indirectly connected biomedical concepts. *J Biomed Semantics* 6 (2015), 28.

[21] B. Wilkowski, M. Fiszman, C. M. Miller, D. Hristovski, S. Arabandi, G. Rosemblat, and T. C. Rindflesch. 2011. Graph-based methods for discovery browsing with semantic predications. *AMIA Annu Symp Proc* 2011 (2011), 1514–23.

[22] Guangxu Xun, Kishlay Jha, Vishrawas Gopalakrishnan, Yaliang Li, and Aidong Zhang. 2017. Generating Medical Hypotheses Based on Evolutionary Medical Concepts. In *Data Mining (ICDM), 2017 IEEE International Conference on*. IEEE, 535–544.

[23] Meliha Yetisgen-Yildiz and Wanda Pratt. 2006. Using statistical and knowledge-based approaches for literature-based discovery. *Journal of biomedical informatics* 39, 6 (2006), 600–611.

[24] Meliha Yetisgen-Yildiz and Wanda Pratt. 2009. A new evaluation methodology for literature-based discovery systems. *Journal of biomedical informatics* 42, 4 (2009), 633–643.

[25] Yating Zhang, Adam Jatowt, Sourav Bhowmick, and Katsumi Tanaka. 2015. Omnia mutantur, nihil interit: Connecting past with present by finding corresponding terms across time. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, Vol. 1. 645–655.