

Supporting Shared Hypothesis Testing in the Biomedical Domain

Asan Agibetov^{1,6}, Ernesto Jiménez-Ruiz², Marta Ondršík^{3,4}, Alessandro Solimando⁵, Imon Banerjee^{1,7}, Giovanna Guerrini^{5*}, Chiara E. Catalano¹, Joaquim M. Oliveira^{3,4}, Giuseppe Patanè¹, Rui L. Reis^{3,4} and Michela Spagnuolo¹

Abstract

Background: Pathogenesis of inflammatory diseases can be tracked by studying the causality relationships among the factors contributing to its development. We could, for instance, hypothesize on the connections of the pathogenesis outcomes to the observed conditions. And to prove such causal hypotheses we would need to have the full understanding of the causal relationships, and we would have to provide all the necessary evidences to support our claims. In practice, however, we might not possess all the background knowledge on the causality relationships, and we might be unable to collect all the evidence to prove our hypotheses.

Results: In this work we propose a methodology for the translation of biological knowledge on causality relationships of biological processes and their effects on conditions to a computational framework for hypothesis testing. The methodology consists of two main points: hypothesis graph construction from the formalization of the background knowledge on causality relationships, and confidence measurement in a causality hypothesis as a normalized weighted path computation in the hypothesis graph. In this framework, we can simulate collection of evidences and assess confidence in a causality hypothesis by measuring it proportionally to the amount of available knowledge and collected evidences.

Conclusions: We evaluate our methodology on a hypothesis graph that represents both contributing factors which may cause cartilage degradation and the factors which might be caused by the cartilage degradation during osteoarthritis. Hypothesis graph construction has proven to be robust to the addition of potentially contradictory information on the simultaneously positive and negative effects. The obtained confidence measures for the specific causality hypotheses have been validated by our domain experts, and, correspond closely to their subjective assessments of confidences in investigated hypotheses. Overall, our methodology for a shared hypothesis testing framework exhibits important properties that researchers will find useful in literature review for their experimental studies, planning and prioritizing evidence collection acquisition procedures, and testing their hypotheses with different depths of knowledge on causal dependencies of biological processes and their effects on the observed conditions.

Keywords: biomedical ontology; ontology mappings; network analysis; hypothesis testing; incomplete knowledge

Background

Diseases and pathologies may be evidenced across multiple biological scales (*e.g.*, cellular, molecular, organic, behavioural) as a set of factors, linked among each other via causal relationships, which constitute the multi-scale pathological cascade reactions. To study the underlying causation mechanism of a certain disease, life science researchers rely on various sources, such as (i) current knowledge (*e.g.* previously published studies from the field), (ii) their data deduced from empirical analysis of laboratory experiments (*e.g.*, gene analysis, immuno-assays, cell viability assays, histology) or other tests (*i.e.* mechanical tests, imaging, gait analysis), as well as on (iii) consultations with other fields (*i.e.* related research areas, hospitals). To effectively make and test (prove or reject) a causality hypothesis life science research studies face two challenges: i) the information used in research processes comes from various sources and is heterogeneous, which makes it hard to organize, analyze, and assess their relevance in the overall disease process, ii) researchers from different fields (*i.e.* molecular biologist, mechanobiologist, orthopaedists *etc.*) investigate the same pathological event from different aspects (biological scales), and might not be aware of the overlaps and the impact of their individual findings in a joint venture of understanding causality mechanisms of pathologies and diseases.

To better convey the idea of causality hypothesis testing we will focus on knee articular cartilage degeneration during the onset of osteoarthritis (OA) to present our use-case scenario. OA is a joint degenerative disease and can be caused due to several factors, such as genetic predisposition, joint overuse, previous injury to the joint. The effect of these factors is hallmarked with a complete joint breakdown and dysfunction,

causing a lot of pain [1, 2]. Based on common knowledge, performed experiments, and diagnosis the causality relation of certain factors to the development of OA might have different degrees of confidence. On the one hand, the degeneration of cartilage, synovial thickening, osteophyte formation and joint space narrowing, are known to be as the most marked features of OA [3, 4, 5, 6]. On the other hand, for some factors we may have lower degrees of confidence in their causality relationship to OA. For instance, while being common in patients with OA, the exact causality relation of inflammation to OA is not completely understood [7, 8]. To handle such scenarios of causality hypothesis testing, we propose to translate what we observe in the biology into a computational framework, which supports the researchers in their hypothesis testing. In such a framework we systematically translate our background knowledge on causality relationships into the representations suitable for the computation, and we quantify confidences in our hypothesis with respect to the amount of evidences that we can supply to the framework.

Hypothesis testing

Schematically, the causality relationships between the factors of diseases can be represented as directed causality networks $H_{0\dots n}$, where factors f_i are represented as nodes and the causality relationships as arcs (f_i, f_j) . For instance, our hypothesis H_0 can state that inflammation contributes to the development of OA, where the inflammation is the cause of biological processes which lead to cartilage degradation (factor f_2 , Figure 1) and finally manifest in joint deformation condition (factor f_3 , Figure 1). To prove such a causality hypothesis we need to evidence the instances of all the participating factors. For example, the factors f_2, f_3 are evidenced as the results of diagnosis of OA done by radiologists and orthopaedists using imaging techniques (*i.e.* magnetic resonance-MRI, X-ray).

*Correspondence: giovanna.guerrini@unige.it

⁵University of Genoa, Genoa, Italy

Full list of author information is available at the end of the article

By studying the literature we can discover that the inflammation can be characterized by the detection of high levels of pro-inflammatory factors in the synovial cavity, and in particular tumor necrosis factor alpha (TNF α) (factor f_1 in Figure 1), was demonstrated to be present in excess during OA [9]. A justification or evidence for the factor f_1 (evidence of f_1 in Figure 1) can be obtained with molecular biological techniques screening the biomarkers of the synovial fluid. Given our knowledge of the participating biological processes (hypothesis H_0) and the supporting evidences (evidences for factors f_1, f_2, f_3) we have a certain level of confidence that the synovial inflammation has been the cause of the development of OA. However, is our hypothesis H_0 complete enough, and are the evidences for factors (f_1, f_2, f_3) enough to support our hypothesis? Have we missed other factors? Have we been complete enough in our characterization of all the participating factors which support the hypothesis that the synovial inflammation has been the cause of cartilage degradation? Is the joint deformation the only consequence of such a pathological cascaded of reactions?

Studying further the causality mechanism of OA, we can refine our initial hypothesis H_0 . In particular, cellular biological studies observed that TNF α facilitates the catabolic processes of the chondrocytes, including the production of matrix metalloproteinases (MMPs), and the production of aggrecanases (members of the ADAMTs family) [10, 11]. The MMPs, especially MMP-13 and aggrecanases are proteases responsible for the degradation of collagen macromolecules and proteoglycans respectively, as evidenced in literature [12]. Collagens and proteoglycans are the main building blocks of articular cartilage. Accordingly, the excess of TNF α in the joint space can be associated to the disruption of biochemical balance in the cartilage. Factors: Loss of collagen and proteoglycan molecules (factors f_4, f_5 in Figure 2), are caused

by the action of matrix degrading proteases, and can be attached to higher scales in the OA processes, such as the mechanical functioning of cartilage. These factors can be evidenced on the tissue level by histology and immuno-histochemistry (evidences of f_4, f_5 in Figure 2). Collaborations with mechano-biological fields allow the detection of the changes in cartilage mechanical properties due to the effect of high levels of MMPs and aggrecanses [13, 14]. It has been shown previously that once the cartilage suffers collagen loss, it is no longer able to withstand the mechanical forces in the knee [15, 16]. Consequently, the cartilage, the trabecular bone beneath it, and all surrounding tissue components suffer damage, which can be evidenced by imaging [17, 18]. Damage to the joint components, will cause pain, joint deformation and loss of function, which is a subject of behavioural scales and can be evidenced by gait analysis [19].

The relationship between inflammation and OA is even more complex, than the example brought above. Nonetheless, collaborations among medical doctors and bench researchers of various fields can reveal the connections between molecular evidence and those observed on organ scale. Accordingly, we can refine our hypothesis by adding new causal relationships.

Shared hypothesis testing framework

In this work we propose a methodology for the translation of biological knowledge on causality relationships of biological processes and their effects on conditions to a computational framework for hypothesis testing. The methodology consists of two main points: hypothesis graph construction from the formalization of the background knowledge on causality relationships, and confidence measurement in a causality hypothesis as a normalized weighted path computation in the hypothesis graph. In this framework, we can simulate collection of evidences and assess confidence in a causality hypothesis by measuring it proportionally to the

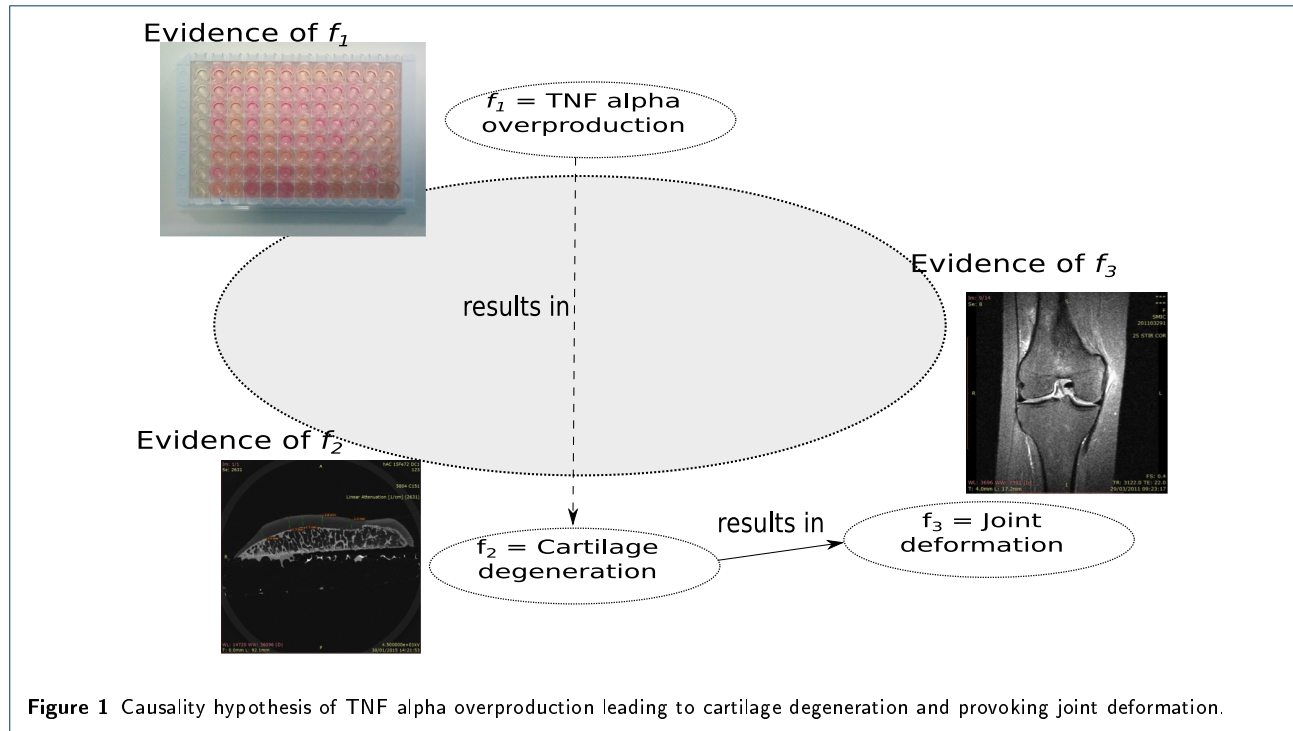


Figure 1 Causality hypothesis of TNF alpha overproduction leading to cartilage degeneration and provoking joint deformation.

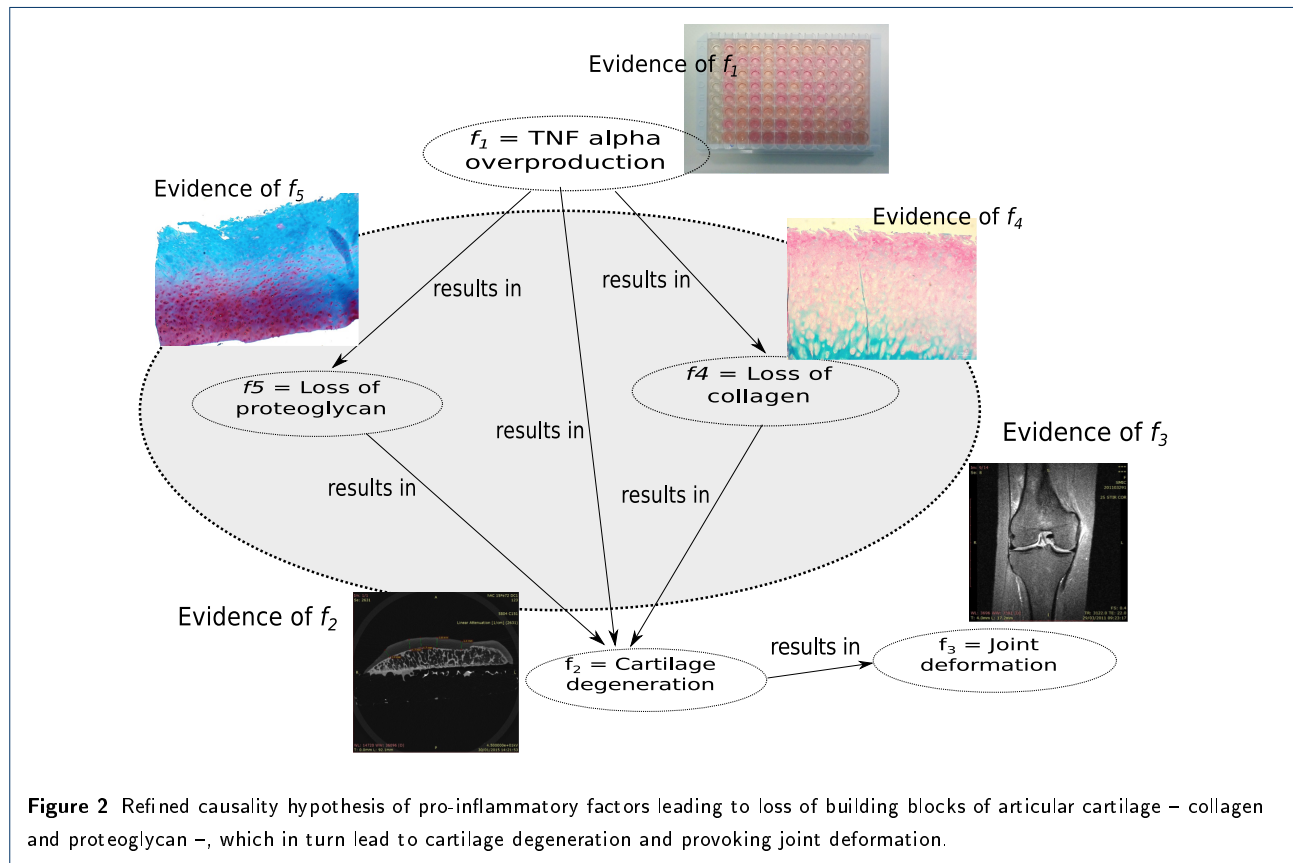


Figure 2 Refined causality hypothesis of pro-inflammatory factors leading to loss of building blocks of articular cartilage – collagen and proteoglycan –, which in turn lead to cartilage degeneration and provoking joint deformation.

amount of available knowledge and collected evidences. We evaluate our method on an example causality hypothesis of factors which cause and, in turn, may be caused by cartilage degeneration during osteoarthritis. The results of the evaluation and the feedback from the domain experts allow us to conclude that our methodology may simulate the execution of evidence collection, and can be used as a means of measuring the confidence in a causality hypothesis with respect to the amount of knowledge on causality relationships among participating factors. Such simulation supports the researchers in the planning and in the prioritization of their next studies by identifying important factors in a causality hypothesis. Our methodology demonstrates robustness towards the addition of potentially inconsistent knowledge by separately representing opposite causality possibilities for complementary biological scenarios.

We would like to emphasize that the contribution of this work is the methodology to extract the causality information from the input ontologies into a hypothesis graph, and perform hypothesis testing on the obtained hypothesis graph. The ontologies and the ontology mappings discussed and provided are created together with the domain experts, and in the context of this work are only meant to serve as proof of concept.

Related work

To the best of our knowledge the proposed methodology to test a causality hypothesis in a collaborative setting with respect to the amount of knowledge available for the framework does not have an equivalent methodology or an implemented system to test against, in its entirety. However, once decomposed, our methodology can be compared on specific steps and modelling choices.

Formalization of background knowledge on a causal hypothesis as ontologies. Our methodology for causality hypothesis testing relies on the formalization of the background knowledge on a hypothesis with ontologies. Indeed, to facilitate knowledge sharing and increase understanding of the method in use, it is common to employ already existing ontologies that are well agreed on in the biomedical community (e.g., Gene Ontology [20]). The most widely used ontology modeling language is the (OWL 2) [21], based on formal logic [22]. The main advantage of using logic over alternative representation mechanisms is that logic provides an unambiguous meaning to ontologies. We assume that the input ontologies to our framework focus on (biological) processes and findings (i.e., laboratory tests) that are or may be linked via a causality relationship, and other (material) entities that (actively or passively) participate in the process or finding. In this work we assume that the input ontologies follow good practices and relevant ontology classes are either subsumed by or annotated with, for example, the concept `Biological_process` (key concept in the *Gene Ontology* [20]) or `Finding` (e.g., common semantic type in the *UMLS semantic network* [23]). We expect the following (object) properties or its potential subproperties as source for causality relationships: `causes`, `results in`, `regulates`, `positively regulates`, `negatively regulates`, `increases levels of` and `decreases levels of`. Most of these properties are available in the *Relations ontology* [24] and are extensively used in biomedical ontologies. We reuse the domain independent categories *Continuant* and *Occurrent*, which are commonly used in the literature (e.g., River Flow Model of Diseases (RFM) [25]) and in upper ontologies (e.g., DOLCE [26] and BFO [27]). For example, processes and findings are typically classified as *occurents*, while material entities as *continuants*.

Graph projection of OWL ontologies. The hypothesis graph construction heavily relies on the graph projection of OWL ontologies. This procedure, at its core, transforms an OWL ontology into its graph representation, by studying the axiomatic structure of the ontology and identifying nodes and edges (arcs) of its equivalent graph representation. Implicitly, Lembo *et al.* [28] use graph projections of OWL QL to propose ontology classification algorithm, which transforms OWL QL ontologies into directed graphs, and computes subsumption relations via transitive closure computation. Analogously, Seidenberg *et al.* [29] use graph representation of ontologies to propose a segmentation algorithm based on subgraph extraction procedure. Some of the proposed methodologies for graph projection of OWL ontologies draw their inspirations from Social Network Analysis (SNA) [30] for the representation of the encoded semantic information in an OWL ontology. SNA is the process of investigating social structures of connected information/knowledge entities through the use of network and graph theories. SNA techniques application to ontology analysis has been pioneered by Hoser *et al.* [31], where standards in SNA community graph metrics based on: node degree, node betweenness and on eigenanalysis of the adjacency matrix, were used to study properties of ontologies. The connection between SNA and ontology analysis have also been studied in a highly cited paper by Mika [32], bridging Social Networks and Semantics. Network partitioning algorithms have been used by Stuckenschmidt *et al.* [33] to identify islands of ontology, a notion comparable to a module of ontology (as used by the graph-based modular extraction community), with the applications to Visual Analytics. Grontocrawler [34] transforms OWL-EL [35] ontologies into networks by defining a rule-based edge production procedure, which takes into account existential and values restrictions on object re-

lations. Formal treatment of rule-based graph projection procedures and their connection to the logical entailment problem for OWL 2 ontologies have been recently proposed [36, 37, 38]. In our work we use Grontocrawler [34] for graph-based ontology projection, enriched with the projection of advanced OWL 2 axioms, as suggested in Soylu *et al.* [38].

Rule-based reasoning with incomplete knowledge in the biomedical domain. Similarly to previous works [39, 40], we focus on graph-based reasoning with incomplete knowledge, by analyzing OWL ontologies, to support researchers in the biomedical domain. In particular, Larson *et al.* [39] propose a method for rule-based reasoning with a multi-scale neuroanatomical ontology, where the authors conclude that OWL is an important technology for merging disparate data and performing multi-scale reasoning. They demonstrate how OWL-based ontologies and rule-based reasoning help infer novel facts about brain connectivity at large scale from the existence of synapses at a micro scale. Oberkamp *et al.* [40] propose a methodology for interpreting patient clinical data (medical images and reports), semantically annotated by concepts from large medical ontologies. They introduce an ontology containing lymphoma-related diseases and symptoms as well as their relations and use it to infer likely diseases of patients based on annotations.

In contrast to Larson *et al.* [39] our graph-based reasoning method relies on network analysis of the final hypothesis graph, which presents an advantage of a full overview of all possible conclusions with the quantification of the confidence measure induced by the number of evidences that have been collected and the final topology of the hypothesis graph. Oberkamp *et al.* [40] focus on the problem of inferring likely diseases in the presence of patient-specific evidences, represented as symptoms, and the similarity of the diseases is then ranked based on their distances to the symptoms. The

focus of our work and the methodology are different. We tailor our causality hypotheses to a single diseases and study causality relationships among the factors, the findings obtained with our methodology may have impact not only in the clinical, patient-specific setting, but can be used in general research. Technically, our methodology for graph projections employs a rich set of OWL 2 axioms, and go beyond the usual taxonomical relationships which can be extracted from the ontologies.

Probabilistic methodologies for reasoning with incomplete knowledge and causality inference, with applications in the biomedical domain. In a more general setting, not necessarily connected to the biomedical domain, there are examples of general theoretical frameworks which marry formal methods (*e.g.*, First-Order Logic) and probabilistic models (*e.g.*, stochastic processes) [41, 42, 43]. Application of those methodologies in biology is studied in Ciocchetta *et al* [44] who tune the Stochastic Process Algebra language PEPA [43] to model biological pathways and complex biological networks, involving stochastic processes. This line of works bridge “uncertainty” and “formal methods” for general frameworks for reasoning with incomplete knowledge in biology, and differently with our methodology is not compatible with OWL ontologies, and thus cannot benefit from OWL reasoning tasks (*e.g.*, classification, alignment).

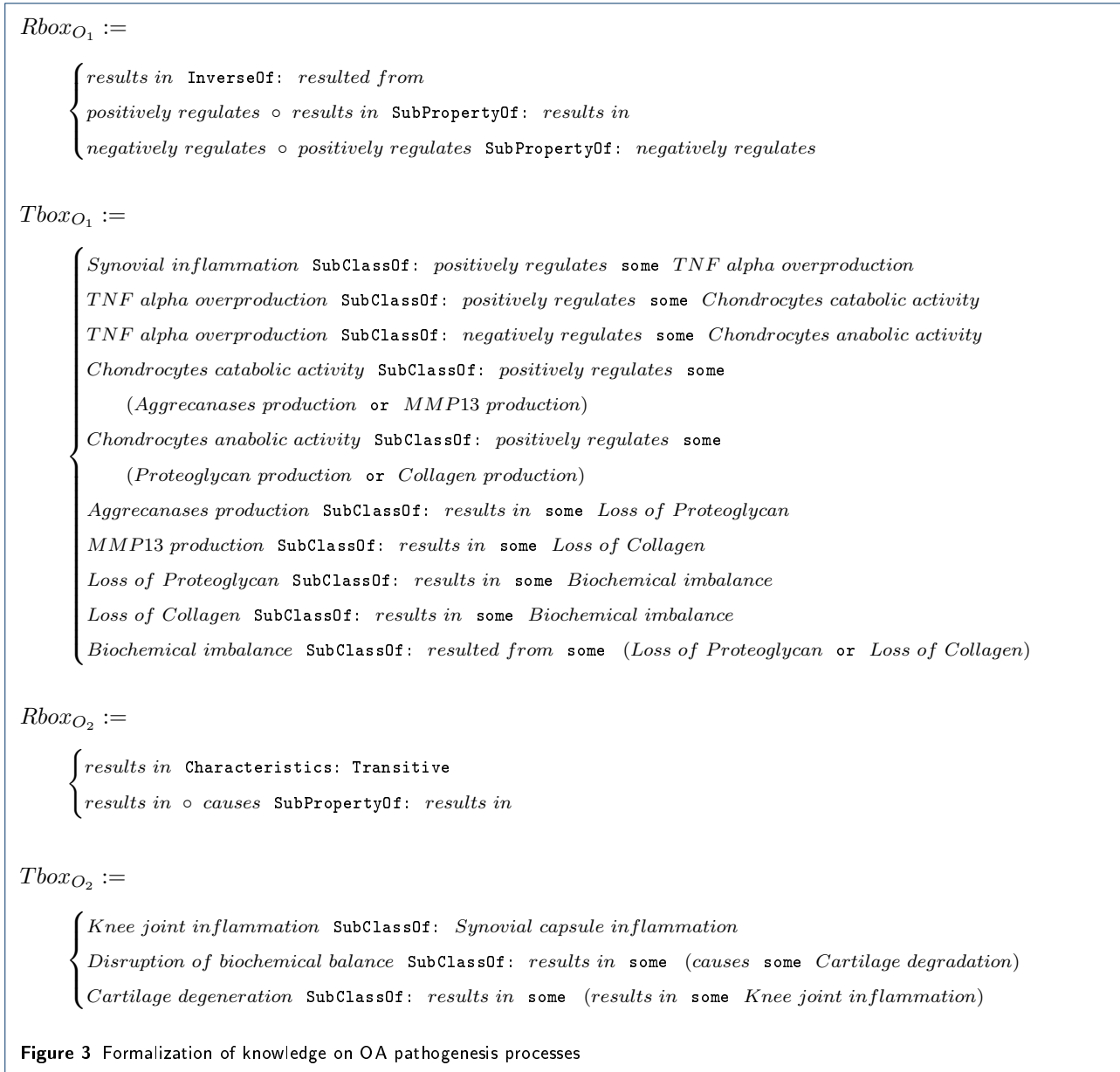
Our work is perhaps similar in spirit to that of Pearl *et al.* [45, 46], where the authors advocate for a paradigmatic shift that must be undertaken in moving from traditional statistical analysis to causal analysis of multivariate data [45, 46]. Pearl *et al.* propose a formal treatment and a unified methodology for the graphical representation of joint probability distributions along with rules for inferring causality directly from such graphical representations. In particular, the directed graphs are introduced as a compact way of

representing conditional independence restrictions for complex multidimensional probability distributions. In contrast, in our work we do not stress the existence of joint probability distributions between the factors of a hypothesis. Rather, we rely on expert knowledge of causality relationship between the factors, already known to the community, such as knowledge graphs which can be obtained from literature sources, and/or can be formalized in an OWL ontology by the domain experts.

Methods

Herein we assume that there exists a universal causality hypothesis H that can be represented as a network of factors with causality relationships, which we call a *hypothesis graph*. The background knowledge on the hypothesis graph H is formalized in an ontology O , which, for instance, may define factors as biological processes and conditions, and the causality relationships may indicate the connections between them. Moreover, we assume that different experts formalize the background knowledge on H in ontologies $O_{i=1\dots n}$, such that each O_i highlights a certain subpart of this hypothesis graph H . Consider $O_1 = \langle Rbox_{O_1}, Tbox_{O_1} \rangle$, $O_2 = \langle Rbox_{O_2}, Tbox_{O_2} \rangle$ in Figure 3, the examples of formalization of the the causality relationships among biological processes that participate in OA pathogenesis, from two different points of view.

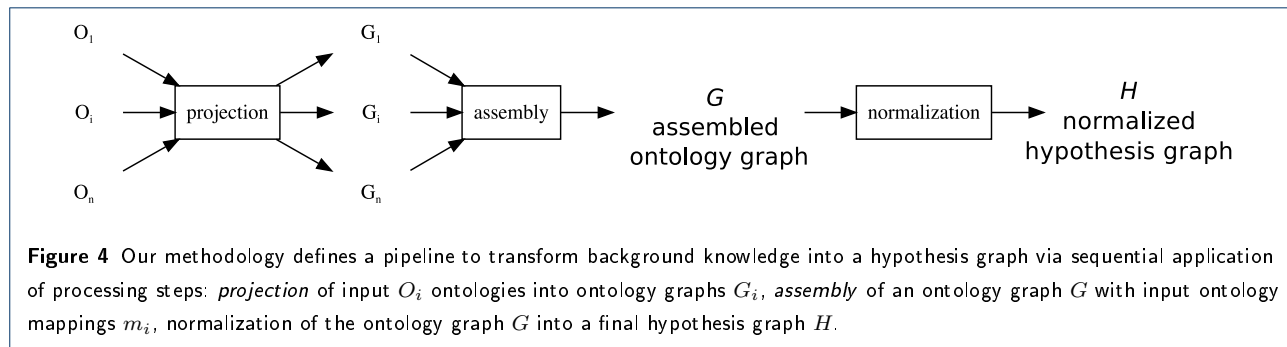
The overlaps among the ontologies O_i may or may not exist and, as the number of ontologies increases, we assume that it is possible to assemble (align) these ontologies. The assembled ontology $\bigcup_i^n O_i = O$ represents the iteratively gathered and formalized biological and biomedical knowledge on the hypothesis graph H . Finally, the causality hypothesis graph H – the network of factors interconnected with causality relationships – can be extracted from the assembled ontology O at any given point in time t_i (H_{t_0}, \dots, H_{t_n}). As a consequence, the *shape* of the causality hypothesis H_{t_i}



depends on the amount of background knowledge formalized in O at t_i . Finally, the hypothesis graph construction from ontologies is performed in a three-step process: (1) projection of OWL 2 ontologies O_1, \dots, O_n into ontology graphs G_1, \dots, G_n , (2) assembly of the ontology graph G from G_1, \dots, G_n , and (3) normalization of the graph G to obtain the hypothesis graph H (Figure 4).

Graph-based ontology projections

The nodes of the ontology-graph are unary predicates and edges are labelled with possible relations between such elements, that is, binary predicates. The key property of this ontology-graph is that every X -labelled edge $e = (v, w)$ is justified by one or more axioms entailed by the ontology which “semantically relates” v to w via X . For example, edges e of the form $A \xrightarrow{\text{broader}} B$ are justified by the OWL 2 axiom: $B \text{ SubClassOf: } A$. We rely on the OWL 2 rea-



soner HerMiT [47] to build the ontology graph (*e.g.*, extraction of classification) to consider both explicit and implicit knowledge defined in the ontology O . In the following, $A, A_{sup}, A_{sub}, B, B_i$ represent classes, while R, S, S_i, R^- represent object properties. Edges e of the form $A \xrightarrow{R} B$ are justified by the following OWL 2 axioms:

- (i) ‘ A SubClassOf: R restriction B ’, where *restriction* is one of the following: *some* (existential restriction), *only* (universal restriction), *min x* (minimum cardinality), *max x* (maximum cardinality) and *exactly x* (exact cardinality). Note that axioms with an union of classes in the restriction (e.g. ‘ A SubClassOf: R restriction B_1 or ... or B_n ’) or an intersection of classes in the restriction (e.g. ‘ A SubClassOf: R restriction B_1 and ... and B_n ’) also justify edges of the form $A \xrightarrow{R} B_i$ with $1 \leq i \leq n$.
- (ii) Nesting (one level) with the same object property: ‘ A SubClassOf: R restriction (R restriction B)’, being R transitive.
- (iii) Nesting (one level) with different properties: ‘ A SubClassOf: R restriction (S restriction B)’, and the role chain axiom of the form: ‘ $R \circ S$ SubPropertyOf: R ’.
- (iv) A combination of range and domain axioms of the form: ‘ R Domain: A ’ and ‘ R Range: B ’.
- (v) Role chain axiom of the form: ‘ $S_0 \circ \dots \circ S_n$ SubPropertyOf: R ’ when the ontology graph already includes the edges $A \xrightarrow{S_0} C_1 \dots C_n \xrightarrow{S_n} B$.

- (vi) ‘ R InverseOf: R^- ’ when the ontology graph already includes the edge $B \xrightarrow{R^-} A$.
- (vii) Top-down propagation of restrictions: ‘ A SubClassOf: A_{sup} ’ when the ontology graph already includes the edge $A_{sup} \xrightarrow{R} B$.
- (viii) Entailment among restrictions: ‘ B_{sub} SubClassOf: B ’ when the ontology graph already includes the edge $A \xrightarrow{R} B_{sub}$.

Assembly of ontology graphs

The ontologies formalizing the hypothesis graph may be created by different group of experts with different modelling (*e.g.*, defining relationships between occurrences, or between occurrences and continuants) and naming conventions. For example, a group may use the concept **Cartilage degradation** (occurent) from SNOMED-CT [48] while another may prefer to use the concept **negative regulation of cartilage development** (occurent) from the GO [20]. Furthermore, other groups would rather use the concept **Cartilage** (continuant) and push the semantics of *degradation* into the ontology property.

Ontology alignment will enable the integration and assembly of the (sub-)ontology graphs in a larger ontology graph. An ontology alignment is composed by a set of ontology mappings. An ontology mapping m between two concepts C_1, C_2 from the vocabulary of two different ontologies O_1, O_2 can be defined as follows: $m = \langle C_1, C_2, r \rangle$, where r is the relation between C_1 and C_2 and, using SKOS vocabulary, it

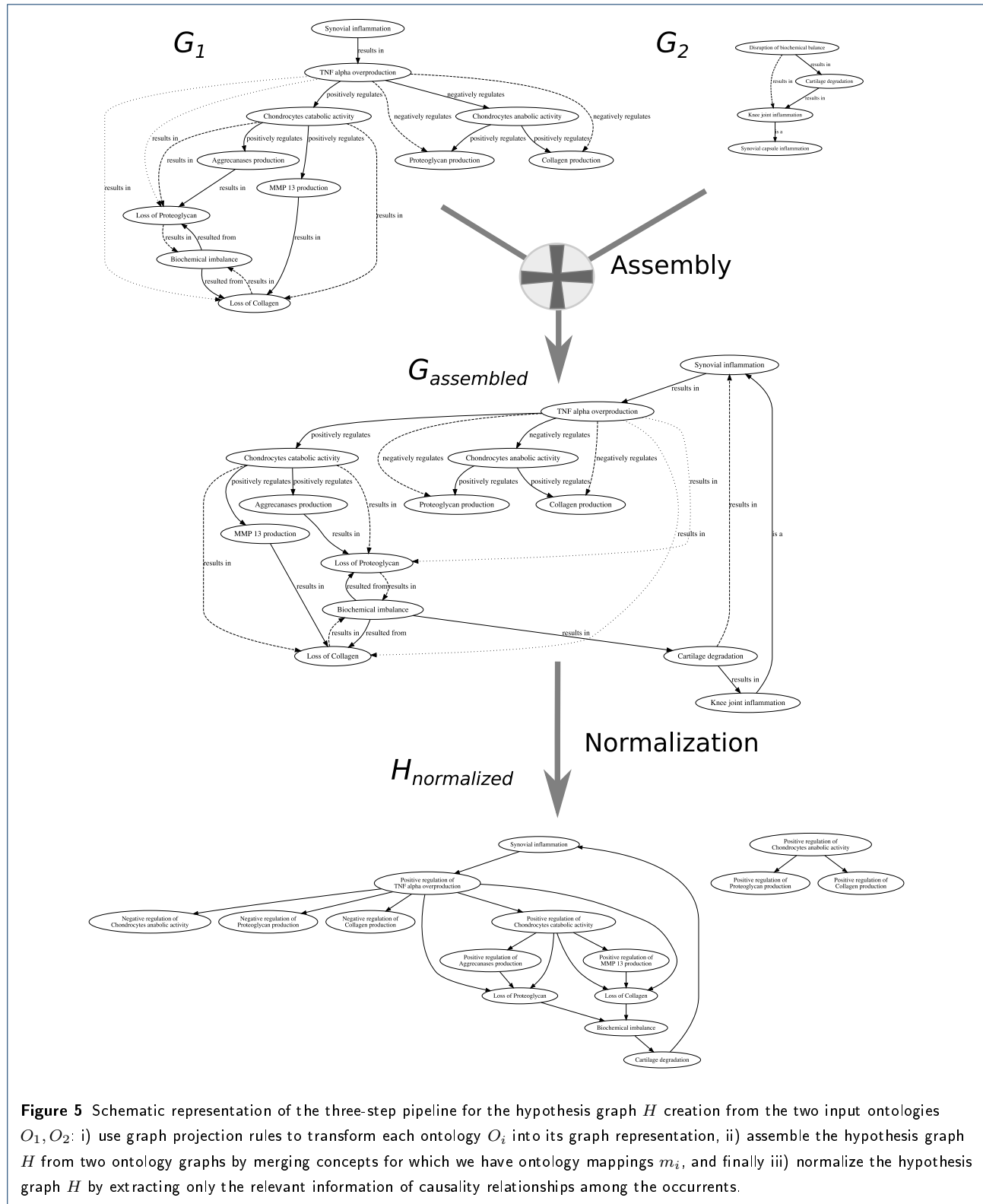


Figure 5 Schematic representation of the three-step pipeline for the hypothesis graph H creation from the two input ontologies O_1, O_2 : i) use graph projection rules to transform each ontology O_i into its graph representation, ii) assemble the hypothesis graph H from two ontology graphs by merging concepts for which we have ontology mappings m_i , and finally iii) normalize the hypothesis graph H by extracting only the relevant information of causality relationships among the occurments.

can be of one of the following types: *skos:exactMatch*, *skos:closeMatch*, *skos:relatedMatch*, *skos:narrowMatch* or *skos:broadMatch*.

Mappings to guide the assembly (*i.e.*, link factors from different hypothesis) can be discovered in online resources like UMLS Metathesaurus [49] and BioPortal [50, 51], or using state of the art ontology alignment systems like LogMap [52] and AML [53]. Mappings in UMLS Metathesaurus or BioPortal typically represent correspondences of the type *skos:exactMatch* and *skos:closeMatch*,^[1] while the output provided by automatic systems will typically provided mappings of diverse type and quality.

If a mapping exists to link two factors f_1 and f'_1 from two different (sub-)ontology graphs, then these two factors are merged into one. The weight of the merged factor will be according to the type of the ontology mapping. In our setting, we assume the following weight values w (ranging from 0 to 1) depending on the mapping type: (1) *skos:exactMatch* mappings are associated with a weight value 1.0, (2) *skos:closeMatch* mappings with 0.75, while (3) *skos:relatedMatch*, *skos:narrowMatch* and *skos:broadMatch* with a weight of 0.5. The weight associated to each (merged) factor will play a key role in our methodology for confidence measurement in a hypothesis.

Normalization of the assembled graph

The final step of hypothesis graph construction is the *normalization* of the assembled hypothesis graph, which pushes the rich semantics of causality relationships (*e.g.*, edges of the type $A \xrightarrow{R} B$) into, possibly newly created, nodes. Generally speaking, the normalization procedure leads to a simplified representation of all the available facts on causality relationships as a directed graph with specific constraints on the types of nodes and edges. Specifically, we aim to build a 1-mode network where all the nodes represent the same fundamental metaphysical type (occurrent), and all the edges represent the simplified causality relationship

defined between two occurrents. This is necessary because the general graph projection step of our pipeline might produce semantic networks of concepts where the concepts and the edges may have different types. For instance, the ontology graph may contain edges representing causality relationships involving both an occurrent and a continuant – two fundamentally different metaphysical types of concepts. Additionally, the semantics of causality relations may reflect complementary effect when we consider causal chains in the hypothesis graph, for instance *negative* and *positive* regulations of biological processes. The hypothesis graph normalization consists in iterative rewriting of the graph, where we filter all edges and rewrite them according to the following patterns:

- (i) $Occurrent \xrightarrow{R} Occurrent$ where R represent the property *results in* or *causes* justifies the edge in the hypothesis graph $Occurrent \mapsto Occurrent$. For example, if the ontology contains the axiom, ‘Chondrocyte catabolism SubClassOf: results in some Collagen degradation’ the ontology graph will include the edge $Chondrocytes\ catabolism \xrightarrow{results\ in} Collagen\ degradation$ and the hypothesis graph will contain the causality relationship $Chondrocytes\ catabolism \mapsto Collagen\ degradation$.
- (ii) $Occurrent \xrightarrow{R} Occurrent$ where R represent the property *positively regulates* or *negatively regulates*. In this case the positive or negative semantics of the property are pushed to a fresh *occurrent* concept. For example, if the ontology projection contains the edge $Chondrocytes\ anabolism \xrightarrow{positively\ regulates} Collagen\ production$, we will add the causal relationship $Chondrocyte\ anabolism \mapsto Positive\ regulation\ of\ Collagen\ production$.
- (iii) $Occurrent \xrightarrow{R} Continuant$ where R represent the property *positively regulates*, *negatively*

^[1]See https://www.bioontology.org/wiki/index.php/BioPortal_

regulates, increases levels of or decreases levels of. For example if the ontology graph includes the edge `TNF alpha overproduction` $\xrightarrow{\text{decreases levels of}}$ `Collagen` the hypothesis graph will include the fresh term `Decreased levels of Collagen` (or `Loss of Collagen`) and the causal relationship `TNF alpha overproduction` \mapsto `Decreased levels of Collagen`.

In Figure 5 we illustrate the whole pipeline of constructing a hypothesis graph H from the two input ontologies O_1, O_2 , defined in Figure 3. The two ontology graphs G_1, G_2 represent the individual extent of background knowledge of the two specialists on causality relationships of factors between synovial inflammation and cartilage degradation (obtained by projecting ontologies O_1, O_2). The assembly of the graphs takes as input the ontology mappings m_1 and m_2 (see Table 1), which have been manually created by the domain experts, to merge the graphs G_1, G_2 . Overall, the graph projection and the graph assembly steps of the pipeline work in couple to entail new causal links among the factors, which we represent in the assembled graph G . For instance, once we align the two graphs we entail the circular causality relationship, which states that `Synovial inflammation` may be, simultaneously, the cause and the effect of `Cartilage degradation`. Notice that before the alignment the two specialists were not aware of this circular relationship. The normalization of the assembled graph G splits the two biological scenarios of chondrocytes' anabolic and catabolic activities, such that the resulting hypothesis graph H contains only unambiguous causality relations among the factors.

Measuring confidence in a hypothesis

Once we obtain the hypothesis graph H , we are ready to form the causality hypothesis and perform evidence-based hypothesis testing. Before we delve into this

topic, we briefly introduce the notation that we use for the hypothesis graphs throughout this work.

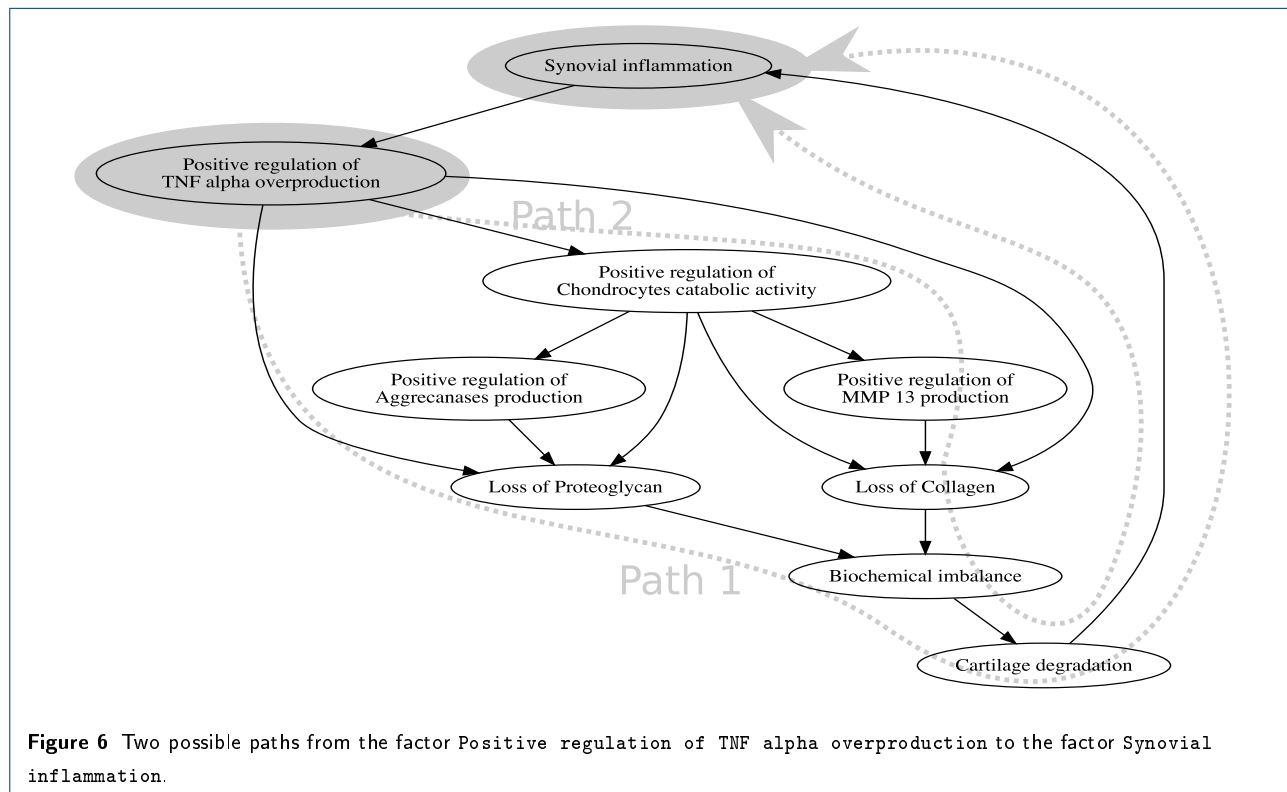
Notation for Hypothesis Graphs. Let $H = (N, A)$ be a directed graph, which we call *hypothesis graph*, with $n_i \in N$ set of nodes. And A is a set of ordered pairs of (s, t) in N , called *arcs*, where s denote the *source* of the arc, and t the *target* of the arc [54]. A path $\pi(s, t)$ from *source* node s to the *target* node t is denoted as $\pi_i(s, t) = (s, n_i, \dots, t)$. We write $\Pi(s, t)$ to denote all possible simple paths in the hypothesis graph from node s to the node t . A simple path is a path which does not have repeating nodes. And we use $\mathcal{I}(s, t) = \{n_i | n_i \in \pi_i, \forall \pi_i(s, t) \in \Pi(s, t)\}$ to refer to all the *interior nodes* which appear in all paths from s to t .

Causality hypothesis. A causal hypothesis asks a question whether some factor (s) has caused another factor (t). There might be a direct causality relationship from s to t , or there might exist an indirect causality relationship, such that s has caused t through some intermediate factors, which might have participated actively or passively to the causality chain from s to t . These causal chains from s to t represent different possibilities of how s might have caused t . We use the notation for hypothesis graph H to represent factors as nodes $f_i \in N$, direct causality relationships as arcs $(f_i, f_j) \in E$, and causality chains as paths $\Pi(s, t)$.

Consider an example causality hypothesis that postulates that $s = \text{Positive regulation of TNF alpha overproduction}$ caused $t = \text{Synovial inflammation}$ in Figure 6. In our example, we do not have a direct causality relationship between these two factors, however there exist 6 different causal chains, i.e., 6 different ways in which s might have caused t . In Figure 6 we present two possible chains of factors (Path 1, Path 2) starting from s and leading to t .

Table 1 Ontology mappings created manually by the domain experts

Mapping m_i	$O_1 : C_1$	$O_2 : C_2$	r	c
m_1	O_1 :Synovial inflammation	O_2 :Synovial capsule inflammation	skos:closeMatch	0.75
m_2	O_1 :Biochemical imbalance	O_2 :Disruption of biochemical balance	skos:relatedMatch	0.5



We are confident in our causality hypothesis – within the domain of the known facts – when we are able to provide evidences to all the factors that participate in causality chains from s to t . $\mathcal{I}(s, t)$ represents the set of nodes in the hypothesis graph H , which correspond to the factors that need to be evidenced, \mathcal{E} is an indicator set which denotes factors evidenced so far, and $\mathcal{C}(s, t, \mathcal{E})$ be the confidence function. Intuitively, confidence in a hypothesis should grow with the number of factors that we are able to evidence, more factors we evidence, more confident we are that s did indeed cause t . Since, we might have several possibilities of s causing t we, first, propose to measure confidence of each causality possibility separately, and then, we propose to measure overall causality hypothesis as a

sum of the confidences of all the known possibilities (Equation 1). To this end, our confidence in a causality hypothesis depends on three parameters: i) source of the causality (s), ii) target of the causality (t), and iii) set of evidenced factors (\mathcal{E}).

$$C_s^t(\mathcal{E}) = \sum_{\pi \in \Pi(s,t)} \sum_{f \in \pi} \mathcal{F}(f), \tag{1}$$

Measuring confidence in a causality hypothesis proportionally to the number of evidenced factors might not be correct, there are two sources of uncertainty that might negatively effect our confidence in the hypothesis, even if we collect all the evidences, and should be reflected in the way we measure confidence in the hypothesis: i) the quality of the evidences, i.e.,

we can surely state that the evidence is not due to errors, and ii) quality of our modelling of the hypothesis. The first source of uncertainty comes from the fact that during our experiments or literature search for the justifications of evidences we might face errors. And the second source of uncertainty comes from the way we model our hypothesis as an assembly of sub-hypotheses, which relies on ontology mappings to merge formalizations of the background knowledge of the hypothesis. During this process we might introduce uncertainty for the matched concepts representing factors of the hypothesis.

To this end, we introduce two functions defined on the nodes of the hypothesis graph, $\phi : N \mapsto [0 \dots 1]$ that associates weights of the confidence in the ontology mapping to every factor, and represents our confidence in the hypothesis modelling, and $\psi : N \mapsto [0 \dots 1]$ associates weights of the confidence in evidence for each factor. Equation 2 represents the contribution function for the hypothesis factors.

$$\mathcal{F}(f) = \begin{cases} 0 & f \notin \mathcal{E} \\ & \text{factor } f \text{ not evidenced} \\ \phi(f)\psi(f) & f \in \mathcal{E} \\ & \text{weighted contribution} \\ & \text{if } f \text{ evidenced} \end{cases} \quad (2)$$

Properties of the confidence function. Confidence in causality hypothesis is defined as a sum of weighted contributions of factors, that participate in causality possibilities. The contributions of factors is a weighted, and most importantly a non-negative, function (Equation 1), thus thus as we add more evidenced factors the value of the function, can only grow. Confidence depends on the evidenced factors, it has its minimum value ($\mathcal{C}_s^t = 0$) when we have no evidences ($\mathcal{E} = \emptyset$), and it has its maximum value when all the factors have been evidenced ($\text{argmax}_s \mathcal{C}_s^t$ when $\mathcal{E} = \mathcal{I}(s, t)$). To

this end, we can normalize our confidence function to the maximum possible confidence value we can obtain, when all the factors have been evidenced, such that the confidence is always measured in the $[0 \dots 1]$ range (Equation 3).

$$0 = \frac{\mathcal{C}_s^t(\mathcal{E} = \emptyset)}{\mathcal{C}_s^t(\mathcal{E} = \mathcal{I})} \leq \frac{\mathcal{C}_s^t(\mathcal{E} \subset \mathcal{I})}{\mathcal{C}_s^t(\mathcal{E} = \mathcal{I})} < \frac{\mathcal{C}_s^t(\mathcal{E} = \mathcal{I})}{\mathcal{C}_s^t(\mathcal{E} = \mathcal{I})} = 1. \quad (3)$$

Results

With the help of our domain experts in biology and biomechanical engineering (multi-disciplinary consortium of the EU FP7 “MultiScaleHuman” project [55]) we have been formalizing the background knowledge around factors participating in the process of cartilage degradation, which can be evidenced across different biological scales. This background knowledge has been captured, as a proof of concept, in an OWL 2 ontology O and has been iteratively validated with our domain experts. This ontology has been designed to contain a significant amount of axioms which go beyond the usual taxonomical relationships in the biomedical ontologies, and instead, model causality relationships with rich ontology concept construction operators including nested OWL restrictions and property chains. During our interviews (t_1, \dots, t_n) with the domain experts we have been updating the background knowledge formalization (O_{t_1}, \dots, O_{t_n}), either with the help of our domain experts or by translating discovered causality relationships from the literature ourselves. Each snapshot of the background knowledge O_{t_i} has been presented as the results of our methodology of hypothesis graph construction H_{t_i} for validation and feedback. To report our results we fix our attention to two specific snapshots of the causality hypothesis, and we refer to them as H_{sub} and $H_{broader}$. H_{sub} has been extracted from the state of the ontology O_{t_i} , which corresponds to the extent of knowledge of the molecular biologist on causality relationships between the

biological processes which lead to cartilage degradation with a focus on cellular and molecular biological scales (H_{sub} is an equivalent hypothesis graph to what we presented as a normalized hypothesis graph in the Methods section). $H_{broader}$ was extracted from the ontology O_{t_j} at time point t_j , which corresponds to the ontology O_{t_i} updated with more knowledge about factors that lead to cartilage degradation, from organ and behavior biological scales. Table 2 summarizes O_{t_i}, O_{t_j} with ontology metrics and descriptions, computed with the Protégé ontology editor.

In Figure 7 we notice that $H_{sub} = \langle N_{sub}, A_{sub} \rangle$ is a subgraph of $H_{broader} = \langle N_{broader}, A_{broader} \rangle$, such that $N_{sub} \subseteq N_{broader}$ and $A_{sub} \subseteq A_{broader}$. The additional knowledge ($H_{broader}/H_{sub}$) is not present in the formalization by the molecular biologist, meaning that he might not be aware about alternative factors that occur during osteoarthritis and might have played a significant role in the causality hypothesis (Figure 7). The subsequent experiments demonstrate how our methodology supports hypothesis testing by quantifying confidence in a causality hypothesis with incomplete evidences, and provides means to compare confidence measures with different depths of knowledge.

Table 2 O_{t_i}, O_{t_j} ontology metrics

Ontology metric	O_{t_i}	O_{t_j}
Axioms	66	151
Logical axiom count	39	92
Declaration axiom count	18	34
Class count	14	30
Object property count	4	4

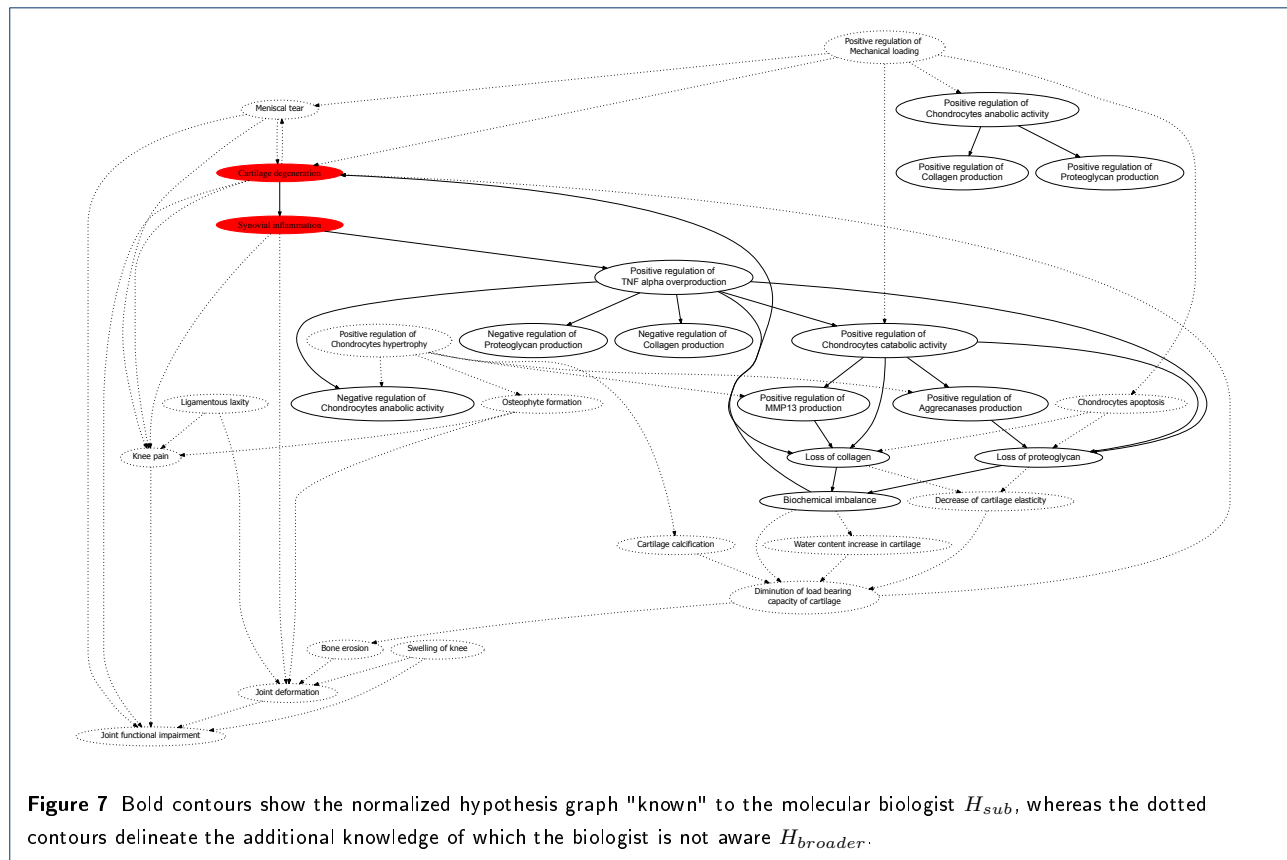
Robustness of the system in presence of complementary causality relationships.

Our methodology is capable of adequately tracking two complementary biological scenarios, where one factor might stand as a cause of two opposite effects. We tested our methodology for hypothesis graph construction with small increments in our knowledge which

might lead to big changes in the shape of the causality hypothesis, and what we can understand from it. In particular, at the time point t_i the knowledge on the hypothesis contained causality path from **Mechanical loading** factor to the **Chondrocytes catabolism** factor. Indeed, the positive regulation of chondrocytes' catabolism by mechanical loading has been demonstrated in the literature [56]. However, it is also known that the mechanical loading can also have positive effect on the **chondrocytes anabolism** (the opposite biological process of catabolism), and thus facilitate proteoglycan and collagen production [57]. Based on the complementary causality effects of mechanical loading on the biochemical balance in cartilage, we can thus hypothesize that mechanical loading might result in both beneficial and detrimental conditions of the joint cartilage. This additional knowledge is reflected in the way our methodology constructs the hypothesis graph. In particular, the normalization patterns (introduced in the Methodology section) split the causality chains starting in mechanical loading, that span two complementary causality possibilities of benign and malign effect on articular joint (Figure 7). Validly, all the possibilities of mechanical loading leading cartilage degradation pass through the factor **positive regulation of chondrocytes catabolism** and we do not have a situation where **mechanical loading** leads to **cartilage degradation** by passing through **positive regulation of chondrocytes anabolism**. Conversely all the causality chains which lead from **mechanical loading** to **collagen** or **proteoglycan** production pass through **chondrocytes anabolism** factor.

Relative confidence measurement

This experiment demonstrates how molecular objectives can measure his confidence in the causality hypothesis according to his knowledge on causality relationships (H_{sub}) and can compare it to the confidence



measure when we add more knowledge $H_{broader}$. We simulate the case where the molecular biologist wants to test a hypothesis that $s = \text{Synovial inflammation}$ has caused $t = \text{Cartilage degradation}$. We treat $H_{broader}$ as a coarse approximation of our universal knowledge on all possible causalities which lead from s to t , and H_{sub} as a personal view of that universal knowledge by the molecular biologist.

Table 3 Statistics of the graphs

Statistic	H_{sub}	$H_{broader}$
Number of nodes $ N $	15	30
Number of arcs $ A $	19	57
Number of possible causal chains from s to t	6	24
Number of possible factors to evidence $ I $	9	12

Table 3 summarizes network statistics of the two graphs. In particular, in the universal hypothesis graph $H_{broader}$ there are 24 possible causal chains which lead from s to t , whereas in the subgraph H_{sub} we have only

6 possible causal chains, which means that the molecular biologist is missing a significant amount of knowledge about the causalities that he is studying. Moreover, in the universal knowledge of causality hypothesis we have 12 ($|I_{H_{broader}}| = 12$) factors that can potentially be evidenced and would contribute positively to the overall confidence of the hypothesis, whereas in the restricted knowledge case we are aware of only 9 ($|I_{H_{sub}}| = 9$) factors which need to be evidenced to obtain the maximum confidence in the same hypothesis that s has caused t . To study the behavior of the confidence function C_s^t in these two cases we perform the following tests: i) study the evolution of the confidence function separately for two graphs, ii) normalize the confidence function with the maximum possible confidence for individual graphs, iii) normalize the two confidence functions with the maximum confidence in the universal graph. Note that, the parameter for the con-

confidence function is the set of evidenced nodes, where each node may have different importance value, as defined by the weighting function \mathcal{F} . To take into account all the possible variability of the confidence function we compute the distributions of the confidence values for a gradually increasing number of evidences. That is, we start with the case where the evidence set is empty, corresponding to the initial phase of hypothesis testing and where our confidence is 0. Then, we compute the distribution of confidences for all evidence sets of size (cardinality) 1, corresponding to different choices of choosing one factor to evidence. For instance, for the universal hypothesis graph $H_{broader}$ we have 12 ways to prove hypothesis by evidencing only one factor (out of 12 possible), whereas for H_{sub} we have 9 factors to choose from. We continue computing confidence distributions until we reach the full evidence set.

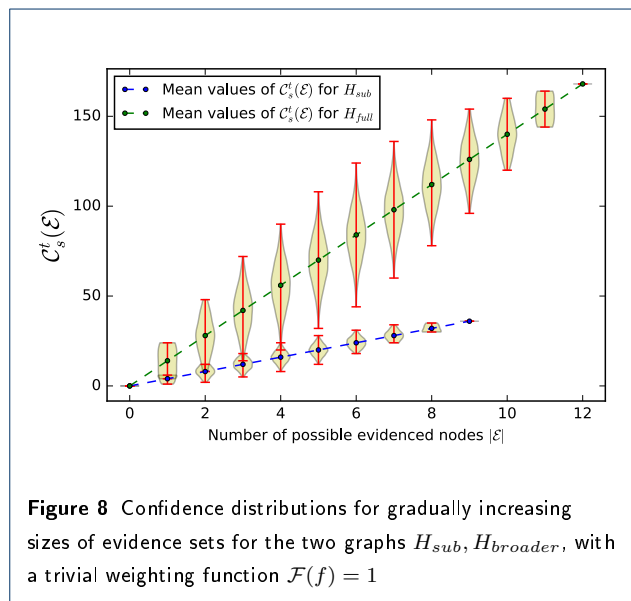


Figure 8 Confidence distributions for gradually increasing sizes of evidence sets for the two graphs $H_{sub}, H_{broader}$, with a trivial weighting function $\mathcal{F}(f) = 1$

Figure 8 represents the distribution of confidences computed with C_s^t (Equation 1) for gradually increasing sizes of evidence sets, with a trivial weighting function of factors – $\mathcal{F} = const 1$ – where every factor has equal contribution to the causality chains. The mean values of the confidence distributions grow linearly as we increase the number of evidences, as expected, the

maximum confidence value obtained in the universal case is bigger than in the restricted case because we take into account more possibilities in the universal case. We now use the individual maximum mean confidence values for each graph to scale our distributions, such that they always stay in the 0..1 range.

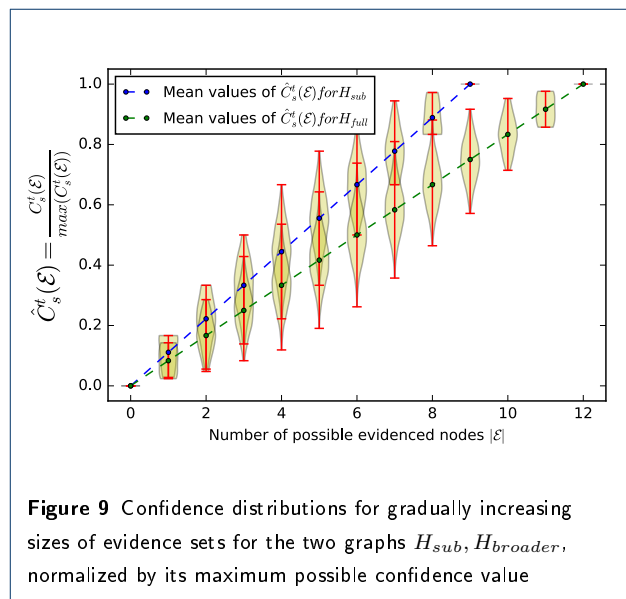
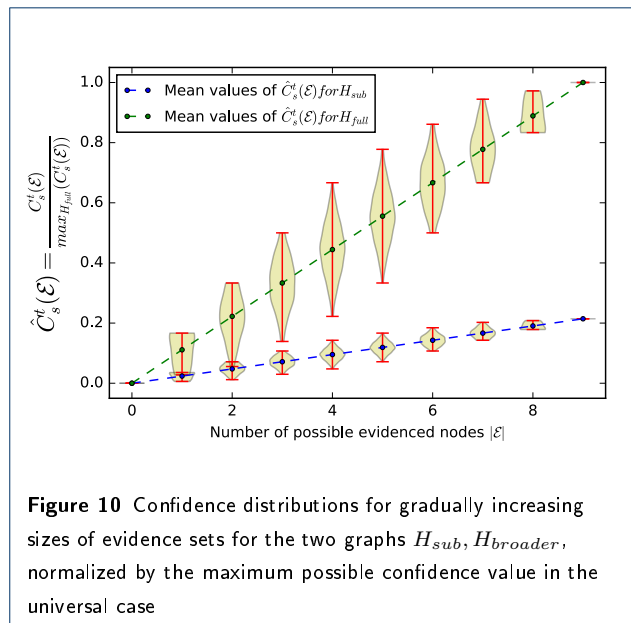


Figure 9 Confidence distributions for gradually increasing sizes of evidence sets for the two graphs $H_{sub}, H_{broader}$, normalized by its maximum possible confidence value

Figure 9 shows the normalized version of the confidence distributions, namely $\hat{C}_s^t = \frac{C_s^t}{\max(C_s^t)}$ for H_{sub} and $H_{broader}$. In particular, it shows that a molecular biologist, relative to his extent of knowledge, obtains the 100% confidence in his causality hypothesis by evidencing all the possible factors which contribute to all the possible ways in which s might have caused t , however, with the same amount of evidence, but taking into account universal knowledge about the causality possibilities, his confidence is less than 100%, which shows that he has missed some important causality possibilities. To quantify this uncertainty, which is proportionate to the amount of missed causality possibilities, we scale both confidence distributions by the maximum confidence value that we may obtain in the universal case.

Figure 10 demonstrates the *relative confidence* of the molecular biologist to the universal causality hypothe-



sis for the same evidenced sets. The x-axis is truncated to evidence sets of size 9, since molecular biologist is only aware of 9 factors which need to be evidenced to prove his hypothesis. If we collect the mean values of the confidence distributions in two vectors x_1, x_2 then we can quantify the error as their Euclidean distance $\|x_1 - x_2\|$. In Table 4 we summarize the errors which quantify the uncertainty in obtained confidence measures with respect to the universal case for different weighting functions \mathcal{F}_i . These weighting functions were chosen as follows: i) \mathcal{F}_1 trivial weighting of importance of factors, ii) \mathcal{F}_2 random weighting of importance of each factor, iii) \mathcal{F}_3 gives more importance to factors which molecular biologist is aware of, whereas those that he is not aware of are given less importance, iv) \mathcal{F}_4 opposite to \mathcal{F}_3 , we give more importance to factors that molecular biologist is not aware of and we decrease the importance of factors that he is aware of. The error variation is intuitive, if we evidence the most important factors, even if we miss other factors and other causality chains, but whose importance to the overall hypothesis is significantly smaller, then we are more *confident* even with a restricted knowledge

of the causality possibilities. Vice-versa, if we evidence less important factors and we miss the important ones, then our confidence is much more compromised.

Table 4 Mean squared error between the confidence distributions for different weighting functions \mathcal{F}

Weighting function \mathcal{F}_i	Error
$\mathcal{F}_1(f) = 1$	2.17
$\mathcal{F}_2(f) = random(0, 1)$	2.09
$\mathcal{F}_3(f) = 1$ if $f \in I_{H_{sub}}$, otherwise 0.1	1.95
$\mathcal{F}_4(f) = 1$ if $f \in I_{H_{broader}}$, otherwise 0.1	2.96

Local importance of factors

Importance of the factors for a causality hypothesis can be deduced from our confidence measure defined on the hypothesis graph. The factors ranked as the most important may help the researchers prioritize their next experiments, studies, and may help in the discovery of the potential collaborations with other scientists. Analogously, the factors that are identified as the least important for a specific causality hypothesis hint on the lack of knowledge about the possibly missing causality relationships, and might represent an opportunity to focus on an underresearched topic. In particular, C_s^t measures our confidence in the causality hypothesis that factor s caused t with a given set of evidenced nodes \mathcal{E} . This function accumulates the weighted contribution of all evidenced nodes in each causality possibility leading from s to t . When we first start proving our hypothesis we do not have any evidence and we have a choice of \mathcal{I} to evidence from. However, do we need to evidence all the factors in the interior of the causality hypothesis \mathcal{I} ? What if we can only obtain an incomplete set of evidences, which factors should we choose? Intuitively, we should first focus on evidencing factors which are most important in our causality hypothesis. But how can we assess the importance of each factor in the causality hypothesis? In this experiment, we propose a general approach to assessing the local importance of factors, independently of

the weighting function \mathcal{F} . To do so we start with a case where we do not have any evidence $\mathcal{E} = \emptyset$, we then rank each factor f_i in the causality hypothesis by its potential contribution to the confidence in the causality hypothesis if it was evidenced $|C_s^t(\mathcal{E} \cup f_i) - C_s^t(\mathcal{E} = \emptyset)|$.

Figure 11 depicts the variation of potential contributions to the overall confidence measure C_s^t for each factor f_i . In particular, we can observe that in both cases: H_{sub} restricted personal view of the hypothesis, and $H_{broader}$ universal causality hypothesis the most important factors are: **Positive regulation of TNF alpha overproduction**, **s =Synovial inflammation**, **t =Cartilage degeneration** and **Biochemical imbalance**. Indeed, to prove that s has resulted in t our best strategy is to focus on evidencing those two factors, however, given our knowledge of causality relationships, we might choose to evidence alternative factors to obtain the same overall confidence in the validity of our causality hypothesis. We also observe that by extracting more knowledge on causality relationships more important factors to our causality hypothesis emerge, i.e., the factors which we did not know about before. For instance, **Decrease of cartilage elasticity** and **Water content increase in cartilage** have relatively low potential confidence contributions (< 0.04) and thus our unawareness of the contribution to causality hypothesis of these factors is not so penalizing. Yet, **Diminution of load bearing capacity of cartilage** is capable of contributing more than 10% of the overall confidence measure C_s^t . It is also interesting to observe that adding knowledge ($H_{broader}$) reduces the importance of **Biochemical imbalance** factor to the point that it is no longer one of the most important factors in the causality hypothesis.

Generalization of the hypothesis configuration

In the previous experiment we identified the most important factors, such that evidencing them would max-

imize our confidence in the causality hypothesis that s resulted in t . We can use the local importance of factors to the hypothesis configuration to target our evidence collection. Suppose we managed to evidence the four most important factors for the hypothesis graph H_{sub} , which we summarize in Table 5.

For the same evidence set \mathcal{E}_{sub} we obtain the normalized confidence of $C_s^t = 0.66$ for H_{sub} and $C_s^t = 0.53$ for $H_{broader}$. Now, we ask ourselves a question "with the same evidence set what other causalities can we prove (with the same confidence)?" If we keep the same evidence set \mathcal{E}_{sub} we are able to prove causalities with a confidence $> 60\%$ as depicted in Table 6. These causalities correspond to very similar causality chains, as our initial causality hypothesis that **Synovial inflammation** has results in **Cartilage degradation**.

Intuitively, Table 7 demonstrates that for the same evidence set, as we add more knowledge ($H_{broader}$) we are able to prove more causality relationships, with a good confidence ($> 50\%$).

Generalization of the hypothesis configuration leads to the scenarios where the seemingly wrong causality relationships, might actually be explained with plausible interpretations. One such example scenario is when we obtain the significant confidence (0.60) in a causality hypothesis that **Cartilage calcification** might result in **Positive regulation of TNF alpha overproduction** (line 1 in Table 7). First, it is tempting to say that this is a wrong hypothesis, and is due to the error in the formalization of the background knowledge on causality relationships. Partly, because calcification of cartilage entails cell apoptosis and thus should cause the decrease of levels of TNF alpha cytokine cells. However, we get the high confidence score in this causality due to the presence of a path from **Cartilage calcification** to **Positive regulation of TNF alpha overproduction** (see Figure 7). This

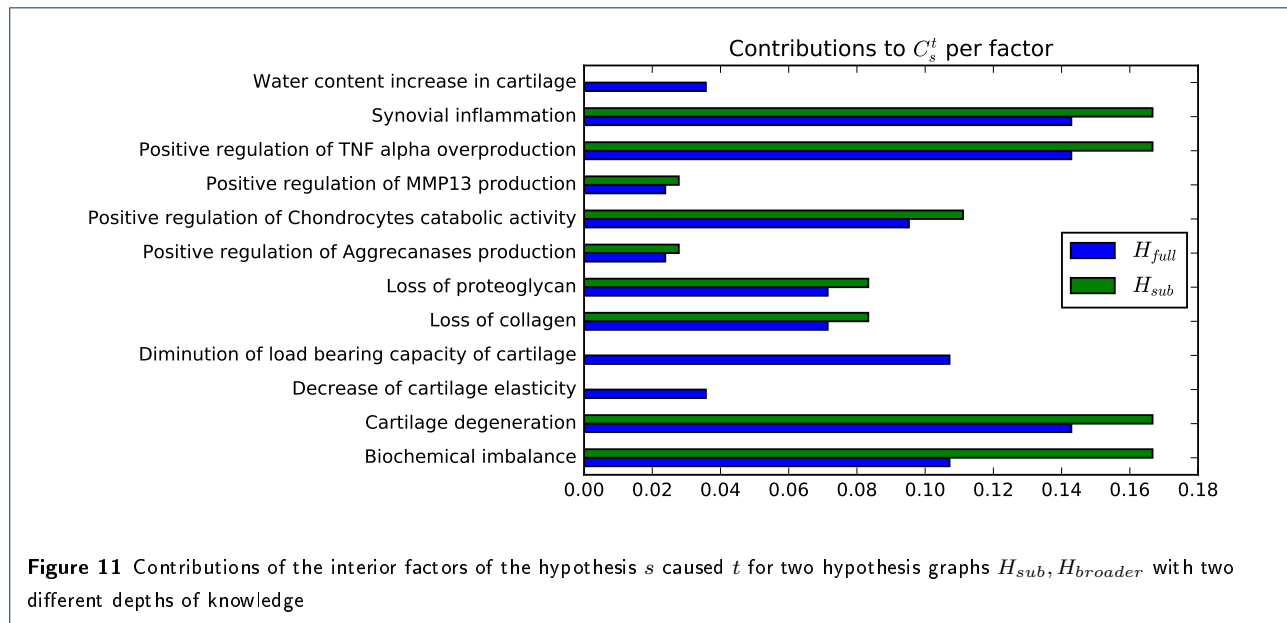


Table 5 4 Most important factors for H_{sub} in the two hypothesis graphs and their relative confidence values in both H_{sub} and $H_{broader}$

Evidence set \mathcal{E}_{sub}	importance for $H_{broader}$	importance for H_{sub}
Biochemical imbalance	0.10	0.16
Cartilage degeneration	0.14	0.16
Positive regulation of TNF alpha overproduction	0.14	0.16
Synovial inflammation	0.14	0.16
	$C_s^t(\mathcal{E}_{sub})$ for $H_{broader}$	$C_s^t(\mathcal{E}_{sub})$ for H_{sub}
	0.53	0.66

Table 6 Other causalities we can prove (> 60% confidence) with the same evidence set \mathcal{E}_{sub}

source s	target t	$C_s^t(\mathcal{E}_{sub})$ for $H_{broader}$	$C_s^t(\mathcal{E}_{sub})$ for H_{sub}
Cartilage degeneration	Biochemical imbalance	0.66	0.66
	Negative regulation of Collagen production	0.75	0.75
	Positive regulation of TNF alpha overproduction	1.00	1.00
Loss of collagen	Positive regulation of TNF alpha overproduction	0.62	0.80
	Positive regulation of TNF alpha overproduction	0.62	0.80
Synovial inflammation	Negative regulation of Chondrocytes anabolic activity	0.66	0.66
	Negative regulation of Collagen production	0.66	0.66
	Negative regulation of Proteoglycan production	0.66	0.66
	Positive regulation of Chondrocytes catabolic activity	0.66	0.66
	Positive regulation of TNF alpha overproduction	1.00	1.00

path represents our knowledge that calcified cartilage will result in degeneration of cartilage tissue, which will provoke synovial inflammation, and we hypothesized that synovial inflammation will result in positive regulation of TNF alpha. After a discussion with

our domain experts we reached the conclusion that, although this causality relationship between calcified cartilage and positive regulation of TNF alpha might seem contradictory, there actually might be a plausible explanation. Namely, while the calcification causes

Table 7 Causalities we can prove ($> 50\%$ confidence), as we add more knowledge, and which we cannot prove with our restricted knowledge of causality relationships

source s	target t	$C_s^t(\mathcal{E}_{sub})$ for $H_{broader}$	$C_s^t(\mathcal{E}_{sub})$ for H_{sub}
Cartilage calcification	Positive regulation of TNF alpha overproduction	0.60	0.0
Diminution of load bearing capacity of cartilage	Biochemical imbalance	0.57	0.0
	Negative regulation of Chondrocytes anabolic activity	0.60	0.0
	Negative regulation of Collagen production	0.60	0.0
	Negative regulation of Proteoglycan production	0.60	0.0
	Positive regulation of Chondrocytes catabolic activity	0.60	0.0
	Positive regulation of TNF alpha overproduction	0.75	0.0
	Synovial inflammation	0.66	0.0
Meniscal tear	Biochemical imbalance	0.57	0.0
	Negative regulation of Collagen production	0.60	0.0
	Negative regulation of Proteoglycan production	0.60	0.0
	Positive regulation of Chondrocytes catabolic activity	0.60	0.0
	Positive regulation of TNF alpha overproduction	0.75	0.0
Water content increase in cartilage	Positive regulation of TNF alpha overproduction	0.60	0.0

tissue death in cartilage, it does so only in a specific region of cartilage. The calcified region, however, will induce the diminution of the load bearing properties of the whole cartilage, and this will provoke the synovial inflammation, which, in turn, will result in excessive levels of TNF alpha in the neighbouring regions of the cartilage (neighbouring to the calcified region).

Prototype

We implemented a prototype (Figure 12) to interactively apply and present the proposed methodology for causality hypothesis testing on the obtained hypothesis graphs. The demo of the prototype is available at <http://hypothtest.plumdeq.xyz/test/>. Source code for the hypothesis testing of the prototype and proof of concept ontologies, as well as the Jupyter Notebooks (reproducible experiments presented in this manuscript) are available on GitHub at <https://github.com/plumdeq/hypothtest> (see Subsection Availability of data and materials).

The interface of the prototype is divided into 4 logical blocks, labeled A , B , C , D in Figure 12.

(A) *Control over the hypothesis configuration.* The users can change the hypothesis configuration in two modes - i) identifying the boundary nodes s, t , ii) se-

lecting the evidenced nodes \mathcal{E} . Each mode is triggered by clicking on an associated button (see Figure 12, label A), and then selecting the specific nodes in the hypothesis graph (Figure 12, label D).

(B) *Hypothesis summary.* A textual summary of a current hypothesis configuration (see Figure 12, label B).

(C) *Local importance of nodes in the hypothesis.* Local importance of each node with respect to the hypothesis configuration.

(D) *Visualisation of the hypothesis graph.* Interactive network visualisation with the force directed layout [58] of the hypothesis graph H . The users can interactively click on the nodes and drag them for a visually better spatial distribution of the network. The boundary nodes are visually distinguished as completely *opaque* nodes in the hypothesis graph (Figure 12), while all other nodes are semi-opaque. Evidenced nodes are visually distinguished as *green* nodes. Consequently, if a node n_i is both *evidenced* and either a *source* or a *target* of the confidence evaluation, then it will be *opaque green*. The backend (server) of the prototype constructs hypothesis graphs, computes

importance measures on each node of the graph, and evaluates confidence in the hypothesis configuration. The frontend (client) is responsible for the interactive visualisation of the hypothesis graph, and serves as a user interface. In particular the user can interactively assign the boundary nodes, and mark nodes as evidenced. The user input is then transmitted to the backend via custom data exchange protocol, based on JSON files. Each time the user changes the configuration of the hypothesis (*i.e.*, evidences/unevidences node or assigns new source or target nodes of the confidence evaluation the hypothesis confidence is reevaluated and the results are sent back to the client.

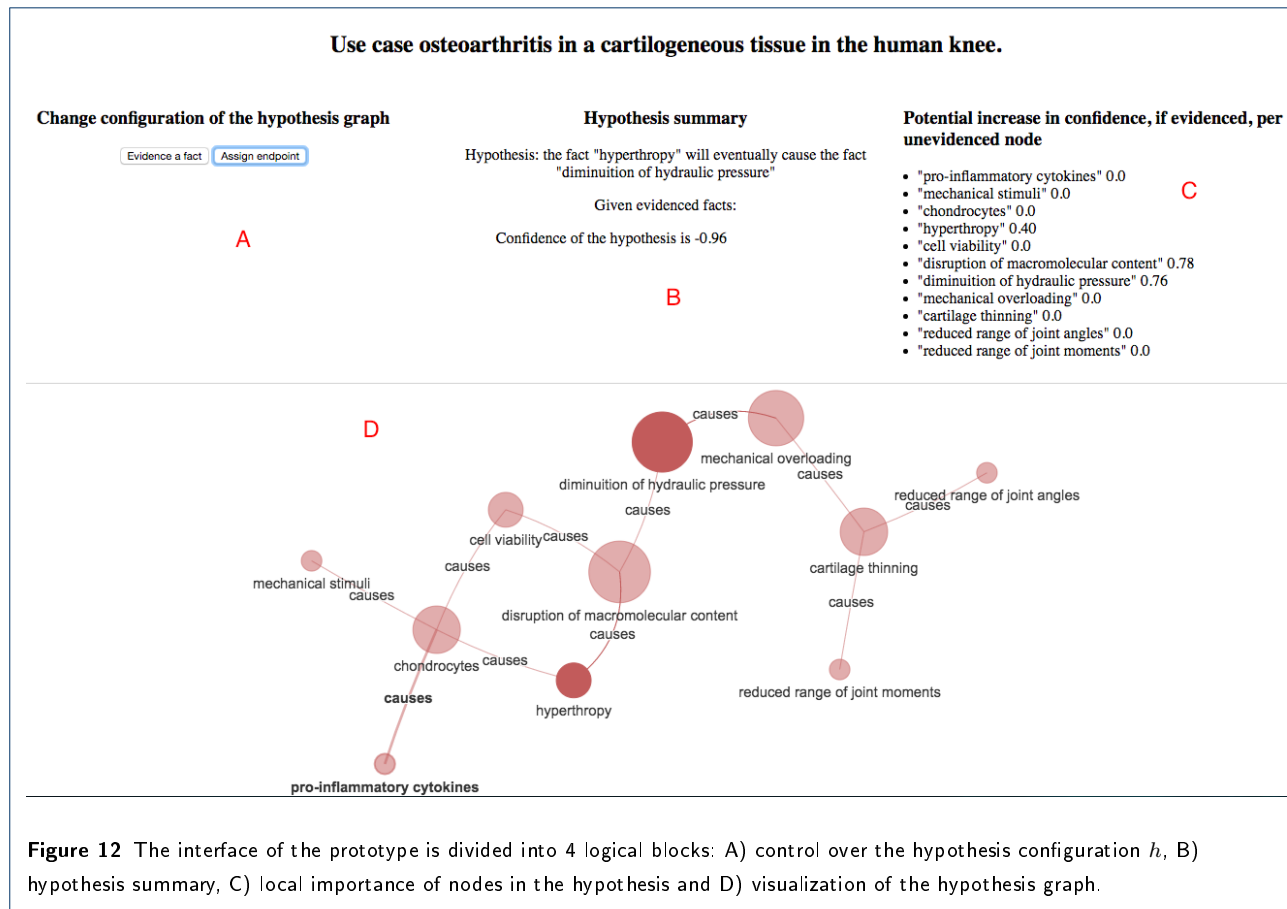
Discussion

We evaluated our methodology on a hypothesis graph which covers our use-case scenario of cartilage degradation during osteoarthritis. The obtained hypothesis graph represents both contributing factors which may cause cartilage degradation and the factors which might be caused by the cartilage degradation. Hypothesis graph construction (see Section Robustness of the system in presence of complementary causality relationships) has proven to be robust to the addition of potentially contradictory information on the simultaneously positive and negative effects, by adequately separating two complementary causality scenarios. By evaluating our methodology for relative confidence measurement (see Section Relative confidence measurement) we have observed the following: i) the more evidences we are able to provide (as $\mathcal{E} \rightarrow \mathcal{I}$) the bigger is our overall confidence function (confidence grows $\mathcal{C}_s^t \uparrow$), ii) our relative confidence to the universal knowledge of the hypothesis (*i.e.*, the difference in confidences) is proportionate to how much knowledge on causal possibilities we lack with respect to the universal causality hypothesis, the less causality possibilities we take into account in our formalization the smaller is our confidence in the causality hypothesis with re-

spect to the universal knowledge of the causality hypothesis, iii) our confidence in the causality hypothesis increases when we evidence more factors favored by \mathcal{F} with respect to the universal formalization of the causality hypothesis, even if we do not have full knowledge of the causality possibilities. The domain experts found that our computational methodology for assessing confidence in a causality hypothesis proportionally to the amount of available knowledge, corresponds to their subjective assessments of confidences in an investigated hypothesis. Moreover, the obtained confidence measures for the specific causality hypotheses have been validated by our domain experts, and, in some cases, have led to new interpretations of the already known causality connections (see Section Generalization of the hypothesis configuration).

Limits, assumptions and dependencies of methodology.

Overall our framework is dependant on the validity, quality and the richness of the modelling, which will induce the final shape and topology of the hypothesis graph and the way the confidence is assessed by using our methodology for confidence assessment. Of course, our methodology has its limits and has its assumptions and dependencies. Main assumptions and dependencies of the methodology for hypothesis testing rely on: i) ontological commitment of the input ontologies O_i that formalize background biological knowledge on causality relationships, ii) biological validity and logical consistency of the formalized knowledge - input to the framework, iii) weighting scheme of factors of the hypothesis that measure the quality of the ontology matching of concepts used to assemble the final ontology, and the confidence of the obtained evidence for a specific factor f_i . Ontological commitment of the modelled realities representing causality relationships among the factors should follow the good design patterns for modelling causalities, for both concepts and relationships that interrelate those concepts. In partic-



ular, we consider the processual perspective of a *disease* as a causal chain structure as in River Flow Model of Diseases [25] as opposed to an object-like perspective of a whole constituting a disease as in Ontology of General Medical Sciences (OGMS) [59]. As has been argued by Rovetto and Mizgouchi [25], the causality in OGMS is unstated, implicit or stated indirectly. The general account of disease in OGMS draws ideas from Scheuermann et al. [60], and distinguishes diseases from disease courses. Diseases in OGMS are treated as dispositions potentially realizable via pathological processes, and have some disorders as their physical basis. In our work, we focus on causality relationships which constitute a disease course, and reason on these relationships by relying on graph analysis techniques. Due to this modelling choice we expect the input ontologies to follow the RFM account of disease as a causal chain

structure. Specifically, our methodology for hypothesis graph construction extracts causality relationships from the assembled ontology such that the final hypothesis graph contains nodes as occurrents, either biological processes, as exemplary modelled in the Gene Ontology [20], or as conditions (abnormal states), according to the guidelines of the RFM. The causality relationships should be compliant with the Relation Ontology [24], which, among other types, covers concurrent and overlapping causality relationships between the occurrent entities, relying on Allen interval algebra calculus for temporal logic [61]. Strategies toward harmonization between disease accounts in OGMS and RFM are brought up in Rovetto and Mizgouchi [25]. Hypothesis graph creation with input ontologies following the OGMS modelling of disease could represent a promising future direction for the community.

Weighting scheme for the factors of the hypothesis graph will largely depend on the context (e.g., studied disease), the quality of the ontology mappings, and the confidence of the obtained evidence. Mappings to guide the assembly (*i.e.*, link factors from different hypothesis) can be discovered in online resources like UMLS Metathesaurus [49] or BioPortal [50, 51], or using state of the art ontology alignment systems like LogMap [52] or AML [53]. Confidence in the obtained evidence will depend on the methodology of the experiment and should be assessed by the executioner of the experiment, which might entail subjective importance weight of the factor and might have subjective consequences on the computation of the overall confidence in the causality hypothesis with our framework.

Conclusions

We have presented a promising and nascent methodology for the translation of biological knowledge on causality relationships of biological processes and their effects on conditions to a computational framework for shared hypothesis testing. Furthermore, we have defined a knowledge-driven, and evidenced-based way of measuring confidence in a causality hypothesis proportionally to the amount of available knowledge and collected evidences. The methodology resumes in two points: hypothesis graph construction from the formalizations of the background knowledge on causality relationships, and confidence measurement in a causality hypothesis as a normalized weighted path computation in the hypothesis graph. Lastly, we have made the source code and materials available to the community on GitHub at <https://github.com/plumdeq/hypothtest> (see Subsection Availability of data and materials).

Herein we took advantage of our domain experts to build a simplified and a tractable version of a causality hypothesis graph of cartilage degradation during to osteoarthritis, and to validate our methodology for con-

fidence assessment of causality hypothesis. The evaluation results, the feedback from our experts, and the lessons learnt from this overall experience allow us to conclude that a methodology for shared hypothesis testing could be incorporated as an invaluable asset to the online biological knowledge graph mining services. In particular, our hypothesis graph construction methodology could be used routinely to enrich biological knowledge graphs (e.g., Knowledge Bio [62]) and online databases (e.g., Gene Wiki [63]) by extracting the causality relationships information from OWL 2 ontologies. Of course, the proposed set of patterns for the normalization of the hypothesis graph will have to be augmented and tuned for a specific studied context. We, for instance, defined graph rewriting normalization patterns to deal with complementary biological scenarios of simultaneously positive and negative regulations of biological processes (see Section Robustness of the system in presence of complementary causality relationships). In fact, the graph rewriting patterns is a general paradigm for the transformation of formalized knowledge on a specific biological pattern into its equivalent graph representation and might open an opportunity for more research and practical contributions from the biomedical community.

Shared hypothesis testing services built on top of the confidence measurement (see Section Relative confidence measurement), and the inference procedures it induces (see Section Generalization of the hypothesis configuration), will enhance the biological knowledge graphs with advanced simulation functionalities for continuous research. These services could support researchers in literature review for their experimental studies, planning and prioritizing evidence collection acquisition procedures, and testing their hypotheses with different depths of knowledge on causal dependencies of biological processes and their effects on the observed conditions. Measuring confidence in a

causality hypothesis relatively to the already discovered causality relationships might serve in the assessment of the fairness of the obtained results, and its significance to the already known results. We believe that the shared hypothesis testing could serve as an important asset for the costless re-enactment of the experiments, and might eventually contribute to the future, purely computational benchmarks for the validation of the experiments.

List of abbreviations

OA: Osteoarthritis **MRI:** magnetic resonance image **SNA:** Social Network Analysis **FMA:** Foundational Model of Anatomy ontology **DL:** description logics **ABox:** description logics ground sentences stating where in the hierarchy individuals belong **OWL:** web ontology language **RDF:** resource description framework

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

We implemented a prototype to apply the proposed methodology for hypothesis testing on an example hypothesis graph. The demo of the prototype is available at <http://hypothtest.plumdeq.xyz/test/>. Source code for the hypothesis testing of the prototype and proof of concept ontologies, as well as the Jupyter Notebooks (reproducible experiments presented in this manuscript) are available on GitHub at <https://github.com/plumdeq/hypothtest>.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was partially funded by the EU Marie Curie, ITN MultiScaleHuman (FP7-PEOPLE-2011-ITN, Grant agreement no.: 289897), the CNR project DIT.AD009.006 Modelling and Analysis of anatomical shapes for computer assisted diagnosis, the BIGMED project (IKT 259055), the HealthInsight project (IKT 247784), the SIRIUS Centre for Scalable Data Access (Research Council of Norway, project no.: 237889), the program Investigador of the Portuguese Foundation of Science and Technology (FCT, IF/00423/2012), the EU project Optique (FP7-ICT-318338), and the EPSRC projects ED3 and DBOnto.

Author's contributions

AA defined most of the theory and did the implementation work. EJR extended the graph projection methodology to a richer subset of OWL axioms and proposed the hypothesis graph normalization methodology. MO composed and validated the biological use-case scenario of cartilage degradation causality hypothesis. All authors contributed to the definition of the framework, and to the writing of the manuscript. All authors read and approved the final manuscript

Acknowledgements

We would like to thank the reviewers for their valuable comments and suggestions which helped us to substantially improve the paper.

Author details

¹Italian National Research Council, Via De Marini 6, 16149 Genoa, Italy. ²University of Oslo, Oslo, Norway. ³3B's Research Group, Biomaterials, Biodegradables and Biomimetics, Headquarters of the European Institute of Excellence on Tissue Engineering and Regenerative Medicine, University of Minho, Caldas das Taipas, Portugal. ⁴ICVS/3B's - PT Government Associate Laboratory, Braga/Guimarães, Portugal. ⁵University of Genoa, Genoa, Italy. ⁶Medical University of Vienna, Center for Medical Statistics, Informatics, and Intelligent Systems, Institute for Artificial Intelligence and Decision Support, Spitalgasse 23, 1090 Vienna, Austria. ⁷Department of Biomedical Data Science, Stanford University School of Medicine, USA.

References

- Sandell, L.J.: Etiology of osteoarthritis: genetics and synovial joint development. *Nature Reviews Rheumatology* **8**(2), 77–89 (2012)
- Hafez, A.R., Alenazi, A.M., Kachanathu, S.J., Alroumi, A.M., Mohamed, E.S.: Knee Osteoarthritis: A Review of Literature. *Physical Medicine & Rehabilitation International* **1**(4), 1–8 (2014)
- Ondr sik, M., Azevedo Maia, F.R., da Silva Morais, A., Gertrudes, A.C., Dias Bacelar, A.H., Correia, C., Gonalves, C., Radhouani, H., Amandi Sousa, R., Oliveira, J.M., Reis, R.L.: Management of knee osteoarthritis. current status and future trends. *Biotechnology and Bioengineering*, (2016)
- Henrotin, Y., Pessesse, L., Lambert, C.: Targeting the synovial angiogenesis as a novel treatment approach to osteoarthritis. *Therapeutic advances in musculoskeletal disease* **6**(1), 20–34 (2014)
- van der Kraan, P.M., van den Berg, W.B.: Osteophytes: relevance and biology. *Osteoarthritis and cartilage / OARS, Osteoarthritis Research Society* **15**(3), 237–44 (2007)
- Muraki, S., Tanaka, S., Yoshimura, N.: Epidemiology of knee osteoarthritis. *OA Sports Medicine* **1**(3), 1–6 (2013)
- Goldring, M.B., Otero, M.: Inflammation in osteoarthritis. *Current Opinion in Rheumatology* **23**(5), 471–478 (2011)
- Kapoor, M., Martel-pelletier, J., Lajeunesse, D., Pelletier, J.-p., Fahmi, H.: Role of proinflammatory cytokines in the pathophysiology of osteoarthritis. *Nature Publishing Group* **7**(1), 33–42 (2011)

9. Aktas, E., Sener, E., Zengin, O., Gocun, P.U., Deveci, M.A.: Serum TNF-alpha levels: potential use to indicate osteoarthritis progression in a mechanically induced model. *European Journal of Orthopaedic Surgery & Traumatology* 22(2), 119–122 (2012)
10. Kunisch, E., Kinne, R.W., Alsalameh, R.J., Alsalameh, S.: Pro-inflammatory IL-1beta and/or TNF-alpha up-regulate matrix metalloproteases-1 and -3 mRNA in chondrocyte subpopulations potentially pathogenic in osteoarthritis: in situ hybridization studies on a single cell level. *International journal of rheumatic diseases* 19(6), 557–566 (2014)
11. Goldring, M.B.: Chondrogenesis, chondrocyte differentiation, and articular cartilage metabolism in health and osteoarthritis. *Therapeutic advances in musculoskeletal disease* 4(4), 269–85 (2012)
12. Troeberg, L., Nagase, H.: Proteases involved in cartilage matrix degradation in osteoarthritis. *Biochimica et biophysica acta* 1824(1), 133–45 (2012)
13. Kempson, G.E.: Mechanical properties of articular cartilage. *The Journal of physiology* 223(1), 23 (1972)
14. Grenier, S., Bhargava, M.M., Torzilli, P.A.: An in vitro model for the pathological degradation of articular cartilage in osteoarthritis. *Journal of Biomechanics* 47(3), 645–652 (2014)
15. Doyran, B., Tong, W., Li, Q., Jia, H., Zhang, X., Chen, C., Enomoto-Iwamoto, M., Lu, X.L., Qin, L., Han, L.: Nanoindentation modulus of murine cartilage: A sensitive indicator of the initiation and progression of post-traumatic osteoarthritis (2016)
16. Rojas, F.P., Batista, M.A., Lindburg, C.A., Dean, D., Grodzinsky, A.J., Ortiz, C., Han, L.: Molecular adhesion between cartilage extracellular matrix macromolecules. *Biomacromolecules* 15(3), 772–780 (2014)
17. Goldring, S.R., Goldring, M.B.: Changes in the osteochondral unit during osteoarthritis: structure, function and cartilage–bone crosstalk. *Nature Reviews Rheumatology* 12(11), 632–644 (2016)
18. Palmer, A.J.R., Brown, C.P., McNally, E.G., Price, A.J., Tracey, I., Jezzard, P., Carr, A.J., Glyn-Jones, S.: Non-invasive imaging of cartilage in early osteoarthritis (2013)
19. Favre, J., Jolles, B.M.: Gait analysis of patients with knee osteoarthritis highlights a pathological mechanical pathway and provides a basis for therapeutic interventions. *EFORT Open Reviews* 1(10), 368–374 (2016)
20. Ashburner, M., Ball, C.A., Blake, J.A., Botstein, D., Butler, H., Cherry, J.M., Davis, A.P., Dolinski, K., Dwight, S.S., Eppig, J.T., *et al.*: Gene Ontology: tool for the unification of biology. *Nature genetics* 25(1), 25–29 (2000)
21. Cuenca Grau, B., Horrocks, I., Motik, B., Parsia, B., Patel-Schneider, P., Sattler, U.: OWL 2: The next step for OWL. *J. Web Semantics* 6(4), 309–322 (2008)
22. Horrocks, I., Kutz, O., Sattler, U.: The even more irresistible SROIQ. In: Tenth International Conference on Principles of Knowledge Representation and Reasoning (KR), pp. 57–67 (2006)
23. McCray, A.T.: An Upper-Level Ontology for the Biomedical Domain. *Comp Funct Genomics* 4(1), 80–84 (2003)
24. Smith, B., Ceusters, W., Klagges, B., Köhler, J., Kumar, A., Lomax, J., Mungall, C., Neuhaus, F., Rector, A.L., Rosse, C.: Relations in biomedical ontologies. *Genome Biology* 6(5), 46 (2005)
25. Rovetto, R.J., Mizoguchi, R.: Causality and the ontology of disease. *Applied Ontology* 10(2), 79–105 (2015)
26. Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening ontologies with dolce. In: Proceedings of the 13th International Conference on Knowledge Engineering and Knowledge Management. *Ontologies and the Semantic Web*, pp. 166–181 (2002)
27. Arp, R., Smith, B., Spear, A.D.: Building Ontologies with Basic Formal Ontology. The MIT Press, Cambridge, US (2015)
28. Lembo, D., Santarelli, V., Savo, D.F.: Graph-Based Ontology Classification in OWL 2 QL. In: The Semantic Web: Semantics and Big Data, pp. 320–334 (2013)
29. Seidenberg, J., Rector, A.: Web Ontology Segmentation: Analysis, Classification and Use. In: Proceedings of the 15th International Conference on World Wide Web, pp. 13–22 (2006)
30. Carrington, P.J., Scott, J., Wasserman, S.: Models and Methods in Social Network Analysis. Cambridge University Press, Cambridge (2005)
31. Hoser, B., Hotho, A., Jäschke, R., Schmitz, C., Stumme, G.: Semantic Network Analysis of Ontologies. In: Proceedings of the 3rd European Conference on The Semantic Web: Research And Applications, pp. 514–529 (2006)
32. Mika, P.: Ontologies Are Us: A unified model of social networks and semantics. *Web Semantics: Science, Services and Agents on the World Wide Web* 5(1) (2011)
33. Stuckenschmidt, H., Klein, M.: Structure-Based Partitioning of Large Concept Hierarchies. In: International Semantic Web Conference, pp. 289–303 (2004)
34. Agibetov, A., Patanè, G., Spagnuolo, M.: Grontocrawler: Graph-Based Ontology Exploration. In: Proceedings of Smart Tools and Apps for Graphics - Eurographics Italian Chapter Conference, pp. 67–76 (2015)
35. Baader, F., Brand, S., Lutz, C.: Pushing the EL envelope. In: In Proc. of IJCAI 2005, pp. 364–369 (2005)
36. Solimando, A., Jimenez-Ruiz, E., Guerrini, G.: Minimizing conservativity violations in ontology alignments: Algorithms and evaluation. *Knowledge and Information Systems* (2017 (in press))
37. Arenas, M., Cuenca Grau, B., Kharlamov, E., Marciuška, Š., Zheleznyakov, D.: Faceted search over RDF-based knowledge graphs. *Web Semantics: Science, Services and Agents on the World Wide Web* 37-38, 55–74 (2016)
38. Soylu, A., *et al.*: Optiquevqs: a visual query system over ontologies for industry. *Semantic Web Journal* (2016 (submitted))
39. Larson, S.D., Martone, M.E.: Rule-Based Reasoning With A Multi-Scale Neuroanatomical Ontology. In: OWLED (2007)
40. Oberkamp, H., Zillner, S., Bauer, B.: Interpreting Patient Data using Medical Background Knowledge. In: Cornet, R., Stevens, R.

- (eds.) ICBO. *CEUR Workshop Proceedings*, vol. 897 (2012)
41. Richardson, M., Domingos, P.: Markov Logic Networks. *Mach. Learn.* **62**(1-2), 107–136 (2006)
 42. Kimmig, A., Bach, S.H., Broecheler, M., Huang, B., Getoor, L.: A Short Introduction to Probabilistic Soft Logic. In: *NIPS Workshop on Probabilistic Programming: Foundations and Applications* (2012)
 43. Hillston, J.: Process algebras for quantitative analysis. In: *20th Annual IEEE Symposium on Logic in Computer Science, 2005. LICS 2005. Proceedings*, pp. 239–248 (2005)
 44. Ciocchetta, F., Hillston, J.: Bio-PEPA: An Extension of the Process Algebra PEPA for Biochemical Networks. *Electronic Notes in Theoretical Computer Science* **194**(3), 103–117 (2008)
 45. Pearl, J.: Causal inference in statistics: An overview. *Statistics Surveys* **3**, 96–146 (2009)
 46. Pearl, J.: *Causality: Models, Reasoning, and Inference*. Cambridge University Press, UK (2000)
 47. Glimm, B., Horrocks, I., Motik, B., Stoilos, G., Wang, Z.: Hermit: An OWL 2 reasoner. *Journal of Automated Reasoning* **53**(3), 245–269 (2014)
 48. Spackman, K.: SNOMED RT and SNOMED CT. promise of an international clinical ontology. *M.D. Computing* **17** (2000)
 49. Bodenreider, O.: The unified medical language system (UMLS): integrating biomedical terminology. *Nucleic acids research* **32**(Database issue), 267–270 (2004)
 50. Fridman Noy, N., Shah, N.H., Whetzel, P.L., Dai, B., Dorf, M., Griffith, N., Jonquet, C., Rubin, D.L., Storey, M.-A.D., Chute, C.G., Musen, M.A.: BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research* **37**(Web-Server-Issue) (2009)
 51. Salvadores, M., Alexander, P.R., Musen, M.A., Noy, N.F.: BioPortal as a dataset of linked biomedical ontologies and terminologies in RDF. *Semantic Web* **4**(3), 277–284 (2013)
 52. Jiménez-Ruiz, E., Cuenca Grau, B.: LogMap: Logic-based and Scalable Ontology Matching. In: *Int'l Sem. Web Conf. (ISWC)*, pp. 273–288 (2011)
 53. Faria, D., Pesquita, C., Santos, E., Palmonari, M., Cruz, I.F., Couto, F.M.: The agreementmakerlight ontology matching system. In: *OTM Conferences*, pp. 527–541 (2013)
 54. Bang-Jensen, J., Gutin, G.Z.: *Digraphs*. Springer Monographs in Mathematics. Springer, London (2009)
 55. EU FP7 MultiScaleHuman: Multi-scale Biological Modalities for Physiological Human Articulation (MSH). <http://multiscalehuman.miralab.ch>
 56. Guilak, F.: Biomechanical factors in osteoarthritis. *Best Practice & Research Clinical Rheumatology* **25**(6), 815–823 (2011)
 57. Nicodemus, G.D., Bryant, S.J.: Mechanical loading regimes affect the anabolic and catabolic activities by chondrocytes encapsulated in PEG hydrogels. *Osteoarthritis and Cartilage* **18**(1), 126–137 (2010)
 58. Fruchterman, T.M.J., Reingold, E.M.: Graph drawing by force-directed placement. *Software: Practice and Experience* **21**(11), 1129–1164 (1991)
 59. *Ontology for General Medical Science (OGMS)*. (Accessed on April 2017): Available at <https://github.com/OGMS/ogms>
 60. Scheuermann, R.H., Ceusters, W., Smith, B.: Toward an Ontological Treatment of Disease and Diagnosis. *Summit on Translational Bioinformatics 2009*, 116–120 (2009)
 61. Allen, J.F.: Maintaining knowledge about temporal intervals. *Communications of the ACM* **26**(11), 832–843 (1983)
 62. Bruskiwich, R., Huellas-Bruskiwicz, K., Ahmed, F., Kaliyaperumal, R., Thompson, M., Schultes, E., Hettne, K.M., Su, A.I., Good, B.M.: Knowledge.Bio: A Web application for exploring, building and sharing webs of biomedical relationships mined from PubMed. *bioRxiv*, 055525 (2016)
 63. Burgstaller-Muehlbacher, S., Waagmeester, A., Mitraka, E., Turner, J., Putman, T., Leong, J., Naik, C., Pavlidis, P., Schriml, L., Good, B.M., Su, A.I.: Wikidata as a semantic framework for the Gene Wiki initiative. *Database* **2016**, 015 (2016)