## Universiteit Leiden Opleiding Informatica

Exploring scientific mobility in co-authorship networks using multilayer temporal motifs

Name: Hanjo Daniël Boekhout
Date: 30/06/2020
1st supervisor: Frank W. Takes
2nd supervisor: Vincent A. Traag

MASTER'S THESIS
Leiden Institute of Advanced Computer Science (LIACS)
Leiden University
Niels Bohrweg 1
2333 CA Leiden
The Netherlands


#### Abstract

Network motifs, small recurring configurations of nodes and edges, provide insight into complex networks by creating a better understanding of their meso-level structure. Here, we look at multilayer temporal motifs in co-authorship networks to gain a better understanding of scientific collaboration, scientific mobility, and how those relate.

In this thesis we make three contributions. First, we extend existing algorithms to efficiently count multilayer temporal motifs that include concurrent edges and to enforce edge attribute exclusivity within motifs. Second, we introduce a systematic categorization of the motifs, such that each category reflects a pattern of behaviour in the domain of scientific collaboration and mobility. Third, we use these categories to infer characteristic co-authorship and mobility behaviours for fields and countries.

We show that, in every field, there exists a geographical clustering of countries based on the importance of mobility. Furthermore, we find that increased team formation within organisations often goes hand in hand with less international collaboration and mobility. Finally, we conclude that the relationship between international collaboration and international mobility exists in both directions, collaboration leading to mobility and mobility leading to collaboration.


## 1 Introduction

Through technological advances and increasing digital communication, the world is becoming more and more connected. Small physical distances are no longer a necessity for interactions to occur. By modelling interactions between entities in complex systems as networks, the field of network science aims to understand these systems, their entities and interactions [1]. Network science has provided new insights into a wide variety of complex systems. From social networks, identifying key persons within them [2], to protein networks, contributing to the understanding of protein structure, folding, stability, function and dynamics [3], to corporate networks, studying corporate governance practices through links of corporate ownership and shared directors [4], and many more. In this thesis, we focus on scientific collaboration networks, specifically co-authorship networks which capture interactions between authors who collaborated on scientific papers.
The study of co-authorship networks, and the study of networks more generally, often focuses on explaining macro-level properties of the network as a whole, using microlevel properties of the nodes, such as node degrees [5, 6]. Some work has been done to identify noteworthy patterns at the meso-level of co-authorship networks [7, 8], often conceptualized as so-called network motifs. A motif is a configuration of nodes and edges, usually only a few, that occurs at a high rate throughout the network [9, 10]. These works studied only static motifs, i.e., motifs that consist of edges on which no order is implied and which all model the same interaction type. However, co-authorship networks are inherently dynamic [11], with new collaborations often resulting from their existing knowledge network, i.e., their past collaborations, either directly or indirectly. Our goal is to capture these dynamics with network motifs in an attempt to reason about and gain a better understanding of scientific mobility: scholars moving between organisations, a frequently studied concept in scientometrics [12]. Scientometrics is the field concerned with the study of the quantitative features and characteristics of science and scientific research. We capture the dynamic evolution of collaborations by imposing a sequential


Figure 1: Example motifs implying mobility. Edge labels imply temporal order, and indicate either O(rganisational), N(ational) or I(nternational) collaborations. Collaboration type indicates the closest proximity between the known organisations of co-authors. For each motif, mobility can be inferred from the change of collaboration type on the parallel edges.
order based on the associated papers' publication year. Additionally, we establish physical distance between collaborating authors by distinguishing between collaborations at the organisational, local, national and international level.
In recent years, methods were proposed to deal with increasingly more complex motifs. Algorithms were introduced that incorporate the evolution of networks over time in temporal motifs, to gain a greater understanding of the dynamic nature of temporal networks, also known as dynamic networks [13]. Furthermore, methods have been proposed to incorporate different types of interactions within motifs, i.e., multilayer motifs [4]. Recently, we proposed a method to efficiently count multilayer temporal motifs in large-scale networks [14]. In this work we use multilayer temporal motifs to study the direct and indirect formation of scientific collaborations based on past collaborations, which we believe will contribute to a more fine-grained understanding of the evolution of such collaborations. By making a distinction into different network layers based on collaborations at the organisational, local, national and international level, we can infer scientific mobility from the configuration of some multilayer temporal motifs. For example, the motifs depicted in Figure 1 imply mobility events through a change in collaboration distance between two authors.
We extracted five large co-authorship networks, covering different fields, from the Web of Science (WoS) database enriched by the Centre for Science and Technology Studies (CWTS). Each network consists of between 4 and 94 million collaborations on papers published in the period 2007-2016. Each pair of authors forms one collaboration edge per paper based on their closest affiliations (organisational, local, national or international layer), with the paper's publication year serving as timestamp. Our goal is to use these timestamps to impose a sequential order. However, using publication years as timestamps leads to many concurrent edges on which we do not want to infer an order. Unfortunately, the algorithms we previously proposed [14] can not properly handle concurrent edges. Furthermore, we wish to avoid counting motifs that are mostly the result of collaboration on a single paper. For example, if a paper involved three authors, all three pairs of authors would be involved in a collaboration. Motifs that include multiple of these collaborations tell us nothing about the the evolution of collaborations over time. Therefore, they are uninteresting in the context of our study and we want to exclude them from the analysis. This too is not possible with the previously proposed algorithms. To overcome these two shortcoming of the existing algorithms, we provide a methodological extension of our motif counting algorithms to handle concurrent edges, i.e., allow for multiple edges to occur within a motif with the same timestamp, and to enforce a type of edge attribute exclusivity, so that in each counted motif every edge is formed from a different paper.
The motif counting algorithms we proposed in previous work [14] had a time complexity of $O\left(m \lambda^{2}\right)$, with $\lambda$ the number of layers and $m$ the number of edges. We show that both proposed algorithm extensions can be accomplished through smart traversal of the edges, adding only a small constant factor to the complexity. Note that the proposed extended
algorithms are also able to efficiently count these motifs in static (multilayer) networks. Furthermore, the attribute exclusivity is applicable not only to co-authorship networks, but to any one-mode network projected from a two-mode network, with the one-mode edge attributes based on node attributes of the projected mode.
Because motifs are a configuration of nodes and edges, their interpretation is dependent on the real-world complex system modelled by the network. Furthermore, the volume of different multilayer temporal motifs we count, makes their combined interpretation exceedingly difficult. Therefore, we systematically assign each motif to categories that represent some real-world meaning relevant to the domain of scientific collaboration and mobility. By studying the prominence of categories in certain fields or countries and studying the interplay between the various categories, we are able to draw conclusions about typical behaviour with respect to scientific collaboration and mobility.
One aspect of scientific mobility that we are especially interested in, is how collaborations lead to scientific mobility, and how scientific mobility fosters collaboration. Studies investigating causes of international mobility $[15,16]$ have found that insertion in international knowledge networks, i.e., international contacts through, for example, co-authorships, plays an important role in the motivations for international mobility. On the contrary, Kato \& Ando [17] concluded that the relationship between international mobility and collaboration is confirmed as going in one direction, mobility resulting in collaboration. The authors state that networks created through international collaboration are not a factor in international migration. Although we are not able to identify the specific causes of individual mobility events, the results from our experiments suggest that the relationship between international mobility and collaboration actually exists in both directions. To sum up, the contributions of this thesis are as follows:

1. we extend existing motif counting algorithms to be able to handle concurrent edges;
2. we extend existing motif counting algorithms to enforce edge attribute exclusivity, such that no two edges in a counted motif can have the same attribute value;
3. we introduce a systematic categorization of the meaning of multilayer temporal motifs in the context of scientific collaboration and mobility;
4. we infer typical behaviour with respect to scientific co-authorship and mobility in general and for specific countries and scientific fields; and
5. we show that the relationship between international mobility and collaboration exists in both directions, shedding new light on the debate in [17].

The remainder of this thesis is structured as follows. First, relevant related and previous work is presented in Section 2. Then, necessary background and definitions are provided in Section 3. The counting algorithms from our previous work and our new methodological extensions are discussed in Section 4. Next, Section 5 describes the network datasets and their extraction from WoS. Then in Section 6, we add meaning to each motif configuration through categorization. Subsequently in Section 7, we perform experiments and interpret results with the use of these categories. Finally, we summarize our results and contributions and discuss future work in Section 8.

## 2 Related work

In this section we first look at work related to the motif counting problem, then work investigating co-authorship networks and studies into scientific collaboration from a network context and finally we consider work that dealt with scientific mobility.

### 2.1 Motif counting

Recently a comprehensive survey on subgraph (motif) counting methods, i.e., motif counting, was performed by Ribeiro et al. [18]. The authors provided a comprehensive review on exact, approximate and parallel methods. However, this work focussed only on methods for simple static motifs and only briefly referenced methods for more complex motifs. One such method was introduced by Paranjape et al. [13] to count a set of temporal motifs. The authors proposed algorithms that were able to efficiently (in $O(m)$ time, with $m$ the number of edges) count these motifs. In [14] we built upon this work, extending the algorithms to count multilayer temporal motifs and handle partial timing. However, this methodology still inferred an order on the 'untimed' edges based on their order in the dataset. Here, we expand on previous work by fully embracing all temporal configurations, allowing for concurrent edges to occur and be counted as such.

### 2.2 Co-authorship networks and collaboration

Kumar [19] provides an extensive review of the literature, up to 2015, on co-authorship networks. Research into co-authorship networks mostly follows three themes. First, there are papers that focus on specific fields or countries, which aim to understand these specific systems by using, for example, centrality measures to find the most prolific or influential scholars [5]. These studies tend to analyse small static networks and focus on microand macro-level properties. Second, there are papers that try to link node-specific social network measures to academic performance [6,20]. Our research falls in the third research theme, studies of collaboration itself, which we shall review below.
As we study collaboration we must realise that co-authorship and collaboration are not the same thing. Melin \& Persson [21] discuss to what extent co-authorship data reflects actual collaboration. The authors state that there is hardly a tendency for collaboration to be under-represented when studying co-authorships. However, we stipulate that this is field dependent as, for example, in social sciences much collaboration is not expressed through co-authorships, but through acknowledgements [22]. Barabási et al. [23] found co-authorship networks to be scale-free and their evolution to be governed by preferential attachment, i.e., new collaborations are more likely to connect to scholars with a high degree of collaborations. Wagner \& Leydesdorff [24] found that the growth of international co-authorships overall could be attributed to preferential attachment (individual scientists linking together searching for recognition and reward). Glänzel \& Schubert [25] found that co-authorship domesticity, the likelihood of collaborations to remain inside a country, to be clearly influenced by country size and country "remoteness" (geographic, linguistic, political, etc.). Furthermore, Wang et al. [26] found that within China geographical distance was still one of the major obstacles for collaboration. These works tend to view collaboration edges as independent, whereas we attempt to find (meaningful) collaboration patterns by considering them within the context of neighbouring collaborations.

In literature on co-authorship networks, the term "collaboration patterns" generally refers to a set of node and path measures that are characteristic for collaboration [27]. However, we use it to refer to network patterns of collaboration edges in co-authorship networks, such as motifs. Krumov et al. [7] analysed the correlation of a small set of single-layered static motifs with citation frequencies. The authors showed that the success of individual authors or publications depends unexpectedly strongly on the meso-level structure of co-authorship networks. Choobdar et al. [8] used motif fingerprints of a set of single-layered static motifs to try and assess similarity among scientific fields. They found that some motifs were overrepresented in some fields, characterizing their collaboration behaviour. Our approach differs from the previous work by Krumov et al. and Choobdar et al. as follows: (1) we consider dynamic, not static, motifs; (2) we consider multiple types of collaboration (organisational, local, national and international), i.e., multilayered motifs; and (3) we consider all motifs of a certain size, rather than a preselected set of motifs.

### 2.3 Scientific mobility

Early research into scientific mobility consisted of, often small-scale, qualitative research into "Brain drain", "Brain gain" and "Brain circulation" [28, 29]. Laudel [30] was the first to propose the use of bibliometric methods to investigate mobility, i.e., using the address field of publications to identify mobility patterns. The advent of author disambiguation methods for large bibliometric databases such as Scopus and WoS [31], allowed researchers to track authors and their affiliations, as listed on their published papers, over time.
Moed et al. [32] concluded that a bibliometric study of scientific migration using Scopus is feasible and provides significant outcomes. This sparked various lines of research. Appelt et al. [33] concluded that collaboration appeared to be a major factor associated with the mobility of scientists. Their analysis showed that the mobility of scientists particularly relied on flows of tertiary-level students in the opposite direction, from destination to origin country. Aman [34] explored the relation between CV data and Scopus data in regard to tracking international mobility of scientists. Aman showed that the majority of scientists under study had a single author ID and that laureates with 'split identities' tended to have a dominant author ID that covered the majority of their publications. Aman concluded that Scopus bibliometric data is suitable to identify a scientist's international mobility. Czaika \& Orazbayev [35] provided an empirical assessment of global scientific mobility over the past four decades. The authors found an increasing diversity of origin and destination countries, a shift of the centre of gravity of scientific knowledge production eastwards, an increase in average migration distances and found visa restrictions establishing a statistically significant barrier affecting international mobility.
Similar as for Scopus, research using WoS was sparked. Notably, Chinchilla-Rodríguez et al. [36] compared the networks of international collaboration and mobility. The authors showed that researchers collaborate internationally to a much higher degree than they become internationally mobile. Chinchilla-Rodríguez et al. [37] compared the flow of mobile researchers and the number of publications in international collaboration. The authors found that there is a significant relationship between the flow of mobile researchers and the capacity for publishing with foreign partners in the more prolific countries, but found mobility to always be lower than collaboration. Furthermore, they found that the more resources available in a country (both scientific and economic) the greater the likelihood of attracting foreign partners and mobilizing human capital.

Unlike these related works, we do not directly obtain mobility information from affiliation data, but we use affiliation data to determine collaboration distances and imply mobility from changing collaboration distances as observed in motifs. Hereby, every mobility event we capture will be directly associated with collaborations, which tells us more about the structure of the mobile author's scientific knowledge network.
Other relevant work, related to scientific mobility, focusses on the motivations for mobility. Guth \& Gill [15] found that the actual moves themselves were often due to 'chance' encounters or opportunities, but found contacts to also play an important role. Leyman [38] demonstrated that researchers that are encouraged by their supervisor to go abroad show more interest in international mobility. Notably, Baruffaldi \& Landoni [16] found that insertion in international knowledge networks and the presence of links with the source country increased the probability of future mobility. On the contrary, Kato \& Ando [17] found that networks created through international collaboration are not a factor in international migration. The authors concluded that the relationship between international mobility and collaboration is confirmed as going in one direction, from mobility to collaboration. Based on the collaboration motifs we find, that lead to or result from international mobility, we try to shed new light on this debate.

## 3 Background, notation and definitions

In this section, we provide definitions and introduce notation used to describe the algorithms discussed in this thesis. We follow the notation and definitions introduced in [13] and build upon the definitions in [14].

### 3.1 Network notation and definitions

The two basic building blocks of any network are nodes and edges. An edge is a directed link between an ordered pair of nodes $(u, v)$, which denotes $u$ as the source node and $v$ as the target node. Given a node set $V$ of size $n=|V|$, a multilayer temporal graph $H=(V, E)$ is defined by a set $E$ containing edges $e_{i}=\left(u_{i}, v_{i}, t_{i}, l_{i}\right)$, for $i=1,2, \ldots, m$, with $u_{i}, v_{i} \in V$, timestamp $t_{i} \in \mathbb{R}^{+}$and layer $l_{i} \in\{1,2, \ldots, \Lambda\}$, with $\Lambda$ the number of layers. Note that for multilayer networks $\Lambda$ must be greater than one, otherwise, if $\Lambda=1$, it is a single-layer network. Furthermore, note that concurrent edges, edges with the same timestamp, are allowed and that parallel edges with the same direction and layer are also possible. The underlying static graph $G$ of a multilayer temporal graph $H$ is the graph formed by ignoring all timestamps and layers and subsequently removing any resulting duplicate edges. Although co-authorship networks are undirected, we assume edges to always be directed for the definitions and algorithms in this thesis. This enables us to define and implement algorithms that can handle both directed and undirected networks, since the undirected results can be obtained through a simple post-processing step, which we describe in Section 6.1.

### 3.2 Multilayer temporal motifs

In our previous work [14], we gave the following definition for multilayer temporal motifs.

Definition 1. A $r$-node, $s$-edge, $\delta$-temporal, $\lambda$-layer motif is a sequence of $s$ edges, $M=$ $\left(\left(u_{1}, v_{1}, t_{1}, l_{1}\right),\left(u_{2}, v_{2}, t_{2}, l_{2}\right), \ldots,\left(u_{s}, v_{s}, t_{s}, l_{s}\right)\right)$ that are time-ordered within a $\delta$ duration, i.e., $t_{1}<t_{2}<\ldots<t_{s}$ and $t_{s}-t_{1} \leq \delta$, and range over at most $\lambda$ different layers, such that the underlying static graph, induced by $M$, is connected and has r nodes.

Note that the definition requires all edges in a motif to occur within $\delta$ time. This requirement gives us control over the period of time between interactions (edges) that we consider short enough to imply a relation between the interactions. For example, in a co-authorship network, co-authorships that are a year apart are very relevant to each other, while in a social network, such as Twitter, the relation between interactions that are a year apart is likely less meaningful. Furthermore, note that $\lambda$ defines an upper limit on the number of layers involved. The definition allows for $\lambda$ different layers in a motif $M$, but also allows fewer layers. This means that, for example, every 3 -node, 3 -edge, $\delta$-temporal, 2 -layer motif is also a 3 -node, 3 -edge, $\delta$-temporal, 3 -layer motif.
Definition 1 induces a strict order on the edges based on the timestamps. Because this ordering is strict, it does not allow for concurrent edges to occur within a motif. We redefine multilayer temporal motifs below, such that it encapsulates concurrent edges.

### 3.2.1 Concurrent edges

To facilitate concurrent edges in our previous definition, we would only have to change the strict order $(<)$ to a partial order $(\leq)$. However, this change would introduce ambiguity as different orderings of concurrent edges could be considered different motifs. Instead, we define a rank-order as

Definition 2. The rank-order of element $x_{i}$ in a set $\left(x_{1}, x_{2}, \ldots, x_{m}\right)$ is an integer $o_{i} \in N^{+}$ (i.e. $o_{i} \geq 1$ ) such that $o_{i}<o_{j}$ if and only if $x_{i}<x_{j}, o_{i}=o_{j}$ if and only if $x_{i}=x_{j}$ and $\min _{j} o_{j}-o_{i}=1$ for $o_{i}<o_{j}$, with $\min _{i} o_{i}=1$,
and redefine a multilayer temporal motif allowing for concurrent edges as follows.
Definition 3. $A$-node, $s$-edge, $\delta$-temporal, $\lambda$-layer motif is a sequence of $s$ edges, $M=\left(\left(u_{1}, v_{1}, t_{1}, l_{1}\right),\left(u_{2}, v_{2}, t_{2}, l_{2}\right), \ldots,\left(u_{s}, v_{s}, t_{s}, l_{s}\right)\right)$ with rank-ordering $o=\left(o_{1}, o_{2}, \ldots, o_{s}\right)$, where $o_{i}$ is the rank-order of timestamp $t_{i}$, such that $t_{s}-t_{1} \leq \delta,\left(l_{1}, l_{2}, \ldots, l_{s}\right)$ range over at most $\lambda$ different layers, and the underlying static graph, induced by $M$, is connected and has r nodes.

This definition covers the full set of multilayer temporal motifs given some values for $r, s$ and $\lambda$. To be able to count these motifs, we must distinguish between different configurations of the edges, their direction, temporal order and layers. We define a multilayer temporal motif configuration as follows.

Definition 4. A multilayer temporal motif configuration, $M_{a, b, c, d}$, of a r-node, s-edge, $\delta$-temporal, $\lambda$-layer motif, is a combination of:
a. a structural configuration, i.e., an assignment of the s edges over the $r$ nodes forming, for example, an (e) edge motif, (s) star motif or (t) triangle motif;
b. a temporal configuration, i.e., an assignment of a rank-order to each of the s edges defining a rank-ordering $o=\left(o_{1}, o_{2}, \ldots, o_{s}\right)$; and
c. a directional configuration, i.e., an assignment of a direction to each of the sedges, with $2^{s}$ possible configurations;
d. a layer configuration, i.e., an assignment of a layer, from $\{1, \ldots, \lambda\}$, to each of the $s$ edges, with $\lambda^{s}$ possible layer configurations.

The static motif configuration of a multilayer temporal motif configuration $M_{a, b, c, d}$ is given by $M_{a, c}$, i.e., the structural and directional configurations. The full set of 2-node and 3 -node, 3-edge, $\delta$-temporal motif configurations is depicted in Figure 2. Here, we only show single-layer motifs, because every $\delta$-temporal $\lambda$-layer motif can be associated with a single $\delta$-temporal motif [14]. For each of the 88 configurations shown in Figure 2, there exist $\lambda^{s}$ layer configurations. Note that, for motif configurations $M_{a, b, c}$, such as $M_{e, 2,1}$, not every layer configuration is unique. After all, interchanging the layers of the concurrent edges of $M_{e, 2,1}$ results in an identical motif. Furthermore, note that the same rank-ordering with the rank-orders assigned to different edges in the same static motif configuration can constitute different temporal configurations, for example, $M_{e, 2,2}$ and $M_{e, 5,2}$.




(a) edge motifs

(c) 3-node triangle motifs

Figure 2: All 2-node and 3-node, 3-edge $\delta$-temporal single-layer motif configurations allowing for concurrent edges. Edge numbers indicate their rank order. Rows have consistent temporal configurations and columns have consistent directional configurations.

Each occurrence of a motif configuration in a multilayer temporal graph $H$ is called an instance and is defined as follows.

Definition 5. An instance of a multilayer temporal motif configuration $M_{a, b, c, d}$ in a multilayer temporal graph $H$, is a sequence $S=\left(\left(w_{1}, x_{1}, t_{1}^{\prime}, l_{1}^{\prime}\right), \ldots,\left(w_{s}, x_{s}, t_{s}^{\prime}, l_{s}^{\prime}\right)\right)$ of $s$ unique edges in $H$ with rank-ordering $o^{\prime}=\left(o_{1}^{\prime}, o_{2}^{\prime}, \ldots, o_{s}^{\prime}\right)$, where $o_{i}^{\prime}$ is the rank-order of timestamp $t_{i}^{\prime}$, such that

1. there exists a bijection $f$ such that $f\left(w_{j}\right)=u_{i}, f\left(x_{j}\right)=v_{i}, l_{i}=l_{j}^{\prime}$ and $o_{i}=o_{j}^{\prime}$; and
2. the edges all occur within $\delta$ time, i.e., $t_{s}^{\prime}-t_{1}^{\prime} \leq \delta$.

Note that this definition requires the sequence $S$ to have the same rank-ordering as the motif configuration, but not the exact same edge ordering. Therefore, we must be vigilant of equivalent edge orderings of the concurrent edges in our counting algorithms. The main problem, for which algorithms are proposed in Section 4, is as follows:

Problem statement. Given values for $\delta$ and $\lambda$ and a multilayer temporal graph $H$, compute the number of instances of every 2-node and 3-node, 3-edge, $\delta$-temporal, $\lambda$-layer motif.

### 3.2.2 Edge attribute exclusivity

In addition to allowing concurrent edges, the second algorithmic contribution we make is edge attribute exclusivity within motifs. That is, we only count motifs that have no common attribute values on their edges. We define an additional edge attribute $p_{i}$ for each edge $(i=1,2, \ldots, m)$. Enforcing edge attribute exclusivity yields the following definition of a multilayer temporal edge-attribute-exclusive motif $M_{\text {excl }}$.

Definition 6. $A r$-node, $s$-edge, $\delta$-temporal, $\lambda$-layer edge-attribute-exclusive motif is a sequence of $s$ edges, $M=\left(\left(u_{1}, v_{1}, t_{1}, l_{1}, p_{1}\right),\left(u_{2}, v_{2}, t_{2}, l_{2}, p_{2}\right), \ldots,\left(u_{s}, v_{s}, t_{s}, l_{s}, p_{s}\right)\right)$ with rank-ordering $o=\left(o_{1}, o_{2}, \ldots, o_{s}\right)$, where $o_{i}$ is the rank-order of timestamp $t_{i}$, such that $t_{s}-t_{1} \leq \delta,\left(l_{1}, l_{2}, \ldots, l_{s}\right)$ range over at most $\lambda$ different layers, the underlying static graph, induced by $M$, is connected and has $r$ nodes and such that for all $i \neq j$ with $1 \leq i, j \leq s$ we have $p_{i} \neq p_{j}$.

Our definition of a motif configuration remains unchanged for edge attribute exclusivity, but we do require a motif instance to adhere to the additional requirement that no two edges in the sequence $S$ may have the same edge attribute value.
The algorithms we provide in Section 4 are able to enforce edge attribute exclusivity for a particular type of edge attributes. These attributes must be directly linked to the edge timestamp, i.e., if the attribute values are equal then the timestamps must be equal as well, but equal timestamps do not need to imply equal attribute values. This will always hold for one-mode networks that are projected from a two-mode network when the one-mode edge attribute uniquely identifies a node in the two-mode network. After all, the timestamp, and all other edge attributes, in the one-mode network originate from the same node in the two-mode network.

## 4 Motif counting algorithms

In this section we present the motif counting algorithms. We build on several existing algorithms $[13,14]$ of which we discuss the basic concepts and functionality in Section 4.1. In Section 4.2, we reformulate and extend the general algorithm in order to count edge motifs (see Figure 2a). A detailed discussion of the extensions that allow for concurrent edges and enforce edge attribute exclusivity is provided. Likewise, we extend existing algorithms and discuss the extensions for star and triangle motifs (see Figure 2b and 2c) in Section 4.3 and 4.4, respectively.

### 4.1 Existing algorithms

Paranjape et al. [13] introduced three algorithms to count temporal motifs with a strict temporal order, a general algorithm and two specialised algorithms for two specific 3-node, 3-edge structural configurations. These algorithms were extended to count multilayer temporal motifs in [14]. The approach for each of the algorithms is to count all motif instances in an input sequence $(S)$ in a single pass, thereby achieving a minimal number of considerations of each edge. The formation of the input sequences and functionality of the counting algorithm differs between the three algorithms. Below, we first compare the format of the input sequences for the three algorithms in Section 4.1.1 and then focus on their functionality in Section 4.1.2.

### 4.1.1 Input sequences

The general algorithm, which focuses on a single undirected static motif configuration, i.e., structural configuration $\left(M_{a}\right)$, at a time, determines a separate input sequence for every instance of $M_{a}$. This is efficient for edge motifs, motifs that consist only of edges between two nodes, because each edge will only belong to one instance of $M_{a}$ and will therefore be added to only one input sequence. On the contrary, Paranjape et al. [13] showed that an edge may appear in a great number of instances of $M_{a}$ for motifs that cover more than two nodes $(r>2)$. The authors concluded that their general algorithm is only efficient $(O(m)$ time) for edge motifs. Our extension of the general algorithm to multilayer temporal motifs in [14], increased the time complexity to $O\left(m \lambda^{2}\right)$. However, for a small number of layers, $\lambda^{2}$ is negligible with respect to the time complexity.
The two specialised algorithms reduce the number of input sequences an edge can appear in. For star motifs, motifs that consist of a center node $u$ and edges to $r-1$ neighbours, this is achieved by grouping together all star motifs with the same center node. Only one input sequence is gathered for each center node $u \in V$ by gathering all edges connected to $u$. Thus, every edge $(u, v)$ is only added to the two input sequences with respectively $u$ and $v$ as center nodes and a time complexity of $O\left(m \lambda^{2}\right)$ is achieved.
For triangle motifs, motifs whose edges form a triangle, the number of input sequences an edge appears in is reduced by assigning each static triangle to a pair of its nodes, i.e., one of its edges. Specifically, each static triangle is assigned to the node pair $u, v$ that is connected by the greatest number of multilayer temporal edges. An input sequence is gathered, for each node pair $u, v \in V$ to which at least one static triangle is assigned, by gathering the edges connecting $u$ and $v$ and the edges connecting them to their common neighbours as determined by the assigned static triangles. Paranjape et al. [13] proved that this reduces the time complexity of counting triangle motifs from $O(m \tau)$ for the general
algorithm to $O(m \sqrt{\tau})$, with $\tau$ the number of static triangles. Therefore, the extended algorithm to multilayer temporal motifs in [14] has a time complexity of $O\left(m \sqrt{\tau} \lambda^{2}\right)$.

### 4.1.2 The delta-timeframe

As mentioned above, all three algorithms count motif instances in a single pass over the input sequence $S$. To accomplish this, the sequence is first pre-processed to sequence $S^{\prime}$ such that $S^{\prime}$ is time-ordered and all layers that are not of interest to a specific study are filtered out. Remember that $S^{\prime}$ can still be strictly time-ordered because the existing algorithms do not yet consider concurrent edges. As such, when we iterate over the edges in $S^{\prime}$ we also move sequentially through time.
As we iterate over the edges in the sequence $S^{\prime}$, at time $t_{j}$ we consider $e_{j}$ the current edge and we know that all motifs that include $e_{j}$ consist of edges in the time window $\left[t_{j}-\delta, t_{j}+\delta\right]$, which we call the $\delta$-timeframe, depicted in Figure 3. Because all motif instances that include $e_{j}$ occur in its $\delta$-timeframe, each edge in the sequence $S^{\prime}$ has to be processed at most three times: (1) when it enters the $\delta$-timeframe; (2) when the edge is the current edge; and (3) when it leaves the $\delta$-timeframe.
The general algorithm only uses the 'pre' segment of the $\delta$-timeframe. Because the algorithm considers a single instance of a structural configuration $\left(M_{a}\right)$ at a time, we have knowledge of all nodes in the input sequence and we can generate all possible combinations of edges up to length $s$, the number of edges in the target motifs. The algorithm maintains a counter for all such combinations that form subsequences of motif configurations ( $M_{a, b, c, d}$ ) under investigation. For example, given motif configuration $M_{s, 4,2}$ on nodes $a, b$ and $c$, as depicted in Figure 4, the set of edge combinations for which a counter is maintained is $\{((a, b)),((a, c)),((c, a)),((a, b),(a, c)),((c, a),(a, c)),((a, b),(a, c),(c, a))\}$. The counters are maintained such that at time $t_{j}$ they indicate how often each of the edge combinations occur within time window $\left[t_{j}-\delta, t_{j}\right]$, i.e., how often they occur in the 'pre' segment. Thus, an edge $e_{x}$ is considered the last edge in the temporal configuration at time $t_{x}=t_{j}$, after which it fulfils the role of earlier edges until $t_{x}<t_{j}-\delta$ and it is removed from the counters. So, the general algorithm has to consider each edge in the input sequence $S^{\prime}$ only twice. Because the number of edge combinations, and thus the number of counters, explodes as the number of nodes under consideration increases, the specialised 3-node, 3-edge star and triangle algorithms are not able to utilise this same counting method. Instead they utilise the full $\delta$-timeframe and consider not the last edge in the temporal configuration at time $t_{j}$, but consider a specific, strategically chosen, edge in the structural configuration as the pivotal edge. Counters are then maintained for all edge combinations of the remaining two edges in the configuration within the full $\delta$-timeframe, such that at time $t_{j}$ all motif instances with the current edge $e_{j}$ as the pivotal edge in the configuration can be counted. However, unlike the general algorithm, which defined a counter for every specific edge combination, knowledge of the exact edges is discarded for the two edge combinations.


Figure 3: $\delta$-timeframe


Figure 4: Example configuration instance


Figure 5: All temporal configurations of star motifs provided no concurrent edges [13]

The counters simply specify specific temporal, directional and layer configurations of the edges. We discuss why this is possible below.
For star motifs, the single edge, e.g., edge $(a, b)$ in Figure 4, is chosen as the pivotal edge. Assuming a strict temporal order, Figure 5 shows the various temporal configurations. Excluding the pivotal edge, the remainder of the structural configuration consists of two parallel edges connecting the center node to the same neighbour. Now, if we discard knowledge of the neighbour to which the parallel edges connect, every combination of the parallel edges with a third edge from the center node forms either an edge motif or a star motif. Because we can count edge motifs in $O\left(m \lambda^{2}\right)$ time using the general algorithm, we can compensate for the edge motifs that we incorrectly counted as star motif by deducting the edge motif counts for center node $u$ to all its neighbours. This is preferable as it reduces the number of counters required, as well as the time complexity, by a factor $n$.
For triangle motifs, where the input sequence is based on a node pair, the edge connecting this node pair is chosen as the pivotal edge. Now, the two remaining edges connect the node pair to a common neighbour and the pivotal edge connecting the node pair requires no knowledge of this neighbour at all. Thus, knowledge of the exact common neighbour can be discarded without issue for counters representing two edge combinations.

### 4.2 Edge motifs

The general algorithm introduced by Paranjape et al. [13] was shown to be efficient only for edge motifs. Recall that edge motifs are motifs that consist of edges connecting only two nodes. Here, we reformulate the general algorithm to focus on counting edge motifs specifically and extend it to handle concurrent edges and enforce edge attribute exclusivity. With three edges and concurrent edges allowed we get four temporal configurations, which we label as shown in Figure 6. Adding edge directionality leads to the set depicted in Figure 2a.
As discussed in Section 4.1.1 an input sequence $(S)$ is gathered for every instance of a static edge, i.e., an input sequence is gathered for every connected node pair $u, v \in V$. Where existing algorithms only time-ordered the input sequence, we now pre-process it into a time-ordered sequence of sets of concurrent edges $S^{\prime}$, such that the rank-order of every edge in a concurrent set of edges is equal. The sequence is further pre-processed to account for edge attribute exclusivity. Note that, for our purposes in this thesis, these attributes are directly linked to the paper from which a co-authorship edge originates. This means that every edge that has the same attribute value, i.e., shares the same origin paper, also shares the same timestamp, i.e., the same publication date. It is exactly this observation that allows us to achieve edge attribute exclusivity within the current approach by grouping the edges with the same attribute value, within the sets of concurrent edges, together. This results in a sequence $S^{\prime \prime}$ of sets of sets of concurrent equal attribute value edges. Each edge can be described by its direction dir, denoting that the edge is from $u$ to


Figure 6: Edge motif temporal configurations
$v(0)$ or from $v$ to $u(1)$, timestamp $t$, layer $l$ and edge attribute $p$. As such, the final input sequences to the edge motif counting algorithm can be defined as

Definition 7. Sequence $S^{\prime \prime}$ is a sequence of sets of sets of edges:
$S^{\prime \prime}=\left(\left(\left\{\left(\left\{e_{1}=\left(\operatorname{dir}_{1}, l_{1}\right), \ldots\right\}, p_{1}\right), \ldots\right\}, t_{1}\right), \ldots,\left(\left(\left\{\ldots,\left(\left\{\ldots, e_{L}=\left(\operatorname{dir}_{L}, l_{L}\right)\right\}, p_{P}\right)\right\}, t_{T}\right)\right)\right)$, such that $t_{1}<t_{2}<\ldots<t_{T}$ and for all $i \neq j$ with $1 \leq i, j \leq P$ we have $p_{i} \neq p_{j}$.

Like the general algorithm, only the 'pre' segment of the $\delta$-timeframe is utilised by our reformulation of the algorithm. However, we shift from counters for exact edge combinations to counters capturing the various temporal, directional and layer configurations, as used by the specialised algorithms. We define the following counters:

- nodes $[d i r, l]$ counts the number of times nodes $u$ and $v$ are connected with direction dir and layer $l$ in the time window $\left[t_{j}-\delta, t_{j}\right)$
- conc_nodes $[d i r, l]$ counts the number of times nodes $u$ and $v$ are connected with direction dir and layer $l$ at a given time $t_{i}$, i.e., the number of concurrent edges
- $\operatorname{sum}\left[\operatorname{dir}_{1}, l_{1}, d i r_{2}, l_{2}\right]$ counts the number of strictly ordered pairs of edges in $\left[t_{j}-\delta, t_{j}\right)$ with the first edge having direction $\operatorname{dir}_{1}$ and layer $l_{1}$ and the second edge direction $d i r_{2}$ and layer $l_{2}$
- pre_partial_sum $\left[d i r_{1}, l_{1}, \operatorname{dir}_{2}, l_{2}\right]$ counts the number of strictly ordered pairs of edges, such that the first edge is in $\left[t_{j}-\delta, t_{j}\right)$ and the second edge is at time $t_{j}$, with the first edge having direction $\operatorname{dir}_{1}$ and layer $l_{1}$ and the second edge direction $\operatorname{dir}_{2}$ and layer $l_{2}$
- pre_conc_sum $\left[d i r_{1}, l_{1}, d i r_{2}, l_{2}\right]$ counts the number of pairs of concurrent edges in $\left[t_{j}-\delta, t_{j}\right)$, with $d i r_{1}$ and $d i r_{2}$ indicating the directional configuration and $l_{1}$ and $l_{2}$ the layer configuration
- conc_sum $\left[d i r_{1}, l_{1}, d i r_{2}, l_{2}\right]$ counts the number of pairs of concurrent edges at a given time $t_{i}$, with $\operatorname{dir}_{1}$ and $d i r_{2}$ indicating the directional configuration and $l_{1}$ and $l_{2}$ the layer configuration
- concurrent $\left[\operatorname{dir}_{1}, l_{1}, \operatorname{dir}_{2}, l_{2}, \operatorname{dir}_{3}, l_{3}\right]$ counts the full motifs of the 'concurrent' temporal configuration within $\delta$ time, with $d i r_{1}, d i r_{2}$ and $d i r_{3}$ indicating the directional configuration and $l_{1}, l_{2}$ and $l_{3}$ indicating the layer configuration of the three edges, respectively
- pre_partial $\left[d i r_{1}, l_{1}, \operatorname{dir}_{2}, l_{2}, d i r_{3}, l_{3}\right]$, post_partial $\left[\operatorname{dir}_{1}, l_{1}, d i r_{2}, l_{2}, d i r_{3}, l_{3}\right]$ and serial $\left[d i r_{1}, l_{1}, d i r_{2}, l_{2}, d i r_{3}, l_{3}\right]$ analogues to the concurrent counter, but each matching their own temporal configuration.

The reformulated and extended edge motif counting algorithm is shown in Algorithm 1. As can be seen on lines 1-6, the approach of this algorithm remains unchanged from that described in Section 4.1.2. Where the existing general algorithm iterated over the input sequence one edge at a time, we now iterate over one set of concurrent edges, in input sequence $S^{\prime \prime}$, at a time (line 2). In fact, this is exactly the same behaviour except that now more than one edge can share the same timestamp. Furthermore, as we iterate over the input sequence the counters are updated at the same points in time as well. The counters
are updated when a concurrent set of edges becomes the current set $\left(\operatorname{coll}_{j}\right)$, line 5 , and when a concurrent set of edges leaves the $\delta$-timeframe, lines $3-4$. Thus, the main change to the algorithm comes from how and which counters are updated at those times.
Because every counter represents a specific combination of a structural and temporal configuration and because no two separate concurrent sets can have edges between them with the same edge attribute value, the extensions to accommodate concurrent edges and enforce edge attribute exclusivity are achieved through the simple addition of the appropriate counters and their update logic. As such, the counters and update logic used for the temporal configuration with no concurrent edges ('serial') remains unchanged from existing algorithms other than requiring to iterate over the edges in the concurrent set. We discuss how the extended algorithm handles concurrent edges in Section 4.2.1 and how edge attribute exclusivity is enforced in Section 4.2.2.

### 4.2.1 Concurrent edges

Because concurrent edges occur at the same time, we process all edges in a concurrent set at the same time. In Algorithm 1, this means that we count all combinations of concurrent edges within a single call to 'DecrementCounts' or 'IncrementCounts'. These combinations are formed in loops at lines 11-16 and 19-28, respectively. Note that, because a single forward pass through the set is performed, each combination of concurrent edges is formed exactly once. Furthermore, by constructing all combinations of concurrent edges in a single forward pass, we prevent counting the same edge as two concurrent edges.
Because we use a single forward pass, the ordering of the edges within a concurrent set now determines the ordering of the directional and layer configurations counted. However, the ordering of the directional and layer configurations among concurrent edges has no meaning for edge motifs. After all, all concurrent edges connect the same two nodes and therefore their ordering is interchangeable, i.e., concurrent edge combinations $((u, v, A),(v, u, B))$ and $((v, u, B),(u, v, A))$, with layers $A$ and $B$, are no different. Note that we are talking about changing the order of both the direction and layer at the same time. Because the ordering of the directional and layer configurations are interchangeable, we would want to count an occurrence of either as an occurrence of both. This is achieved in a post-processing step by adding the counted total of all equivalent directional and layer configuration permutations together and giving their sum as the result for each of them. Note that for concurrent edges in an undirected network, the equivalence of layer configurations is no longer dependent on the directionality. For example, $((u, v, A),(v, u, B))$ and $((u, v, B),(v, u, A))$ are not equivalent in a directed network, but are equivalent in an undirected network. The resolution of these new equivalences as an additional post-processing step, allow us to go from directed to undirected motif count results.

### 4.2.2 Edge attribute exclusivity

Earlier we stated that to realise edge attribute exclusivity with co-authorship networks, where the attribute uniquely identifies the source paper, we only have look out for equal edge attribute edges when dealing with concurrent edges. In Algorithm 1, we achieve this with the addition of the various temporary ('tmp') counters. Looking at lines 11-16 and 19-28, we see that we update these temporary counters, such as 'tmp_nodes', in the inner loops (lines $12-15$ and $20-27$ ), where we loop over edges with the same edge attributes, as stand-ins for real counters such as 'conc_nodes'. The temporary counters are subsequently

```
Algorithm 1: Algorithm for counting the number of instances of all 2-node 3-edge
\(\delta\)-temporal \(\lambda\)-layer edge motifs \(M_{\text {excl }}\). We assume the keys of counters are accessed in
order of length. The ":" notation indicates element-wise operations on those indices.
    Input: Sequence \(\left(S^{\prime \prime}\right)\) of sets of sets of edges, with respectively equal timestamps \((t)\) and
                        edge attributes \((p)\), with \(t_{1}<\ldots<t_{T}\), time window \(\delta\) and \(\forall_{i}: l_{i} \in\{0, \lambda-1\}\) :
                            \(S^{\prime \prime}=\left(\operatorname{coll}_{1}=\left(\left\{\operatorname{coll}_{11}=\left(\left\{e_{1}=\left(\operatorname{dir}_{1}, l_{1}\right), \ldots\right\}, p_{1}\right), \ldots\right\}, t_{1}\right), \ldots\right.\),
            \(\left.\left(\operatorname{coll}_{T}=\left(\left\{\ldots,\left(\left\{\ldots, e_{L}=\left(\operatorname{dir}_{L}, l_{L}\right)\right\}, p_{P}\right)\right\}, t_{T}\right)\right)\right)\)
    Output: Number of 2-node 3 -edge \(\delta\)-temporal \(\lambda\)-layer edge motifs \(M_{\text {excl }}\) in sequence \(S^{\prime \prime}\)
    Initialize all counters to 0 , start \(\leftarrow 1\)
    for \(j=1, \ldots, T\) do
    while \(t_{\text {start }}<t_{j}-\delta\) do
        DecrementCounts(coll \({ }_{\text {start }}\), nodes, sum, pre_conc_sum), start \(+=1\)
    IncrementCounts(coll \({ }_{\mathrm{j}}\), nodes, sum, pre_conc_sum)
    return concurrent, pre_partial, post_partial, serial
    Procedure DecrementCounts(colls, nodes, sum, pre_conc_sum)
    for coll in colls do
            for \(e=(d i r, l)\) in coll do
                nodes \([\) dir,\(l]-=1\)
    for coll in colls do
            for \(e=(d i r, l)\) in coll do
                \(\operatorname{sum}[d i r, l,:,:]\) - = nodes[:,:]
                pre_conc_sum \([:,:, d i r, l]-=\) conc_nodes[:,:]
                tmp_nodes \([d i r, l]+=1\)
            conc_nodes \(\leftarrow\) tmp_nodes
    reset conc_nodes, tmp_nodes
    Procedure IncrementCounts(colls, nodes, sum, pre_conc_sum)
        for coll in colls do
            for \(e=(d i r, l)\) in coll do
                concurrent \([:,:,:,:,:\) dir,\(l]+=\) conc_sum \([:,,:,:,:]\)
                pre_partial[:,:,:,:,dir, \(l]+=\) pre_conc_sum \([.,,,:,:\),
                post_partial[:,:,:,:,,dir,l] += pre_partial_sum[:,,:,:,: \(]\)
                serial[:,,:,:,:,dir,\(l]+=\operatorname{sum}[:,,:,:,:]\)
                tmp_sum[:,:,dir,\(l]+=\) conc_nodes \([:,:]\)
                tmp_pp_sum[:,:,dir,\(l]+=\) nodes \([:,:]\)
                tmp_nodes \([\) dir,\(l]+=1\)
            conc_sum \(\leftarrow\) tmp_sum, pre_partial_sum \(\leftarrow\) tmp_pp_sum, conc_nodes \(\leftarrow\) tmp_nodes
    for coll in colls do
            for \(e=(d i r, l)\) in coll do
                sum[:,:,dir,l] += nodes[:,:]
        for coll in colls do
            for \(e=(d i r, l)\) in coll do
                nodes \([\) dir,\(l]+=1\)
    pre_conc_sum \([:,,:,:,:]+=\) conc_sum \([:,:,:,:\),
    reset conc_nodes, tmp_nodes, conc_sum, tmp_sum, pre_partial_sum, tmp_pp_sum
```

used to update the real counters in the outer loops (lines 16 and 28), thereby updating these counters for the entire set of equal attribute value edges at a time. As a result, larger combinations of concurrent edges formed using these real counters, never include two edges from the same set of equal attribute value edges, i.e., we never count combinations of concurrent edges with the same edge attribute value.
Note that we are able to enforce edge attribute exclusivity without having to store information regarding the attribute as part of any counter nor the input sequence. We only require the addition of a small set of temporary counters and a minimal set of operations.

In short, we are able to deal with concurrent edges and realise edge attribute exclusivity through the simple addition of a few counters and a more systematic loop over the edges in the input sequence $S^{\prime \prime}$. Thus adding only a small constant number of operations per edge, based on the number of additional counters and maintaining time complexity $O\left(m \lambda^{2}\right)$.

### 4.3 Star motifs

Star motifs are motifs that consist of a center node $u$ and edges to $r-1$ neighbours [13, 14]. Given three nodes and three edges and allowing for concurrent edges, there are eight temporal configurations, which we label as shown in Figure 7. The full set of directed star motif configurations is depicted in Figure 2b.
As discussed in Section 4.1.1 an input sequence $(S)$ is gathered for every (center) node $u \in V$. These input sequences are pre-processed into sequences $S^{\prime \prime}$ consisting of sets of sets of concurrent equal attribute value edges, in the same manner as for edge motifs. Each edge in a star motif can be described by its neighbour node $n b r$, its direction dir outward from (0) or inward to (1) $u$, timestamp $t$, layer $l$ and edge attribute $p$. As such, the final input sequences to the star motif counting algorithm can be defined as

Definition 8. Sequence $S^{\prime \prime}$ is a sequence of sets of sets of edges: $S^{\prime \prime}=\left(\left(\left\{\left(\left\{e_{1}=\right.\right.\right.\right.\right.$ $\left.\left.\left.\left.\left.\left(n b r_{1}, d i r_{1}, l_{1}\right), \ldots\right\}, p_{1}\right), \ldots\right\}, t_{1}\right), \ldots,\left(\left(\left\{\ldots,\left(\left\{\ldots, e_{L}=\left(n b r_{L}, d i r_{L}, l_{L}\right)\right\}, p_{P}\right)\right\}, t_{T}\right)\right)\right)$, such that $t_{1}<t_{2}<\ldots<t_{T}$ and for all $i \neq j$ with $1 \leq i, j \leq P$ we have $p_{i} \neq p_{j}$.

As discussed in Section 4.1.2 the single, non parallel, edge in the star motif configuration is chosen as the pivotal edge and counters are formed for all edge combinations for the remaining two parallel edges over the full $\delta$-timeframe. This leads to the following set of counters:

- pre_nodes $[d i r, n b r, l]$ counts the number of times neighbour $n b r$ has appeared in an edge alongside $u$ with direction dir and layer $l$ in the time window $\left[t_{j}-\delta, t_{j}\right)$
- pre_sum $\left[\operatorname{dir}_{1}, l_{1}, d i r_{2}, l_{2}\right]$ counts the number of strictly ordered pairs of parallel edges in $\left[t_{j}-\delta, t_{j}\right)$ with the first edge having direction $\operatorname{dir}_{1}$ and layer $l_{1}$ and the second edge direction $\operatorname{dir}_{2}$ and layer $l_{2}$


Figure 7: Star motif temporal configurations

- pre_conc_sum $\left[d i r_{1}, l_{1}, \operatorname{dir}_{2}, l_{2}\right]$ counts the number of pairs of concurrent parallel edges in $\left[t_{j}-\delta, t_{j}\right.$ ), with $d i r_{1}$ and $d i r_{2}$ indicating the directional configuration and $l_{1}$ and $l_{2}$ the layer configuration
- post_nodes $[d i r, n b r, l]$, post_sum $\left[d i r_{1}, l_{1}, \operatorname{dir}_{2}, l_{2}\right]$, post_conc_sum $\left[d i r_{1}, l_{1}, d i r_{2}, l_{2}\right]$ analogues to the pre counters but for the time window $\left(t_{j}, t_{j}+\delta\right]$
- pre_partial_sum[dir $\left., l_{1}, d i r_{2}, l_{2}\right]$ counts the number of strictly ordered pairs of parallel edges such that the first edge is in $\left[t_{j}-\delta, t_{j}\right)$ and the second edge is at time $t_{j}$, with the first edge having direction $\operatorname{dir}_{1}$ and layer $l_{1}$ and the second edge direction $\operatorname{dir}_{2}$ and layer $l_{2}$
- post_partial_sum $\left[\operatorname{dir}_{1}, l_{1}\right.$, dir $\left._{2}, l_{2}\right]$ counts the number of strictly ordered pairs of parallel edges such that the first edge is at time $t_{j}$ and the second edge is in $\left(t_{j}, t_{j}+\delta\right]$, with the first edge having direction $\operatorname{dir}_{1}$ and layer $l_{1}$ and the second edge direction dir $_{2}$ and layer $l_{2}$
- conc_nodes $[d i r, n b r, l]$ counts the number of times center node $u$ and neighbour $n b r$ are connected with direction $\operatorname{dir}$ and layer $l$ at a given time $t_{i}$, i.e., the number of concurrent edges
- conc_sum[dir $\left.{ }_{1}, l_{1}, d i r_{2}, l_{2}\right]$ counts the number of pairs of concurrent parallel edges at a given time $t_{i}$, with $d i r_{1}$ and $d i r_{2}$ indicating the directional configuration and $l_{1}$ and $l_{2}$ the layer configuration
- mid_sum $\left[\operatorname{dir}_{1}, l_{1}, \operatorname{dir}_{2}, l_{2}\right]$ counts the number of pairs of parallel edges where the first edge is in direction $\operatorname{dir}_{1}$, with layer $l_{1}$, and occurred at time $t \in\left(t_{j}-\delta, t_{j}\right)$ and the second edge is in direction $\operatorname{dir}_{2}$, with layer $l_{2}$, and occurred at time $t^{\prime} \in\left(t_{j}, t_{j}+\delta\right)$ such that $t^{\prime}-t \leq \delta$
- conc $\left[d i r_{1}, l_{1}\right.$, dir $\left._{2}, l_{2}, d i r_{3}, l_{3}\right]$ counts the full motifs of the 'conc' temporal configuration within $\delta$ time, with $d i r_{1}, d i r_{2}$ and $d i r_{3}$ indicating the directional configuration and $l_{1}, l_{2}$ and $l_{3}$ indicating the layer configuration of the three edges, respectively
- post_partial $\left[d i r_{1}, l_{1}, \operatorname{dir}_{2}, l_{2}, \operatorname{dir}_{3}, l_{3}\right]$, post_conc $\left[d i r_{1}, l_{1}, \operatorname{dir}_{2}, l_{2}, \operatorname{dir}_{3}, l_{3}\right]$, post $\left[d i r_{1}\right.$, $\left.l_{1}, \operatorname{dir}_{2}, l_{2}, \operatorname{dir}_{3}, l_{3}\right]$, pre_conc $\left[\operatorname{dir}_{1}, l_{1}, \operatorname{dir}_{2}, l_{2}, \operatorname{dir}_{3}, l_{3}\right]$, pre_partial $\left[\operatorname{dir}_{1}, l_{1}, \operatorname{dir}_{2}, l_{2}\right.$, $\left.d i r_{3}, l_{3}\right], \operatorname{mid}\left[d i r_{1}, l_{1}, \operatorname{dir}_{2}, l_{2}, d i r_{3}, l_{3}\right]$ and $\operatorname{pre}\left[\operatorname{dir}_{1}, l_{1}, d i r_{2}, l_{2}, d i r_{3}, l_{3}\right]$ analogues to the 'conc' counter, but each matching their own temporal configuration.
The extended algorithm is given in Algorithm 2 and 3 shown in Appendix A. Algorithm 2 shows, that the approach of this algorithm remains unchanged from that described in Section 4.1.2. We process sets of concurrent edges when they enter the $\delta$-timeframe on lines $5-6$, when they become the current set on lines 7-9 and when they leave the $\delta$-timeframe on lines $3-4$. Similar to the edge motif algorithm, we simply move from processing one edge to processing a set of concurrent edges at a time.
Again, the extensions to accommodate concurrent edges and enforce edge attribute exclusivity are achieved through the addition of the appropriate counters and their update logic. Note that edge attribute exclusivity is achieved in exactly the same way as was the case for the edge motifs, through the addition of temporary counters. Therefore, we will not further discuss this for star motifs. We explore new complications that arise for concurrent edges in star motifs and explain the extensions made to solve these below.


### 4.3.1 Concurrent edges

For edge motifs we had the convenience that every edge in the motif connected the same pair of nodes. This meant that a pair of concurrent edges and its reverse order, for example $((u, v, A),(v, u, B))$ and $((v, u, B),(u, v, A))$, could be counted using the same counters and update logic. Therefore, all possible orderings of concurrent edges could be counted in a single forward pass and their true total counts could be obtained by resolving for equivalences in post-processing.
Unfortunately, three of the temporal configurations of star motifs ('conc', 'post_partial' and 'pre_partial') have concurrent edges connecting the center node $(u)$ to different neighbours $(v, w)$, where we consider the parallel edges to connect to neighbour $v$. To be able to count these temporal configurations in a single forward pass, we can not assign a specific edge in the configuration as the pivotal edge, as we have done for star motifs up to this point, because this might not be the last edge considered in a traversal of a set of concurrent edges. For example, the concurrent edge combination $((u, v, A),(u, w, B))$ remains equivalent to its reverse order $((u, w, B),(u, v, A))$. However, counting star motifs given the latter order in a single forward pass, means that edge $(u, w, B)$ must be processed before the parallel edge $(u, v, A)$. This requires us to consider one of the parallel edges to $v$ as the pivotal edge instead of the single edge to $w$. This presents a significant algorithmic problem. After all, if one of the parallel edges $(u, v)$ is the pivotal edge, then we require a counter that represents a combination of the two remaining edges in the configuration, which connect to $v$ and $w$ respectively, i.e., a counter that represents a connection of two edges to two different neighbours. As we previously discussed in [14], this would inevitably lead to neighbour loops, which would, in worst case $|n b r s|=n-1$, increase both the time (and space) complexity by a factor $n$. Therefore, the simpler and more efficient solution is to traverse each set of concurrent edges both forward and backward, such that each of the loops covers one of the two possible orders of two concurrent edges. This is far more efficient, because the backward loop (lines 53-69 in Algorithm 3) adds just a small number of additional operations per edge and requires no additional counters. As such, it only adds a small constant factor to the time complexity, instead of a factor $n$, to both time and space complexity.
After introducing the backward loop, there remains one problematic case involving the 'conc' configuration. This temporal configuration consists of three concurrent edges. If a set of concurrent edges is ordered such that we have $((u, v),(u, w),(u, v))$, neither the forward nor the backward loop on its own can prevent one of the parallel edges from being considered the pivotal edge. To allow the middle edge of three concurrent edges to be considered the pivotal edge, we approach the problem in a similar way as the 'mid' configuration. First, during the forward loop, we count all one edge directional and layer configurations and store this in a new counter 'conc_pre_nodes'. During the following backwards loop, this counter keeps track of the number of one edge directional and layer configurations that may be considered the last edge in the order $((u, v),(u, w),(u, v))$. A second additional counter called 'conc_mid_sum' is added, which keeps track of the number of pairs of parallel edges $((u, v),(u, v))$ of which one edge occurs before and the other after the current edge under consideration in the backward traversal. As we traverse the edges in the backward loop, for each set of equal attribute edges we perform the following three actions:

1. as the edges in this set become the current edges, they can no longer be consid-
ered possible last edges in the order $((u, v),(u, w),(u, v))$ and we reduce counters 'conc_pre_nodes' and 'conc_mid_sum' accordingly (lines 54-56);
2. we consider the edges in this set the current edges, i.e., pivotal edges, and update the counter for the 'conc' configuration accordingly (line 59); and
3. as the edges have been fully processed as current edges, they now become preceding edges, i.e., the first edges in the order $((u, v),(u, w),(u, v))$, and we update the counter 'conc_mid_sum' accordingly (lines 66-67).

Like for edge motifs, counting star motifs also requires a post-processing step to account for equivalent directional and layer configurations. Because we directly count every possible ordering of concurrent edges that connect to different neighbours, these equivalences only occur for the three temporal configuration that have concurrent parallel edges ('conc', 'post_conc' and 'pre_conc').

Similar to the existing star motif counting algorithms [13, 14], we drastically reduce the number of counters by discarding the knowledge of the neighbour to which the two parallel edges are connected for counters that represent two edge combinations. As such the algorithm can not ensure, for the pivotal edge, that neighbour $v \neq w$. The number of additional star motifs counted when $v=w$ are exactly the sum of the number of 3-edge edge motifs for $u$ and each of its neighbours. Therefore, as a second post-processing step, we subtract the matching edge motif counts, based on matching temporal, directional and layer configurations, for every neighbour of $u$ from the star motif counts. Note that all directional and layer configuration equivalences should be resolved for both the edge and star motifs before the subtraction is performed.
Although the extensions to accommodate concurrent edges and enforce edge attribute exclusivity have made the star motif counting algorithm more complex compared to existing algorithms, its time and space complexity have remained virtually unchanged. Both complexities increase by only a small constant factor and thus remain $O\left(m \lambda^{2}\right)$.

### 4.4 Triangle motifs

The last structural configuration for which we extend the motif counting algorithm is that of triangle motifs. Triangle motifs are motifs whose edges form a triangle [13, 14]. Given three nodes and three edges and allowing for concurrent edges, there are four temporal configurations, which we label as shown in Figure 8. The full set of directed triangle motif configurations are depicted in Figure 2c.
As discussed in Section 4.1.1 an input sequence $(S)$ is gathered for every node pair $u, v \in V$ to which a static triangle has been assigned. These input sequences are pre-processed into sequences $S^{\prime \prime}$ consisting of sets of sets of concurrent equal attribute value edges, in the same manner as for edge and star motifs. Each edge in a triangle motif can be described

conc

pre_partial

post_partial

serial

Figure 8: Triangle motif temporal configurations
by an indicator uorv, indicating whether it is connected to $u(0)$ or $v(1)$, its neighbour node $n b r$, its direction dir outward from (0) or inward to (1) nbr, timestamp $t$, layer $l$ and edge attribute $p$. As such, the final input sequences to the triangle motif counting algorithm can be defined as

Definition 9. Sequence $S^{\prime \prime}$ is a sequence of sets of sets of edges: $S^{\prime \prime}=\left(\left(\left\{\left(\left\{e_{1}=\right.\right.\right.\right.\right.$ $\left(\right.$ uorv $\left.\left.\left.\left.\left._{1}, n b r_{1}, \operatorname{dir}_{1}, l_{1}\right), \ldots\right\}, p_{1}\right), \ldots\right\}, t_{1}\right), \ldots,\left(\left(\left\{\ldots,\left(\left\{\ldots, e_{L}=\left(\right.\right.\right.\right.\right.\right.$ uorv $\left.\left.\left._{L}, n b r_{L}, \operatorname{dir}_{L}, l_{L}\right)\right\}, p_{P}\right)$ $\left.\left.\}, t_{T}\right)\right)$ ), such that $t_{1}<t_{2}<\ldots<t_{T}$ and for all $i \neq j$ with $1 \leq i, j \leq P$ we have $p_{i} \neq p_{j}$.

As discussed in Section 4.1.2 the edge connecting node pair $u, v$ in the triangle motif configuration is chosen as the pivotal edge and counters are formed for all edge combinations for the remaining two edges connecting to the common neighbour over the full $\delta$-timeframe. The same set of counters as defined for star motifs in Section 4.3, is used with minor adjustments. The one and two edge counters are given an additional index uorv, which indicates whether the first edge is connected to node $u$ or $v$ and, of course, a different set of counters is used for the full motifs matching the temporal configurations in Figure 8:

- conc $\left[k e y_{1}, k e y_{2}, l_{1}, l_{2}, l_{3}\right]$ counts the full motifs of the 'conc' temporal configuration within $\delta$ time, with $k e y_{1}$, $k e y_{2}$ indicating the directional configuration and $l_{1}, l_{2}$ and $l_{3}$ indicating the layer configuration of the three edges, respectively
- pre_partial $\left[k e y_{1}, k e y_{2}, l_{1}, l_{2}, l_{3}\right]$, post_partial $\left[k e y_{1}, k e y_{2}, l_{1}, l_{2}, l_{3}\right]$ and serial $\left[k e y_{1}\right.$, $\left.k e y_{2}, k e y_{3}, l_{1}, l_{2}, l_{3}\right]$ analogues to the 'conc' counter, but each matching their own temporal configuration.

The extended algorithm is given in Algorithm 2 and 4 shown in Appendix A. Unlike for edge and star motifs, the full motif counters use 'key's to indicate a specific directional configuration because we do not know to which node pair each triangle is assigned. Therefore, the algorithm must consider all three possibilities and map those to the same counter (lines 44-56 and lines 73-84). At the end of Algorithm 4 the key map translating the full motif counters to the configurations in Figure 2c is given.
Similar extensions were made to the triangle motif counting algorithm as discussed for edge and star motifs, including directional and layer configuration equivalence post-processing. Note that in Figure 2c only motif configuration $M_{t, 1,2}$, a concurrent circle, lends itself to layer configuration equivalence within the same directional configuration.
As the same extensions are made to the triangle motif counting algorithm as was done for the star motif counting algorithm, here too we have only a small constant increase of our time and space complexity, which thus remains $O\left(m \sqrt{\tau} \lambda^{2}\right)$.

## 5 Data

In this section we discuss the co-authorship datasets used in this work. We discuss how the datasets were obtained from Web of Science (WoS) and define the various network layers in Section 5.1. In Section 5.2, we examine how country specific datasets are extracted from the global datasets.

### 5.1 Extracting multilayer network datasets from WoS

We extracted our five global datasets from the in-house version of WoS at the Centre for Science and Technology Studies (CWTS). The CWTS version of WoS has been enriched with in-house author identifiers based on an improved author disambiguation algorithm [31]. We use these in-house author identifiers to associate authors to their respective oeuvres. Furthermore, this version has enriched organisation information and more consistent and accurate assignment of papers to universities and organisations [39]. Each extracted co-authorship network covers one main field and includes papers published in the period 2007-2016. A ten year period was chosen so that there is an increased likelihood of one or more mobility events to have occurred for each active author. Papers, and by extension co-authorships, are assigned to the fields on the journal level and can be associated with multiple fields. Papers with more than 25 authors are excluded to prevent papers with large author lists from skewing our results. For example, in the field of High Energy Physics publications with hundreds or thousands of authors are not uncommon. Given a mostly similar group of authors, just three such publications would generate such a large number of motifs that the balance of motifs found in the overarching field would be skewed towards motifs representing co-authorship in High Energy Physics. Additionally, for such publications the meaning of authorship with respect to individual contributions and collaboration is different compared to other fields [40]. Note that the 25 author limit is arbitrarily chosen and no robustness checks for different author limits were performed. Co-authorship links are formed for every pair of authors on a paper, provided organisation affiliation information for that paper was present in WoS for both authors. Organisation affiliations can be missing when, for (some) authors, it is not properly indicated on the published paper which authors were affiliated with which of the listed organisations. Each co-authorship link is assigned to a specific layer based on the proximity of the organisations to which the respective authors were affiliated. We define the following set of layers.
O. Organisational co-authorship, both authors were associated with the same organisation.
L. Local co-authorship, the authors were associated with organisations based in the same city.
N. National co-authorship, the authors were associated with organisations based in the same country.
I. International co-authorship, the authors were associated with organisations based in different countries.

Because authors can be affiliated with multiple organisations at a time, multiple coauthorship links in different layers between two authors for the same paper are possible. When this occurs, only the link with the closest proximity ( $O<L<N<I$ ) is included. The publication year of a paper is used as the timestamp of co-authorship links associated with that paper. We use the publication year because the listed publication months in WoS are not always accurate, possibly leading to inaccurate co-authorship links.
Descriptive statistics on the global datasets are provided in Table 1, listing the number of nodes and edges, the number of static edges in the underlying static network, the percentage of edges in each of the layers, the number of papers from which the co-authorship edges are

Table 1: Descriptive global network dataset statistics

|  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: |
| Nodes (in millions) | 1.0 | 7.7 | 4.6 | 3.1 | 1.2 |
| Edges (in millions) | 4.6 | 94.0 | 35.2 | 22.6 | 4.6 |
| Static edges (in millions) | 3.2 | 53.4 | 21.4 | 15.3 | 3.2 |
| O(rganisational) edges (\%) | 55.8 | 67.6 | 65.6 | 61.8 | 63.4 |
| L (ocal) edges (\%) | 3.3 | 3.8 | 2.6 | 2.8 | 2.8 |
| N (ational) edges (\%) | 23.9 | 15.9 | 13.3 | 16.8 | 13.7 |
| I(nternational) edges (\%) | 17.0 | 12.7 | 18.5 | 18.6 | 20.1 |
| Papers (in thousands) | 826 | 5,612 | 3,412 | 2,092 | 950 |
| Inter-disciplinary papers (\%) | 33.4 | 16.1 | 19.2 | 36.1 | 34.8 |
| Missing org-affiliation links (\%) | 22.2 | 26.0 | 15.5 | 20.2 | 17.3 |

formed, the percentage of those papers that are inter-disciplinary, i.e., that are a associated with at least one other field as well, and the percentage of missing organisation affiliations.

### 5.2 Country datasets

Extracting country specific networks from global co-authorship networks is for the most part straightforward. All organisational, local and national edges within a country are included for that country, requiring only a decision to be made regarding the international co-authorships. In a non-motif context, one could choose to simply include the co-authorship links that are directly tied to a given country. However, in a motif context where we always consider multiple connected edges together, co-authorships that are outside a country but directly connect to an international link to that country may very well still be of interest. This begs the question: which edges outside a country are still of interest to that country, within a motif context? Here, we choose to include all co-authorships of every paper of which at least one author is associated with a given country. This means we consider a co-authorship link relevant for a country as long as the paper on which it is based is associated with that country.

## 6 Systematic interpretation of motifs in co-authorship networks

In this section we try to systematically assign meaning to the various motif configurations by mapping them to categories relevant to the domain of co-authorship and scientific mobility. First we discuss how we retrieve undirected motif counts from directed results in Section 6.1. Then, we introduce some collaboration categories in Section 6.2, we talk
about international categories in Section 6.3 and finally we tackle mobility categories in Section 6.4. A summary of the various categories is then given in Table 2. How the categories are mapped onto the full set of motifs is shown in Figure 10.

### 6.1 Directed to undirected results

Co-authorships links are an undirected relation between two authors. As a result, coauthorship networks are inherently undirected as well. Of course, direction could be used to encode additional information, such as seniority, by having co-authorship links point to the more senior author. In the datasets we extracted, direction represents seniority based on their total normalized citation score (TNCS), with organisation-level fractional counting, because for seniority within organisations participation may be more indicative than contributions [41]. Recall that the algorithms discussed in Section 4 count directed motifs. However, due to the great number of layer configurations with four layers, along with the various temporal configurations, we have chosen to analyse only the undirected results from our experiments. Not accounting for equivalent layer configurations, this effectively reduces the number of motif configurations to categorize from $5,632\left(88 * 4^{3}\right)$ to $1,024\left(16 * 4^{3}\right)$.
The set of undirected motifs is depicted in Figure 9. The numbering of the undirected motif configurations used throughout the remainder of the thesis follows the one shown in this figure, i.e., from $M_{1}$ to $M_{16}$. The motif counts of the undirected motifs are directly retrieved from those of the directed motifs in Figure 2. This is done by first resolving layer configuration equivalence for motifs with concurrent edges where equivalence was previously prevented by directionality, such as $M_{t, 1,1}$. After that, counts for equivalent temporal configurations are summed. For star and triangle motifs this translates to adding together the rows as depicted in Figure 2b and 2c.

### 6.2 Collaboration categories

The first set of motif categories that we define, consists of categories that capture the structural configuration. The most obvious distinction to be made is between edge, star and triangle motifs. Each of these structural configurations also has a distinct meaning in the context of co-authorship networks. As such we define three main categories:
CC. Continued Collaboration between two authors (edge motifs);
MC. Multiple Collaborators, i.e., an author that has multiple co-authors (star motifs); and

TC. Team Collaboration, i.e., three authors with each pair having co-authored a paper together (triangle motifs).


Figure 9: The set of undirected motifs, w.r.t the set of directed motifs in Figure 2

Recall that we enforce edge attribute exclusivity, which means that all motifs must consist of co-authorships on three different papers.
For the three node motifs (star and triangle), we define additional subcategories based on specific meaning derived from the layer configurations. Specifically, for star motifs we distinguish between layer configurations that indicate that the co-authors are equidistant, possibly equidistant or not at all equidistant with respect to the central author (center node). This leads us to define two subcategories:

MEC. Multiple Equidistant Collaborators, i.e., an author that has multiple co-authors with the same proximity at the same time; and

MPEC. Multiple Possibly Equidistant Collaborators, i.e., an author that has multiple co-authors whose equal proximity may be prevented by a change in proximity for one of the co-authors.

For triangle motifs we define subcategories of category TC based on the same concept of equidistant co-authorships:

ETC. Equidistant team collaboration, i.e., three authors with each pair having co-authored a paper together with the same proximity;

EP. Equidistant Partner, i.e., two authors, that have co-authored a paper at a local, national or international proximity, have both co-authored a paper with the same partner at the same proximity, which is equal or larger than their own proximity; and

OEP. Organisational Equidistant Partner, i.e., two authors, that have co-authored a paper at an organisational proximity, have both co-authored a paper with the same partner at the same proximity.

Note that ETC is entirely covered by EP and OEP, but that EP and OEP cover more motif configurations than ETC. For both the EP and OEP subcategories, we define two more subcategories: 'cause' (EPC, OEPC), where the link between the two authors comes before the formation of the equidistant partnership in the temporal configuration and is likely the cause of the equidistant partner, and 'effect' (EPE, OEPE), where the link between the two authors comes after the formation of the equidistant partnership and therefore likely follows from having the equidistant partner. An overview of all of these categories, with an example and short description, is given in Table 2a. Their mapping onto the full set of configurations is shown in Figure 10a.

### 6.3 International categories

Because we are interested in the relation between international collaboration and international mobility, the second set of motif categories we define deals with these concepts. We define two main categories:
I. International collaboration, i.e., motifs with at least one edge indicating an international co-authorship; and

IM. International mobility, i.e., motifs where a mobility event is implied by the transition of an international collaboration to an organisational, local or national collaboration, or vice versa.

Table 2：An overview of all collaboration and mobility motif categories．In the examples， edge labels indicate the rank－order and layer of the edges．
（a）Collaboration categories

| Category | Example | Description |
| :--- | :---: | :--- |
| CC | continued collaboration between two authors |  |
| MC | central author with multiple co－authors |  |
| MEC | central author with multiple equidistant co－authors |  |
| MPEC | central author with multiple possibly equidistant co－authors |  |
| TC | team collaboration |  |
| ETC | equidistant team collaboration |  |
| EP | equidistant partner |  |
| EPC | equidistant partner likely caused by collaboration |  |
| EPE | OEP | orgidistant partner likely cause of collaboration |
| OEPC | OEPE |  |

（b）International categories

| Category | Example | Description |
| :---: | :---: | :---: |
| I | ${ }^{(1,1)(2,2)(, 3,1)}$ | international co－authorship |
| IM |  | international mobility，unknown direction |
| IMI |  | incoming international mobility |
| IMO |  | outgoing international mobility |

（c）Mobility categories

| Category | Example | Description |
| :---: | :---: | :---: |
| M |  | mobility event implied |
| CM | 或令 | certain mobility event implied by an edge or star motif |
| MP | 或 | mobility event implied accompanied by a preceding edge |
| MS | 0\％ | mobility event implied accompanied by a succeeding edge |
| PM |  | possible mobility event implied by a triangle motif |
| MTC |  | possible（incoming）mobility event leading to collaboration |
| MSC | $\underset{\substack{-1.0 \rightarrow 0 \\ 10,2, i t}}{\substack{0}}$ | collaboration despite possible（outgoing）mobility event |
| M2 |  | two mobility events implied |
| RFM |  | return or follow mobility |
| VM |  | visit mobility |

From the perspective of individual countries, it is especially interesting whether the international mobility is incoming or outgoing. Therefore, we define the subcategories international mobility incoming (IMI) and international mobility outgoing (IMO). Indicating respectively whether we move from an international collaboration to a closer collaboration or move from a closer collaboration to an international one. Note that the direction of a mobility event cannot be determined when it is implied only by concurrent edges. Furthermore, for triangle motifs a mobility event can only be implied by a contradiction that occurs between all three edges and it can be associated with any of the authors. For example, if we have edges $\left(a, c, t_{1}, \mathrm{O}\right),\left(b, c, t_{2}, \mathrm{O}\right),\left(a, b, t_{3}, \mathrm{I}\right)$ for authors $a, b, c$ and assume a single affiliated organisation per author at a time, then the first two edges imply all authors are associated with the same organisation and the third edge implies that authors $a$ and $b$ are associated with organisations in different countries. Thus, an international mobility event is implied. However, this mobility event can be associated with every author, including author $c$ where author $c$ first has the same affiliation as $a$ and moves to the same organisation as $b$. In this case, the organisation associated with either author $a$ or $b$ must be located outside the country in consideration, yet we can never be sure which. Therefore, we can never determine a direction for the mobility events implied by triangle motifs.
An overview of the international categories, with an example and short description, is given in Table 2b. Their mapping onto the full set of motifs and layer configurations is shown in Figure 10b. Here, the 'non-international' category indicates motif configurations that do not involve any international co-authorship links.

### 6.4 Mobility categories

The third and final set of motif categories we define are mobility categories. The mobility categories either describe a certain type of mobility or describe the context of the edges surrounding the mobility event. We define two main categories:
M. Mobility, i.e., a mobility event is implied by a contradiction in organisational proximity between co-authorship edges; and

M2. Duo-mobility, i.e., two mobility events are implied.
In Section 6.3, we reasoned that we can never determine the direction of mobility events implied by triangle motifs. However, for triangle motifs we cannot be sure a contradiction of collaboration distances even implies a mobility event or if it is an indicator that an author is affiliated with multiple organisations. For example, given the same set of edges as before, $\left(a, c, t_{1}, \mathrm{O}\right),\left(b, c, t_{2}, \mathrm{O}\right),\left(a, b, t_{3}, \mathrm{I}\right)$, we required the assumption of a single affiliated organisation per author at a time to imply a mobility event. If we assume multiple organisation associations are possible for an author, then the author organisation affiliations $a \rightarrow\{A\}$, $b \rightarrow\{B\}$ and $c \rightarrow\{A, B\}$ would fit this motif configuration without implying any mobility event. As we can not be sure that any mobility event implied by triangle motifs is not caused by an author being affiliated with multiple organisations, we divide category M into two subcategories:
CM. Certain mobility, i.e., a mobility event implied by an edge or star motif; and

PM. Possible mobility, i.e., a mobility event implied by a triangle motif.


Figure 10: Mappings of the motif categories listed in Table 2 onto the full set of motifs. The 'duplicate' categories indicate layer configurations that are equivalent to layer configurations listed above them. For configurations where categories overlap, subcategories take precedence. The full hierarchy of the categories is shown in Figure 11.

Note that we are assuming that authors always list all their current affiliations for each paper. After all, for edge and star motifs, mobility is implied from a change in proximity between two co-authorship edges between the same two authors. Since this proximity is set to the minimum of all listed affiliations, for the proximity to change their list of affiliations must change. So, if authors always list all current affiliations, then a mobility event must have occurred when the proximity changes.
For certain mobility (CM), we know that the mobility event is implied by only two out
of the three edges. This means we have either an additional preceding, succeeding or concurrent edge and define two subcategories accordingly:

MP. Mobility Preceding, i.e., a mobility event is implied by a change in proximity between two co-authorship edges which are preceded by a third edge; and

MS. Mobility Succeeding, i.e., a mobility event is implied by a change in proximity between two co-authorship edges which are succeeded by a third edge.

Note that the additional edge can still be either a preceding or succeeding edge when it is concurrent with only one of the edges involved in the mobility event, because the mobility event itself will have occurred somewhere in the time between those two edges.
Despite the fact that we can never be certain about the direction of mobility for triangle motifs, we can imply a causation, i.e., meaning, on the possible mobility. We define two such subcategories for the PM category as follows.

MTC. Mobility To Collaboration, these motifs may imply collaboration as both a possible cause and effect of an incoming mobility event. For example, motif configuration $\left(a, b, t_{1}, \mathrm{~L}\right),\left(a, c, t_{1}, \mathrm{O}\right),\left(b, c, t_{2}, \mathrm{O}\right)$ may imply that collaborations between authors $a, b$ and $a, c$ may have inspired author $b$ or $c$ to move to the same organisation and start collaborating.

MSC. Mobility Sustained Collaboration, these motifs may imply that even after an author has moved further away, the ties to their previous organisation may allow them to establish new collaborations through their former colleagues. For example, this may be implied by motif configuration $\left(a, b, t_{1}, \mathrm{O}\right),\left(a, c, t_{1}, \mathrm{O}\right),\left(b, c, t_{2}, \mathrm{~L}\right)$.

To finish, we define two subcategories of duo-mobility:
RFM. Return or Follow Mobility, an author moving away from the same organisation as their collaborating partner after which they either return to their old organisation or the collaborating partner follows them to the new organisation, i.e., the proximity returns to organisational; and

VM. Visit Mobility, an author first moving to the same organisation as the collaborating partner after which the author either moves back or moves to yet another organisation at the same proximity as before, i.e., the proximity first changes to organisational and then returns to its old state.

An overview of the mobility categories, with an example and short description, is given in Table 2c. Their mapping onto the full set of motifs and layer configurations is shown in Figure 10c. Here, the 'no mobility' category indicates motif configurations that do not appear to imply any mobility event.

## 7 Experiments and results

In this section we discuss our experiments and results. First, we discuss our experimental setup in Section 7.1. Then, we analyse the performance of our new algorithms in Section 7.2. In Section 7.3, we compare the five scientific fields by using the categories defined in the previous section to create a profile for each field. Using this same method in Section 7.4, we compare for each field the 50 largest countries, where country size is based on its scientific output. Finally, in Section 7.5 we provide discussion of our data, methods and results.

of a category $i$ in a given field $j$ and country $k$ with respect to all countries is determined, analogous to Equations 1 and 2, as:

$$
\begin{equation*}
r i_{i, j, k}=\frac{\frac{c_{i, j, k}}{c_{p(i), j, k}}-a v g_{i, j}}{a v g_{i, j}}, \text { with } a v g_{i, j}=\frac{1}{50} \sum_{k=1}^{50} \frac{c_{i, j, k}}{c_{p(i), j, k}} . \tag{3}
\end{equation*}
$$

For each field, we analyse in detail the relative importance of all categories and their interplay to give insight into typical co-authorship and scientific mobility behaviour. As such, we aim to identify what sets each field apart. Within each field, we examine outlier countries that represent unique co-authorship and mobility behaviour and investigate commonalities between countries showing the same behaviour.
We analyse the $r i$ computed for $\delta=10$ years, i.e., the full timespan of the datasets. A shorter timespan, such as $\delta=3$ or 5 years, excludes motifs where the causal link between the co-authorships may be weaker due to the passing of time. Because a shorter timespan can impact the ri of a category, we investigate the robustness of $r i$ and our conclusions in Appendix C. We find that $r i$ is robust for the larger datasets and categories and that conclusions drawn for $\delta=10$ are representative for shorter timespans.
The multilayer temporal motif counting algorithms introduced in Section 4 were implemented as a component of the Stanford Network Analysis Project (SNAP, see [42] for details). Our implementation can be found at [43]. All experiments were run on a single machine with 16 Intel Xeon E5-2630v3 CPUs at 2.40 GHz ( 32 threads) with 512GB RAM. For the previous version of our algorithms we showed that execution at four or eight threads provided optimal performance with respect to runtime [14]. We confirmed this still holds for our extended algorithms and performed all experiments at eight threads. All reported execution runtimes include counting edge, star and triangle motifs but do not include the time required for reading the graph from disk into memory. Before we discuss the motif counts obtained from our experiments, we first discuss the performance of our implementation.

### 7.2 Results - Algorithm performance

In Section 4 we claimed that the time complexities of the extended algorithms increased by only small constant factors with respect to the existing algorithms with time complexities $O\left(m \lambda^{2}\right), O\left(m \lambda^{2}\right)$ and $O\left(m \sqrt{\tau} \lambda^{2}\right)$, respectively. Since we have only four layers $(\lambda=4)$, we expect the algorithms to show linear performance with respect to the number of edges in a network. In Figure 12a, we confirm this by comparing the execution runtimes of all our experiments, i.e., the runtimes for all global and country datasets for all five fields and three $\delta$ values, with respect to the size of the dataset. The figure shows a linear relationship between the size of a dataset and the runtime of our implementation.
In Figure 12b, we show that for datasets with more than ten thousand edges the algorithms process between thirty and fifty thousand edges per second. Additionally, Figure 12c shows that for almost all datasets between three and five thousand nodes are processed per second. As the number of edges in the networks increases, these figures show that the performance seems to converge to around 40,000 edges and 3,000 nodes per second.
Figure 12d shows that as the datasets get larger the density of the networks decreases. This makes sense, because the primary reason for co-authorship networks to be larger is the inclusion of co-authorships on a greater number of papers. Although there are more papers


Figure 12: Performance of our algorithm implementation for all experiments
overall, that does not mean that the average productivity of an author increases, i.e., the average node degree remains constant. In other words, the addition of more papers likely leads to more (new) authors while the average degree of authors need not increase. Thus, the addition of more papers creates far more potential edges than it actually adds and as a result decreases the density. A second reason for co-authorship networks to be larger is the inclusion of a greater number of authors per paper and therefore a greater number of co-authorships. Although this creates fully connected clusters (cliques) of connected authors, the authors in these clusters are less likely to connect to other clusters and each paper potentially adds a lot of new authors. As such, an increase in authors per paper does not need to increase the density. Because the density decreases as the size of the network increases, Figures 12e and 12 f look to simply be the mirror images of Figures 12b and 12 c , respectively.

### 7.3 Results - Field comparison

Based on the motif counting results from our experiments, for $\delta=10$ years, we determined the relative importance of each category defined in Section 6 with the help of Equations 1 and 2. The results are shown in Table 3. Together the relative importance of all categories for a field creates a profile of that field with which we can identify typical co-authorship and mobility behaviour. Below we take a closer look at the profile of each of the five fields and interpret these in the context of what is already known about these fields.

## Social sciences \& Humanities

For the field of Social sciences \& Humanities (SSH) the most prominent observation is that of the equidistant partner categories EP, OEP and their subcategories. In this field, a far greater proportion of teams, i.e., triangles of co-authorships, represents the formation of an equidistant partner with someone outside their own organisation (EP +0.99 ). Conversely, a much smaller proportion of teams represents the formation of equidistant partners with someone at their own organisation (OEP -0.22). We note that this observation can be

Table 3: Field comparison of the relative importance of each category defined in Section 6. The 'mpe' column indicates the number of motifs counted per edge.

explained, to some extent, by the reduced proportion of organisational edges and increased proportion of national edges for this field, as shown in Table 1. Although this provides an explanation for the observation of relatively more EP and relatively fewer OEP motifs, the tendency to co-author to a greater level with authors outside the own organisation and to form equidistant partners with them remains an identifying trait for this field.
The second observation we make for SSH is the negative ri for international categories I and IM. This means that, although we noticed more equidistant partners outside the own organisation, these partners are more likely to work within the same country. Notably, in Table 1 we see that the proportion of international edges for SSH is barely less than that of the fields with a positive $r i$ for category I and as such is not a sufficiently explaining factor. Consequently, fewer motifs are formed on average per international edge. There are multiple possible causes for this phenomenon. First, authors linked to international co-authorships may be less productive than the average author in the field. Second, authors linked to international co-authorships may publish papers that involve relatively fewer coauthors per paper than on average. Third, co-authorships of authors linked to international co-authorships may be spread over a larger knowledge network, i.e., they have relatively fewer co-authorships per co-author. Each of these would result in relatively fewer motifs per edge for authors involved in international co-authorships and explain the phenomenon. A third observation to be made for SSH is that the relative importance of $M_{\text {all }}$, i.e, all motifs implying some mobility, is almost neutral, but that among them a relatively large proportion of motifs imply duo-mobility ( $\mathrm{M} 2+0.36$ ). In part this can be attributed to an increased proportion of edge motifs $(\mathrm{CC}+0.11)$, the only motifs that can imply duomobility. For the most part though, the relatively large proportion of M2 motifs implies that authors who continue to co-author through the years experience more changes of their proximity. In fact, because we see a neutral ri for VM, we can surmise that many of the additional M2 motifs can be categorized as visit mobility while no additional return or follow mobility occurs (RFM -0.21). As such, this third observation may provide insight into possible causes of the positive ri for EP in our first observation for SSH. For example, when an author visits the organisation of a co-author, they may be introduced to that
co-author's knowledge network, including their distant partners. Relationships formed with these partners could subsequently be upheld as they return to their old organisation, creating equidistant partners between authors at different organisations.
The final observation we make for SSH is that among mobility motifs international mobility is under-represented ( $\mathrm{IM}_{m}-0.20$ ), re-enforcing our second observation of a reduced proportion of motifs including international co-authorships.

## Biomedical \& Health sciences

For the field of Biomedical \& Health sciences $(B \& H)$ we see the greatest proportion equidistant team collaborations (ETC +0.31 ). Additionally we see a relatively high proportion of team collaboration representing organisational equidistant partners, but with a very low proportion of the cause and effect subcategories OEPC and OEPE. Together these categories imply that a large proportion of ETC motifs are in fact three authors connected by only organisational links, because, when this occurs, neither a cause nor effect can be determined for the equidistant partnership and therefore relatively few OEPC and OEPE motifs are formed.
The tendency for team formation within organisations forms an identifying trait for this field where the nature of the research often does not lend itself to inter-organisation collaboration and often leads to relatively large groups of authors on papers. We see this reflected in the relative importance of many other categories for this field. For example, the strong negative ri for the EP category as well as the low proportion of international motifs (I -0.41) reflects the smaller proportion of inter-organisation collaboration. Furthermore, the relatively low proportion of CC motifs and the high proportion of ETC motifs reflect the tendency for larger groups of authors, since larger groups form a greater proportion of triangle motifs than edge motifs over the course of several publications compared to smaller groups of authors. Note that the negative $r i$ for the M2 category is directly linked to the negative ri of CC and does not provide a 'new' observation.
As more and larger teams are formed within organisations, relatively fewer motifs are formed that imply a mobility event. B\&H has by far the lowest $r i$ of any field for the $M_{\text {all }}$ category. In part this is yet another result of the larger and more teams at the organisation level, but it might also suggest that scientists are less prone to move between organisations in this field. Additionally, we observe that among the mobility motifs a relatively small proportion of motifs represent international mobility ( $\operatorname{IM}_{m}-0.20$ ).

## Physical sciences \& Engineering

With respect to the collaboration categories, the field of Physical sciences \& Engineering (P\&E) appears to be associated with mostly intermediate ri values. Where other fields have strong positive or negative relative importances for a category, $\mathrm{P} \& E$ is neutral. For example, where SSH has a very large proportion of EP motifs and the other three fields have very small proportions of EP motifs, $\mathrm{P} \& E$ is around average ( $\mathrm{EP}+0.03$ ). However, unlike SSH, the percentage of organisational edges is roughly the same for P\&E and the remaining three fields, B\&H, Life \& Earth sciences and Mathematics \& Computer science. If we compare $\mathrm{P} \& E$ only with these fields, we see a similar pattern emerge for the ETC, EP and OEP categories as observed for SSH. This tells us that P\&E forms comparably more equidistant partners with co-authors outside their own organisation and less with co-authors within their organisation. Notably, P\&E is the only field where we can clearly
observe a difference in the $r i$ of the cause and effect subcategories of EP (EPC -0.10, EPE $+0.16)$. This indicates that a co-authorship between two authors at the local or national level more often follows from them having an equidistant partner at a greater proximity than that their co-authorship pre-dates, i.e., causes, the equidistant partnership. For P\&E we observe a positive ri for international motifs. Herein, it does not differ from the fields Life \& Earth sciences and Mathematics \& Computer science, but shows an equal $r i$ at around the same percentage of international links. At most we can state that $\mathrm{P} \& \mathrm{E}$ forms a greater amount of motifs including international co-authorships than SSH per international edge.
Finally, we see that $\mathrm{P} \& E$ has a relatively large proportion of mobility motifs $\left(\mathrm{M}_{\text {all }}+0.15\right)$. Here, only Life \& Earth sciences observes more mobility motifs and the proportion of international mobility $\left(\mathrm{IM}_{m}\right)$ follows a similar trend between the fields as observed for category I. We note a relatively high proportion of MTC mobility motifs in this field, possibly indicating an increased likelihood of incoming mobility having a direct cause or effect in the knowledge network of the authors. However, as we can not be sure that triangle motifs even imply mobility, we can not definitively conclude this.

## Life \& Earth sciences

Life \& Earth sciences (L\&H) shows, like B\&H, a reduced proportion of equidistant partner motifs (EP -0.32). Unlike B\&H, this is not associated with a greater proportion of ETC and OEP motifs. In other words, there is a relatively larger proportion of triangles of co-authorship that include edges with three different proximities. The occurrence of such motifs requires authors involved in them to either be mobile or be associated with multiple organisations, otherwise an equidistant partnership would be formed. For L\&H we observe a positive $r i$ for mobility motifs $\left(\mathrm{M}_{\text {all }}+0.28\right)$, indicating that mobility is likely the cause of the increased number of triangles with three different proximities.
Like for $\mathrm{P} \& \mathrm{E}$, we see a positive ri for international motifs for L\&H. More importantly, we see that the increased mobility and internationalism also translates to a larger proportion of international mobility motifs ( $\mathrm{IM}+0.42, \mathrm{IM}_{m}+0.13$ ).

## Mathematics \& Computer science

For the field of Mathematics \& Computer science (M\&C) we observe by far the greatest proportion of continued collaboration motifs $(\mathrm{CC}+0.43)$, i.e., edge motifs, which comes at the cost of team collaborations (TC -0.08). Furthermore, among the team collaborations we see more organisational equidistant partnerships than outside the organisation (EP -0.22 , OEP +0.08 ) and observe a greater likelihood for a clear cause or effect of EP motifs $(\mathrm{EPC}+0.18, \mathrm{EPE}+0.18)$. Together this indicates that, in $\mathrm{M} \& \mathrm{C}$, there is a greater trend to continue to co-author within the established knowledge network and organisation and to expand the knowledge network through the sharing of contacts with people at the same organisation rather than outside the organisation. A possible cause, or symptom, of this behaviour is the lower proportion of mobility motifs ( $\mathrm{M}_{\text {all }}-0.05$ ). Less mobility may cause authors to continue to collaborate with the same co-authors or as authors remain more set within their known knowledge network they may see less cause to become mobile. Note that the large proportion of duo-mobility motifs (M2) can be directly explained by the large proportion of CC motifs.

Despite the tendency to continue to co-author within the known knowledge network, we observe the same positive ri for international motifs as observed for $\mathrm{P} \& \mathrm{E}$ and $\mathrm{L} \& E$. Additionally, among the mobility motifs we also observe a positive $r i$ for international mobility $\left(\mathrm{IM}_{m}+0.11\right)$.

## The relationship between (international) mobility and collaboration

In Table 3 we have observed the same trend for all fields. A larger proportion of international motifs translating to a larger proportion of international mobility motifs among all mobility motifs (P\&E, L\&E and M\&C). At the same time, a smaller proportion of international motifs leads to a smaller proportion of international mobility motifs (SSH and B\&H). This trend forms a good indicator of the existence of a relationship between international co-authorship, i.e., international collaboration, and international mobility, but it does not imply a direction for this relationship.
Between categories IMI-IMO, MP-MS and MTC-MSC, the categories that may imply some causation between collaboration and mobility, we see only minor variations in the relative importance. Table 4 shows that neither MP nor MS is more dominant, indicating that collaboration occurs before and after mobility to an equal degree. Additionally, Table 4 shows that a greater proportion of motifs that imply international mobility suggest outgoing international mobility (IMO). It also shows a greater proportion of MSC, mobility sustained collaboration, motifs compared to MTC. This means that more motifs are formed by authors sustaining their old knowledge network after moving abroad than motifs are formed by authors moving closer to people they have previously co-authored with.
Although one might interpret this as evidence of a relationship between international mobility and collaboration in one direction, namely that international mobility leads to international collaboration, as suggested by Kato \& Ando [17], it actually shows that the relationship is bidirectional. After all, a single international co-authorship preceding an incoming international mobility event may be sufficient to establish a causation between the collaboration and the mobility, whereas proof of maintaining the old knowledge network after an international mobility event requires international co-authorships with multiple co-authors from before the international mobility event. In other words, one international mobility event is likely to form international collaborations with more previously organisational co-authors $(\mathrm{O} \rightarrow \mathrm{I})$, than it is to form organisational collaborations with previously international co-authors $(\mathrm{I} \rightarrow \mathrm{O})$. Therefore, a greater proportion of IMO motifs over IMI motifs is to be expected. Thus, we conclude that the relationship between international mobility and collaboration appears to exist in both directions.

Table 4: The proportion of a subset of categories w.r.t. their parent category, for each field

| field | IMI | IMO | MP | MS | MTC | MSC |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: |
| Social sciences \& Humanities | 0.32 | 0.48 | 0.36 | 0.41 | 0.24 | 0.56 |
| Biomedical \& Health sciences | 0.32 | 0.46 | 0.38 | 0.40 | 0.25 | 0.63 |
| Physical sciences \& Engineering | 0.33 | 0.44 | 0.39 | 0.40 | 0.29 | 0.56 |
| Life \& Earth sciences | 0.34 | 0.46 | 0.37 | 0.41 | 0.27 | 0.58 |
| Mathematics \& Computer science | 0.33 | 0.45 | 0.39 | 0.41 | 0.23 | 0.59 |

### 7.4 Results - Country comparison

Up to now we have considered and compared fields in their entirety, but variations in co-authorship and mobility behaviour also occur within fields. To study these variations we computed the relative importance ( ri ) of all categories for the 50 largest countries, where country size is based on scientific output, i.e., the number of papers associated with the country. The results, for $\delta=10$ years, are shown in Tables 5-9 in Appendix B. In this section we first highlight some categories, countries and regions that show specific behaviour within the various fields and then discuss recurring patterns between the categories that recur over multiple fields.

## Social sciences \& Humanities

For Social sciences \& Humanities (SSH) we observe geographical clustering of countries with positive and negative $r i$ for mobility $\left(\mathrm{M}_{\text {all }}\right)$. Figure 13 shows that clusters of countries with negative ri exist in South-Eastern Europe (Italy, Croatia, Hungary, Serbia, Greece and Cyprus), South Asia (Malaysia, Thailand, India and Iran) and parts of the American continents (USA, Mexico, Colombia, Brazil). Note that Singapore forms an exception in South Asia with a positive ri of +1.47 . Clusters of countries with positive ri for $\mathrm{M}_{\text {all }}$ can be found in Northern Europe (Ireland, Great Britain, Sweden and Estonia), Western Europe (Germany, France and Switzerland) and Eastern Europe (Romania, Slovakia, Czech Republic and Russia). Here exceptions occur for The Netherlands and Lithuania. In Figure 14, we see that the Czech Republic has a very high proportion of continued collaboration (CC) and duo-mobility (M2) motifs, with an over four times higher proportion of CC motifs and nearly six times higher proportion of M2 motifs than on average in the field. Among the M2 motifs, they have a higher proportion of visit mobility (VM) than on average. Inversely, the Czech Republic has a smaller proportion of team collaborations (TC), i.e., triangle motifs. These results suggest that in the Czech Republic authors cling much more to their established knowledge network than that they expand it. As a result, a move by one author to another organisation may result in many visit mobility events from their established knowledge network.
Russia shows a high proportion of mobility motifs and among them M2 motifs form a greater proportion than on average, yet to a much smaller degree than the Czech Republic. What makes Russia unique though, is the strong negative $r i$ for return or follow mobility


Figure 13: Relative importance of $M_{\text {all }}$ of countries in Social sciences \& Humanities


Figure 14: Relative importance of some categories and countries in Social sciences \& Humanities
(RFM) and visit mobility (VM). So, not only does Russia have more mobility motifs among the edge motifs, relatively few of them have a clear meaning. Along with a relatively small proportion of MSC motifs, this suggests that, as authors move to a new organisation, they maintain little of their knowledge network from their old organisation. However, they do continue to co-author with a few co-authors regardless of further mobility, but rarely does this inspire visit mobility or return or follow mobility between those co-authors. A similar, albeit weaker, pattern can be observed for Ireland.
Throughout all fields, Japan shows a decreased proportion of international collaboration and international mobility motifs and a decreased proportion of mobility motifs usually accompanies this. However, in SSH, Japan has an average proportion of overall mobility motifs ( $\mathrm{M}_{\text {all }} 0.02$ ) with a very small proportion of international mobility ( $\mathrm{IM}_{m}-0.82$ ). In other words, authors in this field move relatively often between organisations within Japan itself and very rarely internationally. Spain, shows a similar, albeit weaker, pattern.

## Biomedical \& Health sciences

Like for SSH, for Biomedical \& Health sciences (B\&H), we see geographical clustering of countries based on the relative importance of $\mathrm{M}_{\text {all }}$. In Figure 15 we observe a cluster of strong positive $r i$ in Northern Europe with Great Britain, Sweden, Finland and Denmark, but see negative ri throughout the rest of Europe, with a cluster of strong negative $r i$ in


Figure 15: Relative importance of $M_{\text {all }}$ of countries in Biomedical \& Health sciences


Figure 16: Relative importance of $\mathrm{IM}_{m}$ and the difference in relative importance between categories IMO and IMI of countries in Biomedical \& Health sciences

Eastern Europe (Lithuania, Poland, Slovenia, Serbia, Romania, Greece and Turkey). With the exception of Peru, we see negative ri throughout the American continents. On the other hand, South Asia is a mixed bag of strong positive and negative ri. Additionally, positive ri can be observed for Australia and the Middle East.
An observation that can be made for the field of $B \& H$ when you look at Table 6 in Appendix B, is that the proportion of international mobility motifs among all mobility motifs $\left(\mathrm{IM}_{m}\right)$ is closely related to the size of a country, where size is based on scientific output. We see that the largest countries have a tendency to have a smaller proportion of international mobility, whilst the smaller (perhaps less prominent) countries have a larger proportion of international mobility. One might interpret this as there being a brain drain, i.e., scholars moving primarily from less prominent countries to more prominent countries in the field and not the other way around. However, if there was a brain drain, then we would expect to see a greater proportion of outgoing international mobility (IMO) than incoming international mobility (IMI) for the smaller countries and vice versa for the more prominent countries. Figure 16 shows that there is no such phenomenon. In fact, most smaller countries with a large proportion of international mobility also display an above average proportion of M2 motifs with either a strong tendency towards return or follow mobility (RFM) or visit mobility (VM). We may take this as an indicator of brain circulation, where scholars either return to or visit their home country, thereby balancing the incoming and outgoing international mobility.

## Physical sciences \& Engineering

For Physical sciences \& Engineering (P\&E), we see, again, a geographical clustering of countries with respect to the proportion of mobility motifs $\left(\mathrm{M}_{\text {all }}\right)$ in Figure 17. Similar to $\mathrm{B} \& \mathrm{H}$, we see that Peru is the sole country on the American continents with a positive ri. Throughout South and West Asia we see countries with negative ri and we see strong positive $r i$ in the Middle East. However, Iran has negative and Pakistan positive ri. In Europe we see clusters that combine parts of European regions, for example we see a cluster of negative ri for Portugal, Spain, France, Switzerland and Belgium, and a cluster of positive $r i$ for The Netherlands, Germany, Poland and Sweden.
As shown in Figure 18, in P\&E Slovenia is a country with relatively many team collaborations (TC) and an above average proportion of motifs that include an international co-authorship (I), but a low proportion of international mobility motifs (IM). Since Slovenia also has a very large proportion of equidistant partner motifs (EP) among the team collaborations, it stands to reason that authors in Slovenia in this field share relatively many international partners and do so largely whilst associated with different organisations.


Figure 17: Relative importance of $M_{\text {all }}$ of countries in Physical sciences \& Engineering

Furthermore, these inter-organisational co-authorships and shared international partners seem to have a negative impact on the proportion of mobility motifs, and equally so on international mobility. The same pattern, albeit weaker, emerges in this field for Canada. On the contrary, Slovakia sees the same pattern with respect to the proportions of team collaborations, international collaboration and mobility, but has a greater proportion of organisational equidistant partner motifs (OEP) than EP motifs.
Remarkably, Thailand has the greatest proportion of continued collaboration (CC) motifs, yet it has a below average proportion of duo-mobility (M2) motifs, a mobility type which we would expect to find above average given the large proportion of CC motifs. Furthermore, international mobility $\left(\mathrm{IM}_{m}\right)$ motifs occur for Thailand at a far below average proportion with a greater proportion of MSC motifs. As such, we can surmise that authors in Thailand are likely to make fewer moves on average and are more likely to move within their own country whilst sustaining their scientific collaboration networks. This suggests a relatively high level of collaboration between (scholars at) the various scientific research organisations in Thailand that work in this field.

## Life \& Earth sciences

Geographical clustering of countries with respect to the proportion of mobility motifs $\left(\mathrm{M}_{\text {all }}\right)$ can also be observed for Life \& Earth sciences (L\&E). Figure 19 shows clusters


Figure 18: Relative importance of some categories and countries in Physical sciences \& Engineering


Figure 19: Relative importance of $M_{\text {all }}$ of countries in Life \& Earth sciences
with a relatively high proportion of mobility motifs in North America (Canada, USA), in Northern Europe (Great Britain, Norway, Sweden, Finland and Estonia), and Australia. Clusters with a relatively low proportion of mobility motifs appear in South America (Brazil, Argentina), Southern and Eastern Europe (Portugal, Spain, Italy, Austria, Slovenia, Croatia, Serbia, Greece, Romania, Hungary, Slovakia, Poland, Czech Republic). On the other hand, Asia and the Middle East form more of a mixed bag.
A surprising missing country in the clusters of negative tendency towards mobility for this field is Greece, which for SSH and B\&H did show the same tendency as its neighbours in Southern Europe. For L\&E, Greece actually shows an increased proportion mobility $\left(\mathrm{M}_{\text {all }}\right)$ motifs, as well as international collaboration (I) motifs. In fact, Figure 20 shows that throughout most categories Greece shows the opposite tendencies with respect to its neighbouring countries (Turkey, Bulgaria, Serbia, Croatia and Slovenia).
Similar to the contradiction between Greece and its neighbour countries, neighbours Iran and Pakistan display inverse proportions of many categories. For example, Iran shows an above average proportion of continued collaboration (CC) and duo-mobility (M2) motifs and among the duo-mobility motifs they show the greatest tendency towards visit mobility (VM) over return or follow mobility (RFM) in the field. On the contrary, Pakistan shows the reverse tendency with a very small proportion of duo-mobility and among duo-mobility a very low proportion of visit mobility motifs.


Figure 20: Relative importance of some categories and countries in Life \& Earth sciences

## Mathematics \& Computer science

Figure 21 shows that the field of Mathematics \& Computer science (M\&S) also displays some geographical clustering based on the relative importance of $\mathrm{M}_{\text {all }}$, but perhaps less than observed for the other fields. Notably, Northern Europe, which showed fairly consistent clustering for the other fields, forms a mixed bag as well as the American continents. Western Europe (Germany, The Netherlands, France, Switzerland, Austria and the Czech Republic) shows fairly consistent $r i$ around zero, a type of clustering we have not seen before. Additionally, we see more negative ri in Southern Europe (Spain, Italy, Slovenia, Croatia and Serbia) and around the South and East of Asia (Japan, South-Korea, Malaysia, Thailand and India). Strong positive ri are again observed for the Middle East (Pakistan, Israel, Saudi Arabia and Egypt) with the exception of Iran, which is also the only of these countries to show a very small proportion of international mobility motifs instead of a relatively large proportion.
In Figure 22, we see that Serbia has a very high proportion of mobility motifs in the duo-mobility (M2) category. Unlike what we saw for the Czech Republic in SSH, it is associated with a much smaller positive $r i$ for continued collaboration (CC) and Serbia has a relatively small proportion of visit mobility instead of a very larger proportion. Furthermore, we observed a strong positive ri for $\mathrm{M}_{\text {all }}$ for the Czech Republic in SSH, whilst Serbia has a consistently negative ri. Together with several other categories, this suggests that authors in Serbia form larger knowledge networks primarily at their own organisation, as evidenced by high ETC and low OEPC and OEPE, but that among mobile authors there is a far greater amount of continued collaboration, resulting in the observed high proportion of M2 motifs despite the low proportion of $\mathrm{M}_{\text {all }}$.
Thailand is a special case in M\&C. Like in P\&E, it has the highest proportion of CC motifs, yet among mobility motifs the duo-mobility (M2) category is under-represented. Furthermore, like in $\mathrm{P} \& E$, it has a very low proportion of international mobility $\left(\mathrm{IM}_{m}\right)$, a high proportion of MSC, a low proportion of MTC and a low proportion of RFM motifs. Unlike in P\&E though, Thailand has an above average proportion of international collaboration (I) motifs in M\&C, as well as a high proportion of visit mobility (VM). This might suggest that, while our conclusion for Thailand in P\&E holds up for M\&C, the added international collaboration for Thailand in this field is characterized by international knowledge networks formed through foreign scholars that made short term visits to Thailand. Note that this behaviour is also supported by the high tendencies towards organisational equidistant


Figure 21: Relative importance of $M_{\text {all }}$ of countries in Mathematics \& Computer science


Figure 22: Relative importance of some categories and countries in Mathematics \& Computer science
partners where the organisational link, for example, the international visit, is the cause or effect of the equidistant partner (OEPC, OEPE).
A similar pattern can be observed for Malaysia with respect to the mobility categories, but here it represents an entirely different type of behaviour. After all, in Malaysia international collaboration (I) plays a much smaller role, equidistant team collaboration are far more prevalent (ETC) and far fewer organisational equidistant partnerships indicate a cause or effect on the formation of the partnership (OEPC, OEPE). This suggests that although similar mobility behaviour may be attributed to Malaysia, Malaysian authors are more likely to form teams and seek partnerships at their own organisations whereas authors from Thailand are more likely to form partnerships with foreign scholars.
By far the greatest divergence between the relative importances of incoming and outgoing international mobility (IMI and IMO) in any field, is observed for Denmark in M\&C. Denmark is further characterized by the largest proportion of MTC motifs, the smallest proportion of duo-mobility (M2) motifs and one of the largest proportions of MS motifs in the field. Along with an increased likelihood of team collaboration being equidistant (ETC), we can surmise that Denmark in M\&C may retain a (much) greater proportion of its incoming foreign scholars for a longer time than any other country in any of the fields.

## Recurring patterns

For all fields, we see that a high positive (or negative) relative importance for MPEC correlates with a high positive (or negative) relative importance for $\mathrm{M}_{\text {all }}$. In fact, over all 250 country and field combinations, the average difference between the ri for MPEC and $\mathrm{M}_{\text {all }}$ is only 0.06 . The uncertainty of the equidistance in MPEC motifs comes from mobility events that might prevent the equidistance. As such, every MPEC motif implies a mobility event and the MPEC motifs make up approximately $50 \%$ of all mobility motifs in all of our experiments. Therefore, it is to be expected that the relative importance of the MPEC category is reflected in the $\mathrm{M}_{\text {all }}$ category. Despite this, it remains surprising just how close the categories correlate. In Figure 23, we can see the correlation between categories MPEC and $\mathrm{M}_{\text {all }}$ for SSH.
Figure 23 also shows a second pattern that reoccurs throughout all five fields. This pattern is marked by a positive correlation between the ri for the ETC and OEP categories. When they are both positive, we almost always see negative $r i$ for the EP, OEPC, OEPE, I, IM


Figure 23: Relative importance of various categories for the 50 countries in the field of Social sciences \& Humanities with the highest scientific output, with $\delta=10$ years
and $\mathrm{M}_{\text {all }}$ categories. In other words, countries where scholars form more equidistant teams (ETC) and organisational equidistant partnerships (OEP), there are in fact relatively more teams formed within organisations and scholars perform relatively less international collaboration, less international mobility and less mobility overall. Note that this is relative to the number of co-authorships and not relative in time. After all, if in two countries a scholar moves on average once every three years, but during those three years scholars in one of those countries form far more co-authorships, then, in our computation, that country will have relatively less mobility. Furthermore, note that an increased proportion of OEP motifs inherently reduces the proportion of EP motifs and increases the likelihood of negative $r i$ for EP. Additionally, note that teams within organisations are counted as both an ETC and OEP motif, but are not counted as OEPC nor OEPE motifs because, within organisations, previous co-authorships are a lot less likely to be the primary cause of future co-authorships. Thus, this pattern of negative relative importances for EP, OEPC and OEPE follows the logic of increased team collaboration at the organisational level. This pattern informs us that team collaboration within organisations is relatively less conducive to international collaboration and (international) mobility. This relationship can of course exist in both directions. That is to say, if an author is less mobile then they are less likely to find new co-authors outside their organisation, both nationally and internationally, and if an author has fewer co-authors outside their organisation they are less likely to move to another organisation due to connections in their knowledge network. There are of course countries that form the exception to the rule: in B\&H, Russia has an average amount of international collaboration and mobility instead of below average; in P\&E, Russia, South-Africa and Malaysia show average or above average (international) mobility; in L\&E, China and Pakistan show above average (international) mobility and, in M\&C, Poland, Russia and Sweden show (above) average (international) mobility.

When instead, both ETC and OEP are negative, we see positive $r i$ for the same set of categories (EP, OEPC, OEPE, I, IM and $\mathrm{M}_{\text {all }}$ ). This pattern tells us that when teams are formed they are more often formed with authors outside of the organisations and that these teams will more often involve foreign partners. Futhermore, it suggests that increased collaboration between organisations with foreign partners also facilitates relatively more international mobility and vice versa that international mobility allows for the creation of inter-organisational teams.
Exceptions here include: in SSH, Canada and Taiwan with (below) average international collaboration, respectively, and Thailand with below average (international) mobility; in B\&H, Canada with average (international) mobility and Denmark with below average international collaboration and international mobility; in P\&E, Denmark and Isreal with below average (international) mobility and, in L\&E, Denmark and Austria with (below) average (international) mobility, respectively.

### 7.5 Discussion

Here we discuss aspects of the datasets and methodology that may impact how well our results and conclusions reflect the real world.
First and foremost, we must take note of missing data. As previously mentioned in Section 5, authorships for which no affiliation information was present in WoS were excluded. In Table 1, we showed that this makes up around $20 \%$ of all authorships, which means our co-authorship networks are formed from only $80 \%$ of all authorships in WoS. Furthermore, WoS itself is not complete. For example, we know that conference papers play a big role in information diffusion in Computer science, but that conference papers are not included in WoS. Additionally, we know that some countries, such as Brazil, have their own internal publication system that is not included in WoS. The inclusion of this missing data could significantly alter the relative importances observed for the affected fields and countries. However, our datasets still cover a significant number of papers and co-authorships in every field and we expect that the relative importances of the categories obtained for these datasets provides a close approximation of the complete research system.
Second, because we imply mobility events from a change in co-authorship proximity, we may not be able to detect mobility which has no cause in previous co-authorships. As such, we may be underestimating the level of mobility in some countries or fields. Because we have no way to speculate about the amount of undetected mobility for specific countries or fields, we draw our conclusions based only on the mobility we are able to detect.
Third, categories that describe some causation, i.e., EPC, EPE, OEPC, OEPE, MTC and MSC, ascribe a connection between the co-authorships within the motifs that may not exist. For example, an EPC motif may imply that two scholars that co-authored locally formed an equidistant partner nationally because of their earlier co-authorship, i.e., one of the scholars introduced the other to the equidistant partner, but they may very well both have been introduced to this equidistant partner directly or through a third party. In fact, the greater the proximity between the scholars, the less certain we can be of the causation the category defines for individual motifs. However, over an entire co-authorship network an increased proportion of motifs of one of these categories over their counterpart, i.e., EPC over EPE, does imply a greater likelihood of more of such causations occurring. Furthermore, we do not draw conclusions based on just one of these categories, but only based on their interplay with other categories.

Fourth and last, in our conclusions we connect several categories to interpret a certain type of co-authorship or mobility behaviour for a field or country. However, motifs of categories that we describe as related may in fact be entirely unrelated and unconnected within the networks themselves. Thus, the conclusions we have drawn throughout Section 7.3 and 7.4 may, although logically sound, not represent the real-world explanation for the relative importances observed for the categories on which the conclusions were based.

## 8 Conclusion and Future Work

As outlined in Section 1, this thesis provides five contributions in an attempt to better understand scientific collaboration, scientific mobility, and how those relate.
First, we extended multilayer temporal motif counting algorithms from previous work to be able to count motifs that include concurrent edges. Second, we further extended these algorithms to enforce edge attribute exclusivity, so that in each counted motif every edge has a unique attribute value. Theoretically, the extensions to the algorithms added only a small constant factor to the time complexity of the original algorithms, which had time complexities of respectively $O\left(m \lambda^{2}\right)$ and $O\left(m \sqrt{\tau} \lambda^{2}\right)$, where $m$ is the number of links, $\lambda$ the number of layers and $\tau$ the number of static triangles. Using experiments on large-scale co-authorship datasets extracted from Web of Science (WoS), we showed that the extended algorithms have execution runtimes linear with respect to the size of the datasets, processing between thirty and fifty thousand edges per second.
For our experiments, we extracted five large global co-authorship datasets from WoS, each covering one field in the period 2007-2016, and extracted country specific datasets from them. Using our extended algorithms motif counts were computed for each of those datasets. As our third contribution, we introduced a systematic categorization of all 2-node and 3 -node, 3 -edge, $\delta$-temporal, 4-layer motifs that assigns meaning to the motifs in the domain of co-authorship and scientific mobility. By determining the relative importance of each the categories in specific fields or countries based on the computed motif counts, we were able to infer characteristic co-authorship and mobility behaviour. The inferred characteristic co-authorship and mobility behaviours, some of which are listed below, form our fourth contribution.

- For Social sciences \& Humanities (SSH), we found that authors in this field co-author to a greater level with authors outside the own organisation than in other fields. Additionally, they establish more equidistant partners with authors outside the own organisation. We also found that authors continue to collaborate throughout multiple mobility events to a greater degree, which likely aids the formation of equidistant partners with co-authors outside the own organisation. Although SSH has a similar amount of international co-authorships as other fields, it has a reduced proportion of motifs that involve international co-authorships. This indicates that internationally active authors in SSH display different co-authorship behaviour than in other fields. We showed that clusters of countries in South-Eastern Europe and South Asia have relatively less scientific mobility while clusters of countries in Northern, Western and Eastern Europe showed relatively more scientific mobility. We concluded that authors in the Czech Republic cling much more to their established knowledge networks than that they expand it, leading to more continued collaboration regardless of the occurrence of mobility. Finally, we showed that authors in Japan experience an
average amount of mobility, but that a much greater proportion of that mobility is between organisation within Japan instead of international.
- For Biomedical \& Health sciences (B\&H), we found that our results reflected the nature of the type of research conducted in this field, which often lends itself more to large team collaborations within an organisation than it does inter-organisational collaboration. With the exception of Northern Europe, countries in Europe have relatively little mobility within $\mathrm{B} \& H$. The same is true for countries in North- and South-America, with the exception of Peru. We found that for B\&H the largest countries, with respect to scientific output, showed a relatively small proportion of international mobility among all mobility, whilst the smaller (perhaps less prominent) countries have a relatively large proportion of international mobility. Additionally, we showed that the smaller countries have more visit mobility and return (or follow) mobility, indicating the presence of brain circulation where authors return to their home country.
- For Physical sciences \& Engineering (P\&E), we found that authors in this field form comparably more equidistant partners with co-authors outside their own organisation and less with co-authors within their organisation. Notably, we found that in P\&E the equidistant partners are relatively more often the cause of the local or national co-authorship than that those exist before the equidistant partner is established. Throughout South and West Asia we observed relatively less mobility, while observing relatively more mobility in the Middle East. We reasoned that authors in Slovenia share relatively many international partners between authors at different organisations. However, these inter-organisational co-authorships and shared international partners seem to have a negative impact on overall mobility, and equally so on international mobility. Finally, we found that authors in Thailand are likely to make fewer moves on average and are more likely to move within their own country whilst sustaining their scientific knowledge networks, suggesting a relatively high level of collaboration between (scholars at) the various scientific research organisations in Thailand involved in P\&E.
- For Life \& Earth sciences (L\&E), we found that relatively more mobility has lead to team formations (triangle co-authorships) at all different organisations and distances. Additionally, we found that the increased mobility and internationalism also translates to relatively more international mobility. Clusters of countries in North America, Northern Europe and Australia showed relatively more mobility in L\&E, whereas clusters of countries in South America, Southern and Eastern Europe showed relatively less mobility. Unlike in other fields, in L\&E, Greece shows tendencies directly opposite that of their neighbouring countries with relatively more mobility and international collaboration.
- For Mathematics \& Computer science (M\&C), we found that there is a greater trend to continue to co-author within the established knowledge network and organisation and to expand the knowledge network through the sharing of contacts with people at the same organisation rather than outside the organisation. Although this is associated with relatively less overall mobility, we still observe a relatively high proportion of international mobility for M\&C. With the exception of Iran, the Middle East has relatively much scientific mobility in M\&C. We concluded that authors in

Serbia form larger knowledge networks primarily at their own organisation, but that among mobile authors there is a far greater amount of continued collaboration. For Thailand the same conclusion as for $\mathrm{P} \& \mathrm{E}$ holds up, but international collaboration for Thailand in this field appears to be characterized by international knowledge networks formed through foreign scholars that made short term visits to Thailand. The same mobility behaviour is observed for Malaysia. However, Malaysian authors are more likely to form teams and seek partnerships at their own organisations instead of with foreign scholars. Finally, we found that Denmark in M\&C may retain a greater proportion of its incoming foreign scholars than any other country in any other field.

Throughout all fields, we found that countries with increased team formation within organisations display relatively less international collaboration and (international) mobility. Conversely, countries that display an increased amount of inter-organisational team formation show relatively more international collaboration and (international) mobility.
Finally as our fifth contribution, we weighed in on the discussion in literature on the relationship between international collaboration and international mobility. We found that the evidence supports the existence of this relationship in both direction, from collaboration to mobility and from mobility to collaboration.
In future work we would like to consider motifs larger than three nodes and three edges. Although we have shown in previous work that these larger motifs can not be counted as efficiently [14], they may give us further insight into typical co-authorship and mobility behaviour. Additionally, larger motifs could provide a greater insight into the evolution of knowledge networks. We also want to apply these algorithms to different types of networks to show the versatility of a multilayer temporal motif counting approach to gain insight into complex networks, as we have shown for co-authorship networks in this thesis.

## References

[1] Albert-László Barabási. Network science. Cambridge University Press, 2016.
[2] Kousik Das, Sovan Samanta, and Madhumangal Pal. Study on centrality measures in social networks: a survey. Social Network Analysis and Mining, 8(1):13, 2018.
[3] Broto Chakrabarty and Nita Parekh. NAPS: Network analysis of protein structures. Nucleic Acids Research, 44(W1):W375-W382, 2016.
[4] Frank W Takes, Walter A Kosters, Boyd Witte, and Eelke M Heemskerk. Multiplex network motifs as building blocks of corporate networks. Applied Network Science, 3(1):39, 2018.
[5] Roland Molontay and Marcell Nagy. Two decades of network science as seen through the co-authorship network of network scientists. arXiv preprint arXiv:1908.08478, 2019.
[6] María Bordons, Javier Aparicio, Borja González-Albo, and Adrián A Díaz-Faes. The relationship between the research performance of scientists and their position in co-authorship networks in three fields. Journal of Informetrics, 9(1):135-144, 2015.
[7] Lachezar Krumov, Christoph Fretter, Matthias Müller-Hannemann, Karsten Weihe, and M-T Hütt. Motifs in co-authorship networks and their relation to the impact of scientific publications. The European Physical Journal B, 84(4):535-540, 2011.
[8] Sarvenaz Choobdar, Pedro Ribeiro, Sylwia Bugla, and Fernando Silva. Comparison of co-authorship networks across scientific fields using motifs. In Proceedings of the International Conference on Advances in Social Networks Analysis and Mining (ASONAM), pages 147-152. IEEE Computer Society, 2012.
[9] Ron Milo, Shai Shen-Orr, Shalev Itzkovitz, Nadav Kashtan, Dmitri Chklovskii, and Uri Alon. Network motifs: simple building blocks of complex networks. Science, 298(5594):824-827, 2002.
[10] Austin R Benson, David F Gleich, and Jure Leskovec. Higher-order organization of complex networks. Science, 353(6295):163-166, 2016.
[11] Franc Mali, Luka Kronegger, Patrick Doreian, and Anuška Ferligoj. Dynamic scientific co-authorship networks. In Models of Science Dynamics, pages 195-232. Springer, 2012.
[12] John Mingers and Loet Leydesdorff. A review of theory and practice in scientometrics. European Journal of Operational Research, 246(1):1-19, 2015.
[13] Ashwin Paranjape, Austin R Benson, and Jure Leskovec. Motifs in temporal networks. In Proceedings of the Tenth ACM International Conference on Web Search and Data Mining (WSDM), pages 601-610. ACM, 2017.
[14] Hanjo D Boekhout, Walter A Kosters, and Frank W Takes. Efficiently counting complex multilayer temporal motifs in large-scale networks. Computational Social Networks, 6(1):1-34, 2019.
[15] Jessica Guth and Bryony Gill. Motivations in east-west doctoral mobility: Revisiting the question of brain drain. Journal of Ethnic and Migration Studies, 34(5):825-841, 2008.
[16] Stefano H Baruffaldi and Paolo Landoni. Effects and Determinants of the scientific international mobility: the cases of foreign researchers in Italy and Portugal. In Paper for the Triple Helix VIII conference, 2010.
[17] Maki Kato and Asao Ando. National ties of international scientific collaboration and researcher mobility found in Nature and Science. Scientometrics, 110(2):673-694, 2017.
[18] Pedro Ribeiro, Pedro Paredes, Miguel EP Silva, David Aparicio, and Fernando Silva. A survey on Subgraph Counting: Concepts, Algorithms and Applications to Network Motifs and Graphlets. arXiv preprint arXiv:1910.13011, 2019.
[19] Sameer Kumar. Co-authorship networks: a review of the literature. Aslib Journal of Information Management, 67(1):55-73, 2015.
[20] Xiaoli Hu, Oliver Zhen Li, and Sha Pei. Of stars and galaxies - Co-authorship network and research. China Journal of Accounting Research, 2019.
[21] Göran Melin and Olle Persson. Studying research collaboration using co-authorships. Scientometrics, 36(3):363-377, 1996.
[22] Adèle Paul-Hus, Philippe Mongeon, Maxime Sainte-Marie, and Vincent Larivière. The sum of it all: Revealing collaboration patterns by combining authorship and acknowledgements. Journal of Informetrics, 11(1):80-87, 2017.
[23] Albert-László Barabási, Hawoong Jeong, Zoltan Néda, Erzsebet Ravasz, Andras Schubert, and Tamas Vicsek. Evolution of the social network of scientific collaborations. Physica A: Statistical Mechanics and its Applications, 311(3-4):590-614, 2002.
[24] Caroline S Wagner and Loet Leydesdorff. Network structure, self-organization, and the growth of international collaboration in science. Research policy, 34(10):1608-1618, 2005.
[25] Wolfgang Glänzel and András Schubert. Domesticity and internationality in coauthorship, references and citations. Scientometrics, 65(3):323-342, 2005.
[26] Yan Wang, Yishan Wu, Yuntao Pan, Zheng Ma, and Ronald Rousseau. Scientific collaboration in china as reflected in co-authorship. Scientometrics, 62(2):183-198, 2005.
[27] Mark EJ Newman. Coauthorship networks and patterns of scientific collaboration. Proceedings of the National Academy of Sciences, 101(1):5200-5205, 2004.
[28] Oded Stark, Christian Helmenstein, and Alexia Prskawetz. A brain gain with a brain drain. Economics Letters, 55(2):227-234, 1997.
[29] Jacques Gaillard and Anne Marie Gaillard. Introduction: the international mobility of brains: exodus or circulation? Science, Technology and Society, 2(2):195-228, 1997.
[30] Grit Laudel. Studying the brain drain: Can bibliometric methods help? Scientometrics, 57(2):215-237, 2003.
[31] Emiel Caron and Nees Jan van Eck. Large scale author name disambiguation using rule-based scoring and clustering. In Proceedings of the 19th International Conference on Science and Technology Indicators, pages 79-86. CWTS-Leiden University, 2014.
[32] Henk F Moed, M'hamed Aisati, and Andrew Plume. Studying scientific migration in scopus. Scientometrics, 94(3):929-942, 2013.
[33] Silvia Appelt, Brigitte van Beuzekom, Fernando Galindo-Rueda, and Roberto de Pinho. Which factors influence the international mobility of research scientists? In Global Mobility of Research Scientists, pages 177-213. Elsevier, 2015.
[34] Valeria Aman. Does the scopus author id suffice to track scientific international mobility? a case study based on leibniz laureates. Scientometrics, 117(2):705-720, 2018.
[35] Mathias Czaika and Sultan Orazbayev. The globalisation of scientific mobility, 19702014. Applied Geography, 96:1-10, 2018.
[36] Zaida Chinchilla-Rodríguez, Lili Miao, Dakota Murray, Nicolás Robinson-García, Rodrigo Costas, and Cassidy R Sugimoto. Networks of international collaboration and mobility: A comparative study. In Proceedings of the 16th International Conference on Scientometrics \& Infometrics. ISSI, 2017.
[37] Zaida Chinchilla-Rodríguez, Lili Miao, Dakota Murray, Nicolás Robinson-García, Rodrigo Costas, and Cassidy R Sugimoto. A global comparison of scientific mobility and collaboration according to national scientific capacities. Frontiers in Research Metrics and Analytics, 3:17, 2018.
[38] Annik Leyman. Home sweet home? international mobility among flemish doctoral researchers. In Higher Education, Partnership, Innovation, pages 67-74. Budapest: IHEPI, 2009.
[39] Ludo Waltman, Clara Calero-Medina, Joost Kosten, Ed CM Noyons, Robert JW Tijssen, Nees Jan van Eck, Thed N van Leeuwen, Anthony FJ van Raan, Martijn S Visser, and Paul Wouters. The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. Journal of the American Society for Information Science and Technology, 63(12):2419-2432, 2012.
[40] Jeremy P Birnholtz. What does it mean to be an author? the intersection of credit, contribution, and collaboration in science. Journal of the American Society for Information Science and Technology, 57(13):1758-1770, 2006.
[41] Ludo Waltman and Nees Jan van Eck. Field-normalized citation impact indicators and the choice of an appropriate counting method. Journal of Informetrics, 9(4):872-894, 2015.
[42] Jure Leskovec and Rok Sosič. SNAP: A general-purpose network analysis and graphmining library. ACM Transactions on Intelligent Systems and Technology (TIST), 8(1):1, 2016.
[43] Hanjo D Boekhout, Walter A Kosters, and Frank W Takes. Counting multilayer temporal motifs. https://bitbucket.org/Fractals-/count_mult_temp_motifs, Accessed June 4, 2020.

## Appendix A Algorithm psuedocode

## A. 1 Algorithmic framework for star and triangle motifs

```
Algorithm 2: Algorithmic framework for counting of 3-node, 3-edge, \(\delta\)-temporal,
\(\lambda\)-layer star (and triangle) temporal motifs \(M_{\text {excl }}\).
    Input: Sequence ( \(S^{\prime \prime}\) ) of sets of sets of edges, with respectively equal timestamps \((t)\) and
                edge attributes \((p)\), with \(t_{1}<\ldots<t_{T}\), time window \(\delta\) and \(\forall_{i}: l_{i} \in\{0, \lambda-1\}\) :
                    \(S^{\prime \prime}=\left(\operatorname{coll}_{1}=\left(\left\{\operatorname{coll}_{11}=\left(\left\{e_{1}=\left(n b r_{1}, \operatorname{dir}_{1}, l_{1}\right), \ldots\right\}, p_{1}\right), \ldots\right\}, t_{1}\right), \ldots\right.\),
            \(\left.\left(\operatorname{coll}_{T}=\left(\left\{\ldots,\left(\left\{\ldots, e_{L}=\left(n b r_{L}, \operatorname{dir}_{L}, l_{L}\right)\right\}, p_{P}\right)\right\}, t_{T}\right)\right)\right)\)
    Initialize all counters to 0 , start \(\leftarrow 1\), end \(\leftarrow 1\)
    for \(j=1, \ldots, L\) do
        while \(t_{\text {start }}<t_{j}-\delta\) do
            \(\operatorname{Pop}\left(\right.\) pre_nodes, pre_sum, pre_conc_sum, coll \(\left._{\text {start }}\right)\), start \(+=1\)
        while \(t_{\text {end }} \leq t_{j}+\delta\) and end \(<L\) do
            Push(post_nodes, post_sum, post_conc_sum, coll \(_{\text {end }}\) ), end \(+=1\)
        \(\operatorname{Pop}\left(\right.\) post_nodes, post_sum, post_conc_sum, coll \(_{j}\) )
        ProcessCurrent(pre_nodes, post_nodes, mid_sum, pre_sum, post_sum, pre_conc_sum,
        post_conc_sum, coll \(_{j}\) )
        Push(pre_nodes, pre_sum, pre_conc_sum, coll \(_{j}\) )
```


## A. 2 Star motifs

Algorithm 3: Implementation of Algorithm 2 functions for counting 3-node, 3-edge, $\delta$-temporal, $\lambda$-layer star motif ( $M_{\text {excl }}$ ) instances.

```
Procedure Push(node_count, sum, conc_sum, colls)
    conc_nodes[:,:,:], tmp_nodes[:,:,:] \(\leftarrow 0\)
    for coll in colls do
            for \(e=(n b r, d i r, l)\) in coll do
            sum[:,:,dir,l] += node_count[:,nbr,:]
            conc_sum \([:,:, d i r, l]+=\) conc_nodes \([:, n b r,:]\)
            tmp_nodes \([d i r, n b r, l]+=1\)
            conc_nodes \(\leftarrow\) tmp_nodes
        for coll in colls do
            for \(e=(n b r, d i r, l)\) in coll do node_count \([d i r, n b r, l]+=1\)
    Procedure Pop(node_count, sum, conc_sum, colls)
        conc_nodes[:,:,:], tmp_nodes[:,:,:] \(\leftarrow 0\)
        for coll in colls do
            for \(e=(n b r, d i r, l)\) in coll do node_count \([d i r, n b r, l]-=1\)
        for coll in colls do
            for \(e=(n b r, d i r, l)\) in coll do
            sum \([\) dir \(, l,:,:]\) - = node_count \([:, n b r,:]\)
            conc_sum [:,:,dir,l] -= conc_nodes[:,nbr,:]
            tmp_nodes \([d i r, n b r, l]+=1\)
            conc_nodes \(\leftarrow\) tmp_nodes
```

```
30 Procedure ProcessCurrent(pre_nodes, post_nodes, mid_sum, pre_sum, post_sum,
pre_conc_sum, post_conc_sum, colls)
    conc_nodes \([:,:,:]\), conc_sum \([:,:,:,:]\), pre_partial_sum \([:,:,:,:]\), post_partial_sum \([:,:,:,:,] \leftarrow 0\)
    tmp_nodes [:,:,:], tmp_sum[:,:,:,::], tmp_pre_sum[:,:,:,:], tmp_post_sum[:,:,:,::] \(\leftarrow 0\)
        for coll in colls do
            for \(e=(n b r, d i r, l)\) in coll do mid_sum \([:,:, d i r, l]-=\) pre_nodes \([:, n b r,:]\)
        for coll in colls do
            for \(e=(n b r, d i r, l)\) in coll do
                conc[:,:,:,:,dir,l] += conc_sum[:,:,:,::] // \(M_{s, 1, x}\)
                post_partial[:,:,,dir, \(l,:,::]+=\) post_partial_sum \([:,:,:,:\),\(] // M_{s, 2, x}\)
                    post_conc[dir,l,:,:,:,::] += post_conc_sum[:,:,:,::] // \(M_{s, 3, x}\)
                    \(\operatorname{post}[\operatorname{dir}, l,:,:,:,:]+=\) post_sum[:,:,:,:, \(] \quad / / M_{s, 4, x}\)
                    pre_conc[:,:,:,:, dir,\(l]+=\) pre_conc_sum[:,:,:,::] // \(M_{s, 5, x}\)
                    pre_partial[:,:,:,:,dir,\(l]+=\) pre_partial_sum \([:,:,:,:]\) // \(M_{s, 6, x}\)
                    \(\operatorname{mid}[:,:\), dir \(, l,:,:]+=\) mid_sum \([:,:,:,:,] \quad / / M_{s, 7, x}\)
                    pre[:,:,:,:,,dir,\(l]+=\) pre_sum[:,:,:,::] \(/ / M_{s, 8, x}\)
                    tmp_sum \([:,:, d i r, l]+=\) conc_nodes \([:, n b r,:]\)
                    tmp_post_sum [dir \(, l,:,:]+=\) post_nodes \([:, n b r,:]\)
                    tmp_pre_sum [:,:,dir,l] += pre_nodes[:,nbr,:]
                    tmp_nodes \([d i r, n b r, l]+=1\)
                conc_nodes \(\leftarrow\) tmp_nodes, conc_sum \(\leftarrow\) tmp_sum
                post_partial_sum \(\leftarrow\) tmp_post_sum, pre_partial_sum \(\leftarrow\) tmp_pre_sum
        conc_pre_nodes \(\leftarrow\) conc_nodes, conc_mid_sum[:,:,:,:] \(\leftarrow 0\)
        reset conc_nodes, tmp_nodes, conc_sum, tmp_sum, post_partial_sum, tmp_post_sum,
        pre_partial_sum, tmp_pre_sum
        for coll in colls.reverse do
            for \(e=(n b r, d i r, l)\) in coll do
            conc_pre_nodes \([d i r, n b r, l]-=1\)
            conc_mid_sum \([\) dir \(, l,:,:]\) - = conc_nodes \([:, n b r,:]\)
            for \(e=(n b r, d i r, l)\) in coll do
            \(\operatorname{conc}[:,:,:,:\), dir,\(l]+=\) conc_sum \([:,,:,:,:] \quad / / M_{s, 1, x}\)
            conc[:,:,:,:,,dir,\(l]+=\) conc_mid_sum[:,:,:,::] // \(M_{s, 1, x}\)
            post_partial[:,:,dir,l,:,::] += post_partial_sum[:,:,:,::] // \(M_{s, 2, x}\)
            pre_partial[:,:,:,:,dir,\(l]+=\) pre_partial_sum[:,:,:,::] // \(M_{s, 6, x}\)
            tmp_sum \([:,:, d i r, l]+=\) conc_nodes \([:, n b r,:]\)
            tmp_post_sum \([\) dir \(, l,:,:]+=\) post_nodes \([:, n b r,:]\)
            tmp_pre_sum \([:,:\), dir,\(l]+=\) pre_nodes \([:, n b r,:]\)
            tmp_nodes \([\) dir \(, n b r, l]+=1\)
            for \(e=(n b r, d i r, l)\) in coll do
                    conc_mid_sum[:,:,dir, \(l]+=\) conc_pre_nodes \([:, n b r,:]\)
            conc_nodes \(\leftarrow\) tmp_nodes, conc_sum \(\leftarrow\) tmp_sum
            post_partial_sum \(\leftarrow\) tmp_post_sum, pre_partial_sum \(\leftarrow\) tmp_pre_sum
        for coll in colls do
            for \(e=(n b r\), dir,\(l)\) in coll do mid_sum \([\operatorname{dir}, l,:,:]+=\) post_nodes \([:, n b r,:]\)
return conc, post_partial, post_conc, post, pre_conc, pre_partial, mid, pre
```


## A. 3 Triangle motifs

```
Algorithm 4: Implementation of Algorithm 2 functions for counting 3-node, 3-edge,
\(\delta\)-temporal, \(\lambda\)-layer triangle motifs ( \(M_{e x c l}\) ) instances.
\({ }_{0}\) Procedure Push(node_count, sum, conc_sum, colls)
    conc_nodes[:,:,:], tmp_nodes[:,:,:] \(\leftarrow 0\)
    for coll in colls do
            for \(e=(n b r, d i r, u o r v, l)\) in coll do
                if \(n b r \notin\{u, v\}\) then
                sum \([1-u o r v,:,:\), ,dir,\(l]+=\) node_count \([1-\) uorv,\(:, n b r,:]\)
                conc_sum \([1-u o r v,,,:,, d i r, l]+=\) conc_nodes \([1-u o r v,:, n b r,:]\)
                tmp_nodes[uorv,dir,nbr,l] \(+=1\)
            conc_nodes \(\leftarrow\) tmp_nodes
        for coll in colls do
            for \(e=(n b r, d i r, u o r v, l)\) in coll do
                if \(n b r \notin\{u, v\}\) then node_count \([u o r v, d i r, n b r, l]+=1\)
    Procedure Pop(node_count, sum, conc_sum, coll)
    conc_nodes[:,,:,:], tmp_nodes \([:,:,:] \leftarrow 0\)
    for coll in colls do
            for \(e=(n b r, d i r, u o r v, l)\) in coll do
                if \(n b r \notin\{u, v\}\) then node_count [uorv,dir, \(n b r, l]-=1\)
    for coll in colls do
            for \(e=(n b r, d i r, u o r v, l)\) in coll do
                if \(n b r \notin\{u, v\}\) then
                    sum [uorv,dir,l,:,::] -= node_count[1-uorv,:,nbr,:]
                    conc_sum \([1-u o r v,:,:\), dir,\(l]-=\) conc_nodes \([1-u o r v,:, n b r,:]\)
                    tmp_nodes[uorv,dir, \(n b r, l]+=1\)
            conc_nodes \(\leftarrow\) tmp_nodes
    Procedure ProcessCurrent(pre_nodes, post_nodes, mid_sum, pre_sum, post_sum,
    pre_conc_sum, post_conc_sum, colls)
        conc_nodes[:,:,:,:], conc_sum[:,,:,:,:,:], pre_partial_sum[:,,:,:,:,:], post_partial_sum[:,,:,:,:,:]
            \(\leftarrow 0\)
        tmp_nodes[:,:,:,:], tmp_sum[:,:,:,:,:,], tmp_pre_sum[:,:,:,:,:,], tmp_post_sum[:,:,:,:,:] \(\leftarrow 0\)
        for coll in colls do
            for \(e=(n b r, d i r, u o r v, l)\) in coll do
                if \(n b r \notin\{u, v\}\) then
                    mid_sum[1-uorv,:,:,dir,l] -= pre_nodes[1-uorv,:,nbr,:]
```

41
utov $=(n b r==u)$ XOR dir
for $0 \leq i, j \leq 1$ do

$$
\operatorname{conc}[i, j,:,,:, l]+=\text { conc_sum[1-utov,i,:,j,: }] \quad / / M_{t, 1, x}
$$

$$
+ \text { conc_sum }[u t o v, \mathrm{j},:, \mathrm{i},:]
$$

pre_partial $[i, j,:,:, l]+=$ pre_conc_sum $[1-u t o v, i,:,, j,:] \quad / / M_{t, 2, x}$

+ pre_conc_sum $[u t o v, j,:, i,:]$
pre_partial $[i, j,:, l,:]+=$ post_partial_sum $[(u t o v==j), 1-i,:, 0,:]$
pre_partial $[i, j, l,:,:]+=$ post_partial_sum $[(u t o v==i), 1-j,:, 1,:]$
post_partial $[i, j, l,:,:]+,=$ post_conc_sum $[1-u t o v, i,:, j,:,] \quad / / M_{t, 3, x}$
+ post_conc_sum $[u t o v, j,:, i,: ;]$
post_partial $[i, j,:, l,:]+=$ pre_partial_sum $[(u t o v \neq i), 1,:, 1-j,:]$
post_partial $[i, j,:,:, l]+=$ pre_partial_sum $[(u t o v \neq j), 0,:, 1-i,:]$
for $0 \leq i, j, k \leq 1$ do
$\operatorname{serial}[i,:, j, l, k,:]+=$ mid_sum[( $j$ XOR utov), $i,:,, k,:] \quad / / M_{t, 4, x}$ $\operatorname{serial}[i, l, j,:, k,:]+=\operatorname{post} \_$sum $[(i$ XOR utov $), j,:, 1-k,:]$
$\operatorname{serial}[i,:, j,:, k, l]+=\operatorname{pre} \_$sum $[(k==u t o v), 1-i,:, 1-j,:]$
else
tmp_sum[1-uorv,:,:,,dir,l] += conc_nodes[1-uorv,:,nbr,:]
tmp_post_sum [uorv, dir,,,$:,:$ ] $+=$ post_nodes $[1-u o r v,:, n b r,:]$
tmp_pre_sum[1-uorv,:,:,,dir,l] $+=$ pre_nodes $[1-u o r v,,:, n b r,:]$
tmp_nodes[uorv,dir,nbr,l] $+=1$
conc_nodes $\leftarrow$ tmp_nodes, conc_sum $\leftarrow$ tmp_sum
post_partial_sum $\leftarrow$ tmp_post_sum, pre_partial_sum $\leftarrow$ tmp_pre_sum
conc_pre_nodes $\leftarrow$ conc_nodes, conc_mid_sum[:,:,:,:] $\leftarrow 0$
reset conc_nodes, tmp_nodes, conc_sum, tmp_sum, post_partial_sum, tmp_post_sum,
pre_partial_sum, tmp_pre_sum
for coll in colls.reverse do
for $e=(n b r, d i r, u o r v), l$ in coll do
if $n b r \notin\{u, v\}$ then
conc_pre_nodes $[u o r v, d i r, n b r, l]-=1$
conc_mid_sum[uorv,dir,, ,:,:] $-=$ conc_nodes[1-uorv,:,nbr,$:]$
for $e=(n b r, d i r, u o r v, l)$ in coll do
if $n b r \in\{u, v\}$ then
utov $=(n b r==u)$ XOR dir
for $0 \leq i, j \leq 1$ do $\operatorname{conc}[i, j,:,:,, l]+=$ conc_sum $[1-u t o v, i,:, j,:,] \quad / / M_{t, 1, x}$
+ conc_sum[utov,j,:,i,::]
+ conc_mid_sum[1-utov,i,:, j,:]
+ conc_mid_sum[utov,j,:,i,:]
$/ / M_{t, 2, x}$
pre_partial $[i, j,:, l,:]+=$ post_partial_sum $[($ utov $==j), 1-i,:, 0,:]$
pre_partial $[i, j, l,:,:]+=$ post_partial_sum $[(u t o v==i), 1-j,:, 1,:]$ post_partial $[i, j,:, l,:]+,=$ pre_partial_sum $[($ utov $\neq i), 1,:, 1-j,:] / / M_{t, 3, x}$ post_partial $[i, j,:,:, l]+=$ pre_partial_sum $[(u t o v \neq j), 0,:, 1-i,:]$



## Appendix B Result tables

Table 5: Relative importances of the categories for countries in Social sciences \& Humanities for $\delta=10$ years

| country | edges | mpe | CC |  | MEC | PEC | TC | ETC |  | EPC | EP | OEP | OEP | EPE |  | IM |  | MO | $\mathrm{M}_{\text {all }}$ | $\mathrm{IM}_{m}$ |  |  |  |  |  | M MTC MSC M2 R |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| WORLD | 4,609,814 | 452 | 0.01 | 0.90 | 46 | 07 | 70.08 | 0.04 | 0.03 | 0.01 | 0.00 | 0.04 | 01 | 0.01 | 0.33 | 0.09 | 0.03 | 0.04 | 0.15 | 0.09 | 0.15 | 0.13 | 0.05 | 0.06 | 0. | 00 | 0.00 |  |  |
| USA | 2,262,767 | 49 | -0.41 | 0.02 | -0.09 | -0.40 | -0.09 | 0.00 | 0.28 | 0.18 | -0.26 | -0.09 | -0.03 | 0.12 | 0.51 | -0.60 | . 07 | 0.12 | -0.36 | -0.34 | 0.00 | -0.00 | -0.05 | -0.01 | 0.03 | -0.30 | 0.22-0.33 | -0.03 | -0. |
| GBR | 694,382 | 235 | -0.2 | 0.0 | -0. |  | -0.25 | -0.18 | 0.25 | 0.15 |  |  | 0.20 | 0.21 | 0.1 | 0.18 | 0.11 | 0.0 | 0.26 |  | 0.00 | 0.03 | -0.02 | 0.02 | 0.39 | -0. | -0.02-0. |  | 0.16 |
| CAN | 418,643 | 286 | -0.31 | 0.00 | -0.19 | . 13 | 30.08 | -0.31 | 0.21 | 0.50 | 0.26 | -0.27 | 0.26 | 0.22 | -0.08 | 0.21 | 0.13 | 0.14 | 0.31 | -0.02 | 0.00 | -0.02 | 0.04 | -0.06 | 0.25 | -0.30 | -0.27-0.12 | - | 0.42 |
| DEU | 395,401 | 352 | -0.32 | 0.02 | -0.14 | 0.19 | -0.14 | -0.04 | 0.42 | 0.17 | -0.48 | -0.27 | 28 | 23 | -0.16 | 0.0 | 0.01 | 0.12 | 0.22 | 07 | -0.00 | 0.02 | -0.04 | 0.02 | -0.24 | -0.32 | 0.120 .14 | -0.33 | 26 |
| AUS | 348,347 | 227 | -0.26 | 0.02 | -0.10 | 0.06 | -0.12 | -0.03-0. | -0.30 | 0.28 | -0.02 | 0.15 | -0.19 | -0.09 | -0.19 | -0.10 | 0.21 | -0.13 | 0.07 | -0.11 | 0.00 | 0.01 | 0.02 | 0.00 | -0.15 | 0.04 | 0.01-0.27 | -0.21 | 0.23 |
| N | 339,059 | 689 |  | 0.04 |  |  |  |  |  | 0.46 | 05 | 21 | -0.42 | -0.37 |  |  |  |  |  |  | 0.00 |  |  |  |  |  | 0.12-0.49 |  |  |
| CHN | 226,770 | 71 | 0.11 | 0.03 | -0.13 | 0.19 | 9-0.28 | -0.24 | 0.35 | 0.39 | 0.82 | 0.14 | . 27 | 31 | 0.2 | , | 0.2 | 03 | 0.1 | 0.1 | 0.00 | 0.04 | 0.0 | 0.04 | 0.52 | -0.18 | -0.15 | -0 | 0.31 |
| ESP | 222,105 | 216 | -0.3 | -0.00 | -0.11 | -0.07 | 70.11 | -0 | -0.28 | 0.20 | 0.13 | -0.01 | -0.35 | -0. | -0.31 | -0. | -0.16 | 0. | 0.10 | -0.44 | 0.00 | -0.05 | -0.19 | 0.13 | 0.56 | 0.34 | -0.24-0.46 | 0. | 34 |
| ITA | 210,865 | 477 | 0.22 | 0.03 | 25 | -0.77 | 0.26 | 0.31 | 1.54 | -0.42 | 0.07 | -0.66 | 32 | 0.18 | 0.44 | -0.83 | 05 | -0.02 | -0.72 | -0.35 | 0. | -0.01 | 0.02 | 04 | 0. 15 | 12 | 06-0.06 | 0. | -0.43 |
| FRA | 195,373 | 90 | -0.18 | 0.03 | -0.14 | , | 20 | -0.07 | -13 | 0.06 | -0.00 | -0.04 | -0.13 | -0.18 | 0.04 | 0.32 | 0.07 | 0.12 | 0.38 | 0.02 | 0.00 | 0.02 | 0.04 | . 02 | -0.21 | -0.14 | -0.04-0.12 |  | . 37 |
| SWE | 134,514 | 395 | -0.00 | 0.02 | -0.30 |  | , | - | 0.20 | 0.57 | 0.78 | 0.1 | . 55 | 0.51 | 0.34 | 0.6 |  | . 04 | 0.50 | 0.1 | 0.00 |  | -0.07 |  | 0.38 | -0. | -0.37 0.34 |  |  |
| CHE | 119,528 | 186 | -0.33 | 0.01 | -0.02 | 27 | 0.01 | 0.02 | 0.37 | -0.37 | -0.21 | -0.24 | 16 | -0.07 | 0.32 | 0.31 | 0.03 | 0.00 | 0.21 | 0.15 | 0.00 | 0.01 | 0.03 | -0.04 | .10 | 0.00 | 0.01-0.38 | 0.03 | 0.07 |
| BEL | 112,156 | 277 | -0.43 | . 3 | -0.03 | 08 | -0.20 | 0.11 | 0.23 | -0.39 | -0.32 | -0.09 | -0.08 | 0.01 | 0.15 | 0.10 | 0.08 | 0.04 | -0.00 | 0.16 | 0.00 | 0.0 |  | -0.00 | -0 | -0.01 | 0.08-0.35 |  | 0.08 |
| JPN | 102,284 | 295 | -0.51 | 0.06 | -0.02 | 03 | 0.66 | -0.15 | -0.56 | 1.14 | 0.18 | 0.14 | -0.36 | -0.12 | -0.80 | -0.8 | -0.24 | 0.03 | 0.02 | -0.82 | 0.00 | -0.13 | 0.01 | -0.12 | 1.60 | 0.34 | -0.11-0.66 | 0. | 33 |
| NN | 54,248 | 101 | 0.12 |  | -0.19 | -0.3 | -0.04 | -0.51 |  | 1.36 | 0.22 | -0.18 | . 39 | 0.24 | -0.21 |  | 0.17 | -0.01 | 0.16 | . 0 | 0. | . |  | -0.24 | 0.66 |  | -0.3 |  |  |
| KOR | 62,022 | 36 | -0.23 | 0.03 | -0.01 | -0.1 | -0.21 | 0.02 | 0.32 | -0.16 | -0.12 | 0.27 | -0.25 | -0.00 | -0.15 | -0.14 | 0.18 | -0.26 | -0.09 | -0.00 | 0.00 | 0.03 | -0.00 | -0.01 | -0.43 | -0.03 | -0.10-0.51 |  | . 38 |
| ISR | 60,720 | 67 | -0.01 | 03 | -0.19 | 0.13 | -0.23 | -0.33 | 0.24 | 0.60 | 0.83 | -0.22 | 0.28 | 0.27 | 0.16 | 0.07 | 0.25 | -0.11 | 0.18 | -0.03 | 0.00 | 0.01 | -0.14 | 0.13 | -0.18 | 0.02 | -0.05 0.20 | -0.08 | 34 |
| BRA | 82,995 | 52 | -0.39 | -0.01 | -0.12 | -0.39 | 0.1 | -0.25-0. | -0.35 | 1.38 | 0.46 | 0.20 | -0.03 | 0.13 | -0.50 | -0.7 | -0.22 | 0.07 | -0.31 | -0.5 | 0.01 | -0.05 | 0.21 | 0.05 | 0.60 | 0.10 | 0.01-0.7 | -0.13 | 0.07 |
| NOR | 78,131 | 97 | -0.29 | 0.01 | -0.06 | -0.13 | 0,01 | -0.17 | 0.08 | 02 | 0.14 | -0.01 | . 19 | 0.05 | 0.13 | 0.0 |  | 0.07 | -0.08 | 0.05 | 0.00 | 0.00 | 0.01 |  | -0.06 | -0.0 | -0. |  |  |
| ZAF | 60,570 | 169 | -0.32 | 0.04 | -0.19 | 0.83 | -0.25 | -0.51 | 0.05 | 0.03 | 0.34 | -0.22 | 0.40 | 0.62 | 0.50 | 1.2 | 0.05 | -0.26 | 1.03 | 0.17 | 0.00 | 0.03 | -0.11- | -0.03 | -0.40 | -0.18 | -0.37-0.3 | -0, |  |
| FIN | 80,721 | 409 | -0.49 | 0.01 | -0.19 | -0.03 | 0.07 | -0.15 | -0.44 | 0.80 | 0.78 | 0.17 | -0.19 | -0.14 | 0.13 | 0.02 | 0.26 | -0.12 | 0.09 | 0.01 | 0.00 | 0.02 | -0.05 | 0.01 | 9 | -0.05 | -0.16-0.36 | -0. |  |
| D | 74,133 | 155 | 42 | 0.00 | 0.02 | -0.16 | 60.13 | -0.02 | 0.97 | 0.14 | -0.14 | -0.42 | 0.63 | 0.73 | 0.38 | -0.21 | 0.07 | -0.01 | -0.20 | 0.0 | 0.00 | 0.02 | 0.01 | 0.01 | -0.24 | 0.16 | 0.05-0.32 | -0. | -0.00 |
| TUR | 41,516 | 24 | 55 | 01 | . 08 | 0.02 | -0.21 | -0.08 | 0.26 | 0.18 | 0.12 | 0.07 | -0.16 | 0.24 | 0.02 | 0.1 | 0.25 | -0.35 | -0. | 0.0 | 0.00 | 0.03 | 0.1 | -0.1 | -0.3 |  | 0.2 | 0.5 | 0.97 |
| NZL | 50,698 | 1,175 | 0.26 | 0.04 | 0.37 |  | 29 | 0.38 | 1.93 | 0.42 | 0.09 | -0.83 | 0.73 | 0.82 | 0.63 |  | 0.14 | . |  | 0.1 | 0.00 | 0.04 | -0.00 | 0.0 | -0.50 | 0.1 | -0.04 0.12 | 0.0 | 0.42 |
| PRT | 51,362 | 105 | -1 |  | -0.05 |  | -0.01 | -0.24 |  | 0.76 | 0.61 | 0.01 | 0.08 | -0.01 | 0.16 | 0.08 | 0.15 | -0.09 | 0.13 | 0.0 | 0.00 | -0.01 | 0.11 | -0.07 |  | -0.10 | -0.03 0.30 | -0.47 |  |
| AUT | 64,879 | 193 | -0.36 | 00 | -0.06 | -0.09 | 90.10 | -0.06 | 0.43 | -0.17 | -0.02 | -0.14 | 0.38 | 0.12 | 0.31 | -0.10 | 0.09 | 0.18 | -0.16 | 0.14 | 0.00 | 0.02 | 0.04 | 0.05 | -0.21 | 0.06 | 0.05-0.17 | -0.13 | -0.01 |
| SGP | 38,924 | 237 | -0.17 | . 02 | -0.25 | 1.49 | 0.12 | -0.49 | 0.37 | . 18 | -0.21 | -0.18 | 56 | 0.44 | 0.39 | 1.5 | 0.14 | . 11 | 1.2 | . 2 | . 00 | 0.01 | -0.06 | 0.14 | -0.09 | 0.2 | 0.0 | 0.21 |  |
| IR | 48,114 | 412 | 08 | 0.01 | -0.18 | 0.85 | -0.10 | -0.38 | 0.54 | 0.30 | 0.41 | -0.56 | 0.78 | 0.60 | 0.52 | 1.19 | 0.14 | -0.02 | 1.02 | . 16 | . 00 | . 01 |  | 0.03 | 0.15 | -0.24 | -0.48 0.39 | -0. | -0.37 |
| GRC | 38,452 | 90 | 0.13 | 0.03 | 0.29 | -0.70 | 0.25 | 0.40 | -0.30 | -0.67 | -0.73 | 0.34 | -0.84 |  | 0.51 | -0.74 | 0.06 | 0.0 | 0.64 | 0.24 | 0.00 | 0.02 | -0.15 | 0.1 | 0.29 | -0.26 | 0.16 -0.06 | -0.0 | 0.20 |
| POL | 33,372 | 57 | -0.11 | 0.02 | -0.03 | 0.12 | -0.14 | 0.11 | 0.09 | -0.45 | -0.3 | -0.01 | -0.15 | -0.13 | 0.28 | 0.2 | 0.10 | 0.0 | 0.16 | 0.18 | 0.00 | 0.03 | 0.01 | 0.09 | -0.39 | -0.12 | -0.07-0.0 | -0.5 | 0.23 |
| CZ | 20,190 | 128 | 4.53 | , 06 | -0.14 | 1.02 | -0.59 | -0.32 | 0.26 | . 01 |  | -0.13 | 0.27 | 0.20 | 0.49 | 1.41 | 0.11 | 0.10 | 1.07 | 0.2 | . 04 | 0.07 | 0.09 | 0.07 | 0.8 | -0.17 | -0.26 5.7 | - |  |
| CHL | 18,894 | 16 | 0.13 | 02 | 0.00 | -0.03 | 0.18 | 0.06 | 0.75 | 0.47 | 0.45 | 0.32 | -0.37 | -0.42 | 0.17 | -0.01 | 0.02 | -0.12 | 0.02 | 0.0 | 0.00 | 0.04 | -0.1 | 0.21 | 0.5 | -0.13 | $0.24 \quad 0.2$ | 0.45 | 0.06 |
| IRN | 15,516 |  | -0.11 | 0.01 | 0.23 | -0.52 | 20.10 | 0.52 | 0.82 | 0.45 | 0.63 | 0.63 | -0.71 | -0.67 | 0.52 | -0.5 | 0.13 | 0.10 | -0.54 | 0.07 | 0.00 | 0.01 | 0.01 | 0.0 | 0.07 | -0.28 | 0.27-0.2 | 0.0 | 0.28 |
| RU | 19,163 | 196 | 0.26 | 0.00 | -0.17 | . 00 | -0.08 | -0.37 | 0.31 | 0.24 | 0.67 | -0.52 | 0.04 | 0.34 | 0.38 | 1.3 | 0.09 | -0.21 | . 10 | 0.20 | . 0 | 0.00 | 0.07 | 0.0 | 0.0 | 0.44 | -0.38 0.35 | 0.78 |  |
| ROU | 16,342 |  | -0.37 | 0.05 | -0.18 | 1.23 | 0.41 | -0.22 | 0.20 | . | 0.8 | -0.13 | 0.44 | 0.19 | 0.44 | 1.33 | 0.14 | -0.06 | 0.98 | 0.25 | 0.00 | 0.04 | 0.13 | 0.00 | -0.4 |  | 0.20-0.55 |  |  |
| HRV | 22,399 | 15 | -0.31 | 0.01 | 32 | -0.84 | 0.17 |  | -0.26 | 0.71 | -0.56 | 0.36 | -0.60 |  | -0.51 |  | -0.64 | 0.41 | 0.8 | 0.54 |  |  |  | -0.10 | 0.4 | -0.12 | -0.23-0.7 |  |  |
| MYS | 14,706 | 473 | 0.10 | . 03 | 0.40 | -0.92 | 0.22 | 0.60 | . 9 | -0.06 | -0.19 | 0.79 | -0.63 | -0.77- | -0.66 | -0.9 | -0.30 | -0.01 | 0.9 | 0.22 | 0.0 | 0.0 | -0.0 | -0.04 | 0.4 | -0.55 | 0.57-0.45 | -0.0 |  |
| HU | 27,671 | 1,074 | -0.48 | 03 | 0.34 | -0.77 | -0.13 | 0.53 | 0.87 | -0.03 | -0.52 | 0.61 | -0.75 | -0.70 | 0.78 | -0.72 | 0.03 | -0.20 | -0.75 | 0.18 | 0.00 | -0.07 | 0.04 | 0.07 | 0.8 | 0.19 | -0.01-0.34 | 0.75 | 0.33 |
| IND | 14,457 | 10 | -0. | 0.01 | 0.30 | -0.74 | 0.11 | 0.32 | 0.36 | -0.31 | -0.09 | 0.03 | -0.20 | -0.33 | -0.05 | 0.8 | -0.49 | 0.37 | -0.77 | -0.11 | 0.00 | 0.03 | 0.12 | -0.14 | -0.33 | 0.07 | $0.09-0.61$ | -0.7 | 㖪 |
| MEX | 16,471 | 31 |  |  | 0.36 | -0.72 | 0.83 |  |  | . 43 | 0.17 | . 50 | -0.53 | 0.55 | -0.33 | -0.7 |  |  | -0.76 | 0.23 | 0.0 | -0.05 |  | - | 0.60 |  | 0.13-0.14 |  |  |
| SVN | 9,343 | 15 | 62 | 00 | -0.01 | 0.03 | 0.13 | 0.25 | 0.15 | 0.78 | -0.46 | 0.16 | -0.32 | -0.31 | 0.10 | 0.0 | 0.04 | -0.08 | -0.08 | 0.24 | 0.00 | 0.03 | 0.04 | -0.04 | -0.34 | 0.16 | $0.08 \quad 0.07$ | -0.4 |  |
| THA | 13,510 | 171 | -0.48 | 0.12 | 0.04 | -0.76 | 61.28 | -0.18 | 1.50 | 0.41 | 0.90 | -0.59 | 0.73 | 1.04 | 0.52 | -0.79 | -0.19 | -0.04 | -0.76 | -0.06 | 0.00 | 0.02 | 0.07 | 1 | -0.22 | 0.37 | -0.09 0.32 | -0.08 | 0.32 |
| ARG | 8,886 | 23 | 0.24 | 0.03 | 0.11 | 0.02 | 20.24 | 0.20 | -0.23 | 0.86 | -0.80 | 0.19 | -0.28 | -0.42 | -0.06 | -0.04 | 0.05 | -0.04 | -0.12 | 0.17 | 0.0 | -0.00 | -0.20 | 0.13 | 0.05 | 0.06 | 0.290 .03 | 0.45 | 0.18 |
| SRB | 8,739 | 25 | -0.09 | . 06 | 0.55 |  | 58 | 0.83 | . 89 | 0.22 | . 17 | 0.73 | -0.94 | . |  |  | 0.5 | . 30 |  | 0.7 |  | .06 |  | -0.46 | 0.80 |  | 0.78 -0.47 | 1.37 | . 6 |
| LTU | 8,058 | 37 | -0.27 | 0.06 | -0.00 | -0.56 | 6-0.52 | 0.26- | -0.56 | -0.75 | -0.64 | 0.37 | -0.38 | -0.39 - | -0.17 | -0.52 | 0.13 | 0.05 | -0.56 | 0.16 | 0.00 | 0.01 | -0.04 | . 16 | 0.15 | 0.29 | 0.15-0.44 | -0.38 | 0.93 |
| EST | 12,195 | 82 | -0.23 | 0.03 | -0.14 | 68 | 20 | -0.03 | 0.06 | -0.73 | -0.62 | -0.07 | 0.21 | -0.01 | 0.22 | 0.63 | -0.32 | 0.32 | 0.42 | 0.22 | 0.00 | 0.02 | 0.02 | -0.10 | 0.26 | -0.32 | 0.42-0.51 | 0.69 | -0.2 |
| SVK | 7,835 | 160 | -0.14 | 0.02 | -0.20 | 1.41 | 1-0.14 | 0.36 | -0.49 | 0.30 | -0.53 | 0.07 | 0.59 | 0.06 | 0.33 | 1.30 | 0.05 | 0.14 | 0.93 | 0.26 | 0.00 | 0.02 | 0.04 | -0.01 | -0.24 | 0.14 | 0.27-0.12 | 0.34 | 0.26 |
| CYP | 5,278 | 25 | 1.10 | 0.00 | 0.02 | -0.21 | -0.28 | -0.08 | 0.14 | -0.43 | -0.14 | -0.10 | 0.28 | 0.17 | 0.29 | 0.15 | 0.31 | -0.01 | -0.17 | 0.10 | 0.01 | 0.01 | -0.10 | 0.11 | 0.10 | 0.16 | $0.14-1.69$ | -0.26 | 0.72 |
| SAU | 10,191 | 847 | 0.34 | 0.02 | -0.01 |  | -0.26 | -0.25 | 0.36 | -0.53 | -0.27 | -0.32 | 0.55 | 0.98 | 0.58 | 1.14 | 0.20 | -0.17 | 0.83 |  | 0.00 | 0.04 | 0.04 | 0.04 | 0.45 | 0.51 | -0.30 0.09 | -0.51 |  |
| COL | 5,372 |  | 0.60 | -0.01 | 0.38 | -0.63 | -0.09 | 0.40 | -0.86 | 0.22 | -0.69 | 0.60 | -0.66 | -0.75 | -0.46 | -0.62 | 0.17 | -0.02 | -0.66 |  | 0.01 | -0.04 | -0.20 | 0.0 | 0.4 | 0.61 | -0.17 0.97 | -0.03 | 0.32 |

Table 6: Relative importances of the categories for countries in Biomedical \& Health sciences for $\delta=10$ years

| country | edges | mpe | CC |  |  |  | TC | ETC | EP | EP |  |  |  |  |  | IM |  |  |  |  | M | CM | MP | S |  |  | MSC |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| ORLD | 59,336 | 8,188 | 0.01 | 0.9 | 0. | 05 | 08 | 0.06 | 0.01 | 0.00 | 0.00 | 0.07 | 0.00 | 0.00 | 0.23 | ,06 | 0.02 | 0.03 | 0.10 | 06 | 0. | 0.09 | 0.03 | 0.04 | 0.01 | 0.00 | 0.0 | 0.00 | 0.00 | 0.00 |
| USA | 33,928,604 | 7,288 | -0.31 | 0.00 | 0.07 | -0.41 | 0.03 | 0.24 | -0.48 | 0.27 | 0.01 | 0.19 | -0.63 | -0. | -0.53 | -0. | -0.01 | 0.0 | -0.36 | 0.28 | 0. | -0.0 | -0.01 | -0.02 | 0. | -0.2 | 0.15 | -0.3 | -0.07 | 0.00 |
| GBR | 9,921,341 | 4,090 | -0.21 | 0.02 | -0.21 |  | -0.19- | -0 | - | -0.02 | 0.14 | -0.16 | -32 | 0.29 | . 27 |  | 0.09 | 0.01 | 8 | 0.07 | 0.00 | 0.02 | -0.03 | 0.02 | -0.24 | 0.05 | -0. | -0.03 | -0.06 | 0.13 |
| CHN | 10,224,002 | 945 | -0.10 | 0.03 | 0.16 | -0.36-1 | -0.30 | . 24 | 0.74 | 0.48 | 0.34 | 0.27 | -0.58 | -0.58 | -0.46 | -0.36 | 0.04 | -0.01 | -0.35 | 0.06 | 0.00 | 0.03 | -0.01 | 0.04 | -0.31 | -0.01 | -0.01 | -0.16 | -0.06 | -0.10 |
| DEU | 9,683,847 | 937 | -0.36 | 0.02 | -0.09 | -0.30 | 0.21 |  | 0.05 | 0.24 | . 03 | 02 | -0.25 | -0.19 | -0.22 | 0.4 | -0.02 | 0.0 | -0.25 | 0.17 | 0.00 | 0.0 | 0.06 | 0.0 | 0.08 | -0.1 | 0.13 | . 32 | 0.02 | 0.01 |
| JPN | 7,392,21 | 9,663 | -0.31 | 0.01 | 0.32 | -0.71 | 0.19 | 0.46 |  | 1.23 | 0.12 | 0.38 | 0.8 |  |  |  | . 05 | 0.0 | -0.7 | . 6 | 0.0 | 0.04 | -0.0 | 0.04 | 0.49 | -0.2 | 0.28 | -0.67 | 0.4 | 44 |
| ITA | 8,219,690 | 8,87 | -0.34 | 0.00 | -0.05 | -0.46 | 0.00 | 0.10 | -0.27 | 0.25 | -0.05 | 0.12 | -0.34 | -0.2 | 0.31 | -0.56 | 0.00 | -0.05 | -0.39 | -0.2 | 0.00 | -0.02 | 0.0 | 0.0 | 0.2 | -0.03 | 0.05 | -0.36 | 0.1 |  |
| CAN | 5,521,942 | 3,822 | -0.29 | 0.01 | -0.12 | -0.03 | -0.02- | -0.18 | 0.66 | -0.03 | 0.26 | -0.21 | 0.43 | 0.32 | 0.25 | . 0 | -0.15 | 0.04 | 0.01 | 0.02 | 0.00 | 0.00 | 0.01 | 0.03 | -0.02 | -0.04 | -0.01 | -0.39 | -0.17 | 0.02 |
| FRA | 7,210,391 | 5,223 | -0.43 | 0.02 | -0.23 | -0.1 | -0.13-0. | 08 | 0.55 | 0.25 | 0.19 | -0.18 | -0.04 | -0.07 | -0.12 | -0.4 | 0.05 | 0.0 | 0.01 | -0.36 | 0.00 | 0.01 | -0.03 | . | 0.10 | -0.13 | 0.02 | -0.46 | 0.0 | 0.11 |
| AUS | 3,699,021 | 3,559 | -0. | 0.02 | -0.26 |  |  | 31 | 0.09 | 0.50 | 0.49 | 16 | 0.36 | 0.44 | 0.12 |  | 0.05 | 0.1 | 0.55 | 0.0 | 0.0 | 0.02 |  |  | 0.21 |  | -0.09 |  |  |  |
| NLD | 4,974,288 | 10,114 | -0.35 | . 03 | 0.01 | -0.37 | 0.26 | . 17 | 0.40 | 0.16 | 0.12 | 0.16 | -0.46 | -0.47 | -0.22 | -0.3 | 0.31 | 0.3 | -0.35 | 0.07 | 0.00 | 0.02 | 0.0 | 0.11 | -0.1 | -0.17 | 0.1 | -0.46 | 0.10 | 0.03 |
| ESP | 4,703,093 | 4,115 | -0.25 | 0.01 | -0.10 | -0.07 | 11 | -0.02 | 0.61 | -0.03 | 0.04 | -0.17 | -0.09 | -0.05 | 0.01 | -0.25 | 0.14 | -0.0 | 0.00 | 0.19 | 0. | 00 | -0.04 | 0.03 | 0.03 | 0.14 | -0.19 | -0.2 | 0.0 | -0.04 |
| KOR | 3,806,842 | 20,560 | -0.14 | 0.01 | 0.36 | -0.88 | 0.09 | 0.51 | 0.8 | 0.72 | -0.21 | 0.41 | . |  |  |  | 0.08 | 0.02 |  | 0.7 | 0.0 | 0.02 | 0.07 | -0.15 | 0.28 | -0.23 | 0.22 | -0.56 | 0.06 | 0.31 |
| BRA | 2,970,780 | 1,968 | -0.12 | 0.02 | 06 | -0.25 | 0.16 |  | , | 0.75 | 0.18 | 0.16 | -0.43 | -0.3 | - | - 4 | . 24 | 0.1 |  | 0.2 |  | . 01 | 0.0 | -0.0 | 0.10 | -0.1 | 0.0 | . 34 |  |  |
| SWE | 2,549,442 | 5,081 | . 40 | 0.01 | -0.20 | 0.87 | 14 | -0.28 | 1.20 | 0.18 | -0.03 | -0.39 | 1.15 | 1.0 | 0.71 | 1.1 | . 01 | 0. | 0.8 | 0.2 | 0.00 | 0.04 | -0.02 | 0.0 | -0, | -0.02 | -0.11 | 0.9 | 0.09 | 0.04 |
| CHE | 2,648,795 | 3,385 | -0.25 | 0.01 | -0.04 | -0.11 | 0.02 | 0.01 |  | -0.18 | 0.17 | 0.00 | -0.16 | -0.1 | 0.13 | -0.0 | 0.10 | 0.0 | -0.12 | 0.14 | 0.00 | 0.01 | 0.01 | 0.01 | 0.1 | -0.07 | 0.10 | -0.13 | 0.3 | -0.15 |
| TUR | 1,436,652 | 4,239 | 0.07 | 0.05 | 0.47 | -0.91 | . 53 | 0.55 | -0.92 | 0.44 | 0.30 | 0.44 | -0.94 | -0.9 |  |  | 0.01 | -0.04 |  | 0.6 | 0.00 | -0.06 | 0.0 | -0.03 | 0.7 | -0.26 | 0.27 | -0.51 | 0.5 | 0.57 |
| TWN | 1,453,522 | 1,626 | -0.15 | 0.03 | 0.12 | -0.37 | 0.27 | 0.24 | -0.77 | 0.92 | 0.01 | 0.23 | -0.71 | -0.62 |  | . 6 | 0.02 |  |  | . 4 | 0.00 | 0.0 | - 0 | 0.0 | . 0 | -0.05 | -0.02 | . |  |  |
| DNK | 1,725,540 | 4,128 | -0.26 | 0.02 | -0.30 | . 37 | 0.13- | -0.37 | 0.18 | 0.56 | -0.07 | -0.24 | 0.47 | 0.69 | 0.16 | 0.4 | 0.16 | -0.05 | 0.6 | 0.61 | 0.00 | 0.00 | -0.10 | 0.0 | 0.06 | 0.00 | -0.09 | -0.22 | 0.3 |  |
| BEL | 1,942,386 | 2,476 | -0.25 | -0.00 | -0.03 | -0.19 | 0.05 | 0.01 | 1.26 | -0.28 | 0.04 | -0.27 | 0.34 | 0.15 | 0.45 | 0.1 | 0.02 | 0.0 | -0.25 | 0.19 | 0.00 | 0.02 | -0.00 | 0.0 | -0, | -0.03 | 0.04 | . 16 | 14 |  |
| POL | 1,373,790 | 2,653 | 15 | . 03 | 0.12 | -0.65 | 30 | 0.12 | 0.95 | -0.18 | 0.06 | -0.14 | -0.11 | -0.18 | -0.01 | -0.67 | 0.04 | 0.01 | -0.65 | 0.0 | -0.00 | 0. | 0.0 | 0.03 | 0.1 | 0.04 | 0.02 | 16 | 0.16 | 0.06 |
| IND | 747,96 | 1,625 | 0.23 | 01 | 0.45 | -0.77 | . 06 | 0.47 | 0. | 13 | 0.28 | 0.38 | -0.86 |  | -0.73 | 0.77 | 0.04 | 0.0 | -0.7 | 0.18 | , | 0.04 | . 0 | 0.04 | 0.4 | 0.10 | 0.13 | 0.5 | -0.0 |  |
| AUT | 1,675,734 | 4,776 | -0.30 | 0.01 | 0.00 | -0.06 | . 04 | 0.06 |  | 0.21 | 0.03 | -0.01 | -0.21 | -0.2 |  |  | 0.12 | 0.09 | -0.08 | 0.16 | 0.00 | 0.00 | 0.00 | . 0 | 0.0 | -0.09 | 0.06 | -0.2 | -0.0 |  |
| ISR | 1,130,866 | 2,291 | -0.26 | 0.02 | -0.10 | -0.2 |  | 0.06 | 0.85 | 0.05 | 0.71 | -0.19 | -0.10 | 0.11 | 0.3 | 0.21 | 0.34 | -0.41 | -0.19 | 0.05 | 0.00 | 0.03 | 0.1 | -0.1 | -0.31 | 0.07 | -0.23 |  | . 10 | 0.20 |
| GRC | 1,098,046 | 23,721 | -0.34 | 0.02 | 0.38 | -0.6 | 17 | 0.39 | -0.65 | -0.13 | -0.11 | 0.27 | -0.82 |  | 0.63 | - 6 | 0.17 | 0.1 | -0.61 | 0.09 | 0.00 | -0.01 | 0.0 | 0.06 | 0.09 | -0.04 | 0.17 | -0.42 | 0.23 | 0.16 |
| NOR | 1,088,074 | 2,636 | -0.23 | 0.02 | -0.08 | -0. | . 13 | -0.10 | 0.29 | 01 | 0.23 | -0.07 | 0.11 | 27 | 0.18 | -0.11 | 0.22 | . 10 | -0.11 | 0.08 | 0.00 | 0.01 | 0.0 | 0.06 | -0.1 | 0.22 | 0.1 | -0.19 | -0.10 | 0.18 |
| IRN | 496,041 | 2,020 | 03 | 0.02 | . 09 | -0.05 | 0.20 | -0.02 | - | -0.25 | 0.19 | 0.01 | -0.06 | -0.04 | 0.12 | -0.01 | 0.13 | -0.05 | -0.06 | 0.14 | 0.00 | 0.02 | 0.0 | 0.0 | -0.20 | -0.12 | -0.07 | 0.3 | -0.35 | -0.28 |
| FIN | 1,182,799 | 5,556 | -0.20 | 0.01 | -0.28 |  |  | -0.41 | 0.39 | 0.49 | -0.01 | -0.24 | 0.83 | 1.00 | 0.24 | 0.60 | 0.06 | 0.02 | 0.56 | . 11 | 00 | 0.0 | 0.0 |  | 0.10 | -0.09 | -0.20 | 0.0 |  | . 03 |
| IRL | 773,394 | 1,736 | 0.09 | 0.01 | 0.01 | -0.26 |  | -0.03 | 0.26 | 0.04 | 0.23 | -0.04 | -0.01 | -0.01 | -0.14 | 0.31 | -0.26 | 0.27 | -0.25 | 0.01 | 0.00 | 0.01 |  |  |  | -0.33 |  | -0.12 |  | 0.05 |
| PRT | 720,738 | 1,702 | 0.01 | 01 | 0.12 | -0.51 | 0.14 | 0.22 | 0.52 | 04 | -0.62 | 0.22 | -0.50 | -0.44 | 0.40 | 0.61 | 0.06 | -0.02 | -0.5 | 0.15 | . 00 | 0.03 | 0.0 | 0.01 | 0. | 0.07 | 70.09 | -0.15 | 0. | 0.04 |
| CZE | 896,87 | 2,515 | -0.21 | 00 | -0.05 | -0.44 | 0.05 | 0.0 | 0.04 | - | -0.28 | 0.04 | -0.07 | -0.01 | -0.12 | -0.63 | -0.07 | 0.02 | -0.40 | -0.34 | 0.00 | -0.03 | . 0 | -0.04 | 0.3 | 0.28 | 0.0 | -0.33 | 0.0 | 0.08 |
| RUS | 484,802 | 893 | 0.27 | . 01 | 0.10 | -0.07 | 08 | 0.20 |  | 0.47 | . 08 | 0.17 | -0.41 | -0.40 | -0.03 | 0.00 | 0.07 | -0.05 | -0.10 | 0.20 | . 00 | 0.03 | -0.0 | . | . | 0.0 | 0.04 | 0.34 | -0.2 | 0.06 |
| SGP | 620,466 | 8,460 | -0.20 | 0.05 | -0.32 |  |  | 0.54 |  | 0.10 | -0.23 | -0.32 | 1.09 | 1.43 | 0.87 | 2.4 | 0.21 | -0.07 | 1.91 | 0.27 | . 00 | 0.05 | 0.1 | -0.04 |  | 0.14 | -0.18 | -0.19 | -0.27 |  |
| ZAF | 453,672 | 2,811 | 0.45 | 0.02 | -0.21 | 1.41 | 0.22 |  |  |  | 0.40 | -0.12 | 0.60 | 0.45 | 0.67 | 1.6 | 0.12 | 0.0 | 1.27 | 0.26 | 0.00 | 0.04 | -0.05 | 0.06 | 0.4 | 0.12 | -0.02 | 0.54 | -0. | 0.49 |
| NZL | 417,422 | 1,584 | . 35 | 0.00 | -0.10 | . 08 | 0.00 | 20 | 0 | 09 | 0.20 | 0.25 | . 58 | 67 | 0.48 |  |  | -0.05 | 0.10 | 0.19 | 0.00 | 0.02 | 0.0 | -0.0 | -0.2 | -0.11 | -0.21 | 0.27 | -0.2 |  |
| HUN | 653,117 | 4,772 | -0.24 | . 04 | 0.11 | -0.20 | 0.41 | 0.02 | 1.91 | -0.27 | -0.05 | -0.46 | 0.58 | 0.39 | 0.50 | 0.1 | -0.02 | 0.02 | -0.24 | 0.17 | 0.00 | 0.00 | -0.00 | 0.02 | -0.02 | -0.01 | -0.13 | -0.01 | -0.2 | -0.13 |
| TH | 431,396 | 13,049 | 0.06 | 0.02 | -0.30 | 2.58 | 20 | -0.56- | -0.26 | -0.76 | 0.77 | -0.26 | 1.42 | 1.2 | 0.88 | 2.68 | 0.22 | -0.01 | 2.06 | 0.3 | 00 | 0.03 | 0.0 | 0.04 | 0.3 | 0.26 | 0.04 | 0.19 | 0.05 | 0.38 |
| EGY | 318,860 | 380 | 0.84 | -0.02 | -0.21 | 1.23 | 0.08 | -0.42 | 0.01 | 0.13 | 0.62 | -0.28 | 0.54 | 0.60 | 0.59 | 1.6 | 0.01 | -0.16 | 1.34 | 0.22 | . 00 | 0.00 | 0.01 | 0.04 | 0.03 | 0.03 | -0.17 | 1.24 | -0.53 | -0.21 |
| MEX | 238,621 | 186 | - | . 01 | 0.06 | -0.23 | . | . 09 | - | . 08 | 0.18 | 0.12 | 0.10 | 0.08 | 0.08 | -0.2 | 0.02 |  | -0.26 | 0.0 |  |  | -0.0 | 0.0 | 0.24 | 0.03 | 0.07 | 0.35 |  |  |
| MYS | 197,485 | 88 | 0.70 | 0.00 | 0.06 | -0.15 | 0.09 | -0.02 | -0.37 | 0.07 | 0.18 | 0.15 | 0.03 | -0.07 | -0.02 | -0.11 | 0.01 | -0.13 | -0.23 | 0.25 | 0.00 | 0.01 | 0.0 | 0.00 | -0.1 | 0.08 | 0.01 | 0.51 | 0.01 | 0.44 |
| SA | 322,965 | 1,496 | 0.16 | 0.02 | 0.05 | 0.26 | 0.24 | 0.04 | . 64 | -0.27 | 0.18 | 0.12 | -0.36 | -0.35 | 0.10 | 0.41 | 0.12 | -0.15 | 0.23 | 0.24 | 0.00 | -0.02 | -0.01 | 0.03 | 0.2 | 0.08 | -0.06 | 0.05 | -0.20 | 0.01 |
| CHL | 288,859 | 1,069 | 0.02 | . 01 | -0.11 | 0.86 | 0.06 | -0.18 | -0.42 | -0.23 | 0.02 | -0.03 | 0.09 | 0.07 | 0.29 | 1.1 | 0.18 | 0.0 | 0.90 | 0.19 | 0.00 | 0.01 | 0.10 | -0.06 | -0.1 | -0.16 | 0.00 | -0.06 | -0.03 | -0.25 |
| ROU | 234,054 | 1,034 | . | . 04 | 0.19 | -0.78 | 0.38 | 0.39 | - | 0.52 | 0.57 | 0.27 | -0.66 | -0.62 | 0.39 | -0.80 | 0.00 | -0.0 | -0.78 | 0.01 | -0.0 | 0.01 | 0.03 | 0.02 | 0.0 | 0.11 | -0.07 | 0.07 | 0.0 | 0.04 |
| ARG | 195,311 | 181 | 0.63 | -0.04 | 0.20 | -0.62 | 0.34 | 0.21 | 0.74 | 0.25 | -0.02 | 0.32 | -0.32 | -0.40 | 0.30 | -0.65 | 0.06 | -0.0 | -0.63 | 0.04 | 0.00 | 0.04 | 0.01 | -0.08 | 0.50 | 0.02 | 0.16 | 0.85 | 0.26 | 0. |
| SRB | 196,159 | 969 | 0.08 | . 03 | 0.18 | -0.82 | 0.29 | 0.21 | 0.78 | 0.06 | 0.68 | 0.38 | -0.28 | -0.35 | 0.33 | -0.86 | 0.04 | . 0 | 0.83 | 0.08 | 0.00 | -0.04 | -0.15 | 0.11 | 0.4 | 0.22 | 0.03 | -0.27 | 0.19 | 0.07 |
| HRV | 208,904 | 858 | . 17 | . 02 | 0.02 | 0.03 | 0.20 | 0.05 | -0.01 | -0.25 | -0.26 | -0.00 | -0.23 | -0.25 | 0.06 | -0.00 | 0.13 | -0.11 | 0.13 | 0.04 | 0.00 | 0.00 | -0.05 | 0.05 | -0.00 | 0.01 | -0.23 | -0.04 | -0.27 | -0.23 |
| SVN | 165,229 | 619 | 0.18 | 0.03 | 0.07 | -0.61 | 0.28 | 0.14 | 1.20 | 0.76 | 0.59 | -0.21 | 0.31 | 0.22 | 0.32 | 0.63 | 0.05 | 0.02 | -0.62 | 0.07 | 0.00 | 0.01 | 0.02 | -0.01 | 0.06 | -0.04 | 0.11 | -0.03 | 0.28 | 0.0 |
| PAK | 124,525 | 1,621 | -0.19 | 0.04 | -0.17 | 1.05 | -0.36- | -0.28 | 0.99 | 0.21 | 0.75 | -0.35 | 0.52 | 0.41 | 0.57 | 1.21 | 0.17 | 0.18 | 0.87 | 0.27 | 0.00 | 0.05 | 0.16 | -0.10 | -0.58 | -0.12 | -0.01 | -0.34 | 0.28 | -0.14 |
| SVK | 163,795 | 1,453 | 0.65 | -0.01 | -0.20 | . 66 | 0.02 | -0.40 | 0.35 | -0.63 | -0.67- | -0.09 | 0.97 | 0.6 | 0.57 | 1.70 | 0.11 | 0.05 | 1.29 | 0.28 | 0.00 | 0.02 | -0.06 | 0.02 | 0.18 | 0.18 | 0.06 | 1.07 | 0.02 | 0.46 |
| LTU | 91,789 | 244 | 0.12 | -0.02 | 0.05 | -0.63 | 0.16 | 0.06 | 0.31 | -0.82-0. | -0.76 | 0.01 | 0.16 | 0.23 | 0.18 | -0.58 | 0.18 | -0.27 | -0.63 | 0.24 | 0.00 | -0.05 | 0.08 | 0.06 | 0.53 | 0.52 | 2 -0.16 | 0.72 | 0.19 | 0.22 |
| NGA | 54, | 179 | 0.42 | -0.01 | -0. | -0.21 | 0.0 | -0.20 | 0.0 | 06 | 0.12 | 0.00 | 0.44 | 0.34 | 0.31 | -0.13 | 0.0 | -0.05 | -0.22 | 0. | -0.00 | -0.02 | 0. | 0. | 0.20 | 0.20 | -0.15 | 0.47 | 0.50 | -0.35 |

Table 7: Relative importances of the categories for countries in Physical sciences \& Engineering for $\delta=10$ years


Table 8: Relative importances of the categories for countries in Life \& Earth sciences for $\delta=10$ years


Table 9: Relative importances of the categories for countries in Mathematics \& Computer science for $\delta=10$ years


## Appendix C Robustness of relative importance

All results discussed in this thesis are based on $\delta$ set to 10 years, i.e., the full timespan of the datasets. In order to confirm the robustness of these results, we compared the relative importances (ri) obtained for $\delta=3,5$ and 10 years, shown in Table 11. The changes in ri of each category from $\delta=3 \rightarrow 10$ for the global datasets are shown in Table 10. In this table, we see only very small ri differences, with a largest reported difference of 0.09 . None of the ri differences are large enough to change the conclusions drawn for any of the fields. The changes in $r i$ of the categories for countries within the respective fields are shown in Tables 12-16. Tables for $\delta=3 \rightarrow 5$ and $\delta=5 \rightarrow 10$ were not included because the changes in ri observed in those tables are simply smaller changes in exactly the same positive or negative direction. For most countries and categories only small changes in ri can be observed. However, for some smaller countries, such as Mexico and Columbia in SSH, we see large changes for the smaller categories M2, RFM and VM. Only Thailand in M\&C observes large enough changes in the relative importances of the categories to require reconsideration of the conclusions drawn in Section 7.4. We had concluded from, amongst others, a low proportion of RFM and high proportion of VM motifs, that the added international collaboration for Thailand in this field is characterized by international knowledge networks formed through foreign scholars that made short term visits to Thailand. Contrary to results for $\delta=10$ years though, for $\delta=3$ years we observe a positive $r i$ for RFM and a small negative $r i$ for VM. Remarkably, this actually strengthens our conclusion. After all, visit mobility from foreign scholars and its effects are less likely to be captured fully in a three year period, whereas after ten years we can observe the strength of the collaboration relationships build through visit mobility from their continued collaboration. In other words, the substantial increase of the proportion of visit mobility (VM) motifs from a $\delta=3$ to 10 years, highlights that much international collaboration can be associated with knowledge networks formed through visit mobility.
We conclude that for larger datasets the relative importances computed for $\delta=10$ years are robust and that for smaller datasets a check is required for variations for different $\delta$. We found that the conclusions drawn in Section 7.3 and 7.4 hold up under this check.

Table 10: Difference in relative importance for $\delta=3 \rightarrow 10$ of each category for all fields


Table 11: Field comparison, relative importances for all fields for $\delta=3,5$ and 10 years


Table 12: Relative importance differences of the categories for countries in Social sciences \& Humanities for deltas $3 \rightarrow 10$


Table 13: Relative importance differences of the categories for countries in Biomedical \& Health sciences for deltas $3 \rightarrow 10$

|  |  | mpe | CC |  |  |  | ETC |  |  |  |  |  |  |  | IM |  |  |  |  | M |  |  |  |  |  |  |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| ORLD | 5,336 | 237.1 | -0.00 0.01 | -0.01 | 0.00 | . 0 | . | 00 |  | -0.00-0.0.0 | . 01 | -0.00 | -0.00 |  | 00 | 0.00 | 0.00 |  | 0.00 | 0.01 | 0.0 | 00 | 0.00 | -0.00 | -0.00 | 0.0 |  | 0.00 |
| USA | 33,928,604 | 4104.3 | -0.01-0.00 | 02 | -0.01 | 0.02 | 0.02 -0 | -0.03 | 0.05 | -0.04 | 0.01 | -0.07 | -0.05 | -0.03 | -0.06 | 0.01 | 0.00 | -0.0 | -0.08 | 0.0 | -0. | -0. | -0.0 | 0.0 | -0.1 | -0.0 | -0.02 | -0.04 |
| GBR | 9,921,341 | 2156.5 | -0.00-0.00 | -0.01 | 0.07 | -0.00 -0.0.000 | -0.03 | 0.06 | -0.08 | -0.07- | -0.03 | 0.09 | 0.08 |  | 0.08 |  | -0.01 | 0.06 |  | -0.00 | -0. | 0.03 | 0.01 | 0.01 | 02 | -0.010 | -0.02 |  |
| CHN | 10,224,002 | 472.0 | 0.00-0.00 | 02 | -0.03 | 0.00 | 0.02 | -0.02 | 0.09 | -0.01 | 0.01 | -0.02 | -0.02 | -0.00 | -0.03 |  | -0.01 | -0.04 | 0.02 | 0.00 |  | 0.01 | 0.02 | -0.03 | 0.02 | 0.01-0.02 | 0.0 |  |
| DEU | 9,683,847 | 93 |  | -0.01 |  |  | 02 | 02 |  | . 05 | . 02 | -0.02 | 0.00 |  |  | -0.01 | . 02 |  | -0.0 |  |  | 0.05 | 0.02 | . 0 | -0.09 | 0.07 |  |  |
| JP | 7,392,211 | 86. | -0.02 0.00 | -0.00 | 0.05 | 0. 01 | -0.02 | 0.01 | 0.05 | -0. | -0.03 | -0.01 | 0.00 | 0.01 | 0.01 | 0.05 | -0.03 | 0.05 | -0.0 | 0.00 | . 01 | -0.02 | 0.0 | -0.0 | -0.0 | -0.02-0.04 |  | . 13 |
| ITA | 8,219,690 | 4835 | -0.01 0.00 | 01 | -0.02-0.0.0 | -0.01-0.0.0.000 | -0.00 | 0.03 | 0.05 | -0.05-0.0 | -0.01 | -0.02 | 0.00 | -0.01 | -0.03 | 0.02 | 0.00 | -0.00 | -0.05 | 0.00 | 0.00 | 0.00 | . 02 | 0.0 | -0.02 | 0.01-0.02 | 0.04 | 0.08 |
| CAN | 5,521,942 | 1823.5 | -0.00 0.00 | 02 | -0.10-0.0 | -0.02 | 0.04 | -0.02- | -0.04 | 0.02 | 0.03 | -0.13 | -0.11 | -0.05 | -0.12 | -0.02 | 0.00 | -0.09 | 0.03 | 0.00 | 0.00 | 0.01 | 0.01 | 0.02 | -0.00 | -0.01-0.04 |  | 02 |
| FRA | 7,210,391 |  | 00 | -0.02 |  |  | -0.02-0.0.010 | 00 | 0.05 | 0.06 | 0.00 | -0.02 | 04 |  | 0.0. | -03 | 0.03 |  | 0.07 | 0.0 |  |  |  |  |  |  |  |  |
| AUS | 3,699,02 | 1808.5 | -0. | -0.00 | . 05 | 0.01 | -0.01 | 01 | -0.03 | -0. | -0.02 | 02 | 02 | 0.0 | 0. | -0.05 | 0.04 | 0.0 | 0.0 | 0.0 | -0.0 |  | . 00 | 0.00 | 0. | 0.01-0.04 |  | . 06 |
| NLD | 4,974,288 | 5262. | -0.00-0.00 | -0.00 | 04 | -0.02-0.0.000 | -0.00 | 0.01 | -0.06 | -0.05-0. | -0.02 | -0.01 | -0. | 0.04 | 0.05 | -0.09 | 0.07 | 0.04 | 0.01 | 0.00 | 0.00 | 0.01 | 6 | -0.0 | -0.08 | $0.02-0.0$ | -0.01 | -0.02 |
| ESP | 4,703,093 | 2077.8 | 0.02-0.00 | -0.01 | -0.01 | 0.01 | -0.01 | 0.08 | -0.04 | -0.05-0.0 | -0.01 | 0.02 | 0.04 | 0.00 | -0.01 |  | -0.03 | -0.00 | -0.00 | 0.00 | 0.00 | -0.0 | 0.01 | 0.01 | 0.04 | -0.02 0.01 |  | . 01 |
| KOR | 3,806,842 | 105 | 0.02 0.00 |  | 0.03 | -0.00 -0.0.000 | 03 | ,02 | 0.05 | -0.05-0. | -0.03 | 0.00 | . 02 | 0.0 | -0.00 |  | . 02 | 0.03 | -0.10 | 0.00 | -0.01 | 0.0 | -0.12 | . 1 | -0.1 | 0.07-0.0 |  |  |
| BR | 2,970,78 | 1045 | -0.02 0.00 | 0.00 |  | -0.03 |  | -0.02 | 0.20 | 03 | -0.0 | -0.04 | -0.02 |  | -0.0 | -0.06 | 08 |  | . 0 | 0.00 |  |  |  | -0.0 | -0.0 | $0.02-0.1$ |  |  |
| SWE | 2,549,442 | 2570.7 | -0.02-0.00 | -0.01 | -0.02 | 0.03 | -0.00 | 0.03 | -0.05 | -0.06 | 0.00 | . 2 | 0.02 | 0.02 | 0.05 | -0.09 | 0.0 | 0.00 |  | -0.00 | -0.01 | -0.00 | -0.00 | 0.0 | -0.00 | 0.020 .0 | 0.05 | -0.17 |
| CHE | 2,648,795 | 1578.0 | -0.02 0.01 | -0.01 | -0.00 | -0.05-0.0.000 | -0.01 | 0.10 | -0.01 | -0.01-0. | -0.02 | 0.03 | 0.05 | 0.04 | -0.02 | 0.00 | 0.0 | -0.01 | -0.0 | -0. | 0.0 | 0. | 0.02 | -0.1 | -0.03 | -0.02 0.0 | -0.02 | . 01 |
| JR | 1,436,652 | 1419.4 | -0.02 0.01 | -0.02 | 0 | 01 | 03 | , 01 | 0.16 | -0.00- | . 02 | . 01 | 0.01 | 0.0 | 0.01 | 0.1 | -0.07 | 0.0 | 0.02 | 0.00 | 0.01 | -0.0 | 0. | 0.0 | -0.0 | $0.01-0.0$ | 0.0 | 13 |
| TW | 1,453,52 | 869.9 | 01 | 0.04 | 00 | 05 |  | 0.03 | 0.07 | 0.01 |  | -0.11 | -0.11 |  | -0.03 |  | 0.01 | -0.0 | -0.0 | 0.00 |  | -0.03 |  | 0.01 | -0.0 | 0.04-0.03 |  |  |
| DNK | 1,725,540 | 2299.7 | -0.01-0.00 | -0.03 | 16 | 0.00 | -0.06 | 0.07 | 0.13 | -0.12- | -0.05 | . 8 | 0.16 | -0.02 | -0.01 | 0.09 | -0.0 | 0.21 | -0.06 | 0.00 | 0.00 | 0.06 | 0.02 | -0.05 | 0.04 | -0.03 0.02 | 0. | 0.28 |
| BE | 1,942,386 | 216.9 | 0.010 .00 | -0.01 | 0.03 | -0.00-0.0.000 | -0.01 | -0.02- | -0.07 | -0.02 | 0.02 | , 2 | -0. | -0.01 | 0.05 | -0.04 | 0.0 | 03 | 0.0 | 0.00 | 0.00 | -0.01 | 0.02 | 0.00 | -0.05 | 0.05-0.01 | 0.0 | . 06 |
| P | 1,373,790 | 1500.5 | $0.01-0.00$ | 0.02 | -0.01 | 03 | $0.02-0$. | -0.02-0.010 | . 05 | -0.05 | . 03 | -0.04 | -0. | -0. | -0.01 | -0.02 | 0.02 | -0.01 |  | -0.00 | -0.00 | 0.04 | 0.02 | 0.02 | 0. | -0.01 0.0 |  | 01 |
| IND | 747, | 846.7 | 0.03-0.00 | 0.02 | 0.01 | 0.02 | -0.00 | . 02 | 0.17 | 12 | -0.00 | 01 | -0. |  |  |  | 03 | 0.0 |  |  | 0.01 | 0.04 | 0.03 |  |  | -0.02 0.07 | -0.16 | 0.21 |
| AUT | 1,675,734 | 332.8 | -0.02 0.00 | 0.01 | -0.06 | -0.01 | . | . 05 | -0.05 | 0.02 | 0.02 | -0.03 | -0.04 | -0.02 | -0.06 | -0.04 | 0.01 | -0.05 | -0.00 | 0.00 | -0.00 | 0.01 | 0.00 | 0.04 | 0.0 | -0.00-0.02 | -0.01 | 0.03 |
| ISR | 1,130,866 | 1155.2 | -0.02 0.01 | -0.05 |  | -0.06 - | -0.00-0.0.000 | -0.07 | 0.02 | 0.08 | 0.03 | -0.00 | -0. | 0.01 | -0.03 | 0.06 | 0.01 | 0.01 | -0.04 | 0.00 | 0.00 |  | 02 | -0.02 | , 4 | 0.030 .05 |  | -0.04 |
| GRC | 1,098,046 | 12780.9 | -0.00-0.00 | -0.01 | 02 | 0.02 | -0.01-0. | -0.05 | 0.04 |  | -0.00 | 0.00 | -0.01 |  | 0.02 | -0.09 | 0.07 | 0.02 | 0.00 | 0.00 | -0.00 |  | . 03 |  | -0.02 | 0.01-0.06 | -0.0 | 0.14 |
| NOR | 1,088,07 | 1431.8 | $0.01 \quad 0.00$ | . 01 | 0.01 | 02 | 00 |  | 0.00 |  | -0.01 | -0.04 | 0.03 |  |  |  | -0.05 | 0.01 | 0.02 | 0.00 |  | -0.03 |  | -0.01 |  | -0.04 0.05 | -0.07 |  |
| IRN | 496,041 | 64. | 0.01-0.00 | 0.01 | . 00 | -0.01 | 0.01 | -0.10- | -0.14 | 0.13 | 0.03 | 0. 01 | -0.05 | -0.03 | 0.02 | 0.05 | -0.01 | -0.02 |  | 0.0 | -0.00 | 0.0 | -0.02 | 0.0 | -0.0 | 0.04-0.0 |  | -0.01 |
| FIN | 1,182,799 | 3132.2 | -02-0.01 | -0.01 | -0.05 | 0.04 | -0.03 | 0.10 | 0.13 | -0.08- | -0.03 | 07 | 0.10 | -0.02 | -0.02 | -0.00 | . 00 | -0.02 |  |  | . | - |  | 0.0 | -0.02 | 0.020 .0 | -0.06 |  |
| IRL | 773,394 | 822.6 | 0.00 | 00 | 0.04 | -0.04 | -0.00 | 0.04 | -0.10 | -0.03-0. | -0.02 | 01 | -0.01 |  | 0.05 | -0.07 | . 04 | 0.03 | 0.03 | 0.00 |  | -00 | . 03 | 0.0. |  | -0.0 |  | 00 |
| PRT | 720,73 | 841.9 | -0.01 0.01 | -0.02 | 0.04 | -0.04-0.0.0.000 | 03 | . 5 | 0.12 | -0.05-0 | -0.03 | 0.02 | 0.05 |  | 0.03 | 0.05 | 03 | 0.0 | 0.01 | 0.00 |  | -0.02 | 0.00 | . 04 | 0.03 | -0.04 0.03 |  |  |
| CZE | 896,872 | 343.6 | -0.02 0.00 | -0.01 | . 0 | 0.02 | 00 | . 06 | 0.07 | 0.02 | . 02 | -0.03 | -0.03 | -0.0 | -0.07 | -0.04 | . 04 | -0.0 | -0.09 | 0.00 | 0.0 | . 0 | 0.03 | 0.02 | 0.04 | -0. |  | -0.07 |
| RUS | 484,802 | 454.9 | 0.030 .00 | -0.01 | 01 | -0.02 | 0.01 | -0.08-0.0 | -0.06 | 0.04 | 0.02 | -0.03 | -0.00 | -0.00 | 0.01 | 0.05 | -0.0 | 0.00 | 0.01 | 0.00 | 0.00 | 0.03 | . 0 | 0.04 | 0.05 | 0.0 | 0.080 | 0.03 |
| SGP | 620,466 | 4313.9 | -0.12 | -0.00000 | -0.00 | -0.02-0.0.000 | -0.01 | 0.09 | 0.04 | -0.10- | -0.01 | . 06 | 0.21 |  | 0.13 | 0.04 | -0.0 | 0.03 |  | 0.00 | 0.00 |  |  |  |  | 0. |  |  |
| ZAF | 453,672 | 145 | -0.05-0.00 | -0.00 | . 04 | 0.02 | 0.00 | 0.01 | -0.07 | -0.06-0 | -0.00 | 0.10 | 0.01 |  | 0.09 | -0.00 | . 0 | 0.0 |  | 0.00 | -0 | 0.03 | 0.0 |  | 0.01 | -0.00 | - |  |
| NZL | 417,42 | 693.5 | - | 02 | 08 | -0.00 | . 05 | 0.02 | -0.0 | -0.06 | . 04 | -0.08 | -0.10 | -0.0 | - 0.1 | -0.02 | 0.02 | -0.12 |  | 0.0 | . 0 | 0.0 | . 02 | 0.04 | 0.0 | 0.030 .0 | 0.10 | 0.16 |
| HUN | 653,117 | 2135.9 | 01 | 0.02 | -0.06 | 0.09 | 0.01 | 0.18 | -0.06 | -0.02-0.0 | -0.01 | . 03 | - | -0.01 | -0.07 | -0.01 | -0.02 | -0.06 | 0.00 | 0.0 | -0.00 | 0.00 | 0.01 | 0.0 | 0.0 | 0.00-0.0 | -0.0 | 0.02 |
| THA | 431,396 | 8229.8 | 01 | -0.00 | 0.11 | 0.06 | -0.01 | 0.01 | -0.07 | -0.12- | -0.01 | . 18 | 0.06 |  | 0.15 | 0.05 | -0.02 | 0.04 |  | 0.00 | 0.00 | -0.07 | 0.02 | . 0 | . 0 | -0.01 0.17 | 0.03 |  |
| EGY | 318,86 | 131.0 | 070.00 | -0.01 | -0.07 | -0.02 | 0.00 | 0.14 | 0.05 | 0.30 | -0.02 | 0.04 | 0.06 | 0.04 | -0.03 | 0.01 | -0.02 | -0.07 |  | 0.00 | 0.01 | 0.0 | 0.01 | -0.06 | 0.0 | -0.05 0.0 | 0.02 | 0.04 |
| ME | 238,62 |  | 0.0300 |  |  |  |  | . 15 | 0.09 | 0.29 |  | -0.14 | -0.07 |  | 0.08 | -0.0 | 02 | -0.0 |  | 0.00 | 0.00 |  |  | . 0 |  | 0.010 .0 |  |  |
| MYS | 197,485 | 9.1 | 0.04-0.00 | 0.00 | -0.04 | -0.01-0.0.0.00 | -0.01 | 0.05 | 0.15 | 0.22 | -0.02 | -0.02 | 0.03 | 0.02 | -0.02 | 0.01 | -0.06 | -0.04 | 0.04 | -0.00 | -0.00 | 0.0 | 0.01 | 0.02 | 0.01 | -0.03 0.18 | -0.0 | 0.03 |
| SAU | 322,965 | 69. | 01 | 0.05 | -0.14 | 0.10 | 0.08 | 08 | -0.17 | -0.01 | 0.04 | -0.11 | -0.09 | -0.08 | -0.15 | -0.01 | -0.04 | -0.16 | 0.03 | 0.00 | -0.01 | 0.05 | 0.01 | 0.0 | 0.05 | $0.00-0.02$ |  | 0.02 |
| CHL | 288,85 | 7. | -0.09 0.00 | 0.00 | -0.06 | -0.01 | 0.02 | 11 | -0.06 | 0.05 | 0.01 | -0.09 | -0.10 | -0.06 | -0.0 | -0.10 | 0.07 | -0.03 | 0.01 | 0.00 | -0.00 | 0.10 | -0.07 | 0.03 | -0.08 | 0.06-0.19 | 0.1 | 0.17 |
| ROU | 234,054 | 6.4 | 0.060 .01 | -0.00 |  |  |  |  | 0.01 |  |  | -0.03 | -0.00 | -0.01 |  | -0.01 | . 01 | . |  | 0.0 | 0.00 | 0.03 |  | 0.02 | 0.01 | 0.010 .02 |  | . |
| ARG | 195,311 | 101.0 | 0.040 .00 | -0.01 | 05 | 0.00 | -0.03 | . 10 | 0.10 | 0.22 | 0.01 | 0.07 | 0.04 | 0.03 | 0.05 | 0.05 | -0.00 | 0.04 |  | 0.00 | -0.00 | 0.01 | -0.03 | 0.05 | -0.02 | -0.01 0.29 | 0.05 | 0.06 |
| SRB | 196,159 | 434.3 | $0.03-0.00$ | 0.02 | . 02 | 0.06 | 0.00 | -0.03 | 0.12 | 0.30-0.0 | -0.01 | 0.02 | -0.02 | -0.00 | 0.01 | -0.05 | 0.05 | 0.02 | -0.02 | 0.00 | 0.01 | -0.11 | 0.08 | 0.15 | 0.06 | 0.00-0.12 | 0.03 | 0.0 |
| HRV | 208,904 | 391.2 | -0.05-0.00 | 0.03 | -0.08 | 0.05 | 0.04 | 0.05 | -0.17 | -0.17- | -0.00 | -0.03 | 0.00 | -0.02 | -0.09 | 0.03 | -0.04 | -0.09 | -0.00 | 0.00 | 0.00 | -0.02 | 0.02 | 0.04 | 0.12 | -0.03-0.13 |  | 0.01 |
| SVN | 165,229 | 283.5 | 0.060 .00 | -0.03 | -0.01- | -0.00 | . 01 | -0.20 | 0.04 | . 00 | 0.08 | -0.15 | -0.12 | -0.10 | -0.01 | 0.01 | 0.05 | -0.01 | -0.01 | 0.00 | 0.00 | 0.00 | 0.00 | 0.01 | -0.05 | 0.01-0.10 |  | 0.07 |
| PAK | 124,525 | 742.6 | -0.03-0.00 | -0.01 | . 00 | . 00 | -0.02 | 0.22 | 0.04 | 0.15 | -0.05 | -0.02 | -0.01 | 0.01 |  | -0.04 | 0.03 | 0.03 | 0.03 | 0.00 | -0.00 | 0.11 | -0.09 | 0.00 | -0.02 | -0.06-0.02 |  | 0.21 |
| SVK | 163,795 | 914.5 | 0.10-0.00 | -0.02 | 0.27 | 0.00 | -0.07 | 0.06 | -0.16 | -0.18-0.0 | -0.04 | 0.35 | 0.16 | 0.13 | 0.32 | 0.04 | 0.02 | 0.19 |  | 0.00 | 0.00 | -0.06 | 0.01 | 0.04 | 0.02 | $0.02 \quad 0.37$ | -0.04 | 0.10 |
| LTU | 91,789 | 109.7 | -0.03 0.01 | -0.01 | -0.07 | -0.06 | 0.01 | -0.17- | -0.01 | -0.01 | 0.07 | -0.05 | -0.05 | -0.05 | -0.07 | -0.02 | -0.02 | -0.07 |  | 0.00 | 0.00 | 0.04 | 0.02 | 0.01 | 0.07 | -0.03 0.02 | -0.08 | 0.03 |
| NGA | 54, | 88.7 | -0.08 0.01 | 0.01 | -0. | -0.05 | 0.00 | 0.02 | 0 | 0.01 | 0.01 | 0.10 | 0.09 | 0. | . 1 | -0.00 | 0.01 | -0.11 | 0.02 | 0.00 | 0.00 | -0.02 | - | -0.02 | 0.07 | -0.05-0.16 |  | 0.12 |

Table 14: Relative importance differences of the categories for countries in Physical sciences \& Engineering for deltas $3 \rightarrow 10$


Table 15: Relative importance differences of the categories for countries in Life \& Earth sciences for deltas $3 \rightarrow 10$

| country | edges | mpe | CC | MEC | MPEC | C | ETC | EP | EPC | EPE | OE | OEPC | OEPE | I | IM | IMI | IMO | $\mathrm{M}_{a}$ | $\mathrm{IM}_{m}$ |  | M | MP | M |  | MTC | MSC | M2 RFM VM |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| WORLD | 22,557,784 | 368.9 | -0.00 0.01 | 0.00 | . 01 | $1-0.01$ | -0.00 | -0.00 | 0.00 | -0.00 | -0.01 | -0.00 | -0.00 | 00 | 0.00 | 0.01 | 0.01 | 0 | 0.00 | 01 |  |  |  | -0.00 |  | 0.00 | 0.00 | 0.00 | 0.00 |
| USA | 6,996,167 | 265.6 | 0.040 .00 | 0.01 | 0.00 | -0.03 | 0.04 | 0.05 | -0.04 | -0.09 | 0.00 | -0.05 | -0.03 | 0.01 |  | -0.01 |  | -0.01 |  | -0.00 | 0.00 |  | -0.00 | -0.07 | -0.04 | 0.01 | 0.08 | 0.06 | 0.02 |
| CHN | 4,149,83 | 274 | . 00 -0. |  | 0.04 | -0.0 | -0.01 | 0.0 | -0.03 | -0, | -0.01 | 0.01 | 0.02 | 0.02 | 0.03 |  | 0.0 | 0.01 |  |  |  |  | 0.0 |  |  | 0.01 |  |  |  |
| GBR | 2,397,358 | 447.8 | 0.02-0.01 | -0.01 | 0.13 | 30.04 | -0.04 | -0.02 | -0.07 | -0.08 | -0.01 | . 13 | 09 | 0.06 | 0.16 | 0.05 | -0.0 | 0.10 | 0.04 | 0.00 | -0.00 | -0.05 | 0.01 | . 0 |  | -0.02 | 0.12 | -0.04 | 0.08 |
| DEU | 2,334,606 | 214.3 | -0.01 0.00 | -0.01 | 04 | -0.03 | -0.00 | -0.03 | -0.08 | -0.02 | 0.00 | -0.03 | . 01 | , | 0.03 | 0.03 | -0.0 | 0.04 | -0.00 | 0.00 | 0.00 | 00 | 0.00 | -0.03 | -0.00 | -0.01 | -0.01 | -0.03 | 01 |
| BRA | 1,535,282 | 253.8 | -0.01 0.00 | 0.00 | -0.00 | -0.01 | -0.03 | 0.08 | -0.04 | -0.11 | -0.04 | -0.00 | 0.05 | -0.03 | -0.07-0.0.0 | -0.01 | -0.02 | -0.00 | -0.08 | 0.00 | -0.00-0.0.0 | -0.02 | 0.01 | 0.04 | -0.06 | 0.02 | -0.15 | -0.00 | -0.06 |
| FR | 2,016,547 | 0.4 | $0.01-0.00$ | -0.01 | 0.01 | -0.00 | -0.03 | -0.04 | 0.03 | 0.03 | . 00 | . 05 | 04 | 0.02 | 0.02 | 0.00 | . 0 | 0.01 | 0.01 |  | 00 |  | -0.04 | 0.02 |  | -0.03 | 0.0 | -0.04 |  |
| CAN | 1,202,79 | 417.0 | 0.00-0.00 | 01 | 08 | 8 -0.0 | -0.02 | 0.08 | -0.03 | 0.1 | -0.02 | 04 | 0.08 | 0.06 | 0.11 | 0.06 | -0.03 | 0.06 | 0.03 | 0.0 |  | 0.02 | 0.01 | -0.0 | 0.04 | -0.0 | 0.04 |  | . 14 |
| ESP | 1,426,076 | 421.5 | 0.01-0.00 | -0.00 | -0.02 | 20.03 | 0.01 | -0.03 | 0.13 | -0.02 | -0.00 | 20 | -0.0 | -0.01 | -0.02 | 03 | 0.02 | 0.01 | -0 | 0.00 | -0.00-0.0 | -0.01 | 0.00 | 0.07 | -0.06 | 0.02 | 00 | 0.05 | 11 |
| JPN | 1,334,389 | 4.0 | 0.020 .00 | -0.01 | 0.07 | -0.02 | -0.02 | 0.05 | 0.06 | -0.13 | -0.03 | -0.02 | -0.01 | -0.00 | 0.02 | 0.11 | 0.1 | 0.07 | -0.04 | 0.0 | 0.01 | 0.0 | 04 | -0.07 | -0.09 | 0.02 | 0.01 | -0. | . 07 |
| ITA | 1,456,927 | 242.6 | 0.04-0.00 | 0.04 | -0.07 | 70.03 | 0.06 | -0.06 | 0.14 | -0.08 | 0.02 | -0.10 | -0.06 | -0.07 | 0.09 | 0. 09 | 0.0 | -0.07 | -0.03 | 0.00 | -0.01 | 0.06 | -0.05 | 0.08 | -0.09 |  | -0.06 | 0.07 | . 17 |
| AUS | 998,525 | 236.3 | 0.00-0.00 |  | -0.01 | -0.02 | 0.0 | -0.06 | -0.20 | 0.00 | . 00 | 0.02 | -0.10 |  | . 02 | . 00 |  | -0.02 | 0.01 |  |  |  | 0.0 | 0.04 |  |  | . 0 | -0.01 |  |
| KOR | 792,561 | , | -0.07 0.00 | -0.01 | 0.06 | -0.04 | -0.04 | 0.11 | 0.02 | 0.03 | -0.05 | -0.02 | 0.05 | -0.01 | -0.03 | -0.01 | -0.02 | 0.05 | -0.08 | 0.00 | 0.00 | 0.04 | -0.03 | -0.02 | -0.06 | 0.02 | -0.05 | 0.17 | -0.01 |
| NLD | 934,686 | 5.8 | 0.03-0.00 | -0.01 | -0.01 | $1-0.02$ | 0.02 | -0.13 | -0.02 | -0.02 | 0.03 | -0.05 | -0. | -0.01 | 0.01 | -0.06 | 0.04 | 0.01 | 0.01 | 0.00 | -0.00- | -0.02 | 0.02 | 0.00 | 0.02 | 0.01 | . 03 | 0.04 | -0.04 |
| IND | 355,780 | . 7 | $0.04 \quad 0.01$ | 3 | 0.08 | -0.02 | 0.07 | -0.06 | -0.05 | -0.06 | . 00 | -0.10 | -0.09 | -0.06 | 0.06 | . 01 | 0.0 | 0.06 | 0.01 | -0.00 | 0.01 | 0.02 | 0.02 | -0.1 | -0.04 | -0.01 | 0.16 | -0.04 | . 17 |
| HE | 736,891 | 192.7 | 0.07-0.00 | 0.01 | -0.08 | 8 -0.01 | -0.00 | 0.07 | 0.02 | -0.03 | 0.00 | . 3 | 00 | . |  | 07 |  | -0.08 |  |  |  | 0.01 |  |  |  | 0.00 | 0.05 |  |  |
| POL | 377,023 | 05 | -0.04 0.00 | -0.01 | -0.01 | 10.01 | -0.01 | 0.06 | 0.00 | -0.06 | -0.01 | . 02 | 03 | 0.01 | -0.02 | 0.09 | -0.07 | -0.02 | 0.00 | 0.00 | 0.01 | 0.01 | 0.00 | -0.05 | 0.04 | 0.04 | -0.08 | -0.01 |  |
| S | 634,280 | 87.3 | -0.03 0.00 | -0.02 | 0.00 | 0 -0.01 | -0.03 | 0.07 | -0.00 | 0.05 | 0.00 | 16 | 0.09 | 0.03 | 0.02 | -0.01 | 0.02 | 0.03 | -0.00 | 0.00 | -0.00 | -0.00 | 0.02 | -0.00 | 0.01 | -0.02 | -0.04 | 0. 03 | 0.05 |
| TU | 272,943 | 117.8 | 0.020 .00 | -0.05 | 13 | 3 -0.03 | -0.08 | -0.10 | 0.19 | 0.08 | . 00 | 11 | 03 | 0.04 | 0.1 | -0.11 | 0.09 | 0.16 | -0.01- | -0.00 | 0.00 | 0.0 | -0.07- | -0.03 | -0.1 | 0.06 | . 14 | 0.2 | . 33 |
| BEL | 563,77 | 438.8 | . $01-0.00$ | 01 | 0.05 | 50.01 | 0.05 | 0.02 | 0.07 | 0.00 | 0.02 | . 03 | 0.04 | 0.04 | 0.06 | 0.04 | -0.0 | 0.04 |  |  | 0.0 |  | . 0 | . |  | -0.02 |  | -0.09 |  |
| RUS | 338,810 | 223.6 | $0.05-0.00$ | 05 | -0.04 | 40.05 | 0.09 | -0.16 | -0.05 | 0.06 | . 06 | -0.11 | -0.10 | -0.08 | 0.05 | 0.08 | -0.01 | -0.06 | . 0 | 0.00 | 0.0 | -0.04 | 0.05 | -0.03 |  | -0.00 | . 23 |  |  |
| IR | 189,304 | 119.9 | -0.04-0.00 | 0.03 | -0.09 | 90.03 | 0.07 | -0.12 | -0.11 | -0.03 | 0.05 | -0.15 | -0.10 | -0.09 | . 0 | -0.09 | 0.01 | -0.08 | -0.00 | 0.00 |  | -0.04 | 0.06 |  | -0.0 | 0.00 | -0.12 | 0.01 | 01 |
| D | 488,052 | 167.0 | 0.00-0.00 | -0.01 | -0.02 | 20.01 | -0.03 | -0.16 | -0.06 | 0.10 | 0.03 | 0.03 | 0.00 | 0.0 | . 01 | 0.00 | -0.01 | -0.01 | . 00 |  | 01 | 0.01 | 01 | 0.07 | 0.03 |  | 0.00 | 0.03 | 0.13 |
| TWN | 322,5 | 80.0 | -0.07 0.00 | 02 | . 00 | -0.03 | -0.06 | 0.10 | . 04 | . 00 | . 04 | . 07 | 0.10 | 0.00 | 0.03 | . 0 | . 0 | 0.01 | 0.04 |  | -0.00 | 0.0 | . 00 | 0.00 | -0.0 |  | -0.03 | 0.0. | 0. 01 |
| PRT | 403,225 | 23. | -0.02 0.00 | 03 | -0.04 | -0.00 | 0.05 | 0.01 | 0.04 | -0.16 | . 00 | -0.07 | -0.08 | -0. | . 05 | 0.02 | -0.0 | -0.05 | 0.00 | . 00 | 0.00 | 0.01 | 0.00 | 0.00 | -0.03 | 0.02 | -0.21 | 0.32 |  |
| ZAF | 277,071 | 399.3 | -0.10 0.00 | 0.02 | -0.05 | $5-0.03$ | 0.05 | -0.08 | -0.06 | -0.05 | 0.02 | -0.07 | -0.08 | -0.04 | -0.09 | 0.03 | 0.03 | -0.09 | 0.01 |  | -0.00 | 0.01 | 0.03 | 0.03 |  | 0.05 | -0.51 |  |  |
| NOR | 401,917 | - | 0.050 .00 | -0.03 | 0.07 | -0.03 | -0.03 | -0.07 | 0.01 | 0.08 | 0.01 | 0.04 | 0.02 | 0.01 | 0.08 | 0.05 | -0.01 | 0.08 | 0.01 | 0.00 |  | 0.01 | 0.01 | -0.02 |  | -0.03 | 0.16 | -0.01 |  |
| AUT | 403,50 | 156.5 | $0.02 \quad 0.01$ | 01 | 04 | -0.09 | -0.0 | 0.13 | 0.14 | 0.14 | 20 | . 6 | 0.11 | 0.0 | 0.06 | 0.0 | -0.02 | 0.02 | 0.06 | 00 | . 01 | . 01 | 0.0 | 0.09 | 0.0 | 0.05 |  | 03 |  |
| CZ | 331,33 | 228.6 | 0.00-0.01 | 02 | -0.02 | 20.09 | 0.04 | -0.11 | -0.02 | -0.01 | 02 | 0.08 | -0.08 | -0.05 | 0.03 |  | . 0 | -0.01 | 0.03 | 0.00 | 0.01 | -0.05 | 0.0 | 0.1 |  | -0.0 | 0.02 | 0.07 | 02 |
| FI | 395,462 | 546.8 | -0.04-0.00 | -0.03 | 0.04 | 40.01 | -0.04 | -0.07 | 0.07 | -0.01 | -0.03 | -0.03 | 0.01 | 0.00 | 0.00 | 0.09 | -0.08 | 0.08 | 0.05 | 0.00 | . 0 | 0.08 | 0.00 | 0.06 | 0.13 | -0.06 | -0.12 | 0.02 | . 08 |
| ME | 192,441 | 64.5 | 0.020 .00 | -0.00 | 0.11 | 1 -0.04 | 0.01 | 0.07 | -0.01 | -0.13 | -0.03 | -0.07 | -0.01 | 0.03 | 0.11 | 0.12 | -0.04 | 0.09 | 0.04 |  |  | . 02 | 0.03 | -0.08 |  | 0.02 | 0.11 | -0.07 | 0.15 |
| GR | 246,55 | 141.9 | -0.02 0.00 |  | 0.08 | 8 -0.04 | -0.07 | -0.12 | 0.13 | 0.17 | . 02 | 03 |  | 0.01 |  | 0.03 | . 0 |  | -02 | 0.00 |  |  | 0.00 | -0.02 |  | 0.00 | 0.0 | -0.08 |  |
| AR | 135,099 | 4.0 | 0.01-0.00 | -0.00 | 0.02 | 20.03 | -0.03 | 0.15 | -0.21 | -0.16 | -0.04 | 06 | 0.04 | 0.05 | 0.03 |  | -0.02 | 0.0 |  | - 00 | -0.00 | . 03 | 0.01 | 0.02 | -0.01 | 0.0 | 0.1 | -0.09 | 01 |
| ISR | 204,154 | 75.3 | -0.00 0.01 | -0.01 | -0.01 | -0.05 | 0.03 | 0.07 | 0.11 | 0.03 | 0.01 | -0.08 | -0.04 | -0.03 | -0.05 | 0.02 | -0.00 | -0.02 | -0.03 | 0.00 | 0.0 | 0.02 | 0.01 | -0.13 | -0.04 | -0.03 | -0.04 | 0.02 | -0.12 |
| THA | 186,695 | 4106.1 | 0.03-0.01 | -0.00 | 0.04 | 40.06 | -0.02 | 0.06 | -0.03 | -0.04 | -0.01 | 0.13 | 0.07 | 0.06 | 0.09 | 0.05 | 0.02 | 0.01 |  | 0.00 | 0.00 | 0.06 | 0.01 | 0.04 | 0.03 | -0.02 | 0.16 | -0.07 |  |
| NZ | 147,677 | 9.1 | $0.01-0.00$ | 0.05 | -0.03 | 30.01 | 0.07 | -0.07 | -0.10 | -0.01 | 0.03 | -0.10 | -0.11 | -0.04 | 0.02 | -0.00 | -0.01 | -0.03 | 0.02 |  | . 0 | 0.03 | 0.01 | -0.01 |  | 0.00 | -0.09 | 0.03 |  |
| MYS | 123,291 | 96.4 | $0.01-0.00$ | 0.01 | 01 | 1 -0.00 | -0.02 | 0.03 |  | 0.00 | -0.03 | -0.01 | 0.01 | -0.01 | -0.02 |  | 0.00 | . 01 | 0.06 | 0.00 | 0.00 | 0.00 | 0.02 | 0.07 | -0.08 |  | 0.04 |  | -0.01 |
| IRL | 192,161 | 225.8 | -0.08-0.00 | 0.01 | -0.12 | 20.01 | 0.01 |  | -0.01 | -0.06 | -0.03 | -0.02 | -0.01 | -0.04 | -0.12 | -0.09 | 0.04 | -0.07 | -0.06 | 0.00 | -0.01 | 0.06 | -0.03 | 0.09 | -0.12 | 0.03 | -0.07 | -0.0 | . 15 |
| C | 127,798 | 50.0 | -0.03 0.01 | 0.03 | -0.24 | -0.07 | 0.06 | -0.00 | 0.02 | 0.12 | 0.05 | -0.12 | -0.07 | -0.07 | -0.26 |  | -0.0 | -0.22 | -0.03 | 0.00 | 0.00 | 0.02 | 0.01 | -0.03 | -0.00 | -0.02 | -0.10 | -0.04 | . 13 |
| HUN | 156,584 | 110.6 | -0.00 0.00 | -0.00 | 0.04 | 4-0.02 | 0.00 | 0.05 | -0.16 | -0.11 | -0.01 | 0.05 | -0.01 | 0.04 | , 0 | 0.03 | 0.0 | 0.03 | 0.02 | -0.00 | 0.00 | -0.07 | 0.03 | -0.05 | -0.03 | 0.05 | 0.10 | -0.12 | 0.17 |
| PAK | 94,225 |  | 0.08-0.00 | 00 | -0.04 | -0.01 |  | 0.04 | . 13 |  |  | -0.02 | -0.02 | -0.02 |  | . 09 |  | -0.04 | 0.02 | . 00 | 0.00 | 0.03 |  |  | -0.12 |  | -0.02 |  | 0.23 |
| EGY | 96,241 | 150.5 | 0.100 .00 | . 00 | -0.04 | 40.03 | -0.01 | 0.12 | -0.09 | -0.00 | -0.00 | 0.05 | 0.11 | 0.06 | 0.04 | 0.00 | -0.0 | -0.03 | 0.04 | -0.00 | 0.00 | 0.05 | -0.02 | -0.02 |  | -0.05 | 0.05 | -0.01 | 0.09 |
| SA | 145,007 | 294.8 | 0.07-0.01 | 0.01 | -0.03 | 30.06 | -0.01 | 0.13 | -0.15 | 0.01 | 0.00 | 0.12 | 0.17 | 0.06 | -0.01 | -0.01 | -0.03 | -0.06 |  | 0.00 | -0.00 | 0.08 | -0.01 | 0.02 | 0.08 | -0.03 | 0.08 | 0.07 | -0.00 |
| SGP | 118,386 | 81.1 | -0.09-0.01 | 0.04 | -0.21 | 10.10 | 0.02 | 0.30 | -0.11 | 0.17 | -0.00 | 0.17 | 0.07 | 0.05 | -0.18 | -0.04 | -0.04 | -0.20 | 0.03 | 0.00 | 0.00 | 0.02 | 0.01 | 0.03 | 0.02 | -0.04 | -0.19 | -0.06 | -0.00 |
| ROU | 80,139 |  | 0.10 |  | 07 | -0.00 | -0.09 |  | 0.00 | -0.17 | -0.03 | 13 | 0.12 |  | . 09 | 0.01 | -0.01 | . 06 |  | . | . 00 | 0.0 |  | 0.02 |  | 0.05 | 0.20 |  |  |
| SRB | 85,290 | 184.4 | 0.02-0.00 | 0.04 | -0.00 | 00.07 | 0.04 |  | -0.07 | 0.08 | -0.03 | -0.07 | -0.10 | -0.02 | -0.01 | 0.02 | -0.03 | -0.00 | -0.05 | 0.00 | -0.01 | 0.00 | 0.01 | 0.22 | -0.11 | 0.04 | 0.03 | 0.17 | -0.31 |
| SVK | 74,869 | 62.4 | -0.03 0.01 | -0.02 | -0.00 | -0.03 |  | -0.05 | 0.06 | 0.16 | 0.02 | . 03 | -0.06 | -0.00 | 0.01 | 0.03 | 0.03 | -0.00 | 0.02 | 0.00 | 0.01 | 0.02 | 0.01 | -0.08 | -0.02 | 0.01 | 0.05 | 0.01 | -0.07 |
| HRV | 61,990 | 63.9 | -0.01 0.01 | 0.00 | -0.00 | -0.01 | 0.00 | -0.06 | 0.11 | 0.18 | 0.01 | -0.01 | -0.04 | -0.02 | -0.00 | -0.09 | 0.05 | -0.00 | -0.00 | 0.00 | 0.01 | 0.03 | -0.06 | 0.01 | -0.06 | 0.01 | -0.28 | -0.19 | 0.01 |
| SVN | 55,977 | 114.4 | 0.10-0.00 | 0.03 | 0.02 | 20.04 | 0.00 | -0.04 | 0.21 | 0.12 | -0.00 | -0.02 | -0.02 | 0.02 | 0.01 | 0.07 | 0.03 | 0.03 | 0.08 | 0.00 | 0.02 | 0.00 | 0.07 | -0.19 | 0.02 | -0.03 | -0.09 | -0.01 |  |
| BGR | 52,172 | 45.5 | $0.08 \quad 0.01$ | -0.03 | 0.11 | 1 -0.04 | -0.00 | -0.00 |  | -0.04 | -0.01 | 0. 03 | -0.03 | 0.03 | 0.14 | 0.09 | -0.02 | 0.10 |  | 0.00 | 0.00 | 0.01 | -0.02 |  | 0.05 | -0.04 | 0.40 | -0.0 |  |
| EST | 75,920 | 143.7 | -0.08-0.00 | -0.01 | -0.14 | 40.06 | 0.02 | -0.12 | 0.02 | 0.07 | 0.05 | -0.15 | -0.08 | -0.04 | 0.10 | -0.02 | -0.05 | -0.12 | 0.03 | 0.00 | -0.01 | 0.03 | -0.02 | 0.08 | 0.02 | -0.03 | -0.16 | 0.02 | -0.05 |

Table 16: Relative importance differences of the categories for countries in Mathematics \& Computer science for deltas $3 \rightarrow 10$


