Alma Mater Studiorum - Università di Bologna

DOTTORATO DI RICERCA IN

DATA SCIENCE AND COMPUTATION

Ciclo 33

Settore Concorsuale: 01/B1 - INFORMATICA

Settore Scientifico Disciplinare: INF/01 - INFORMATICA

EMBEDDED REPRESENTATIONS OF SOCIAL INTERACTIONS

Presentata da: Maddalena Torricelli

Coordinatore Dottorato

Andrea Cavalli

Supervisore

Ciro Cattuto

Co-supervisore

Laura Sartori

Esame finale anno 2022

Alma Mater Studiorum
Università di Bologna

Fondazione I.S.I.

Dottorato di Ricerca
in
Data Science and Computation

Ciclo XXXIII

Settore Concorsuale: 01/B1 - INFORMATICA
Settore Scientifico Disciplinare: INF/01 - INFORMATICA

Embedded Representations
of
Social Interactions

Candidata:
Maddalena Torricelli

Supervisore:
Prof. Ciro Cattuto

Co-Supervisore:
Prof. Laura Sartori

Cordinatore Dottorato:
Prof. Andrea Cavalli

Final Exam - 2022

"E dai dai dai"
Renè Ferretti

"Quando moriremo, nessuno ci verrà a chiedere quanto siamo stati credenti,
ma credibili."
Rosario Livatino

"Mah."
Io, un giorno sì e l'altro pure

"Should have burned this place down when I had the chance."
Michael Scott, "The Office"

Abstract

SOCIAL interactions have been the focus of social science research for a century, but their study has recently been revolutionized by novel data sources and by ideas and methods from computer science, network science, and complex systems science. The study of social interactions is crucial for understanding complex societal behaviours, linking the individual and collective scales.

Social interactions are naturally represented as graphs or networks, which have emerged as a unifying mathematical language to understand structural and dynamical aspects of socio-technical systems. Networks are, however, highly dimensional objects, especially when considering the scales of real-world systems and the need to model the temporal dimension. Hence the study of empirical data from social systems is challenging both from a conceptual and a computational standpoint. A possible approach to tackling such a challenge is to use dimensionality reduction techniques that represent network entities – nodes, edges, sub-graphs – in a low-dimensional feature space, preserving some desired properties of the original data. Low-dimensional vector space representations, in particular, also known as network embeddings, have been extensively studied, also as a way to feed network data to machine learning algorithms.

Network embeddings were initially developed for static networks and then extended to incorporate temporal network data. In this Thesis, we focus on dimensionality reduction techniques for time-resolved social interaction data

modelled as temporal networks. We introduce and characterize a novel embedding technique that models the temporal and structural similarities of events rather than nodes. Using empirical data on social interactions, we show that this representation captures information relevant for the study of dynamical processes unfolding over the network, such as epidemic spreading. We then turn to another large-scale dataset on social interactions: a popular Web-based crowdfunding platform. We show that tensor-based representations of the data and dimensionality reduction techniques such as tensor factorization allow us to uncover the structural and temporal aspects of the system and to relate them to geographic and temporal activity patterns. Based on this, we provide a comprehensive anatomy of a crowdfunding system.

Acknowledgements

NESSUNO avrà il proprio nome su queste pagine di ringraziamenti, ma ci siete proprio tutti. Riconoscetevi oppure no, ignorate pure le righe dedicate a voi e attribuitevi quelle che in realtà sarebbero per altri. Un po' come se fosse l'oroscopo del giovedì dell'Internazionale.

A chi mi ha insegnato qualcosa, anche piccola, e mi ha messo a disposizione il suo prezioso tempo per spiegarmela.

A chi ha creduto in me, e me l'ha detto.

A chi mi ha alleggerito o riempito il cuore, a seconda delle necessità.

A chi ha condiviso le sue storie, i suoi percorsi e le sue cazzate con me e mi ha concesso di fare altrettanto.

Alla gentilezza degli sconosciuti e ai nuovi luoghi.

Al binge watching, al bere e al mangiare bene (e tanto, e non sempre così bene).

A chi ha cucinato per me.

A tutti gli sforzi inutili fatti, a tutte le lacrime versate, a tutta la rabbia e la stanchezza, alla disperazione immotivata.

A me, che nonostante e grazie a tutto questo, sono ancora eroicamente viva.

Ai suonatori un po' sballati
Ai balordi come me
A chi non sono mai piaciuta
A chi non ho incontrato
Chissà mai perché
Ai dimenticati, ai playboy finiti
E anche per me
A chi si guarda nello specchio
E da tempo non si vede più
A chi non ha uno specchio
E comunque non per questo non ce la fa più
A chi a ha lavorato
A chi è stato troppo solo
E va sempre più giù
A chi ha cercato la maniera
E non l'ha trovata mai
Alla faccia che ho stasera
Dedicato a chi ha paura
E a chi sta nei guai
Dedicato ai cattivi
Che poi così cattivi non sono mai
Per chi ti vuole una volta sola
E poi non ti cerca più
Dedicato a chi capisce quando il gioco finisce
E non si butta giù
Ai miei pensieri, a com'ero ieri
E anche per me
E questo schifo di canzone non può mica finire qui
Manca giusto un'emozione, dedicato all'amore
Lascia che sia così
Ai miei pensieri, a com'ero ieri
E anche per me
Ai miei pensieri, a com'ero ieri
E anche per me

L. Bertè, "Dedicato", 1979

Index

1

Introduction

SOCIAL interactions are the fabric of society. These interactions form the basis for social structure and culture and therefore are a key object of social analysis. Historically, the exchange of ideas and the interactions between different cultures have been at the basis of the evolution of culture and the structure of human society.

With the growing pervasiveness of digital devices and online social networks, there has been an enormous increase in data on individual behaviours and social interactions. Internet, email, and social media have now entered people's daily use. From navigating the Web to online purchases, from dating apps to photo-sharing apps, our everyday life is now instrumented by devices capable of quantifying and logging most human preferences, choices, and behaviours.

These digital traces of human behaviours have opened new possibilities for measuring and studying social interactions. They have driven the establishment and growth of new interdisciplinary research domains bordering with the social sciences, such as computational social science, making it possible to answer old research questions using new data sources and study new problems enabled by access to fine-grained behavioural data. Computational social science [1, 2, 3] has emerged as a field at the intersection of social sciences and computer science, embracing novel data sources and tackling social science research questions with methods originating in computer science and complex systems science, such as machine learning and agent-based simulations.

The connectivity brought forth by digital platforms and online social networks also changed how information propagates in society, yielding new opportunities but also unforeseen challenges. On the one hand, people in different corners of the planet can connect quickly and easily, collaborating on planetary-scale socio-cultural artefacts such as Wikipedia [4]; on the other hand, this new connectivity and the governance of digital platforms have been recognized as a challenge for democratic systems, e.g., by allowing echo chambers, the spread of misinformation and more. In light of this complexity, knowledge in computational social science will play an increasing role in policy- and decision-making at all levels of society. Indeed, data-driven approaches have proved valuable for policy design [5, 6], early detection of social risk [7, 8], identification of gaps or biases that affect specific minorities [9], and more.

Many of the novel data sources are relational in nature, as they involve interactions between individuals. Interactions find their natural representation in the mathematical representation of graphs or networks. Graphs of social interactions - social networks - have been intensely studied because of their capability to encode the structural and dynamical complexity of social interactions and have become the representation underpinning both fundamental and applied research in socio-technical systems.

The term *social network* is now established in the current lexicon; however, the study of social networks is far from new: social networks have been the focus of social science research for a century [10, 11, 12, 13, 14, 15]. Social network analysis seeks to understand networks, focusing on their actors and their relationships. Social network analysis has a long history: its origins lie in the first studies by Jacob Moreno on "sociometry" [16] and by Fritz Heider on the triad equilibrium analysis [17]. With the use of methods from the newborn graph theory [18], social network analysts acquired new tools for analyzing structures [19, 20]. Subsequent seminal contributions, such as those by Granovetter [21], established the use of graph concepts to describe the emergent role structures in social systems. Towards the middle of the XXth century, the main lines of research in social network analysis have dealt with studying the sub-structures composing a network and their internal relations [22, 23]. The convergence of traditional sociology with quantitative research fields has given life to a fertile and growing interdisciplinary community that uses a common

language to study the dynamics of social interactions. In 1977 the sociologist Barry Wellman created the International Network Society of Social Network Analysts (*INSNA*) to gather scholars interested in social networks analysis. As a sign of different communities converging, it is worth remarking that last year, in 2021, the main *INSNA* conference, the Sunbelt Conference, was jointly held – for the first time – together with the Network Science Conference, the flagship conference of the Network Science Society that caters to the computer science, network science, and complex systems communities.

All networks, in particular social networks, are high-dimensional objects: the number of relationships we can observe among N individuals scales as N^2, hierarchical structures are common, and higher-order correlations are increasingly considered important. In many cases of interest, social network data also have a temporal dimension, greatly adding to its complexity, both from a methodological and computational standpoint. Networks are thus conceptually, analytically, and computationally challenging objects of study.

To tackle the aforementioned challenges, dimensionality reduction techniques for network data have come to play an important role. Low-dimensional representations that preserve specific features or relations of the original data can help make sense of structures, roles, and similarity relations between nodes and links and help uncover the relationships between network structure and dynamics. Low-dimensional vector-based representation of the network, in particular, is critical to enable the application of machine learning methods to predictions or classification tasks based on network data.

The topic of low dimensional representation of networks has gained massive popularity in the scientific literature. Many of the most recent approaches involve node, edge, or graph embedding. Initially designed for static networks, embedding methods have subsequently been extended to temporal networks, a research area that this Thesis will focus on. Specifically, we will focus on time-resolved networks of social interactions, represented as temporal networks, which in turn can be represented as three-way tensors, with one tensor axis used for the temporal dimension. We will provide an overview of the state of the art of temporal networks modelling, including node embedding techniques and tensor decomposition techniques, and we will introduce a new embedding technique for *events* in a temporal interaction network. As an application,

we will test our method on empirical networks of human close-range proximity interactions. We will further study the relation between low-dimensional vector-space representation of events in the temporal network and dynamical processes unfolding over the temporal network, focusing in particular on simple epidemic processes.

We will then change scale and study human interactions in large-scale, real-world systems for which time-resolved fine-grained data is available. Specifically, we will study data from a Web-based crowdfunding platform, where different kinds of actors are at play, and different types of action can occur.

We will provide a first overview of the anatomy of a crowdfunding system, studying both temporal evolution and network structure. We will use tensor-based representations to uncover patterns in the empirical data and gain insights into the crowdfunding system.

The Thesis is structured as follows:

- In Chapter 2, we will give a general overview of social interactions and how they have been studied and formalized over the centuries. We will focus on time-resolved interactions, using time-varying graphs and tensors to represent them. We will review recent network embedding techniques, with a particular focus on embeddings for temporal network data.

- In Chapter 3, we will present the core of our work, which can be here described as a new method of temporal network embedding for studying dynamical processes. We use empirical data of time-resolved contacts of individuals, and we represented them as temporal interactions networks. We then introduce a weighted event graph representation for temporal networks, a novel type of high-order description of interactions that allows us to map each event in a vector space in which other events will be close to it based on these similarity principles. The final result is the embedding of events of the original temporal network. The temporal network embedding leads us to an informative representation of the original temporal network that captures its essential features impacting the diffusion process. This representation should be as well as compact as possible to meet computational or interpretative needs. This method

proves useful to model paradigmatic processes like infectious disease dynamics.

- In Chapter 4, we will turn to study a large-scale social system, a web-based crowdfunding platform, based on real, fine-grained data. We analyse the interactions between borrowers and lenders on this platform, alongside studying the platform's trends and evolution. We find that dimensionality reduction techniques such as Non-Negative Tensor Factorization yields insights into the structure and dynamics of the system.

- Chapter 5 will summarize our work and contributions and illustrate open challenges and directions for future research.

2

Low-dimensional representation of temporal social networks

I N this Chapter, we will follow the silver thread that connects our interest in the study of social interactions to an effective representation of them which allows us to explore, investigate, analyze and understand the dynamics that regulates human societies.

2.1 The quantitative study of social interactions

In sociology, social interaction is a dynamic, changing sequence of social actions between individuals or groups. Social structures and cultures are founded upon social interactions. By interacting with one another, people design rules, institutions, and systems they seek to live. With symbolic interactionism, the reality is seen as social, developed interaction with others. Social interactions can be occasional or routine, dictated by the need to establish a cooperative or competitive bond; indeed, their study was the common starting point of sociology, psychology, economics, ethnology, and other social sciences.

Social interactions have been at the center of research in several disciplines ranging from economics, physics, sociology [24], [25], [26], [27]. Studying

human interactions proved to be necessary to analyze the complexity of society, the individual or collective behaviour of human beings, and their dynamics.

The main effort in defining the concept of social interaction is due to the German sociologist Max Weber. By "social action" (or interaction) [28], we mean the process in which individuals give meaning to their way of acting by formulating reciprocal conjectures on the sense that another actor would attribute to this way of acting, and based on these conjectures, they orient their attribution of meaning. There is, therefore, a reciprocal attribution of significance to the interaction between the same actors.

Let us think for a moment about how many connections an individual can have in his life. Relationships can be emotional or professional; they can be established and disappear over time. Furthermore, looking back on Weber's definition, the interconnection between two individuals brings with it a recognized common sense that adds a further layer in the understanding and interpretation of the relationship itself. All these characteristics, added to the fact that millions of individuals exist in our societies, make the *social network* system (this is how the set of all professional, friendship and family ties of an individual is defined) a complex system.

In physics, complex systems are systems whose behaviour is intrinsically challenging to model due to the dependencies, competitions, relationships, or other types of interactions between their parts or between a given system and its environment. Here complexity is a keyword: in general, the diversity of concepts and theories that explain social action depends on the complexity of the interactive processes themselves.

Looking at the social interaction system through the lens of the physics of complex systems, we cannot fail to mention network science. An intricate network encodes the interactions between the system's components behind any complex system. It is thus of primary importance to develop a deep understanding of the networks behind complex systems if we want to understand them. Here, then, network science comes to our aid.

2.2 Networks: a natural representation of social interactions

Network science is an academic field that studies complex networks such as social networks, considering distinct elements or actors represented by nodes (or vertices) and the connections between the elements or actors as links (or edges). Network science draws on theories and methods from mathematics, computer science, statistics, and sociology. The first papers relative to this topic date back to the mid-18th century, when Euler submitted the problem of the Seven Bridges of Königsberg [29].

However, we can say that the field has established itself and made remarkable progress at the end of the last century, with contributions, among others, by Paul Erdős and Alfréd Rényi [30], Duncan Watts and Steven Strogatz [31], Albert-László Barabási and Reka Albert [32, 33].

Network science, therefore, responds to the fundamental need in dealing with social interactions: their representation. Networks are a natural way to represent them. Moreover, they are a useful tool that allows the formalization of physical-mathematical concepts or the study of complex dynamics, which can then be used to answer research questions such as those of social science. We can model a wide range of systems in nature or society as graphs of vertices coupled by edges, so through networks. The network structure allows us to understand the behaviour of dynamic systems. In many cases, however, the edges are not continuously active. Contacts among people in a real-world situation, for instance, are a set of instantaneous edges. Like network topology, the temporal structure of edge activations can affect the dynamics of systems interacting through the network, from epidemic spreading on the network to information diffusion. In this context, we will focus on temporal networks [34] as an invaluable tool to represent and study real-world systems and dynamics.

2.3 Time-resolved social networks

A time-varying network, also known as a temporal network, is a network whose links are active only at specific points in time. Each link carries information

on when it is active, along with other possible characteristics such as weight.

As mentioned above, in real systems, the network structure of social interactions may change in time. This kind of network is a thus useful tool to model a wide variety of real-world interactions. Just think of the importance of including time in modelling and analyzing phenomena such as social relationships to understand the impact and the relevance that may have modelling these types of interactions through temporal networks [35], [36], [37], [34]. In particular, temporal networks are useful for understanding and analyzing spreading processes, such as epidemics, opinion diffusion, or information spreading [38], [39], [40]. Information spreading processes are central to human interactions, and the way information spreads through society has changed significantly over the past decade with the advent of online social networking.

With temporal networks, indeed, we have a representation that considers both the time evolution and the network structure of the system under exam.

Using and analyzing them to represent social interactions may yield insights into the structure of social systems. Many approaches have been proposed to perform the analysis so far. As we mentioned above, the challenge of representing the system under exam with an adequate tool is also common to temporal networks. In fact, in cases where the time scale of topological changes of the network is not too much slower than the network dynamics, temporal networks could provide a useful framework. Among these cases, though we can recall three common representations for time-varying network data [34], depending on how negligible the time of the individual interactions is in comparison with the characteristic timescale for the evolution of the network. We can have contact sequences if the duration of interactions is negligible; we can represent the network as a set $\mathcal{C}$ of contacts. We will have a set $\mathcal{C}$ composed by elements as (i,j,t), where i and j are the nodes and t the time of their interaction. For the benefit of future discussion, we define the tuple (i, j, t) as an *event*. In fact, in the following Chapter 3, we will deal with this type of contact network. We can then have interval graphs if the duration of interactions is non-negligible. Each edge in the network will correspond to a set of intervals over which the edge is active. Time-varying networks can also be represented as a series of static networks, one for each time step, called *snapshots*.

Understandably, each of the representations mentioned above is closely linked to the type of system it describes. We propose here: in approaching the study of the temporal network, how do we manage to capture the important information? How do we identify its key features that impact the dynamic processes (such as an epidemic) that take place on top of it? The answer once again is in the network representation, and below we will talk about techniques that will help us find compact yet informative representations of temporal networks.

2.4 Low dimensional representation: embedding of time-varying networks

Networks are a general language for describing complex systems of interacting entities, as we explain above. In the real world, a network always contains complex information, leading to high complexity in computing and analyzing tasks.

Given the rich real-life applications of network mining and the surge of representation learning in recent years, many researches have focused on extracting network's relevant structural information. The methods that use graph node representation in vector space have gained traction from the research community. These particular representations are called *network embedding* and have become the focal point of increasing research interests in both academic and industrial domains.

Network embedding aims to transform one network into a low-dimensional vector space that benefits the downstream network analysis tasks. It is an effective method to learn low-dimensional representations of nodes, which we can apply to various real-life applications.

The approach known as data embedding aims to learn data representation in low dimensional spaces; it is a data-driven method as it encodes data into a generic representation, independent from downstream machine learning tasks. The early applications of data embedding focused on text mining [41]. To understand how a network embedding works, the comparison with the world of texts and documents embedding applies (see Word2Vec [41]: this embedding

technique is the one we adopted for our model - see Chapter 3). When embedding a text, every word is projected into a d-dimensional vector space based on similarity relationships according to the context in which it is inserted (that is, words with similar contexts will be close to each other in the target vector space). In the same way, nodes are embedded following a principle of similarity (i.e., considering their neighbourhood as their context). We can have a visual explanation of this parallel in Figure 2.1.
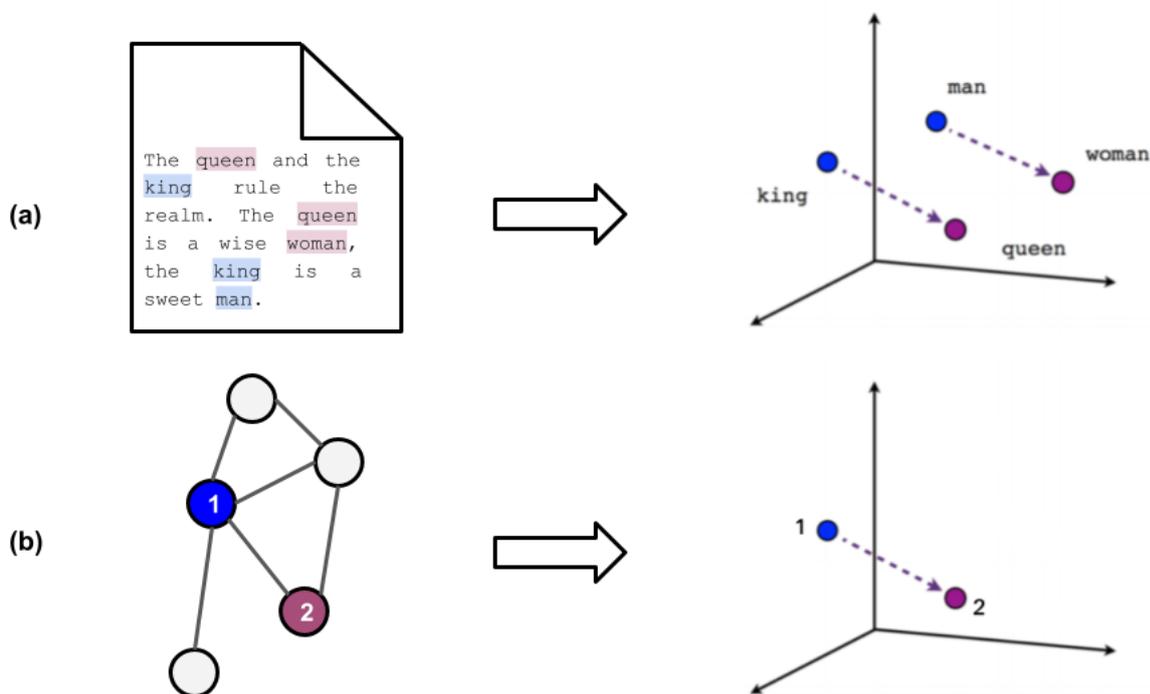
Figure 2.1: How does text embedding works (a) in comparison to network embedding (b). Just as the words close in vector space should be similar in meaning, the representations of the learned features should reflect the similarity between the nodes in the network in the representations of the learned features.

These embedding techniques have been mainly used in the past for static networks [42, 43]. Most common methods use random walk sampling [44, 45] or graph convolution [46, 47] to capture the local structural context of network nodes. Although significant progress has been made on this problem in recent years, several critical challenges remain, such as adequately capturing temporal information in evolving networks. Only recent methods for temporal networks have been developed, but mainly for tasks such as link prediction, node classification, and clustering [42, 43, 48, 49, 50].

Last years have seen a surge in graph representation learning for temporal

networks; literature deals with time incorporation in several ways. Recently a few *dynamical network embedding* methods [51, 52, 53, 54] have been developed to consider dynamical changes in the structure in the learned network representations. At the base of many methods, there is the modification of the standard representation of the temporal network, whether it is in the form of a list of events, a tensor [55], or a supra-adjacency matrix [56]. All of these methods, like DyANE [56], Online Node2Vec [53], STWalk [51], or the one proposed by Singer et al. [52] commonly aim to solve a node embedding problem. They do it by locally sampling the temporal-structural neighbourhood of nodes to create contexts, which they feed to a Skip-gram learning architecture borrowed from the text representation literature [41]. As a solution, they build a sequence of correlated/updated embeddings of network snapshots, which consider the short-term history of the network backwards in time.

However, managing a high number of hyperparameters for controlling the sampling random walk process and the embedding itself might be a problematic limitation to get around. Moreover, node embedding may not reflect the dynamical evolution of temporal interactions. Talking then about a possible prediction task, taking into account only past and present interactions in the embedding can crucially limit it. In contrast, the consideration of future interactions can significantly improve this task.

In the next Chapter, we will propose a new temporal network embedding method that we call *weg2vec* (weighted event graph to vector) [57], which aims to tackle all these shortcomings mentioned above. In case of temporal networks, the recently proposed *event graph* representation [58, 59] defines a higher-order description by identifying relations between events (see Section 2.3), which are adjacent, i.e. not simultaneous and share at least one ending node. We can consider event graph representation as a temporal network extension of the line-graph representation of static networks. It is a useful description as it can condensate the causality information (that has a big impact on spreading processes). At the same time, it is a fast and computationally cheap solution [58] to condense the key features of the temporal network. We approached the techniques of temporal network embedding, intending to study good representations of temporal networks, with particular attention to the modelling of spreading processes.

3

Embeddings for temporal social networks

I N this Chapter we will address the discussion of the temporal network embedding method seen in Section 2.4, named *weg2vec*. As seen above, this method differs from the others introduced in Section 2.4 because it projects events, not nodes, in a low dimensional vector space. This choice is supported by many reasons, including using a handful of hyperparameters and an efficient description of high-order correlations in the network, an undoubted advantage in the study of spreading processes.

3.1 The computational framework

In this Section, we will describe to the Reader each step we made in designing our temporal network embedding method *weg2vec*. To help to visualize the methodological approach we followed, we report a schematic presentation of the methodological backbone of our method in Figure 3.1.

Briefly, we projected our temporal contact networks (see Section 3.1.1) into weighted event graph (we will treat this topic in more details in Section 3.1.2). We then use this static representation of the network to operate a

neighbourhood sampling strategy and sample a set of *contexts* (if we want to use the word embedding nomenclature) for each event (see Section 3.1.3 for further details). Finally, we used these contexts as input for Word2Vec [41], the embedding method we employed to obtain an event embedding of the original temporal network (see Section 3.1.4 for more details).

Once the embedding has been created, we used it for various tasks that we will discuss in Section 3.3. The core of our work is undoubtedly the study of spreading processes on the network, for which we have employed embedding in a prediction task of the outcome of the epidemic (see Section 3.3.1).
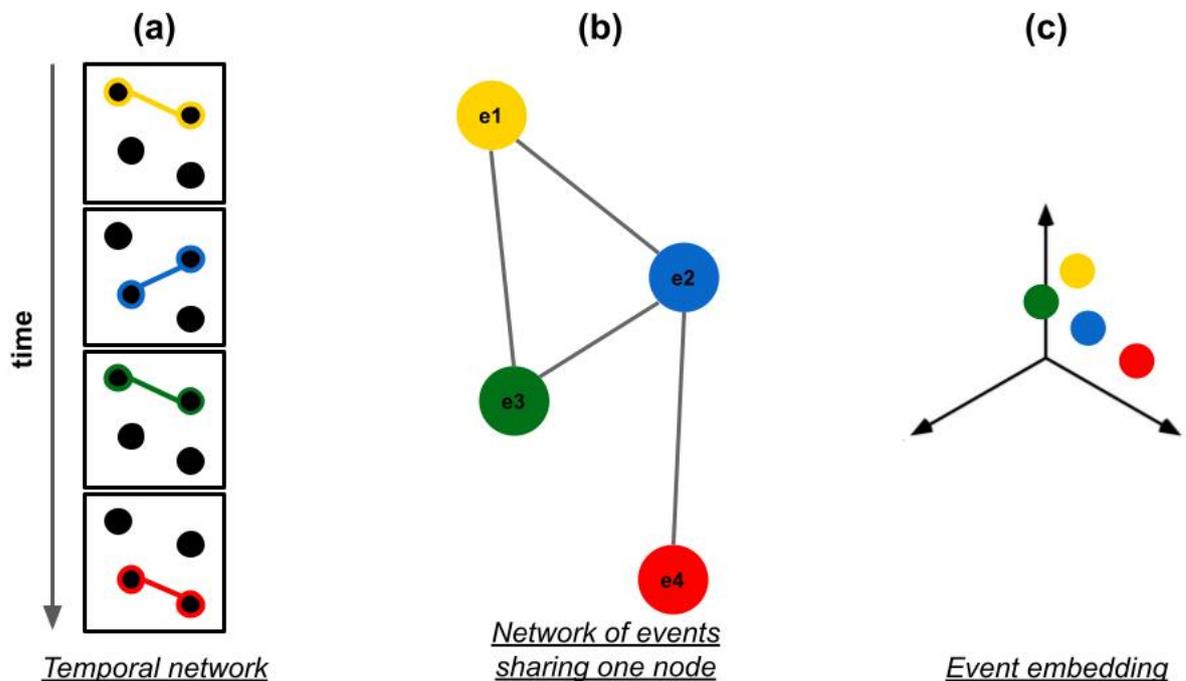
Figure 3.1: Schematic presentation of the methodological approach of our temporal network embedding method. It takes a temporal network (represented here as successive time snapshots) (a) and then projects it into a weighted event graph (b), where nodes are adjacent events. Samples of contexts for each event are then used as an input to a Skip-gram model to get an embedding of events from the original network (c).

3.1.1 Temporal network data

To demonstrate the performance of the *weg2vec* method for the embedding of events in real networks, we used temporal network datasets from the SocioPatterns project [60]. SocioPatterns is an interdisciplinary research collaboration

formed in 2008. It presents a collection of data on physical proximity and
face-to-face contacts of individuals in numerous real-world environments across
several countries. We can find another important example in the *Reality Mining Project*, in which data collected from mobile phones have been employed
to study the structures in the behaviour of both individuals and organizations
[37]. This project joins many other publications and projects in collecting and
studying proximity data, mobility data, and tracking of mobile devices. We
can find an important example in the *Reality Mining Project*, in which data
collected from mobile phones have been employed to study the structures in
the behaviour of both individuals and organizations [37].

 We concentrated on four different settings, a *conference* [61], *hospital* [62],
primary school [63] and *high school* [64], where we could expect particularly
different interaction dynamics and in turn different final outcome of the simulated spreading process. All the networks here presented are undirected, i.e.,
all the edges are bidirectional. General details of the datasets are summarised
in in Table 3.1.

Network	Nodes	Events	Temporal Interval
conference	113	10457	$\sim$ 2.5 days
hospital	75	13650	$\sim$ 4 days
high school	327	36015	$\sim$ 2 days
primary school	236	35921	$\sim$ 8 hours $*$

Table 3.1: Main features of the empirical temporal networks we analyze in
our paper.
$*$ We concentrated only on certain periods of the high school and primary
school networks in order to have a consistent number of events across all
datasets.

3.1.2 Temporal networks as weighted event graphs

Let us consider a temporal network

$$G_T = (N, E_T, T), \tag{3.1}$$

where E_T denotes a set of events (temporal edges) among nodes in N at
times $t \in T$. As defined in Section 2.3, we refer to an event $e = (i, j, t)$

as an interaction between two nodes $(i, j) \in N \times N$ at a given timestamp $t \in T$. The time aggregation of interactions in G_T over T maps the underlying structure into a static graph $G = (N, E)$ defined over the same set of nodes N, which are connected if they interacted at least once. For simplicity here we assume that events are undirected and no self events/links exist, i.e. for any event (i, j, t) or link (i, j), $i \neq j$.

We define two events $e_1 = (i, j, t_1)$ and $e_2 = (k, l, t_2)$ to be *adjacent* $(e_1 \rightarrow e_2)$ if they share at least one node $(\{i, j\} \cap \{k, l\} \neq \emptyset)$ and $t_2 - t_1 > 0$. The concept of adjacency if fundamental: it introduces a directed relation between events, related to an orientation respecting their order in time. Using this notion we can formally define a static directed network representation $D = (E_T, E_D)$ of any temporal network, where original events in E_T are defined as nodes and they are connected by directed links $e_D \in E_D$ if they are adjacent $e_D = e_1 \rightarrow e_2$. The obtained network is a weighted and directed acyclic graph called the *weighted event graph*, defined earlier in Section 2.4. We may consider it as a temporal line graph: in graph theory, the line graph of an undirected graph G is another graph l(G) that represents the adjacencies between edges of G [65]. To simplify our representation, if a given event has multiple future adjacent events with the same pair of nodes, we only consider the earliest one for it.

We have enriched our event graph with link weights that reflect temporal/structural information coded in the original structure. The first type of weight we consider is relative to the time difference between events. We defined it as $w_{path} = \frac{1}{(1 + |t_2 - t_1|)}$, which is a measure inversely proportional to the absolute time difference between adjacent events at t_1 and t_2. This definition of the weight allows us to include the temporality of interactions such as long decay in social activities.

We then define a second weight for adjacent events (links of the event graph) based on the total number of co-occurring events on the underlying adjacent links in the static network. Specifically, the $w_{co}(e_1, e_2)$ co-occurrence weight counts the number of δt-adjacent events in G_T appearing on a given pair of adjacent links $l_1(i, j)$ and $l_2(k, l)$ in the static graph G. By definition, the events which correspond to the same links in the underlying network G will have the same w_{co}. The temporal network data we analyzed (see Section

3.1.1) are sequences of snapshots aggregating temporal interactions. In these systems, we compute w_{co} for adjacent links as the number of co-occurrence of related events within a single snapshot.

As a result of all this process, we finally obtain a static representation for the temporal network under exam, a weighted event graph. We will therefore use this static network to define an context for each event (see Section 3.1.3), which will subsequently serve as input to Word2Vec (see Section 3.1.4).

3.1.3 Neighbourhood sampling strategy

We describe in Section 2.4 the parallel between text embedding and network embedding. We observed that the process through with nodes are projected in a low dimensional vector space is equivalent to the one embedding words (see Figure 2.1).

In the same spirit of text embedding techniques [41] based on the Skip-gram model, we built an event embedding method, which samples neighbourhoods for events from the weighted event graph representation to map them to a lower-dimensional space.

Using again the comparison with texts, we need to assign a *context* to an event e_k to then project it into a vector space. To do this, as for words we use their context to embed them, here we sample the local neighbourhood set N_k of an event e_k. N_k is the set of its first in- (past) and out- (future) neighbours (also called its *predecessors* and *successors* from now on). The sampling is done according to probabilities determined by the two types of weights of the links that connect the actual event to its neighbours. The probability $p(e_r)$ of picking an event e_r from the combined neighbourhood set N_k of the central event e_k is given by :

$$p(e_r) = \alpha F(w_{path}(e_k, e_r)) + (1 - \alpha)F(w_{co}(e_k, e_r)) \qquad (3.2)$$

where α is a coefficient between 0 and 1 scaling the contribution of the two types of weights and F is a normalised weighted function defined as:

$$F(w) = \frac{w(e_k, e_r)}{\sum\limits_{n \in N_k} w(e_k, e_n)}. \qquad (3.3)$$

We sampled nb number of random contexts; each of them contains s events - we call s the length of the context. We will show in Section 3.2 how we selected the hyper-parameters nb, s and α (see Section 3.2.2), alongside with the embedding dimension d (see Section 3.2.1).

We want to emphasize that we use as the local neighbourhood of an event its predecessors and successors. As mentioned in Section 2.4, taking into account only past and present interactions in the embedding can be a limit for the performance in embedding prediction tasks. Considering future events in the neighbourhood of a central event can significantly improve performance in this regard. It is, however, a clear difference with many examples of temporal network embedding techniques shown above. As we will see later (see Section 3.3.1.iii), this choice has proved to be successful in some cases regarding the comparison with other methods.

3.1.4   Embedding of temporal networks

Before discussing the last part of our methodological approach, the embedding, we would like to give the Reader some intuitions about the embedding method we used for the analysis, Word2Vec. We will not enter any technical detail, but it is worth examining it to make the procedure we have followed so far more straightforwardly.

3.1.4.i   Word2Vec: words to vector space

Nobody can deny the importance of words in the evolution of the individual and, in general, of humankind: we use words to learn, we communicate using words. With the advent of more and more advanced technologies, the digital analysis of textual corpus has had more and more success, and with it, the interest in finding computational methods that allow the understanding of words and context. The vectorization of words is nothing more than this: the representation of words as vectors by an algorithm that "learns" the word itself starting from its innate meaning and the context in which it is placed.

Word2Vec is based on this concept. Created by Mikolov and his group at Google in 2013 [41], it represents a set of methods that have a contextual understanding of a word and are therefore able to map it into a vector rep-

resentation, extracting features of the word by its context. It uses a neural
network architecture that can predict the context of a word from the word
itself, or vice versa, depending on whether you use the bag-of-word version
(first case) or the Skip-gram version (second case). In this discussion, we will
examine the Skip-gram model, the one we used for our analyses.

In Figure 3.2 we show a simple description of the neural network architecture of Word2Vec. For each word in the corpus under exam, you may have the
context for each word. The word is given as input to the network after being
encoded using the one-hot encoding [66]. If we think of all the n unique words
in the corpus as a n-dimensional vector, the one-hot encoding of the ith word
will be a n-dimensional vector of zeroes except the one in the ith position.

The 1-layer neural network is trained on that: the one-hot vector multiplied
by the weights matrix of the neural network change the hidden layer, which is
then passed to a softmax function. The softmax is applied on the scalar products of pairs of (words, context) that are in turn taken from a one-hot vector
multiplied by the weight matrix. This softmax function returns the likelihood
of observing the input word along with the context words. The algorithm needs
to maximize this likelihood, tuning itself using backpropagation.

When the likelihood mentioned above is maximized, the resulting hidden layer is the vector representation of the word we have trained Word2Vec
against.

3.1.4.ii *weg2vec* embedding

We shed some light on the Word2Vec architecture in the previous Section
because we used it to obtain our final temporal network embedding. Once the
events context is ready, built as described above, we passed it as input to the
Skip-gram model. We thus obtain a projection of all the events into a lower-
dimensional vector space. Each event has now a $d - dim$ vector representation
obtained as shown in Section 3.1.4.i.

As the first glimpse of our compact representation of temporal networks,
we show the embeddings for the four empirical networks used in the analyses
in Figure 3.3. Here each event is represented as a point in the 3-dimensional
embedded space with colour indicating the time at which they occurred in the
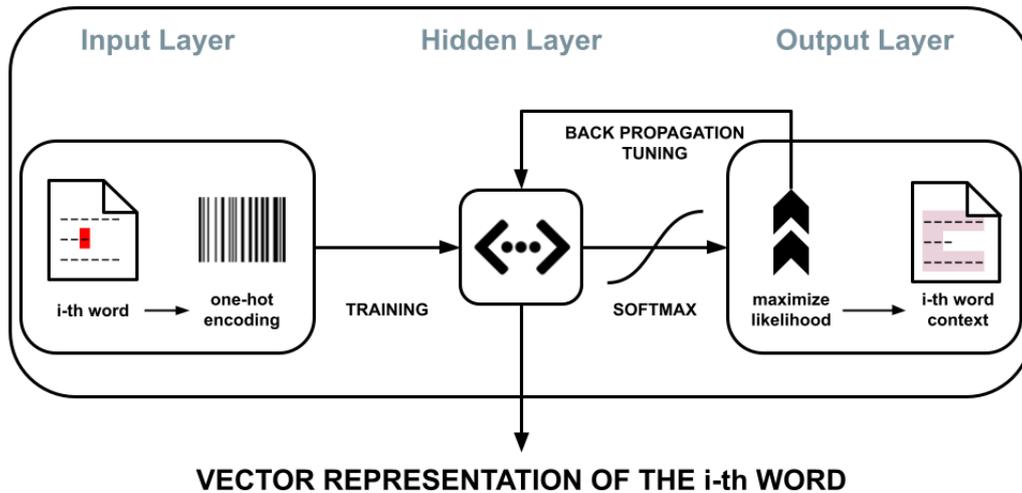
Figure 3.2: Schematic presentation of the Word2Vec neural network. The neural network trains on the one-hot-encoded word, changing the hidden layer. It then tunes itself until the probability distribution given by the softmax function is maximized, using backpropagation. The resulting hidden layer is the vector representation of the word we trained Word2Vec against.

original temporal network. It is interesting to note how the gradient change of colours indicates that these embeddings capture the time ordering of the events in large part.

We will discuss in Section 3.2 how we selected the parameters for the events contexts; the dimension for the embedding has been chosen here for illustrative purposes. We can observe how our method already guarantees the ability to capture the temporal order of events even if at low embedding dimensionality.

We will explain how this compact representation helped us predict the final outcome of an epidemic. In the following Sections, we would like to present to the Reader a showcase of the *weg2vec* embedding properties.

3.1.4.iii Local properties of the embedding

How well is the embedding learning local properties of the temporal networks? Is it able to capture essential key features as time ordering and structure? We answered these questions by measuring the correlation between the time difference observed in the temporal network and the euclidean distance observed in the embedding among pairs of events.

We selected separately 10000 pairs of linked events and 10000 pairs of
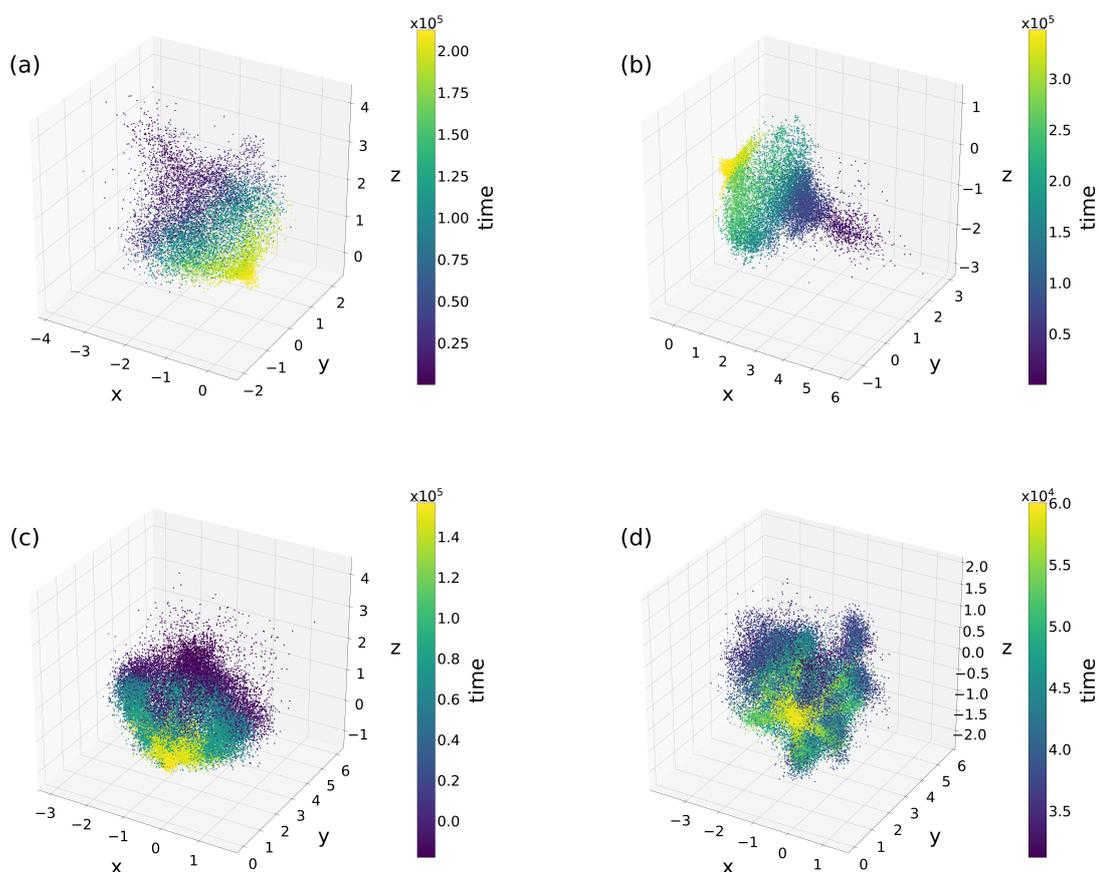
Figure 3.3: 3-dimensional embeddings of the conference (a), hospital (b),
high school (c) and primary school (d) networks. x,y and z axes indicate
event coordinates, while colour shows the time at which the event occurs.
hyper-parameters were set to $\alpha = 0.5$, $nb = 10$ and $s = 10$.

random events to do this analysis. To assign a Euclidean distance to each
pair of events, we computed the average distance over 10 realizations of the
embedding. We set the context parameters s and nb both to 10 and α to 0.5.
The embedding dimensions were set to their optimal values - we will discuss
about dimensionality selection in Section 3.2.1.

We found that in both cases, for connected and random pairs of events, the
correlations between time difference and euclidean distance are positive. This
means the embedding well captures the time ordering. Besides, the correlation
between linked adjacent events is significantly higher than between randomly
selected pairs of events. These results, shown in Table 3.2, demonstrate that
the embedding also captures the local structure.

To confirm these observations, we made another additional analysis. We
evaluated the median of the distribution of distances among both random and

linked events for each value of measured time difference. This relation is shown as a scatter plot in Fig.3.4. There we observe that the centres of the point clusters are positioned below the diagonal in all the cases. It indicates that the distances among linked events are smaller than those among random events.

PCCs / Data	RE	LE
Conference (d=20)	0.36±0.01	0.52±0.02
Hospital (d=14)	0.69±0.02	0.73±0.01
High School (d=26)	0.40±0.02	0.61±0.01
Primary School (d=24)	0.34±0.01	0.59±0.01

Table 3.2: Pearson's correlation coefficients (PCCs) obtained comparing the time difference and the euclidean distance among randomly selected pairs (RE) and pairs of linked (LE) events. We set the context parameters s and nb both to 10 and α to 0.5. The optimal embedding dimension were chosen as explained in Section 3.2.1. The results are obtained over 10 realizations of the embedding for each network.

3.1.4.iv   Tensor decomposition application

As an extra application to test the capability of incorporate network information effectively of our embedding, we show in this Section its employment in a tensor decomposition analysis. This example will demonstrate the ability of our embedding in incorporating not only the network temporal features but also the structural ones.

In the previous Chapter (see Section 2.3), we saw how time networks are an excellent tool for representing natural systems that evolve. The great advantage of their use is facilitating the study of dynamic processes on the network and their impact on it. However, there are other possible representations for time-varying systems with other specific advantages. One of these, which we will discuss in this Section, is the tensor representation.

As defined in literature [67], an N-rank tensor is an object with N indexes, which, in general, has the form $\mathcal{T}_{a_1 \cdots a_n}$. Depending on the rank, we will have different types of tensors. Speaking of common low-rank tensors, 0-rank tensors are scalars, 1-rank tensors are vectors, and 2-rank tensors are matrices.
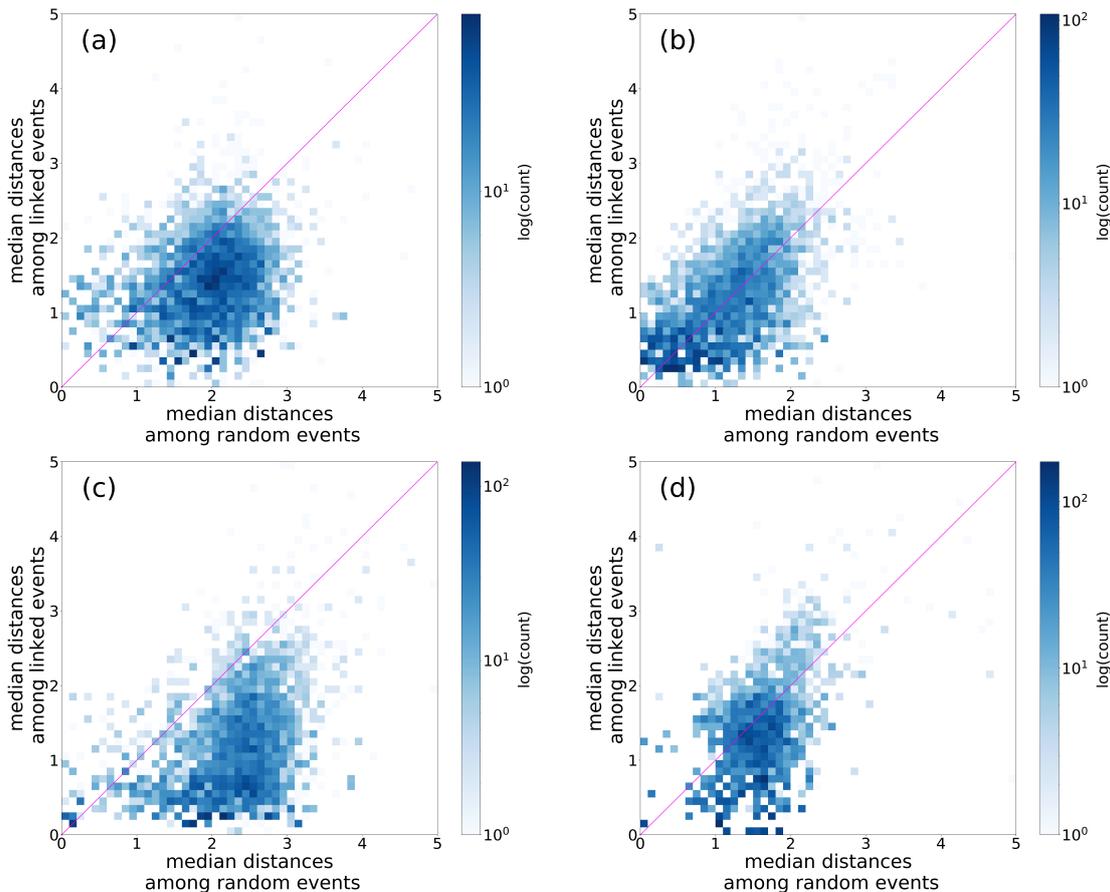
Figure 3.4: Median of the distribution of distances among random and linked events in the conference (a), hospital (b), high school (c), and primary school (d) networks. Each point corresponds to a unique time difference. hyperparameters were set to $\alpha = 0.5$, $nb = 10$ and $s = 10$. The optimal embedding dimension were chosen as explained in Section 3.2.1. The results are obtained over 10 realizations of the embedding for each network.

In our case, we will treat 3-rank tensor, which represents the time-dependent adjacency matrix (or, in other terms, time-dependent 2-rank tensor).

Suppose we have n individuals interacting during a time interval T. In the case of static networks, we can represent the interactions with the so-called adjacency matrix. The adjacency matrix $A_{n\ timesn}$ is a square matrix $n \times n$ used to represent this finite graph. The matrix element $a_{i,j}$ indicates whether the pair of nodes i, j is connected ($a_{i,j} = 1$) or not ($a_{i,j} = 0$) in the graph.

Because we have a time dependence in our case, we will have a T adjacency matrices sequence, one for each time $t \in T$. In other terms, for each snapshot $t \in T$, we will have an adjacency matrix $A_{n\times n}(t)$ which will contain the information of the interactions between the n nodes of the network at that given

instant t. We have thus built a 3-rank tensor whose dimensions are $n \times n \times t$ representing our real time-varying system (see Figure 3.5 (a) and (b)).

The tensor representation of a temporal network paves the way for us to study the temporal network's topological structure and understand the connections of its temporal component. Just as in static networks, this analysis is feasible through the use of community detection methods, so for temporal networks, we can go and trace mesoscale structures that are nothing but community-activity structures of temporal networks.

It is worth mentioning that exists an important difference between community detection and tensor decomposition. On the one hand, community detection techniques are able to find cohesive groups in static networks. To extend these methods to temporal networks, we need to consider them as a series of static snapshots. This temporally-aggregated representation may overlook essential features of the system. On the other hand, tensor decomposition is intrinsically temporal and allows to simultaneously identify communities and to track their activity over time.

The above-mentioned $n \times n \times t$ tensor can then be factorized into a sum of r rank-1 tensors, i.e., the number of mesoscale structures we search into the temporal network.

Tensor decomposition [55, 68] may be seen as a generalization matrix decomposition, which has found application in statistics, computer vision, linguistics, and many other fields. To test our embedding, here we investigate the use of a latent factor decomposition technique, the non-negative tensor factorization. It aims to extract the mesoscale structures of temporal networks (that we can interpret as community-activity structures). In particular, we focused on non-negative tensor factorization [69, 70], since it is a powerful tool for learning representation that leads to more interpretable results and models [71, 72].

Canonical tensor decomposition of a 3-dimensional tensor aims to write a tensor $\mathcal{T} \in \mathbb{R}^{N \times N \times T}$ in a factorized fashion. In other words, the tensor $\mathcal{T}$ can always be expressed as a sum of rank-1 tensors in the form:

$$\mathcal{T} = \sum_{r=1}^{R_{\mathcal{T}}} \mathbf{a_r} \cdot \mathbf{b_r} \cdot \mathbf{c_r} \qquad (3.4)$$

i.e., as the sum of outer products of three vectors, where the smallest value of $R_\mathcal{T}$ for which such a relation can hold is the rank of the tensor $\mathcal{T}$. Again, the three vectors $\mathbf{a_r}, \mathbf{b_r}, \mathbf{c_r}$ can be re-written as matrices $\hat{\mathbf{A}} \in \mathbb{R}^{N \times R_\mathcal{T}}, \hat{\mathbf{B}} \in \mathbb{R}^{N \times R_\mathcal{T}}$ and $\hat{\mathbf{C}} \in \mathbb{R}^{T \times R_\mathcal{T}}$.

This representation help us delineating the factorization we want to implement. In fact, the non-negative decomposition of the tensor will be represented in terms of $\hat{\mathbf{A}}, \hat{\mathbf{B}}, \hat{\mathbf{C}}$ as $[[\hat{\mathbf{A}}, \hat{\mathbf{A}}\hat{\mathbf{B}}, \hat{\mathbf{C}}]]$. The final goal is to approximate the tensor with a r number of rank-1 tensors (or *components*) smaller than the rank of the original tensor - that we will call $r_{original}$. This is equivalent to:

$$\min_{\{\mathbf{A},\mathbf{B},\mathbf{C}\}} ||\mathcal{T} - \mathbf{A}, \mathbf{B}, \mathbf{C}||_F \qquad (3.5)$$

subject to the non-negativity condition $\mathbf{A} > 0, \mathbf{B} > 0, \mathbf{C} > 0$ $\qquad$ (3.6)

where $\mathbf{A}, \mathbf{B}, \mathbf{C}$ indicate the approximate decomposition (instead the ˆ notation indicated the exact decomposition), while the notation $||||_F$ stands for the Frobenius norm defined as: $||M||_F^2 = \sqrt{\Sigma_{ijk} |m_{ijk}|^2}$.

We transformed then the 3-dimensional problem of Equation 3.5 into 2-dimensional sub-problems by unfolding the tensor $\mathcal{T}$ through a process called *matricization*. The matricization is an operation which transforms a tensor into a matrix. In our case, the mode-i matricization consists in linearizing all the indices of the tensor except i. In our case this yields three modes: X_1, X_2, X_3. Each element corresponds to one element of the tensor T , i.e., each mode contains all the values of the original tensor. Thanks to matricization, the factorization problem can be reframed in terms of individual factorizations of the three modes. In other words, minimizing the $||\mathcal{T} - [[\mathbf{A}, \mathbf{B}, \mathbf{C}]]||_F$ is equivalent to minimizing the difference between each of the modes and their respective approximation in terms of $[[\mathbf{A}, \mathbf{B}, \mathbf{C}]]$.

Without entering in the mathematical details (see [55] for more information), Equation 3.5 can be rewritten in terms of the three modes. Since we focus on non-negative factorization (see Equation 3.6), we impose a condition of non-negativity on all the elements of the three modes.

In the case of temporal networks, the so-called *factor matrices* $\mathbf{A}, \mathbf{B}, \mathbf{C}$ give access to different interpretations. $\mathbf{A}$ and $\mathbf{B}$ provide the community structure

of the network, while $\mathbf{A}$ gives the temporal activity of each community. In our case, as exposed in Section 3.1.1, we have undirected networks, so their adjacency matrix represented on each tensor slice is symmetric, and we can solve the problem imposing $\mathbf{A} = \mathbf{B}$ in general (see [73] for further details on this point).

Figure 3.5 (c) shows the final result of the application of the tensor decomposition on an undirected temporal network. Here are represented the factors $\mathbf{A}$ and $\mathbf{C}$; for undirected networks we can solve the problem imposing $\mathbf{A} = \mathbf{B}$. $\mathbf{A}$'s rows correspond to nodes, while $\mathbf{C}$'s ones to discrete-time intervals; each row in the two matrices corresponds to one extracted component.
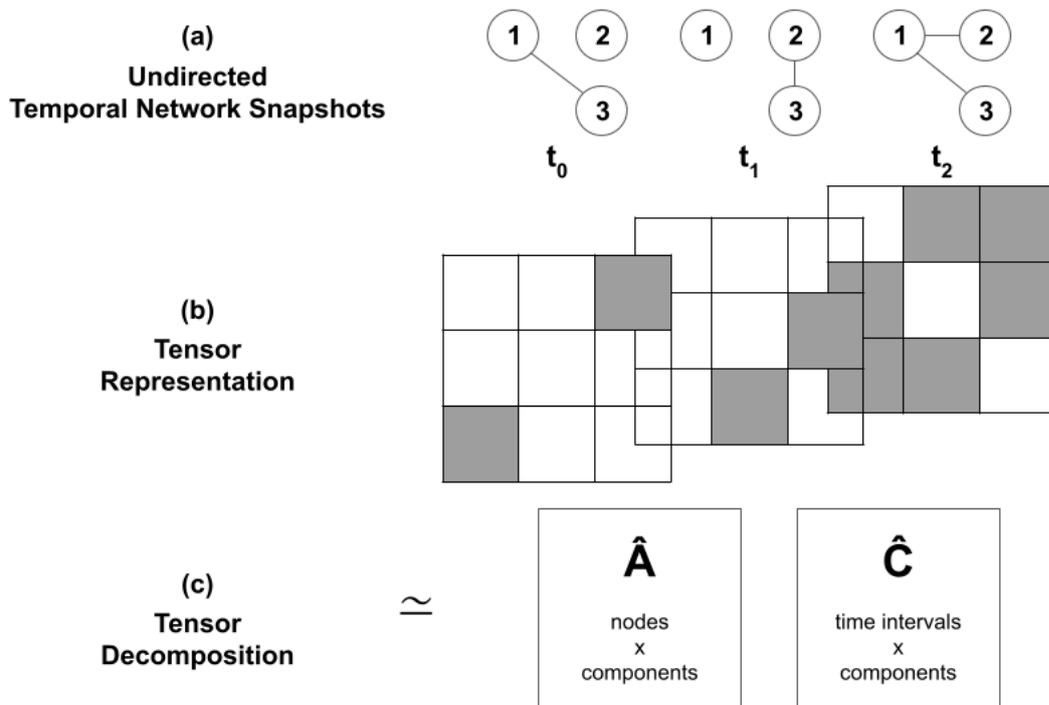
Figure 3.5: Schematic presentation of the tensor decomposition process. From the temporal network snapshots (a) we derive the tensor representation of the network (b) and the applying non-negative tensor decomposition to this object we obtain the factors $\mathbf{A}$ and $\mathbf{C}$ (c).

We used the core consistency metric to find the optimal number of mesoscale structures r [74]. It is based on scrutinizing the *appropriateness* of the structural model based on the data and the estimated parameters of gradually augmented models. A model is appropriate if adding other combinations of the same components does not improve the fit considerably. The core consistency heuristic is used to choose the largest rank according to which the model

is still sufficiently appropriate. In practice, we operate different tensor decompositions for different rank values (ranging from 2 to 20 for all our temporal networks) to estimate the best value for it.

After having identified the rank $r < r_{original}$ through the core component metric and therefore the tensor decomposition has been performed, we have finally obtained the factors mentioned above of our tensor for each temporal network. This technique allowed us to group our events into mesoscale structures. We now have each link at a given time in our temporal network (so each event) corresponding to a 1-rank tensor. We can thus assign a mesoscale structure to each event. The mesoscale structure assigned to each event is the one for which the corresponding link at the corresponding time has the highest value.

Figure 3.6 shows the same 3-dimensional embedded representations as in Figure 3.3 but with colours representing the membership to mesoscale structures detected by the tensor decomposition method applied on the original temporal network. It is interesting to notice how the distribution of colours is not random, but similar colours are somewhat grouped in space. It suggests that our embedding can capture some of these mesoscale structures as well as incorporate the temporal information as shown in Figure 3.3.

3.2   Hyper-parameters selection

In this Section, we will discuss the hyperparameter selection. We can say that the small number of parameters allows us to control the embedding straightforwardly. As mentioned in Section 2.4, having a lot of parameters to adjust for the simulations can be a limitation in terms of understanding and tuning the model. The feature of our embedding method of including a handful of parameters can tackle this lack.

We will treat separately the discussion on the embedding dimension selection and the choice for the parameters that control the contexts.
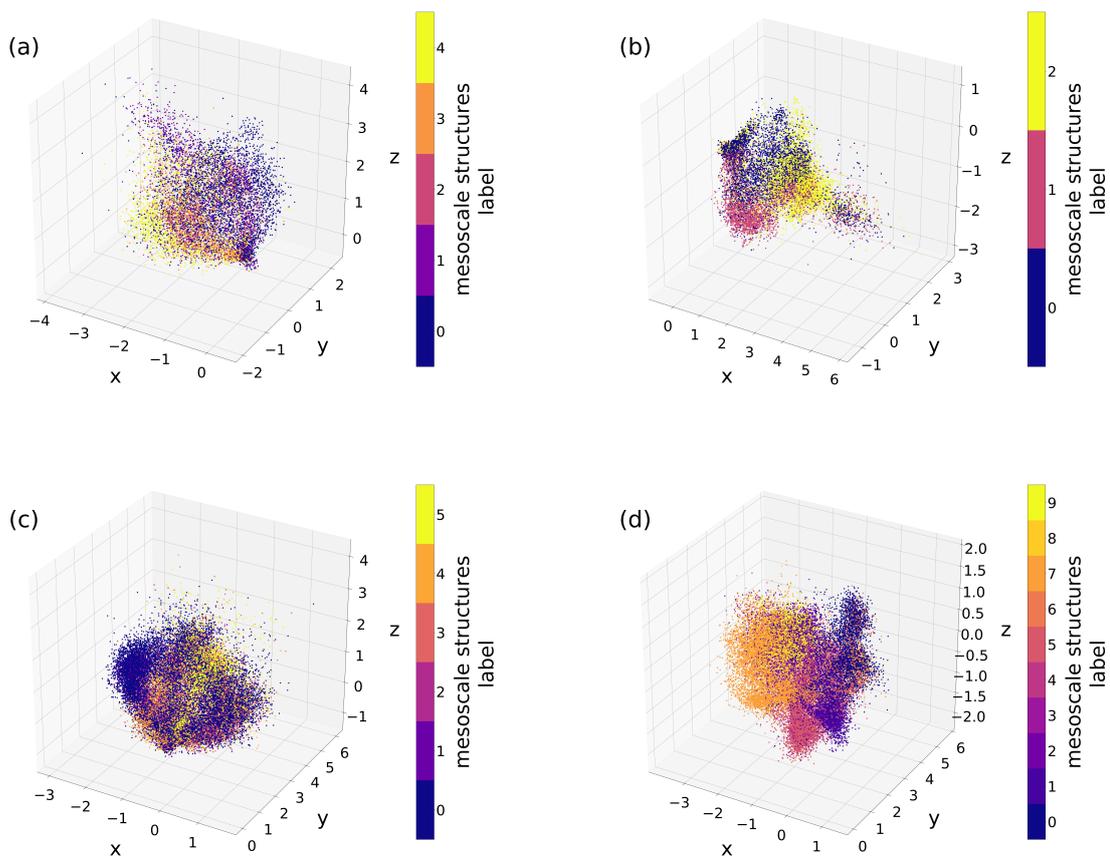
Figure 3.6: 3-dimensional embeddings of the conference (a), hospital (b), high school (c) and primary school (d) networks. x,y and z axes indicate event coordinates, while colour shows the mesoscale structure membership found using a tensor factorization method, respectively set to find 5 (a), 3 (b), 6 (c) and 10 (d) of these structures. hyper-parameters were set to $\alpha = 0.5$, $nb = 10$ and $s = 10$.

3.2.1 Embedding dimension: the selection

One of the most crucial hyperparameters of our method is the dimensionality of the embedding. In choosing the optimal size, we face a significant challenge. How do we manage to keep the size of the embedding low enough to guarantee a compact representation of the original temporal network but simultaneously set it so that the embedding can capture the key features of the network? This trade-off underlies our embedding size selection criterion. If the dimension is lower than the optimal value, relevant latent correlations in the temporal structure may be neglected. If the embedding size is overestimated, on the contrary, we may obtain a highly redundant embedded space.

We test here the consistency and robustness of our embedding technique

in terms of this parameter. We guess that by increasing the number of dimensions, once the embedding reaches and overpasses the optimal number, it should incorporate redundant information. Talking about the Euclidean distance between events, this should not be altered after a certain threshold, so it should stabilize. We checked this assumption using an entropy measure capturing the fluctuations of pairwise Euclidean distances; we tested it on several realizations of embedding with the same dimension.

To measure the entropy over the distributions of the euclidean distances between pairs of event coordinates in each temporal network embedding, we build 10 embeddings for 50 different dimensions (from 2 to 100 at step 2), setting $\alpha = 0.5$. The hyperparameters nb and s were set to 10 and 10 respectively. For each embedding, we then divide our dataset into 10 consecutive samples of 1000 consecutive events each, both to avoid computing all the pairwise distances (which would be very costly computationally) and at the same time to have a representative set of events.

Specifically, we compute the euclidean distance of each pair of events in the samples for each of 10 different but same-dimensionality embeddings. We then bin the distances into $k = 10$ bins ranging from the global maximum and the global minimum values over all the possible dimension and realizations of the embedding for the same network and measure the entropy over these sets of distances as

$$H = \sum_{k} p_k \log p_k, \qquad (3.7)$$

where p_k represents the probability associated with the k_{th} bin. This method thus provided us an indicator of the stability of the embedding. In Figure 3.7 we show our results. The blue curve represents the entropy values with respect to the number of embedding dimensions, averaged on the 10 samples as described above, and the shaded surrounding area shows the variance among the 10 samples. The vertical dash line corresponds to the dimension at which the embedding stabilizes. To determine this point, we looked for the best fit of a horizontal line on the average entropy curve and took the value of the first interception of the curve with its fit.

As we expected, the entropy decreases as we increase the number of dimensions due to the stabilization of the distribution of pairwise Euclidean distances.
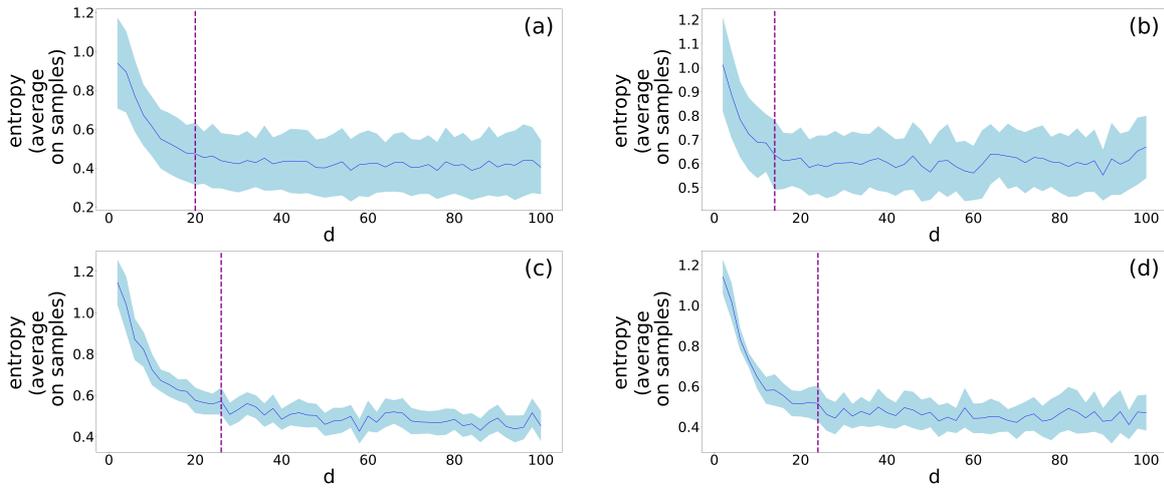
Figure 3.7: Entropy values with respect to d number of embedding dimensions for the conference (a), the hospital (b), the high school (d) and the primary school (b) networks at $\alpha = 0.5$. The dash line represents the value ($d = 20$, $d = 14$, $d = 26$ and $d = 24$ respectively for (a), (b), (c) and (d)) of the optimal embedding dimension in which stability is reached. The blue line and the shaded area represent respectively the average and the variance among the samples we used for the analysis.

It revealed to be a good hint on the optimal size at which the embedding well captures the local properties of the networks.

We show the results of the analyses in Table 3.3.

Network	Optimal Dimension
Conference	d=20
Hospital	d=14
High School	d=26
Primary School	d=24

Table 3.3: Optimal dimension for the four empirical temporal networks, obtained with our entropy method. We set the context parameters s and nb both to 10 and α to 0.5.

3.2.2 Events context parameters: the selection

In this Section, we will explain how hyperparameters of the context sampling may impact the final embedding and the information it incorporates. As mentioned before, the core task of this work on temporal network embedding is

predicting the final outcome of an epidemic spreading on top of the temporal network using its compact representation. We will explain this application in more detail in Section 3.3.1. Here, for what we will treat in this Section, it is sufficient to say that we can associate a simulated epidemic size to each event in the embedding, and we can thus infer this size using event embedding coordinates. For this task, it is understandable that changing the value of the hyperparameters may affect the performance of the prediction. In the following lines, we will test the performance of our embedding method in inferring epidemic size by tuning its hyperparameters.

How do context hyperparameters impact the prediction score on different real networks? Figure 3.8 shows the r^2 scores computed for the empirical temporal networks with respect to the length s and number nb of contexts sampled for each event. For these computations we fixed $\alpha = 0.5$ and the embedding dimensions to their optimal values (see Table 3.3). We can observe that increasing the length of the context has the same effect as increasing the number of contexts on the r^2 score. As we increase them, a plateau of r^2 emerges where the prediction becomes invariant of these parameters beyond statistical fluctuations. Consequently, choosing a large enough value for both of these hyper-parameters would be optimal for the prediction task.

From now on, we will fix the context parameters to $s = 10$ and $nb = 10$ based on the evaluation we outlined above.

At last, we investigated the influence of the embedding dimension and the α sampling balance parameter. As shown in Figure 3.9, both increasing the number of embedding dimensions and α lead to better performances in predicting the spreading outcome. As we can observe, increasing d, each case reaches a plateau. On the other hand, we observed somewhat stronger dependencies on α. While for the conference and the hospital networks, the more one increases α, the better the prediction gets, for the primary school and the high school networks, the score reaches a plateau and becomes less sensitive to the change of α. In general, as we increase the α parameter, the performance improves. When we increase dimensions, after a given value, the improvement is marginal or may even decrease (in case Figure 3.9(b)). In terms of dimensions, the scaling of r^2 initially shows rapid improvements of the prediction task. Still, after a certain number of dimensions, the gain is only
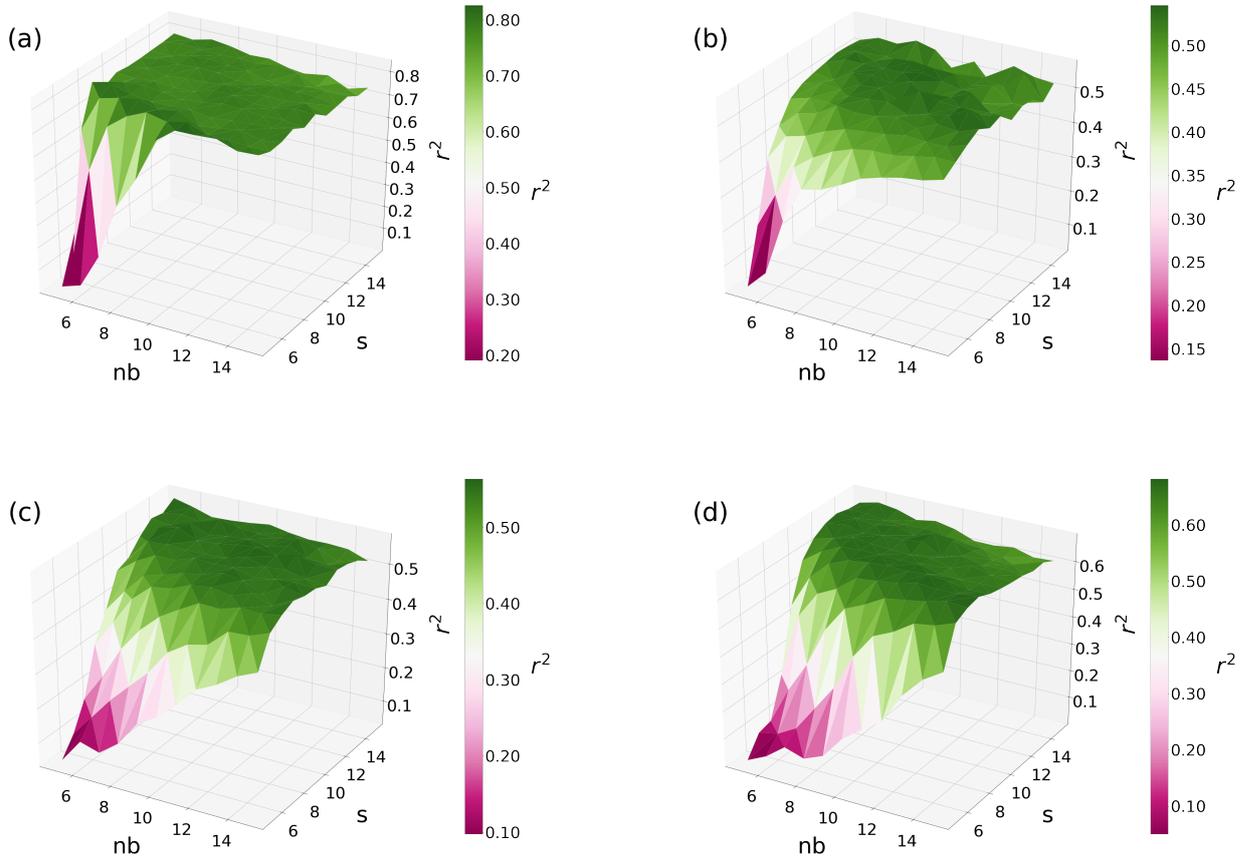
Figure 3.8: R-squared values, r^2, dependency on the nb number (x-axis) and s size (y-axis) of sampled contexts. The results are shown for the conference (a), for the hospital (b), for the high school (c) and for the primary school (d) network. Colors and z-axis code the obtained average r^2 score values for given nb and s parameter pairs computed over 10 realisations. α was fixed to 0.5; we set $d = 20$ for Figure (a), $d = 14$ fo Figure (b), $d = 26$ for Figure (c) and $d = 24$ for Figure (b) - see Figure 3.7.

marginal, indicating an optimal dimension number for training, in agreement with our entropy analysis (see Section 3.2.1).

If we consider lower values of α, the similarity we capture between the event between adjacent events is mainly based on the co-occurrences, which are more relevant in school networks where participants might be active simultaneously (e.g., in breaks between classes). This argument only moderately applies to a conference or hospital where simultaneous interactions typically happen in smaller groups or not at all. Higher values of α imbalance the sampling to contain more information about temporal paths, which indirectly codes co-occurrence frequencies. It gives the advantage to the model to learn both types of similarities and predict the epidemic outcomes with higher precision.
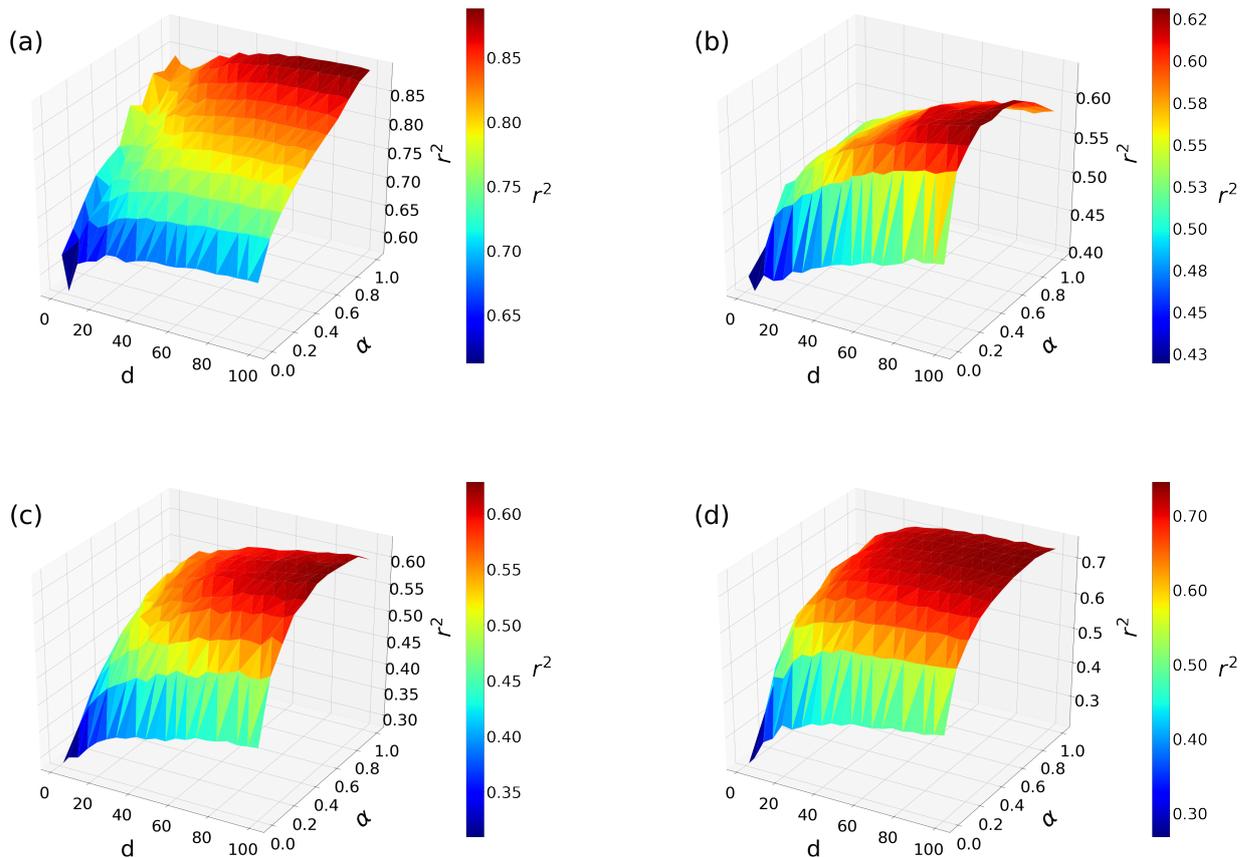
Figure 3.9: R-squared values, r^2 , as the function of the d number of embedding dimensions (x-axis) and α sampling balance (y-axis) parameters. The results are shown for the conference (a), for the hospital (b), for the high school (c) and for the primary school (d) network. Colors and z-axis code the obtained average r^2 score values for a given d and α parameter pairs computed over 10 realisations. Other hyper-parameters were fixed to $nb = 10$ and $s = 10$.

From now on, to have a balance in capturing both temporal and structural information through the embedding, we will fix $\alpha = 0.5$, based on our evaluation made above.

3.3   An application: predicting epidemic outcomes

In this Section, we will discuss the core application of our temporal embedding method, which consists of studying spreading processes occurring on top of the network. It is crucial to highlight how finding the optimal representation for capturing key features of a temporal network is very important for this task. Temporal features of networks (e.g., the discontinuous increase and decrease

of interactions activity, called the *burstiness* of interactions [75]) influence the dynamics of network processes [76]. On the other side, the temporal structure of the network has an impact on the spreading process on top of it [77, 76]. Studying spreading phenomena on a temporal network using embedding can help shed light on such a complex and interesting diffusion phenomenon and give us a new perspective on which structural components and temporal characteristics play an important role in epidemic processes on networks.

In the following Sections, we will have an overview of epidemic spreading processes, with particular attention to the compartmental models. Finally, we will discuss the results of applying our embedding technique to the study of diffusion phenomena. We will look at these results to help us better understand how effective embedding is in the representation of our network.

3.3.1 Epidemic spreading

The Greek physician Hippocrates, the father of medicine, is the first person known to have examined the relationships between disease and environmental influences. To him goes the credit for distinguishing between diseases that interested a population (epidemic) from those circumscribed to a population (endemic). It undoubtedly also earned him the title of the first epidemiologist in history. Since then, epidemiology studies have always been at the centre of interest in human studies. To cite the most famous cases, we pass from Fracastoro's "De contagione et contagiosis morbis" in 1543 to the well-known map by John Snow showing the clusters of cholera cases in the London 1854 epidemic - one of the first data visualization cases in history.

In the contemporary era, thanks to the significant steps forward made by medicine, genetics, molecular biology, epidemiological studies have also been enriched and have broadened their horizon. They became the cornerstone of public health, helping to form political decisions and evidence-based practice for preventive health care. Modern studies use advanced statistics and machine learning techniques to create predictive models as well as to define treatment effects [78].

In this context, it is worth mentioning the importance and impact that compartmental models have had in the mathematical modelling of infectious

diseases. They are a model that simplifies the analysis of epidemic processes by
operating within compartmentalization, hence their name - of the population
subject to the epidemic. Basically, by labelling the population with compart-
ments (such as Susceptible or Infectious), the compartmental model describes
the epidemic spreading from one compartment to another.

The conception of compartmental models occurs at the beginning of the
twentieth century, with contributions from various scholars such as Kermack
and McKendrick [79] and Kendall [80]. These models served as a starting
point to multiple studies of epidemic models.

To introduce our discussion, we will now see a specific and straightforward
case of a compartmental model, the SI model, used as an epidemiological
model for our analyses.

3.3.1.i SI: Susceptible-Infectious spreading process

The simplest epidemic models are based on the assumption that we can divide
the population into compartments, each representing a phase of the disease [81,
82, 83, 84, 85]. The one we used for our analysis, the Susceptible-Infected(SI)
model, which foresees that once a healthy node (belonging to compartment
S) is exposed to the infection, it will become infected (compartment I) with
a given rate β and will never return to the original healthy state. In terms of
temporal network structures, the infection can spread from a selected infected
node via temporal interactions and can reach all other nodes via connected valid
temporal paths. The dynamics of I in a SI model are also known as logistic
growth. If there are no vital processes (birth and death), every susceptible will
eventually become infected. The SI model can be written in terms of ordinary
differential equations:

$$\frac{dS}{dt} = -\frac{\beta SI}{N}, \tag{3.8}$$

$$\frac{dI}{dt} = \frac{\beta SI}{N} \tag{3.9}$$

where $N = S + I$ is the total population. Formulating these equations we
made some assumptions. Given that the total population N does not change,
thus $\frac{dS}{dt} + \frac{dI}{dt} = 0$. Starting then with I infected individuals, each of them

will infect $\beta\frac{S}{N}$ susceptible individuals per unit time. So the variation of I will depend on the number of infected I, on the fraction of susceptible $\frac{S}{N}$ and finally on the transmission rate β, which basically impacts on the speed of the epidemic. This dynamic is exactly shown in Equation 3.9. Equation 3.8 derives from the assumption that $\frac{dN}{dt} = 0$.

In our specific case, the SI model has been implemented in such a way that $\beta = 1$. It becomes thus a deterministic process from the spreading point of view. Figure 3.10 shows a schematic representation of the SI model in (a), while in (b) is shown three different simulations of an epidemic spreading on a random network of 1000 nodes, with varying numbers of initially infected nodes and different β. We can observe how the epidemic reaches all the nodes of networks, regardless of the initial number of infected nodes or the transmission rate (i.e., the logistic growth mentioned above). The only variation among the simulations is the speed of the epidemic in reaching all the nodes. Both initially I and β impact on the rate: the higher $I(t = 0)$ or β are, the faster all the nodes will be infected.
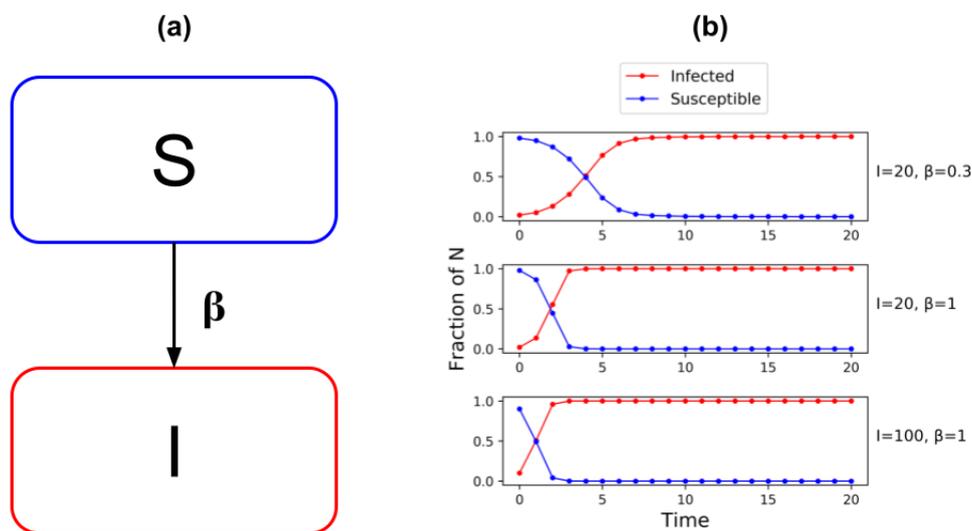
Figure 3.10: Schematic representation of a SI epidemiological model. In Figure (a) is shown the spreading of the simulated epidemic from compartment S to compartment I, according to a β transmission rate. Figure (b) shows different simulations of an epidemic spreading on a random network of 1000 nodes. From top to bottom: a case with 20 initial infected nodes and $\beta = 0.3$, a case with 20 initial infected nodes and $\beta = 1$ and a case with 100 initial infected nodes and $\beta = 1$.

3.3.1.ii Prediction of epidemic outcomes: results for empirical temporal networks and randomized models

As we have shown in Figures 3.3 and 3.6, our model is capable of capturing both the temporal ordering of events and the underlying mesoscale structures of the temporal network. In this Section, we want to go a step further: our embedding technique could provide more information on embedded events. We have seen how our embedding techniques can embed events following a principle of similarity (i.e., considering their neighbourhood as their context). We can see how this impacts the level of the nodes: in this perspective, they are embedded to be part of overlapping time paths. Since the temporal paths are closely linked with the spread of diffusion processes on the network (see Section 3.3.1.i), then we can use the information incorporated by the embedding to predict the final outcome of a simulated epidemic on the network.

We have so far introduced and explained the SI model because it is of our interest to understand how we valued the efficiency of our embedding method in predicting the outcomes of an epidemic. The process is straightforward: taking each event as the starting seed of the epidemic, we simulate the epidemic spreading on the original temporal network. We then assign the final epidemic size to each seed event.

We have thus an epidemic size for each event. We store this information into a dataset containing embedding coordinates of each event and the associated epidemic size that will be the target to be predicted. We also add the square of each coordinate and the euclidean distance from each event in the network and the first event in time. Along with the embedding coordinates, they will be used as regressors. In other terms, we assume a linear relationship between the epidemic size (y) and the embedding coordinates ($\vec{x}$) defined by the so-called *regression equation*

$$y = \beta_0 + \sum_{i=1}^{r} \beta_i x_i \qquad (3.10)$$

where $x_1, ..., x_r$ are the predictors (the embedding coordinates, their square, and the euclidean distance) and $\beta_0, ..., \beta_r$ are the regression coefficients. For training, we operate with a 10-fold cross-validation, i.e., we first randomly

partitioned the original sample into 10 equal-sized sub-samples and retained a single sub-sample as the validation data for testing the model while using the remaining 9 sub-samples for the actual training. We repeat this process 10 times to train the embedding best to learn the coordinates of each event in the network. We computed the r^2 scores (coefficient of determination) between the predicted and simulated epidemic sizes as the goodness of the prediction. It helps to understand which amount of variation in y can be explained by the dependence on $\vec{x}$ using the particular regression model.

Larger r^2 indicates a better fit and means that the model can better explain the output variation with different inputs.

Note that our aim to investigate the final outcome of an epidemic differs from the one pursued with DyANE [56]. In our case, we are not interested in the node's status by time but in the final outcome of the epidemic originated by a specific event.

We know that several temporal and structural correlations interweave real temporal networks. They have complex and various impacts on the spreading processes that occur on top of them. For example, *local temporal correlations* which emerge on same-link events may induce a bursty behaviour. In contrast, *higher-order temporal correlations* may lead to temporal motifs in the network. Moreover, *structural correlations* and *weight-structural correlations* are responsible for any non-random connection pattern in networks (as communities, or non-random distribution of strong and weak ties).

To shed some light on the effects of these correlations on the spreading process, we used three types of randomized reference models (RRM) [86]. When we eliminate combinations of temporal and structural correlations, we can identify which are determining for the prediction task. The first RRM is the *snapshot shuffling*, which randomizes the timestamps of events: the aim here is to eliminate any temporal correlation between events. The second RRM we used is the *timeline shuffling*. We took the timeline of events of a specific link in the temporal network, and we switched it with the timeline of another randomly selected link. Doing this, all the correlations between the underlying structure and timelines are eliminated. The last RRM is the *link shuffling* method. We took the static aggregated network underlying the temporal one to randomize its links, and then we reassign the original timelines of events to

the new links randomly. In this way, this shuffling destroys any structural and structural-temporal correlations in the network.

To further understand the effects of the different RRMs on the epidemic outcome, for each model, we generated five different randomized network realizations and simulated the spreading process starting from each event in the networks to obtain the distribution of final epidemic sizes.

We summarized the prediction results for the original and the RRM networks in Table 3.4, where we depict the observed average r^2 values with their standard deviation computed over the embedding realizations. We fixed the context parameters s and nb both to 10 and chose the optimal embedding dimension for each real network detected as we explained in Section 3.2.1.

r^2 / Data	Original	Snapshot	Timeline	Link
Conference (d=20)	0.79±0.01	0.53 ± 0.04	0.66 ± 0.03	0.57 ± 0.01
Hospital (d=14)	0.53±0.03	0.11 ± 0.02	0.35 ± 0.06	0.50 ± 0.04
High School (d=26)	0.56±0.02	0.23 ± 0.01	0.53 ± 0.02	0.76 ± 0.04
Primary School (d=24)	0.68±0.02	0.12 ± 0.01	0.31 ± 0.02	0.55 ± 0.02

Table 3.4: R-squared values, r^2, obtained by comparing the simulated and predicted epidemics outcomes using embedding of the real empirical temporal networks and of the randomised model. We set the context parameters s and nb both to 10. The optimal embedding dimension were chosen as found in Section 3.2.1.

We can observe that these results show, in general, worse performance in predicting the final epidemic size of the randomized model embeddings with respect to the original network embeddings. In some cases, though, the performance of the RRMs is slightly better. We are going to get a better sense of this in the following.

On the one hand, since some correlations have been eliminated from RRMs, which might be determinant for the prediction task, we observe a decrease of the r^2 score. It is the case of the snapshot shuffling method, which consistently leads to a significant drop in performance. It suggests that temporal correlations (local or higher-order) are very important for the spreading process: it means that our embedding can capture these dependencies successfully. Timeline shuffling seems to perform better than the snapshot shuffling method but

worse than the original network. It suggests that while the embedding can capture structural correlations, local temporal correlations might be better predictors than weight-structural correlations.

On the other hand, RRMs also appear with a less complex structure, which may help the prediction. The link shuffling method is an example: it performs the best among the RRMs, sometimes even better than the original dataset. We can observe that the local temporal dynamic is the most important feature of the temporal network. Still, by removing structural correlations, the system becomes homogeneous and easier to predict.

Another explanation for the fluctuations in the results observed with computing the analysis on different settings is the diversity among the various epidemic size distributions. Here we look at these distributions measured for simulated epidemic processes seeded from every event in the empirical networks. As shown in Figure 3.11 for each dataset, we find that these distributions vary in different ways for different networks. While they are concentrated for the (a) conference and (c) high school datasets, they are more homogeneously distributed for the (b) hospital and (d) primary school data.

According to the RRMs, we present below a showcase for the different methods. Figure 3.12 and Figure 3.13 show the epidemic size distributions respectively for the snapshot shuffling RRM and for the timeline shuffling RRM. As we saw in Table 3.4, these models performed worse in predicting the final outcome of the epidemic than the original. However, they did not alter the final outcomes of the spreading processes considerably as the obtained distributions are remarkably similar to the ones measured on the empirical network (see Figure 3.11). This finding means that the results obtained in the prediction task are independent of the shape of the epidemic size distribution for these two cases. Not only that, they confirm that both the local temporal correlations and the structural correlations that are missing in the networks generated by these models are determining factors in epidemic processes (the first have more impact on them than the second one). They, therefore, cause a worse performance in predicting its final outcome.

In contrast, for the link shuffling RRM, the distributions become very narrow around relatively large values. This particular shape of the distributions, in this case, explains why this method performs the best, sometimes even better than

the original network, in the prediction task. Predicting a narrow outcome
of a process is a considerably easier task than predicting a process with an
outcome of high variance. In this case, peaked distributions correlate with
higher predictive performance.
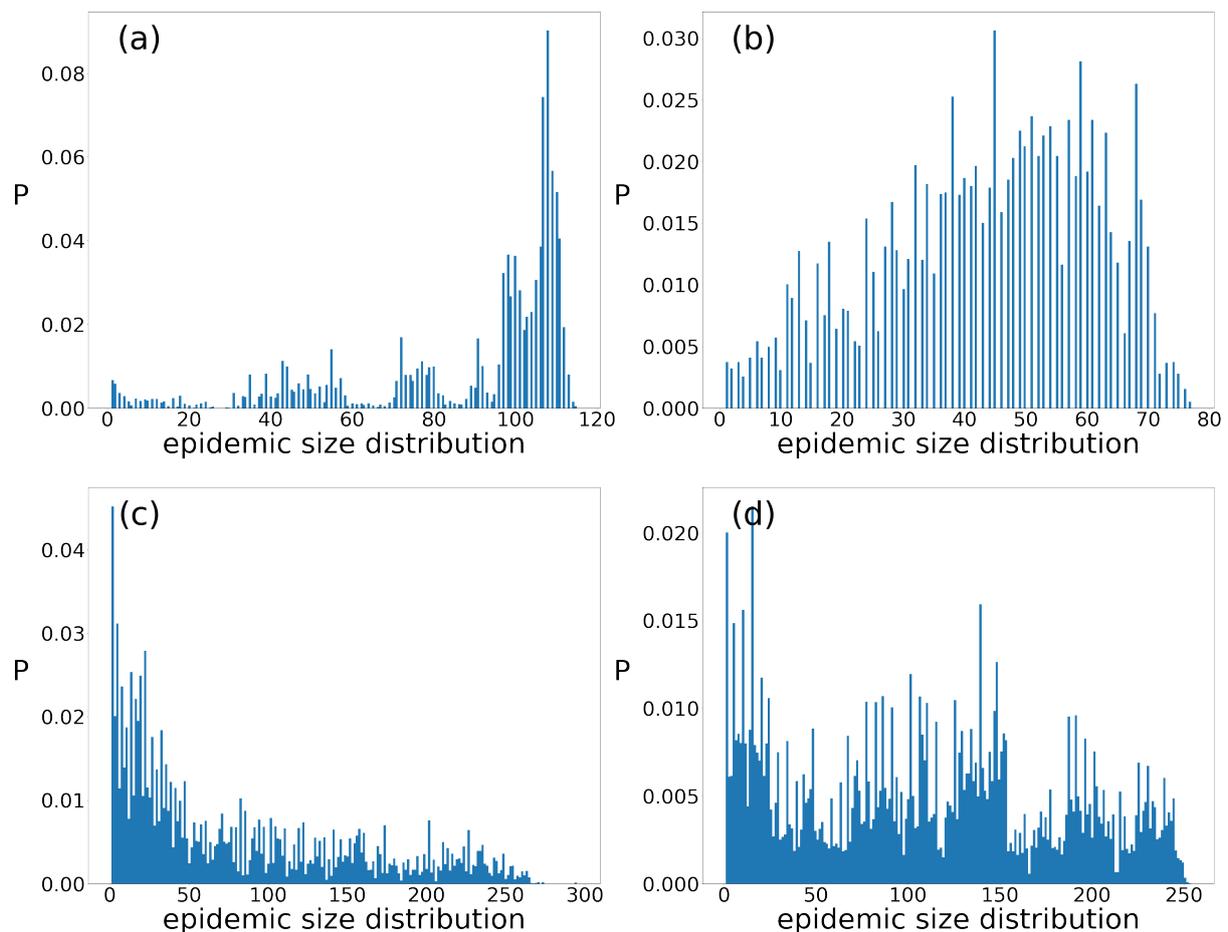
Figure 3.11: Epidemic size distribution of the (a) conference, (b) hospital,
(c) high school, and (d) primary school original temporal networks.

As a general conclusion, we demonstrate that the embedding successfully
captures both temporal and structural network features. The fact that tempo-
ral and structural features can be entangled impacts embedding performance
in the predictive task but not on what the embedding can learn. If we had a
network with many communities but whose nodes have non-correlated activity,
our model could underperform in a prediction task. However, it can provide
precise predictions if structural and temporal correlations code redundant in-
formation.
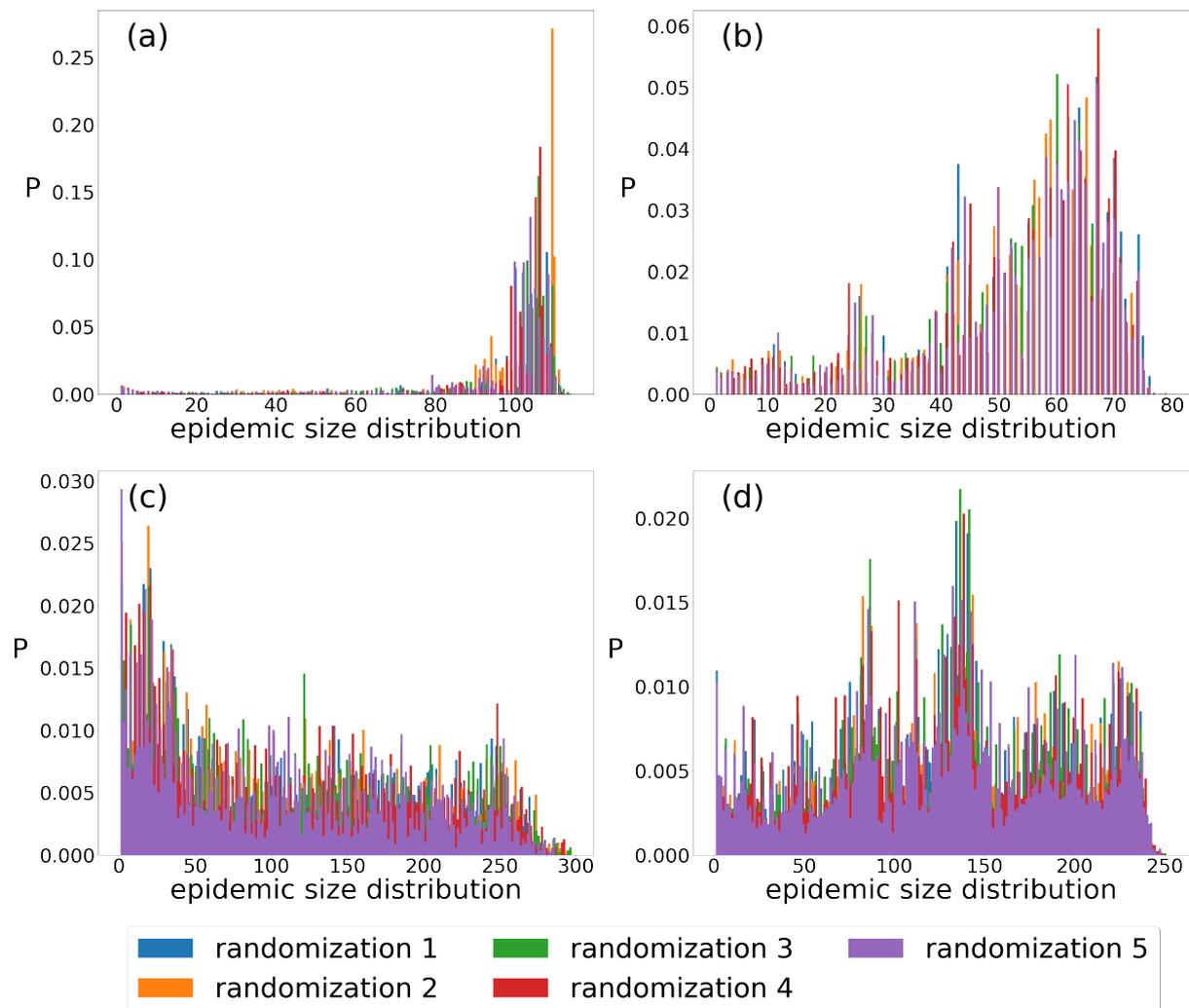
Figure 3.12: Epidemic size distributions on *snapshot shuffled* RRM networks for (a) conference, (b) hospital, (c) high school, and (d) primary school networks. Colors assign different random realization of the actual network model.

3.3.1.iii   Comparison with other methods

In literature, there are a few other recently proposed temporal network embedding methods, as we discuss in Section 2.4. We consider here two of the most promising ones, the STWalk [51, 87], and the Online-Node2vec embedding methods [53]: we will dedicate this Section to compare their predictive performances to our *weg2vec*.

Before looking at the results of this comparison, we will introduce the Reader to the main features of the two methods, which also allows us to understand how these impact their predictive performances.

Both methods are thought to build node embeddings for dynamic graphs using the Skip-Gram model, which introduces a significant difference to our
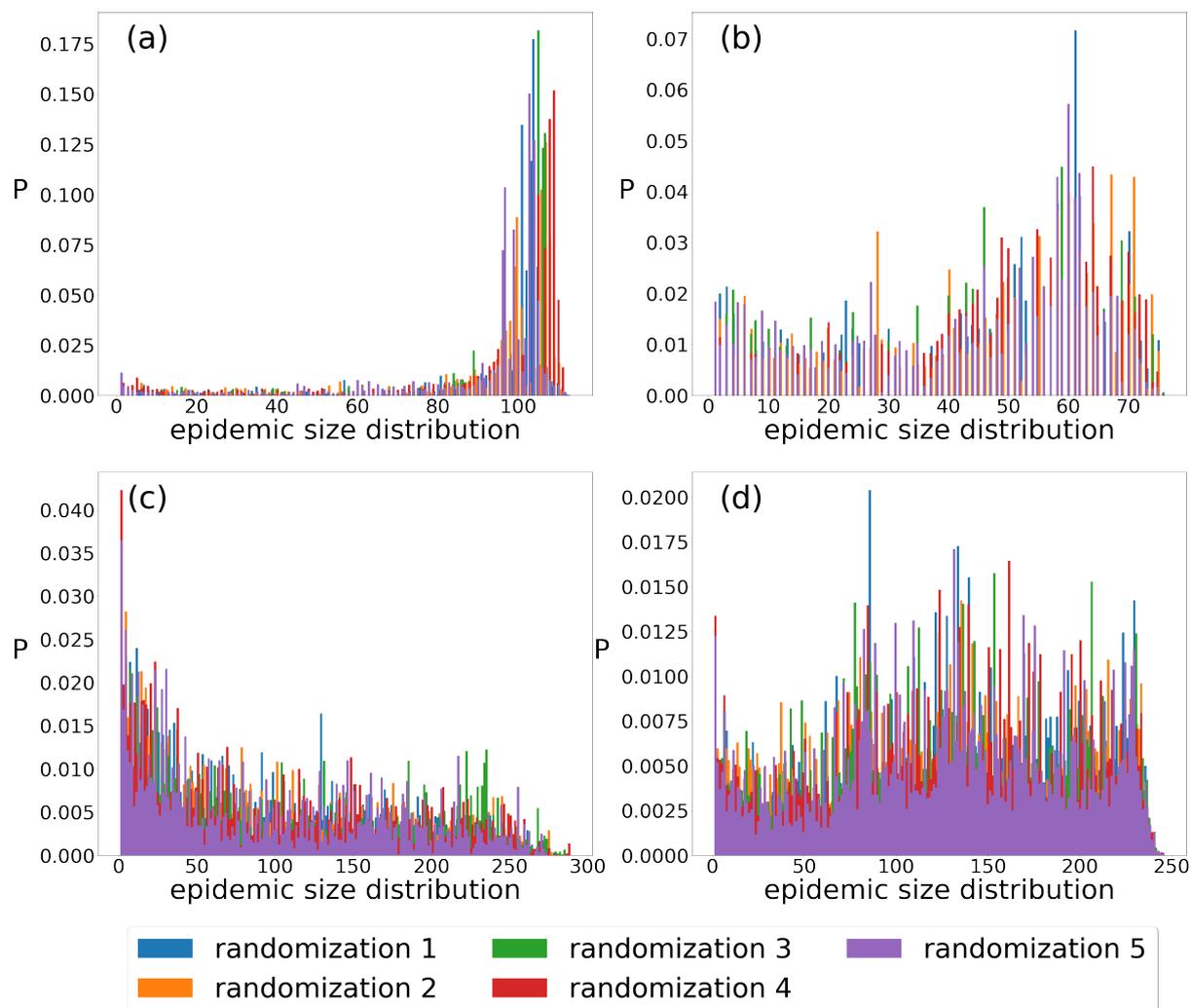
Figure 3.13: Epidemic size distributions on *timeline shuffled* RRM networks
for (a) conference, (b) hospital, (c) high school, and (d) primary school net-
works. Colors assign different random realization of the actual network model.

event embedding method.

In particular, STWalk is designed to learn trajectory representations of
nodes in temporal graphs by operating with two graph representations, a graph
at a given time step and a graph from past time steps. It performs random
walks respectively called *space-walk* and *time-walk*, to sample contexts to input
for the Skip-Gram embedding. The authors propose two variants of STWalk,
different in the way the context is built. In STWalk1, *space-walk* and *time-
walk* are performed as part of a single step on a combined graph, while in
STWalk2, *space-walk* and *time-walk* are done separately.

The second method, Online-Node2vec, is a node embedding method up-
dating coordinates each time a new event appears in a temporal network. It
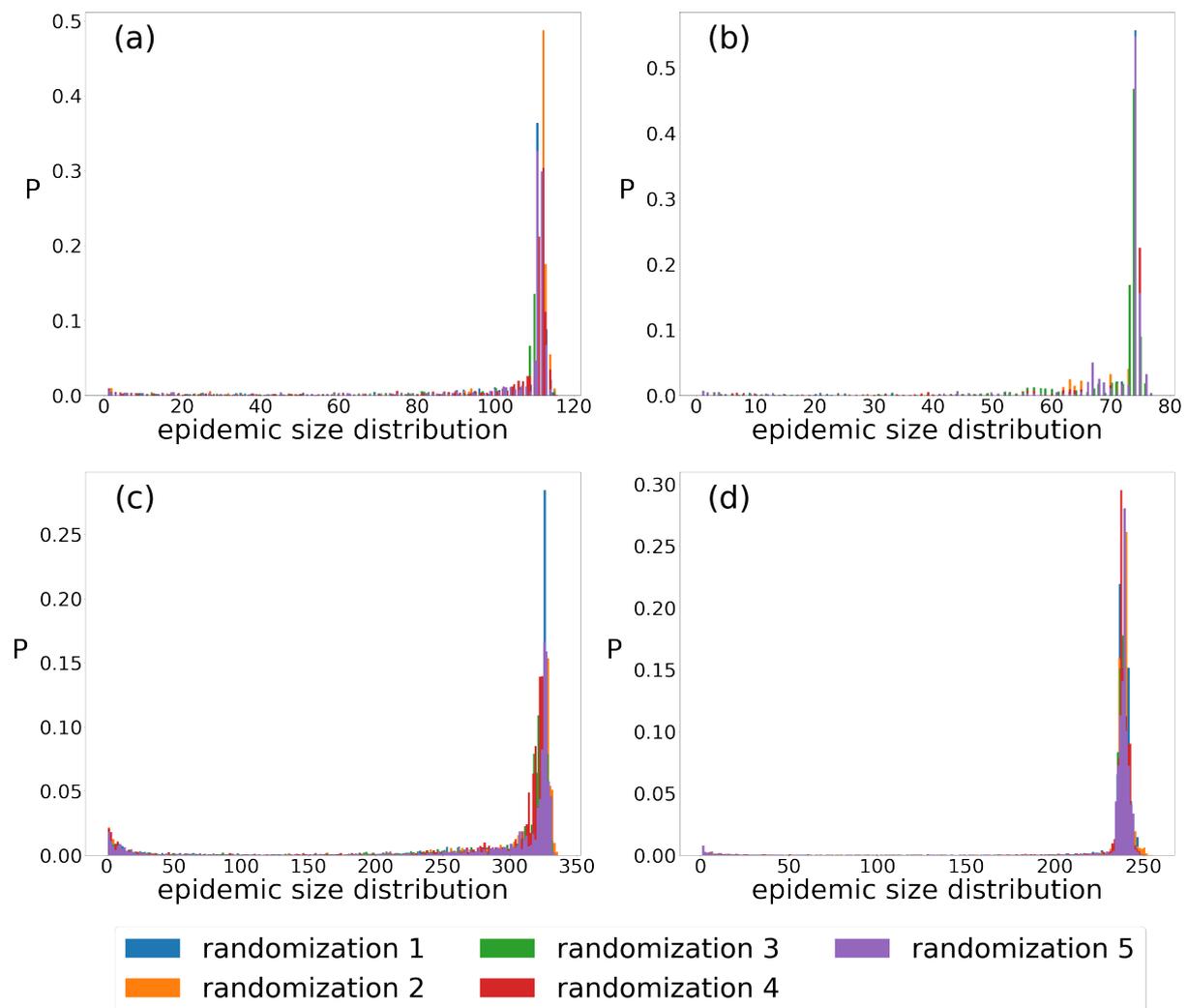
Figure 3.14: Epidemic size distributions on *link shuffled* RRM networks for (a) conference, (b) hospital, (c) high school, and (d) primary school networks. Colors assign different random realization of the actual network model.

also applies random walks to generate contexts, possibly using two strategies, the *Temporal Walk* algorithm, and the *Temporal Neighbourhood* algorithm. In the *Temporal Walk* algorithm [88] a temporal path-based centrality metric is used to capture the similarity between nodes by projecting nodes on the same temporal path close to each other in the embedding. In the *Temporal Neighbourhood* algorithm [89], node similarity is inferred via a fingerprinting method, which projects nodes with similar neighbourhoods close to each other.

To compare the performance of the different methods, we test all of them on our four empirical networks introduced above. The context parameter nb and s have been set to 10 and 10 for all cases to give them the same amount of information to learn and for a fair comparison of outcome. Further, we fix

the balance parameter α to 0.5. We then compute the average r^2 scores of simulated spreading outcomes as we vary the embedding dimensions. Since STWalk and Online-Node2vec use only the past and the present as a basis for the context of the node, we run the simulation for our methods using only the predecessors for each event as well (see Section 3.1.3). Finally, as previously, we estimate the epidemic size by using the coordinates of the actual embedding in a linear regression model (see Section 3.3).
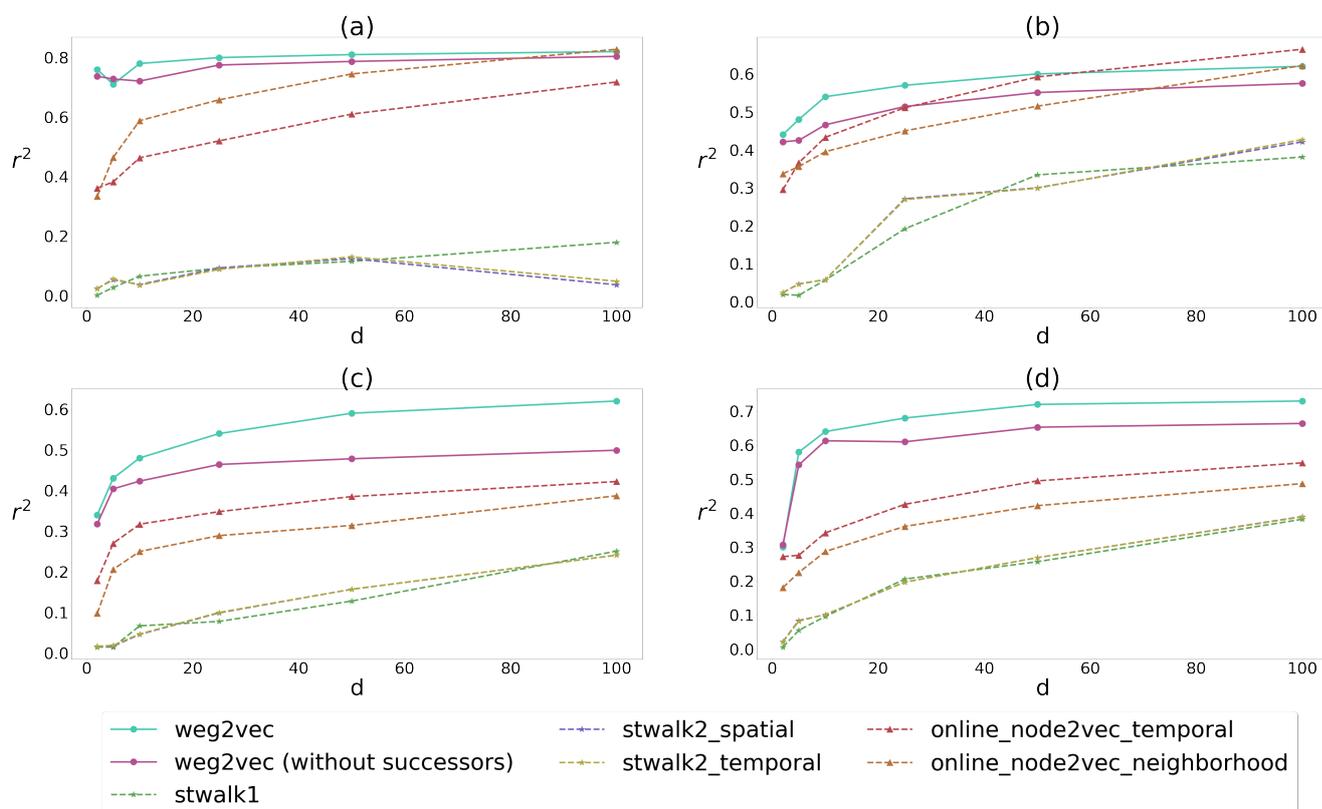
Figure 3.15: Comparison of STWalk, Online-Node2vec and our embedding methods in predicting spreading outcomes on empirical networks in different settings as (a) conference, (b) hospital, (c) high school, and (d) primary school. Results shown are r^2 scored obtained from linear regression on coordinates in embedding spaces with various dimensions computed for each method and empirical temporal networks.

According to the results in Figure 3.15, our method outperforms all the other methods on any of the networks for a broad range of dimensions. The performance improves if we also consider the successors and predecessors in building the context, as expected. The exception is the hospital network, where our method gets slightly lower scores with respect to Online-Node2vec for dimensions 50 or larger. In general, we can explain the difference in the scores

due to the advantage of event embedding instead of node embedding. Indeed, working with events becomes a natural solution if we look at epidemic spreading as a phenomenon mediated by temporal interactions. Specifically, STWalk may get lower scores because this embedding method codes higher-order correlations among nodes: this more complex information may be less relevant or noisy to be learned. The relative worse performance of Online-Node2vec can be because information of the temporal and neighbourhood information is considered separately instead: this can lead to limited information and thus limited prediction capacities.

3.4   Final Remarks

Embedding of networks has recently drawn a lot of attention (see Section 2.4). In fact, it proved to be an efficient tool to resolve tasks such as link prediction or node classification while providing a lower-dimensional representation of networks.

Our work stands as a novelty in a field of network embedding that has not yet been explored, temporal network embedding. The real point of difference from the literature and the strength of our embedding method is that it incorporates events rather than nodes. This peculiarity has proved successful in the analysis of epidemic processes and related prediction tasks.

The simplicity of our embedding method, which relies on the sampling of neighbourhoods on a higher-order static representation of the temporal networks, using the Skip-Gram model and tuning through a handful of hyperparameters, makes it easy to handle. We have shown that our event embedding is particularly efficient to provide compact representations of a temporal network, capturing its essential features such as its time ordering and its underlying mesoscale structures.

Along with this, we show that the embedded representations code the essential information of the original network to get good performances in predicting the outcome of the spreading process. This ability has also been tested against other methods and has always been confirmed.

We have also created a method to establish the optimal size of the embedding, which has allowed us to obtain a compact representation of the temporal

network without affecting the quality of the information encoded by the embedding.

Future works would be worth exploring other sampling strategies that decouple the purely structural properties, i.e., the presence of the communities in the aggregated network, from the temporal properties. Another important follow-up of this work would be applying this embedding technique to solve questions such as detecting critical events in misinformation spreading.

4

Anatomy of a crowdfunding platform

As we discussed and illustrated in the previous Chapters, in treating high-resolution human behavioural data, a good representation of the system is fundamental. So far, we have dealt with social contacts temporal networks. In this Chapter, we will introduce a different source of data. We will turn to a social interactions system that describes a community interacting with the rest of the world, its economy, and its social structure. We scale to a system that spans years of interactions, where the individual actions are resolved in terms of fine time granularity and many descriptive metadata. With respect to the temporal social networks, we treated before, in which we dealt with a limited number of nodes in a temporal interval relatively contained, here we approach data that covers interactions of people all around the world and in almost two decades.

We will discuss the analysis of transactions on an online crowdfunding platform, Kiva. As a global online crowdfunding platform, this platform hosted transactions from users and lenders around the world for nearly two decades. We chose to analyze Kiva because the interactions on it can be phrased naturally in terms of temporal networks.

Kiva records with great finesse a rich series of actions performed on the relative platform. In fact, there have been loan requests from over 1 billion

users for a time period of almost 15 years. To make sense of such an amount of information, a representation of the data that uses some reduction in dimensionality is crucial. A methodological choice of dimensionality reduction and the representation of interactions via temporal networks prove opportune in exploring this particular type of social interaction.

Crowdfunding may offer insights and a good starting point of discussion about social interactions and human behaviour issues [90, 91, 92], since it pertains to how people build relationships and influence each other for a higher objective as solidarity.

In fact, analyzing the actions of each individual who participates in the funding of a project of his interest represents a precious opportunity to study specific behaviours or social activities through the techniques and models of network and data science. Understanding what stimulates people to be part of a project of solidarity towards others, how and to what extent they operate, finds in the study of crowdfunding an invaluable source of information and a way to shed light on some crucial aspects of human behaviour in the constraints of society.

The interest in this particular type of social interaction between individuals originates the line of research that we will introduce in the following chapters. The relationship between entrepreneurs and supporters is an event-type of interaction, different from state-type ties such as friendship [93]. Therefore, it may not repeat itself over time and is essentially based on a bond of trust between the borrower and lender.

In the following, we will show various aspects of this interaction and more general analyses on the crowdfunding phenomenon and its impact on social interactions, whose representation and study are the silver thread of our research Thesis.

4.1   What is crowdfunding?

Crowdfunding is the practice of funding a project or venture by raising small contributions from many individuals, typically via the web; it is a form of alternative finance.

The crowdfunding model generally has three actors involved: the project

initiator(s) who propose(s) the idea or project to be funded (the *borrower(s)*), the supporters of this idea (individuals or groups, the *lenders*), and a moderating organization (the *platform*) that intervenes to mediate the transactions [94]. In the case of Kiva, we will present the fourth type of actor, the *Field Partner*, who stands between the so-called borrower and the platform itself.

Crowdfunding has an ancient history. Interestingly, the concept of crowdfunding is much earlier than the use of the word "crowdfunding" itself, which was established in 2006. We can cite as an example the book. Since books came into circulation, people subsidized them through crowdfunding models. They could be printed and distributed only if sufficient people were interested in their purchase and were thus available to pre-buy the books. Another famous example of crowdfunding in the past centuries was the history of the Statue of Liberty building. Édouard René Lefebvre de Laboulaye, a French professor of law and progressive politician, passionately supported the reasons for the Union in the American Civil War. In 1865 he proposed the idea of a gift that celebrates the brotherhood between the two nations, a memorial for the centenary of the two revolutions (French and American) that immortalize them as symbols of justice and freedom. In 1875 the Franco-American association, promoter of the statue's construction, took an important decision. The statue would be paid for by the French, while the pedestal by the Americans. Immediately, the problem arose of how to pay for everything. The Americans reached their quota after much effort and many complications. Crowdfunding came into play in this context. In 1885, an advertisement in The World newspaper, held by Joseph Pulitzer, managed to fetch sufficient funds for its construction.

As the crowdfunding model matured, more and more companies engaged in this business model, first in the United States of America and later all over the world [95]. The story of Kiva, for instance, begins in late 2005.

Crowdfunding is based on two different models. On one side, there is the rewards crowdfunding: entrepreneurs (or *borrowers*) operate a presale of a product or a service to launch their business. In the other model, the equity crowdfunding one, the supporters (or *lenders*) receive shares of a company in exchange for the funds.

4.2   Anatomy of a crowdfunding platform: Kiva

The online crowdfunding platform that we have chosen to study is that of Kiva, born in 2006. This platform has been increasingly successful over the years worldwide. It contains information on microcredit transitions that can help us shed light on the complexity of the study of social interactions.

4.2.1   History and impact

Kiva is a U.S. nonprofit founded in 2005. It has its base in San Francisco, but it has offices all around the globe.

The idea of Kiva belongs to Matt and Jessica Flannery, its founders, inspired by the Nobel Peace Prize Muhammad Yunus, founder of the Grameen Bank.

A woman in Uganda made the first loan in 2005, and she used it to expand her fishmongering business off the coast of Lake Victoria.

Since then, over 14 years, Kiva has crowdfunded more than 1 billion in loans for entrepreneurs, farmers, students, educators, and more in over 90 countries worldwide. Even the former U.S. President Bill Clinton conferred recognition to Kiva in 2007. In his book "Giving: How Each of Us Can Change the World" [96] he covered Kiva and the work the organization is currently doing and has done in the past. Quoting a passage: "People with a very modest amount of money can make a huge positive impact all around the world."

4.2.2   Academic studies of the Kivas' system

The Kiva phenomenon has also succeeded in research, ranging from socio-economic analyses to methodological and quantitative analyses.

The research topics of which Kiva has been the subject range from management [97] to game theory [98], from natural language processing [99] to social sciences [100, 101]. For example, the platform has been studied as the objective of general examination of the growing phenomenon of crowdfunding as an alternative method for raising finance [102, 103].

A variety of gender studies that evaluated the impact of alternate microcredit systems on the empowerment of women take into account the Kivas' platform as an example [104, 105]. Always related to gender studies, Kiva is

at the centre of the attention of an article in which there was discrimination in donating to female borrowers based on their ethnicity and attractiveness, which was not the case for male borrowers [106].

Even the textual data relating to Kiva have aroused interest in the literature. On the one hand, we tried to understand the impact of messages between lenders to push other lenders to donate [107]. On the other hand, we looked for a proxy in the textual information to understand the reasons of people participate in crowdfunding [99]. Some studies have used the texts have also to understand how grouping loans can impact their fundraising time [108], or how loans textual information can lead lenders to donate [109]

Many articles have discussed the importance and efficiency of the Kivas recommendation system [110, 111, 112]. They range from finding the final recommendation for lenders by studying their behaviour on the platform [113] to more general socio-economic studies to preserve an equitable distribution of capital between different countries in the face of well-known user prejudices [114].

A macro topic that various lines of research have touched upon is the studies of lending teams or groups in which lenders come together based on a communion of ideas and intentions to find loans of common interest to be subsidized. The studies have been challenged general analyses of the lenders network [115, 116], the intercultural contacts among them [117], the impact of team competition on donations [118].

4.3 How Kiva works: the platform and the data

This Section will review the main steps of how the fundraising of a loan occurs on Kiva; next to each step, we will show the data we have available on this process. In Figure 4.1 we show a schematic presentation of the Kiva platform and the process for a loan to be fundraised.

The first step consists of a borrower applying for a loan, and it can happen through two models: partner and direct. Local Field Partners manage the partner loans. Field Partners are local organizations (such as nonprofit organizations, microfinance institutions, schools, and more) working in communities to vet borrowers, provide services and administer loans. Field Partners are lo-
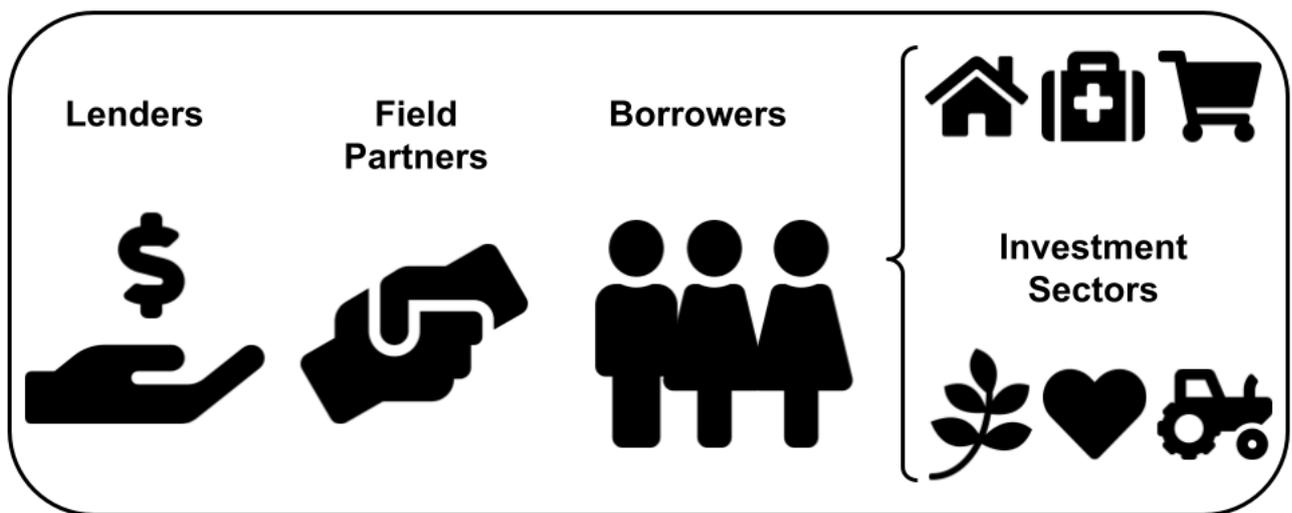
Figure 4.1: How Kiva works. Individual borrowers or groups of borrowers rely on the help of Field Partners to manage the application of a loan into a specific sector of investment. In most cases, the Field Partner prepays the loan, which is then funded by lenders and refunded by the borrower(s).

cal organizations. They act as brokers between borrowers and lenders through Kiva. The flow of activities on Kiva is almost entirely in the hands of the F.P.s.

In contrast, borrowers apply directly through the Kiva website for direct loans. After the loan has gone through the approval process and passed it, it enters the fundraising period — the timing can vary from loan to loan. The loan is thus posted to Kiva for lenders to support. Note that for most Field Partner loans, the money is pre-disbursed. The money is disbursed when the fundraising is complete, and lenders have entirely crowdfunded the loan for the direct ones.

The data available to us reflects this entire procedure. We collected different datasets[1], which contains loans metadata, lenders metadata, Field Partners metadata, and the data about the transactions. Our data range almost 15 years (from late 2006 to early 2020) and includes a high number of worldwide countries. The data collected about 1,4 billion loans managed by just under 500 Field partners and supported by more than 2 billion lenders. Speaking of the loan metadata, for each loan, we have various kinds of information. The time it has been posted and then possibly fundraised, its amount, the lenders who supported it, the sector of investment it belongs to, and the possible Field Partner are only some features of the data.

---

[1]Data are available at: https://www.kiva.org/build/data-snapshots .

Moreover, we also have information about the loan borrower(s), the country of origin, gender, spoken language, and a brief personal description. We have little and limited information about the lenders, such as their country of origin, occupation, motivation in lending, and time of their first subscription to Kiva. Even for the Field Partners, the metadata is not very detailed. We have access to countries where they operate, the number of loans and the relative amounts they managed, and some indicators regarding their activity, such as the rating.

4.3.1 Temporal evolution of the platform

In this Section, we will show some exploratory and preliminary analyses of the datasets at our disposal. It is necessary to understand the impact of Kiva in the world over the years and the volume of information we are analyzing.

The basin of Kivas' users (whether they are borrowers, lenders, or Field Partners) has been enriched year after year. In Figure 4.2 we can see how the numbers of loans and lenders evolved with passing years. We can see that all these numbers increased year after year, a signal of Kivas' platform's efficiency. Another indicator of the growth and diversity of the Kiva user group is undoubtedly the evolution in the number of languages spoken by users. In Figure 4.3 we can see how, over the years, the platform achieved a consistent linguistic diversity from an initial English-language user base.
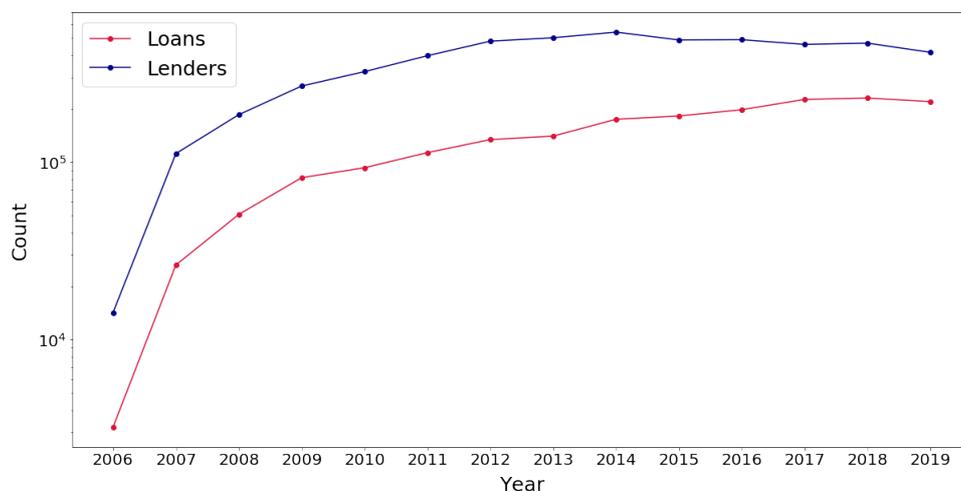
Figure 4.2: The evolution of the number of loans and lenders on Kiva in time. The time interval under consideration is from 2006 until 2020.

In Figures 4.4 and 4.5 are reported the world maps representing the volume
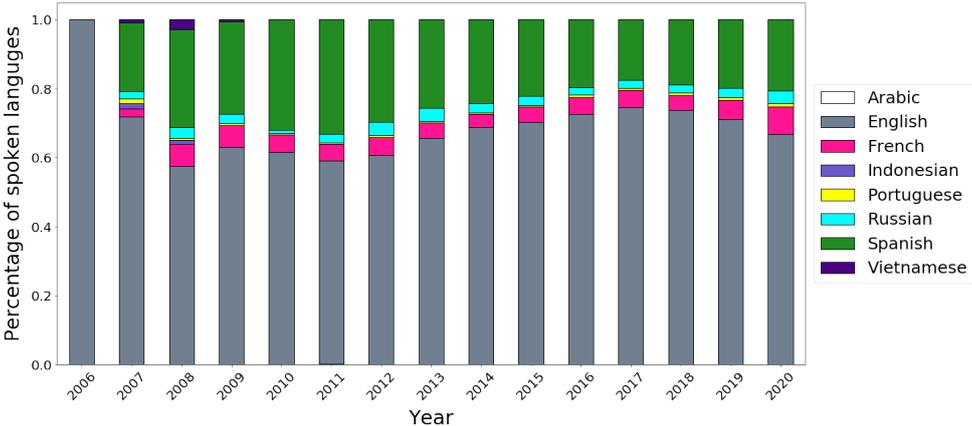
Figure 4.3: The evolution of the number of languages spoken on Kiva in time. The time interval under consideration is from 2006 until 2020.

of loans and lenders in a specific country, respectively. The colour shows the abundance of each of these actors for each country. We compared the 2006 world map with the 2020 world map; we can observe how great has been the growth in the use of the Kivas' platform around the world.

In terms of the gender distribution of the borrowers, we can observe how Women mostly use Kiva. $\tilde{7}0\%$ of the borrowers are women or groups composed of women (see Figure 4.6). Kiva has made women's investments on its platform a flagship, pushing more and more women to empower themselves from an entrepreneurial and financial perspective over the years. In this sense, Kiva stands as a solution to a problem encountered in many countries worldwide, namely the impossibility of accessing banking or financial services [119, 120, 121]. Women make up $\tilde{5}5\%$ of the world's unbanked population, meaning they have no access to banking or insurance products. For many of these, almost 1 billion women globally have no access to financial services; they have thus no access to loans. One of the core reasons why women face this problem is due to the lack of formal identity. In a context like Kiva, women can invest and find a way to finance their businesses without any difficulties, accessing loans otherwise impossible.

We need to consider another aspect of the Kiva platform in terms of user participation and fundraising timelines. How long does it take before the lenders fully subsidize a loan? The fundraising times distribution shown in Figure 4.9 answers this question. We took into account only the funded loans. Considering for each funded loan the time it took to get paid in full compared
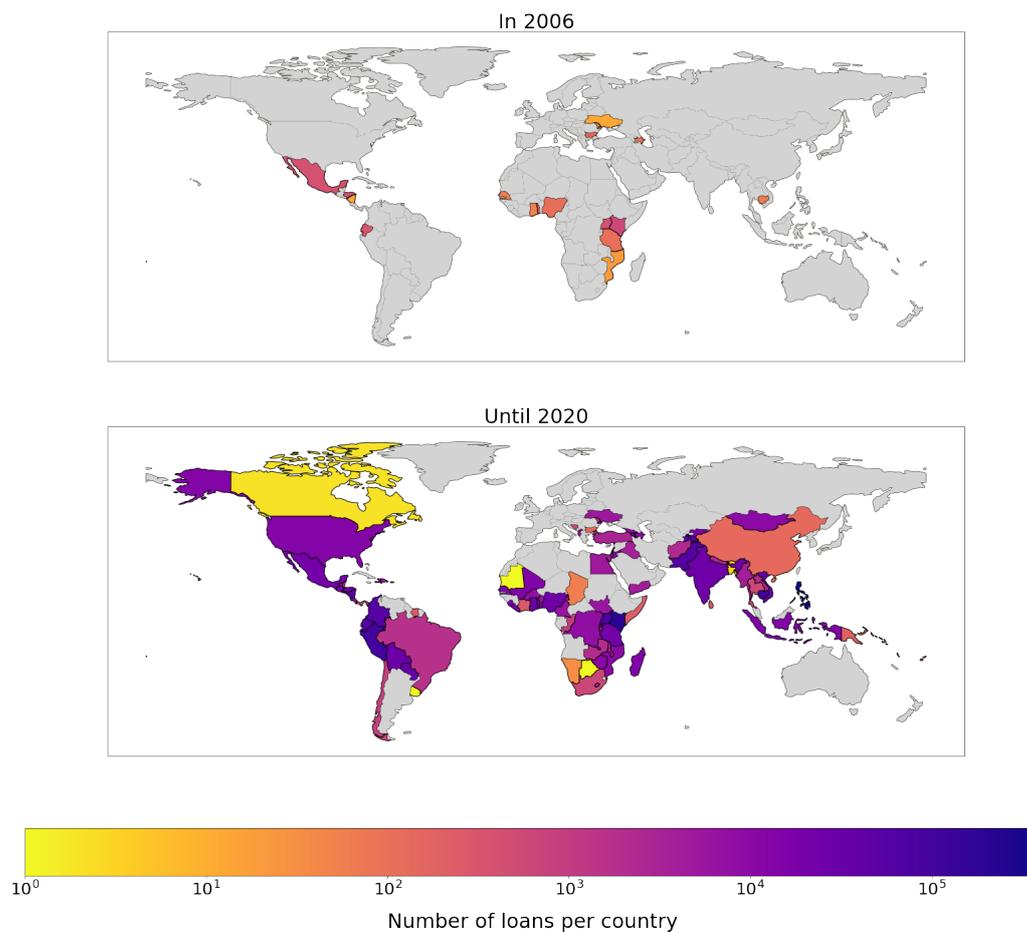
Figure 4.4: Comparison of the number of loans per country on Kivas' platform in 2006 and until 2020. The top Figure map represents the number of borrowers who have asked for a loan in 2006, according to their countries of origin. The bottom Figure map shows the ones who asked for it until 2020, according to their countries of origin. The colour indicates the number of loans each country has asked for in the respective time intervals.

to the total time available (variable), we can see that the fundraising of most of the loans occurs in less than half the time allowed to them. Only for a small part, the fundraising occurs at the last moment. This finding may mean that from a lender recommendation system point of view, Kiva works efficiently.

We want to dedicate the final part of this Chapter to an overview of what we believe to be a fundamental feature of the data on Kiva: the investment sectors. Kiva divides its loans into different sectors, ranging from agriculture to retail to health and education.

The starting point of our analysis on Kiva consists of giving a good representation of the interactions on the platform, looking specifically at the investment activity in the various sectors. In the next Chapter, we will explain the

In 2006
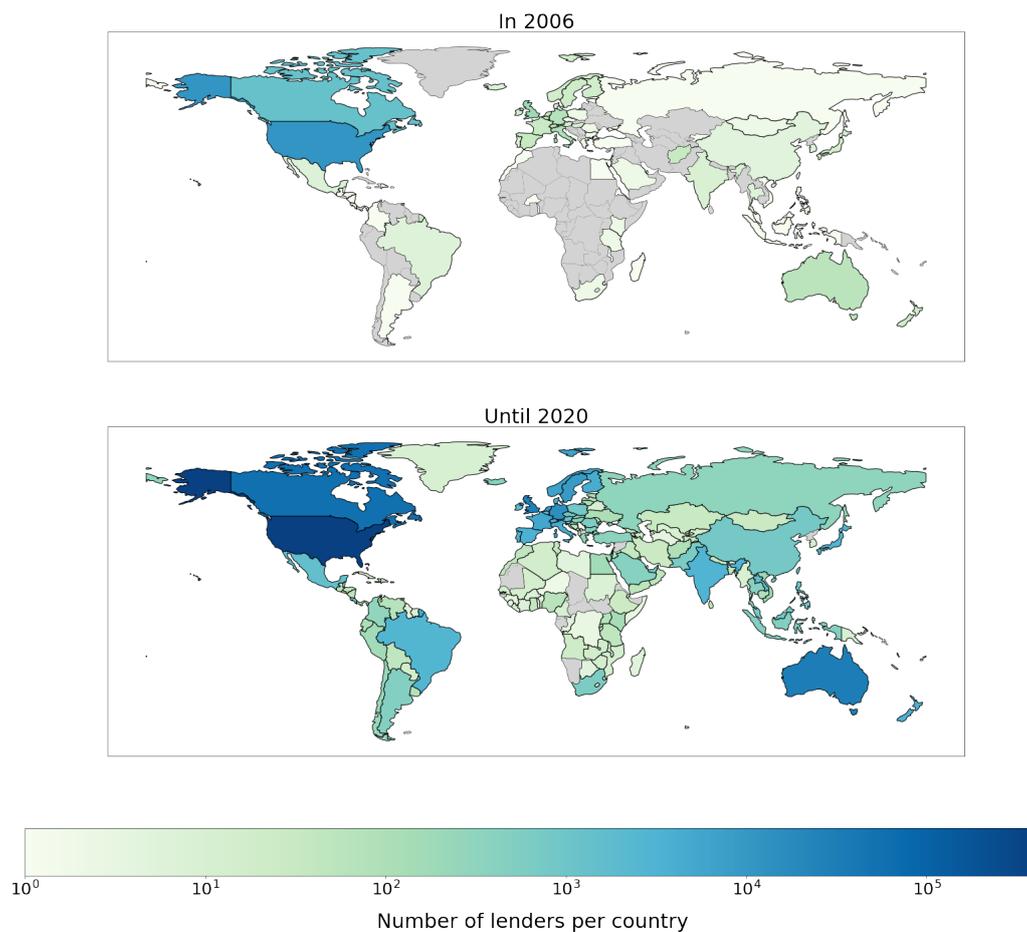
Until 2020

Figure 4.5: The map represents the number of lenders who have funded a loan in 2016 (top Figure) and the ones who fund it until 2020 (bottom Figure), according to their countries of origin. The colour indicates the number of lenders each country has in the respective time intervals.

Figure 4.6: Borrowers gender distribution on Kiva. We show in magenta the percentage of female borrowers (individuals and groups), in cyan the percentage of male borrowers (individuals and groups), and in yellow the percentage of mixed-gender borrowers (groups). The time interval under consideration is from 2006 until 2020.

silver thread of the analyses; in this Section, we want to give some information on the activity in the Kivas' sectors from a more general perspective.
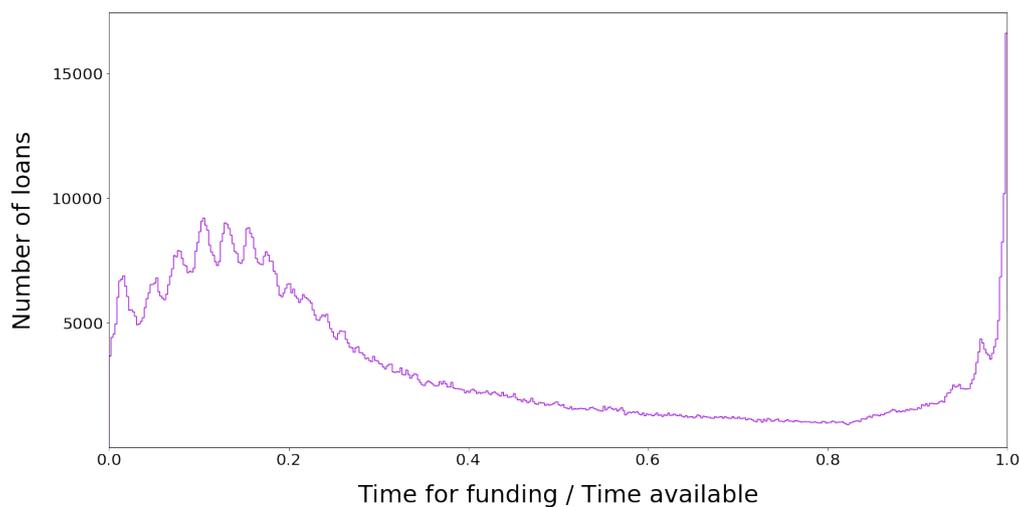
Figure 4.7: Distribution of the times for funding a loan. We considered the time for funding a loan with respect to the total time available for that loan in the analyses. We took into account only the funded loans.
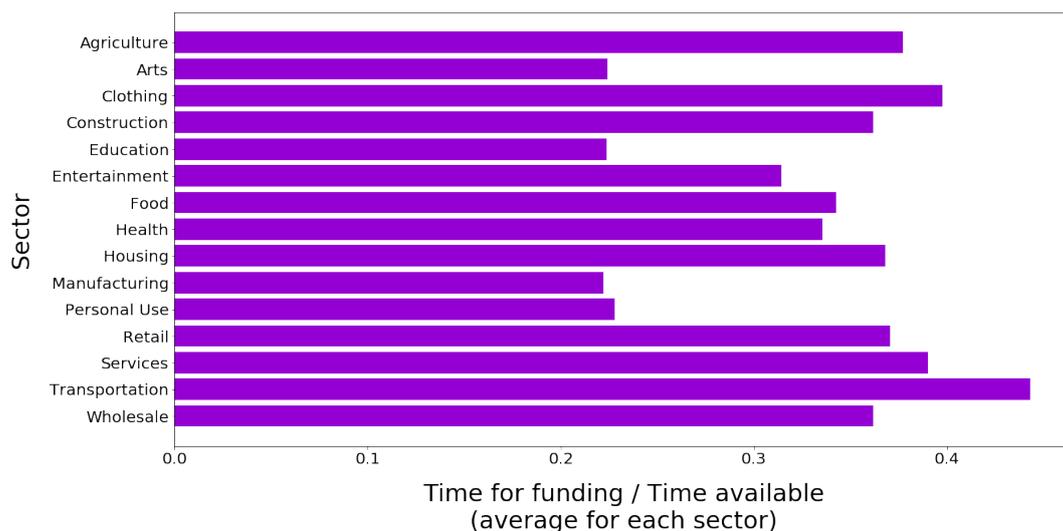
Figure 4.8: Distribution of averages of the times for funding a loan accordingly to its sector. We considered the time for funding a loan with respect to the total time available for that loan in the analyses. We took into account only the funded loans.

Figure 4.10 shows the volume of activity (i.e., the number of loans) in each sector of investment of Kiva. We aggregated the measure over the years. As a result, we can observe how diverse the number of loans is: the agriculture, retail, and food sectors are the most populated. In some way, this result is an expected result: these three investment categories are essential globally, even outside of microcredit platforms and alternative market forms.

One somewhat counterintuitive thing is the evolution of the activity in the
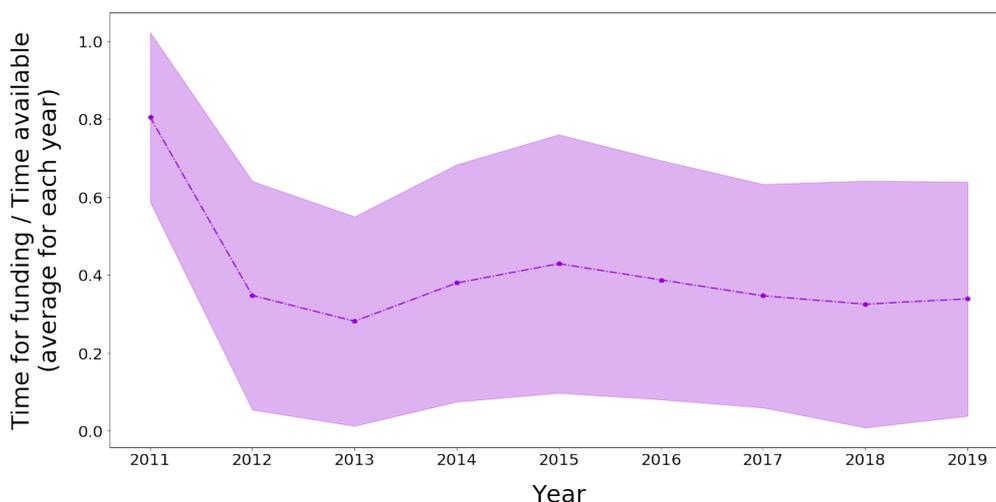
Figure 4.9: Distribution of averages of the times for funding a loan accordingly to the year it has been posted on Kiva. We considered the time for funding a loan with respect to the total time available for that loan in the analyses. We took into account only the funded loans.

sectors over time. Figure 4.11 illustrates a bar plot for each sector. It indicates the percentage over the total of loans operated in that sector from 2006 to 2020. As we can see, some sectors that we could consider a not primary necessity, such as personal use or art, slightly increased in volume over the years.
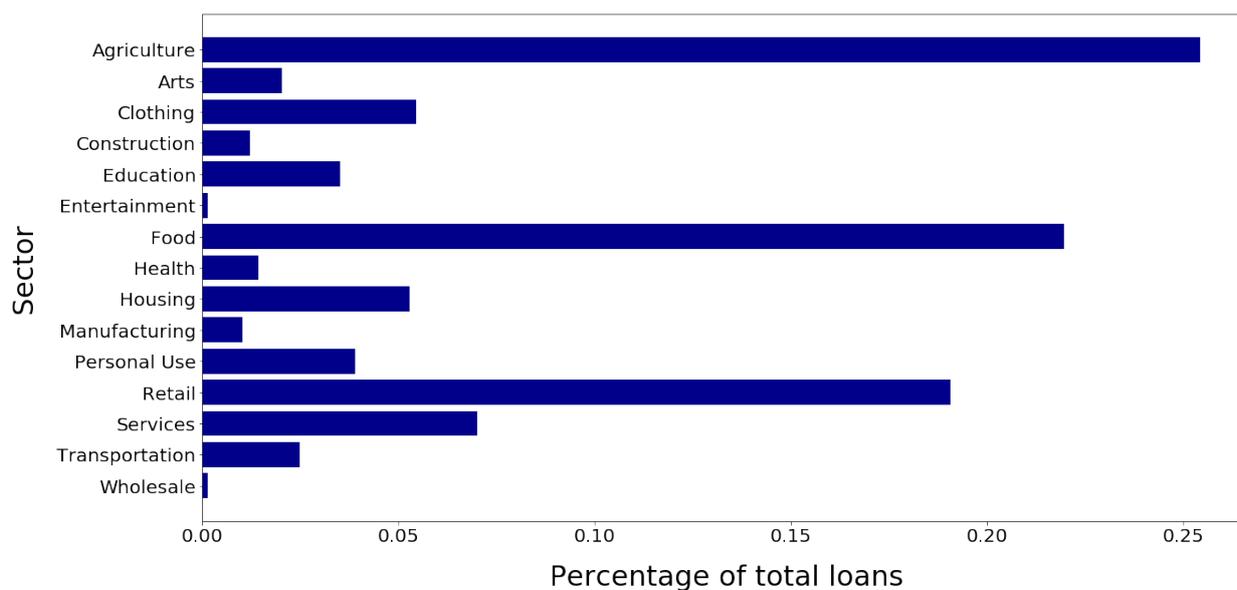
Figure 4.10: The bar plot represents the percentage of loans according to a specific sector, with respect to the total number of loans. The time interval under consideration is from 2006 until 2020.

With this entirely exploratory analysis on Kiva activity in mind, we will work
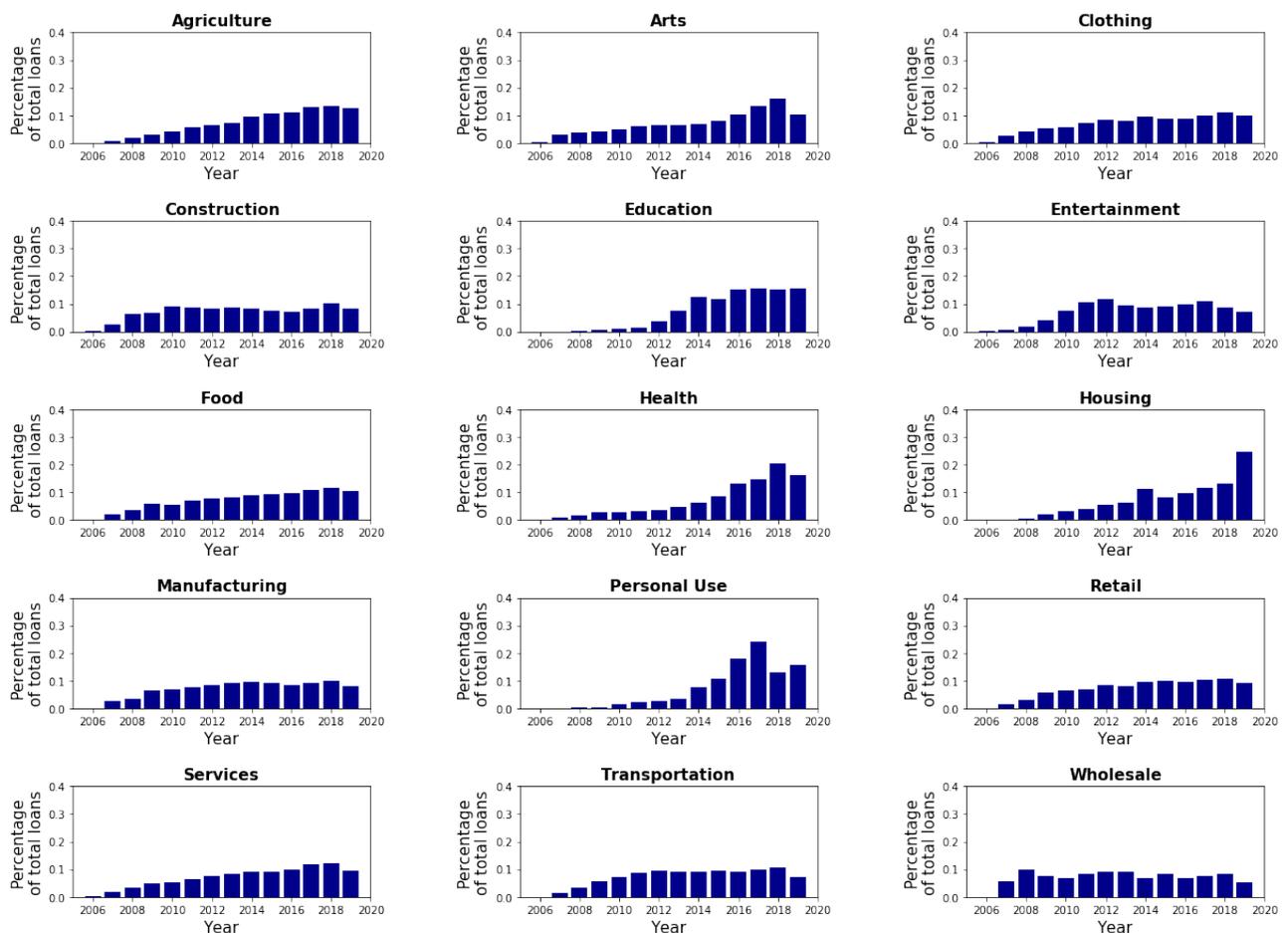
Figure 4.11: The mosaic shows a bar plot for each sector of activity on Kiva. Each bar plot represents how the percentage of total loans relative to a specific sector evolves in time (from 2006 until 2020).

in the next Chapter and a compact but efficient representation of this activity over time to capture the system's key features.

4.3.2 Data limitations: from micro-scale to meso-scale level of analysis

In the light of what we have seen, we can meanwhile observe the following. First, as briefly mentioned in Section 4.3, Kiva has limitations and a lack of detail for some metadata.

Kiva shows some lack of detail in the metadata of the lenders mainly. First, the lenders' dataset presents a few features: a limit for studying their behaviour on the platform. Second, although the loans metadata are detailed and numerous, the most relevant characteristics to be studied are not evenly distributed in the dataset, creating unbalanced classes, an issue for a possible

classification/forecasting task. We can see, for instance, in Figure 4.6 the not uniform distribution of the gender of the borrowers. Another example of unbalanced classes feature is the loan status: almost 95% of the loans labels are "funded". Finally, there is also a substantial imbalance in the platform's disbursement methods: most borrowers rely on a Field Partner ($\tilde{9}9\%$) while the remaining 1% apply directly independently.

However, we can summarize the main limitation for our analyses in the first perspective in which we treated the data. Studying Kiva at the level of borrowers and lenders, that we will call *micro-scale* level, therefore, without the superstructure of Field Partners, we found that this kind of fine-grained level of our analysis led to mediocre results in prediction tasks, among others. In general, we observed a lack of signal data because the analyzes are too fine-grained.

Going instead to the Field Partners level, that we will call *meso-scale* level, thus grouping the borrowers on their Field Partners, allows us to extract a stronger signal from the data. This change of perspective works in two ways: it will enable us to evaluate Field Partners as the leading players on Kiva. Moreover, it represents the first step in the sense of dimensionality reduction in the representation of the system.

4.3.3   Micro-mechanisms and macro-processes: the Coleman's Boat

To better understand the concept of change of perspective of the analyzes introduced at the end of the previous Section, we would like to offer a clarifying parallel with the so-called Coleman's Boat [122].

The Coleman's Boat is an intellectual tool designed by the American sociologist James Coleman to provide an explanatory understanding of the social phenomena in general.

Sociology is interested in how the social world works. In particular, many central questions in sociology are related to relations between macro mechanisms and micromechanisms in society. For example, suppose we think of states, communities, organizations in general as macro-social reality and at individual or small groups as micro-social reality. In that case, we can easily argue how the macro-level impacts the micro-level. The individuals, who are,

with their interactions, the crucial components of macro social realities, are continuously influenced in their behaviours by what happens at a macro level. In contrast, we can explain a social phenomenon by studying the intentional actions of the individuals who produce it [123, 124]. Since the social phenomenon is the variable that we need to explain, we necessarily start from a relatively simple model of individual action: and here, Coleman's Boat comes into play.

Following the scheme proposed in Figure 4.12, we can see how a social phenomenon occurring on the macro-level (1) impacts the micro-level (2). Moreover, by shedding light on what happens on this level, i.e., on the individuals and their behaviour (3), we can thus understand how the aforementioned social phenomenon will impact society in general (4).
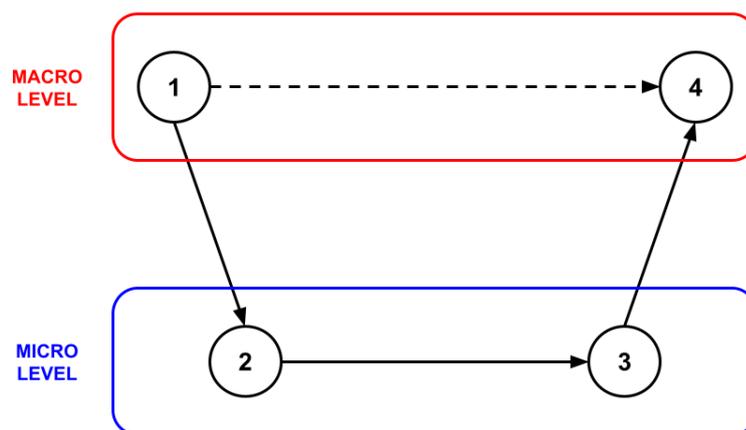
Figure 4.12: The Colemans' Boat intellectual tool. It is the exemplification of how micro-mechanisms can shed light on macro-processes [122]. The under-exam social phenomenon occurring on the macro-level (1) impacts the micro-level (2). By shedding light on what happens on this level, i.e., on the individuals and their behaviour (3), we can thus understand how the aforementioned social phenomenon impacts society (4).

4.4 A perspective change: Field Partners

As mentioned in Section 4.1, in Kiva, we have three types of actors: borrowers, lenders, and Field Partners. To bring Coleman's Boat pattern (see Section 4.3.3) back to our Kiva system, we intend to explain the evolution and observed

trends of the Kivas' platform. It is a macro-level phenomenon, and we get a sense of it by studying the behaviour not at the level of lenders and borrowers (micro-level) but by investigating the activity of the Field Partners. In this case, the Field Partners act as the meso-level of the system: we may describe them as "super-borrowers" or "meta-borrowers".

Going back to the representation of the system, and therefore of the interactions that occur, a "good" representation intervenes in managing and describing dynamics at a non-macro level to understand the macro-level.

However, before talking about the representation of interactions, we must understand which interactions we are discussing.

Once we passed to the mesoscale level of analysis, as we said in the previous Section, we faced the study of the activity of the Field Partners in the Kiva sectors. As discussed in Section 4.3.1, these analyses will become central in our approach to social interactions on Kiva and their optimal representation - we will tackle their discussion in the next Chapter.

For the moment, we will show exploratory analyzes similar to those shown above regarding the intervention of Field Partners in the Kiva sectors.

As a starting point, Figure 4.13 reported the world map representing the number of Field Partners in a specific country, aggregated over the years. The colour shows the abundance of each of these actors for each country. It may give us an idea of how many Field Partners are present in each area and, consequently, how important their role is on the Kivas' platform.

Figure 4.14 indicates the number of Field Partners operating in a given sector; we aggregated these values over time. As we can see, we have a more uniform distribution than that found in Figure 4.10. Therefore, it may help us to at least partially eliminate the bias resulting from any possible unbalanced classes discussed in Section 4.3.2. Finally, we show the same information shown in Figure 4.15 for completeness but analyze the activity of the Field Partners in each sector in the various continents.

4.4.1   Field Partners and business models: configurations

As a final example of a preliminary analysis, we propose studying how the activity of Field Partners changes over time. We want to study whether there
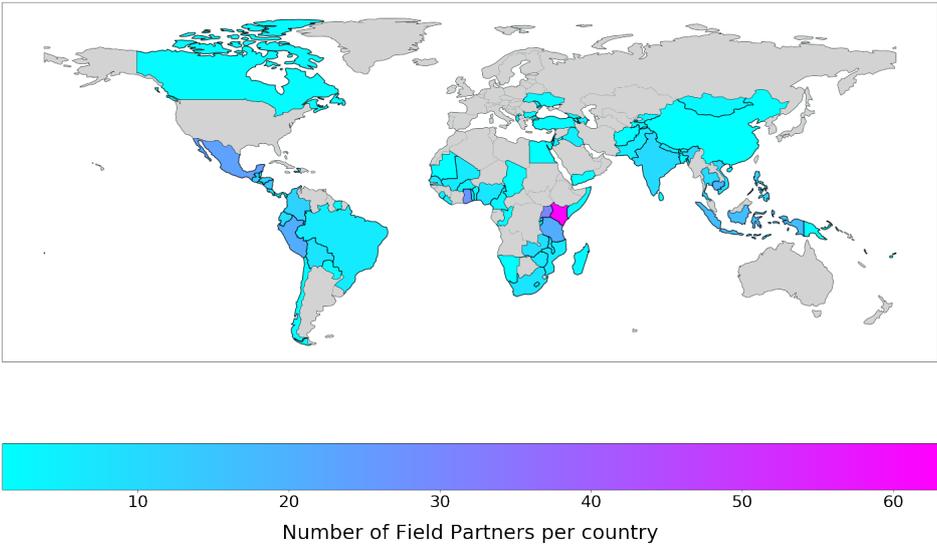
Figure 4.13: The map represents the number of Field Partners who have operated on Kiva from 2006 until 2020, according to their countries of origin. The colour indicates the number of Field Partner each country have in this time interval.
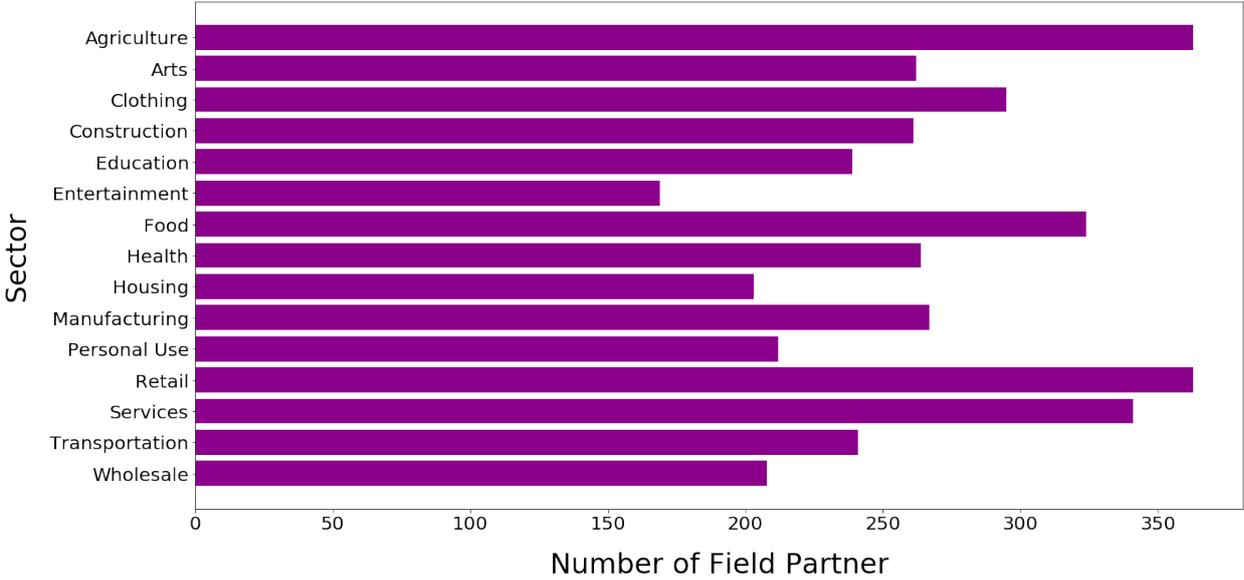
Figure 4.14: Distribution of the Field Partners among the Kiva sectors of activity. We count each Field Partner in every sector he operated in. The time interval under consideration is from 2006 until 2020.

is some "consensus activity" (i.e., a set of "ideal" business models to adopt) that the Field Partners achieve over time in their collaboration with Kiva's borrowers. Understanding if it exists can give us clues on what the behaviour of Field Partners is and consequently outline the evolution of the platform itself. Year after year, each Field Partner on Kiva adopts an action strategy that is
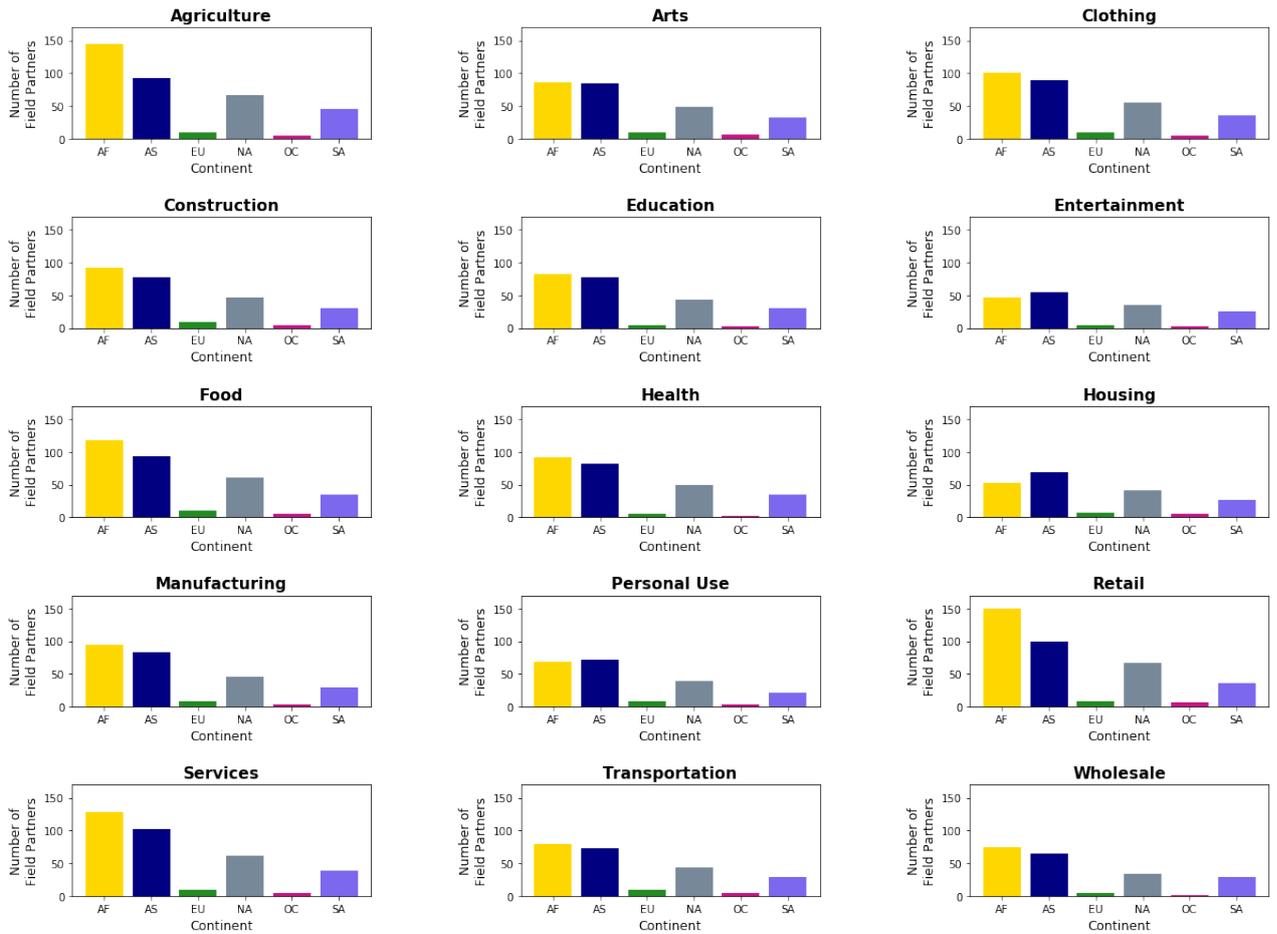
Figure 4.15: Distribution of the Field Partners among the Kiva sectors of activity, considering the continent of the loans they managed. We count each Field Partner in every sector he operated in. The time interval under consideration is from 2006 until 2020. The continent abbreviations stand for AF=Africa, AS=Asia, EU=Europe, NA=North America, OC=Oceania, SA=South America.

reflected in the choice to work on specific sectors or not. The homogeneity in these activities can give us a hint about a possible "consensus activity".

The tool we want to use in this analysis is the configurations. Given a Field Partner in a specific year y, we define as *configuration* c the activity of the Field Partner in the $N = 15$ Kivas' sectors of investments:

$$c_{FP}(y) = \vec{u} = (u_1, ..., u_N) \qquad (4.1)$$

where u_i is equal to 1 or 0 if there is or is not at least one loan in the ith sector for the given year y.

Let us see in detail how we calculated the number of configurations and

the relative entropy.

Figure 4.16 shows an example of a set of configurations for a specific year. For simplicity, we have reduced the problem to 5 Field Partners and 5 sectors. If each vector of white and black boxes represents the activity of each Field Partner in the sectors of Kiva, i.e., a configuration, we can distinguish 3 unique configurations among them, labelled A, B, C as shown in Figure 4.16.
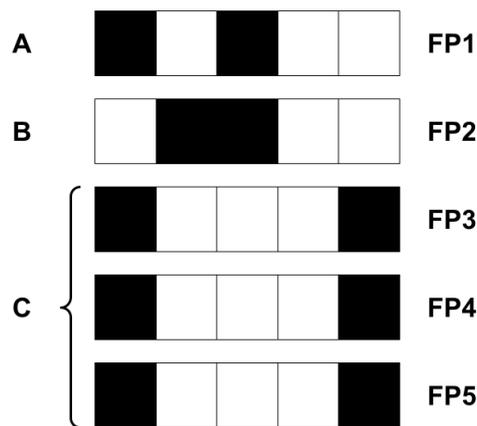
Figure 4.16: Example of a set of configurations for a year. Having 5 active Field Partners and 5 sectors of investment for simplicity, we have 3 unique configurations with a relative entropy H(year) ≈ 0.86.

The probability distribution of this set of configurations is $\left[\frac{1}{5}, \frac{1}{5}, \frac{3}{5}\right]$, and the associated entropy is $H(year) \approx 0.86$, as calculated as described above.

Once we define a configuration for each Field Partner, we will have a set of configurations for each year from different Field Partners. Note that the number of active Field Partners may change year after year.

Given a set of Field Partners configurations for a year, we count how many times each configuration occurs. In general, we assume the maximum number of configurations that one can see is equal to the number of Field Partners active in that specific year. Theoretically, it should be equal to $2^N = 2^{15}$, but the number of Field Partners is in general much lower than this, and it constitutes a good upper bound for our analysis. We guess that if there is a "consensus activity", the observed configurations should be less than the possibilities (so the number of active Field Partners). Moreover, this number should decrease year after year.

To get a better sense of it, we also measured the entropy of each set of configurations, year after year. Given the counting of Field Partners configu-

rations for a year y, we evaluate the relative Shannon entropy H of the set of configurations, defined as below:

$$H(y) = -\frac{\sum_c p_y(c) ln(p_y(c))}{ln(\sum_c)} \tag{4.2}$$

where the sum is over the configurations c of the specific year y; $p_y(c)$ is equal to the probability of seeing the configuration c to occur in the specific year y. If some configurations are used more often than others, we may guess that the entropy should decrease year after year.
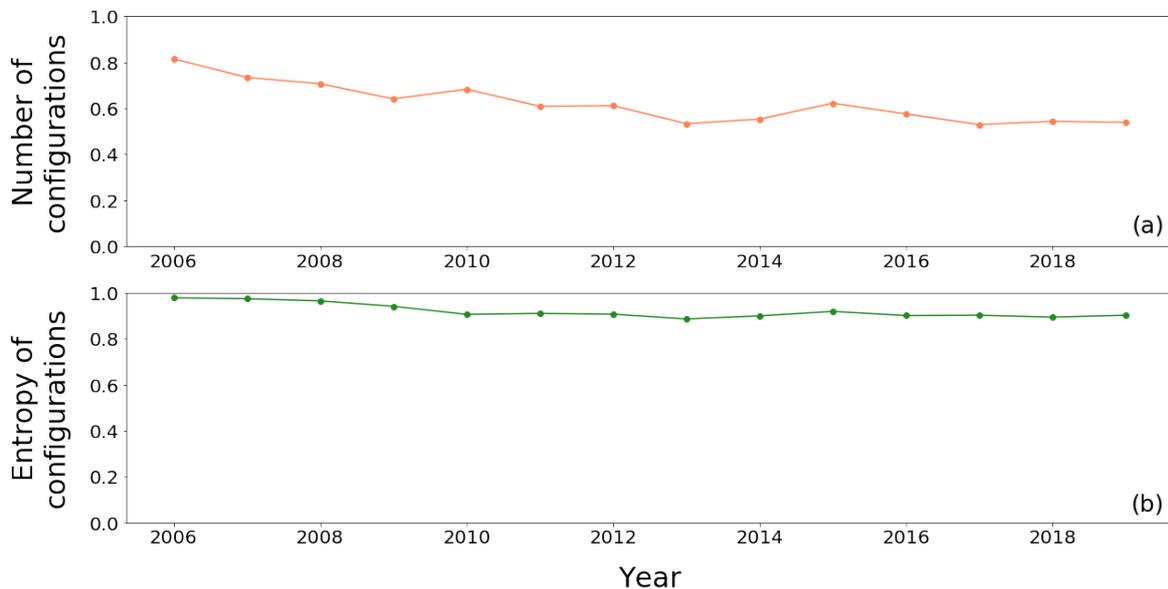
Figure 4.17: Analaysis on Kivas' Field Partners activity configurations. In Figure (a) we show the evolution of the number of configurations as a percentage of the maximum number of configurations (i.e., the number of active Field Partners), year after year. In Figure (b) we show the evolution of the Shannon entropy of configurations, normalized on the logarithm of the number of configurations in each year.

In Figure 4.17 we observe a decreasing number of active configurations in (a). Moreover, although it is weak, we observe a decrease in the entropy of configurations in (b), so we can assume that some are used more often. We should note that the findings are independent. We could have an increase of configurations with a decrease of configurations entropy and vice versa.

To prove it, we show with an example how the growth/decrease in the number of configurations in a given year is independent of the growth/decrease of the relative entropy.

Figure 4.18 shows a comparison between two different sets of configurations corresponding to two different years. The case at year y (left) is the same as in Figure 4.16. The case at year $y+1$ is slightly different: the number of unique configurations decreases because Field Partner 2 adopted the same "business model" of Field Partner 1. In this case, the probability distribution of the configurations in year $y+1$ is $[\frac{2}{5}, \frac{3}{5}]$. It gives an entropy $H(y+1) \approx 0.97$. This value is bigger than $H(y)(\approx 0.86)$: it demonstrates that the increase in the number of unique configurations is independent of the increase or decrease of their relative entropy.
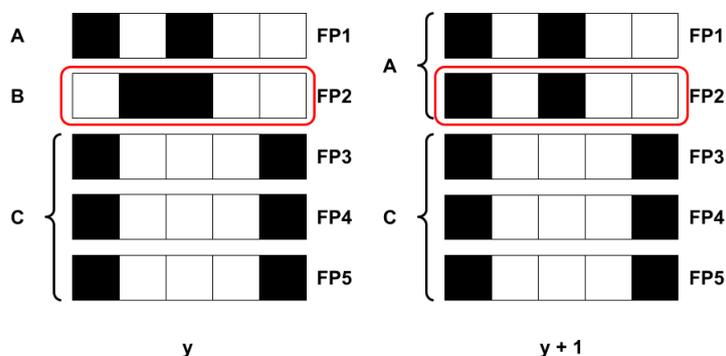
Figure 4.18: Example of a set of configurations for different years. Having 5 active Field Partners and 5 sectors of investment for simplicity, we have 3 unique configurations in year y and 2 in year $y+1$. The relative entropies are $H(y) \approx 0.86$ and $H(y+1) \approx 0.97$. The red square highlights the difference between the two set of configurations.

In general, we observed a trend of slightly increasing homogeneity among Field Partners activity. The decreasing number of configurations is independent of the changes in configurations entropy. Therefore, this finding strengthens our hypothesis that there is a set of "ideal" business models for Field Partners.

Therefore, our question is: can we find a good representation of this activity of the Field Partners in the sectors of Kiva that can condensate the information to explain phenomena like the behaviour outlined above?

What matters most to us is studying how to represent Field Partners' activity over time in the various sectors. Once again, embedding and matrix and tensor decomposition techniques will be our tools to find a solution to this task and better understand the particular social interactions on this crowdfunding platform. In the next Chapter, we will explain how a good representation of the

Field Partners interactions on Kiva can help us make sense of the platform's evolution over time. We will work once again with time-varying systems, adopting tensor decomposition techniques to shed light on the dynamics in the Kiva sectors.

4.5 Representation of the Field Partners interactions

In the previous Section, we observed a trend in the Field Partners activity. Year after year, the number of possible configurations adopted by them decreases: this means that they restricted their business models to a limited number (see Figure 4.17(a)). Even the configuration entropy confirms this trend: not only the number of configurations drops, but there are configurations more frequent than others (see Figure 4.17(b)). This increasing homogeneity in the Field Partners activity makes us think a common business model could exist among them. With this assumption in mind and in the spirit of this Thesis silver thread, an adequate representation of the problem is essential. The similarities among Field Partners activities in time make us hypothesize that their cause may lie in some higher-order interaction correlations among Field Partners interactions. The concept of a "common business model" here is central: it summarizes both the evolution in time and the common points of Field Partners activity.

Chapter 2 proposed a variety of systems and applications for which using a temporal network representation can be a perfect solution. We can easily include our analysis here in the number. In fact, the interactions between Field Partners find a natural representation in a time-varying network.

We saw in Section 2.2 that a possible representation for a temporal network is a tensor, and that this representation is optimal for combining community detection and temporal activity (see Section 3.1.4.iv). Using temporal decomposition to study this problem can let us understand the connections between the temporal component and the topological structure of a time-varying network.

For these reasons, we decided to adopt the tensor as our suitable representation and the tensor decomposition technique as our tool to study this problem.

In the following Sections, we will explain how we decided to represent the

Field Partners interactions and all the analyses related to them.

4.5.1   Field Partners temporal network: data selection

We decided to use a Field Partner-Field Partner temporal network to study higher-order correlations among Field Partners interactions. For computational reasons and to remove noisy information related to infrequent interactions, we cut on the number of Field Partners. We selected the most active Field Partners on the Kivas' platform aggregated on time using the elbow method heuristic [125, 126]. To be more precise, in our dataset, we can associate the number of managed loans to each Field Partners. In this case, we decided to aggregate the number of loans all over the years. Sorting the number of loans in descending order, we aim to find a point (i.e., the Field Partner relative to this number of loans) beyond which the variation will become negligible. In this context, using the "elbow" as a cutoff point reveals to be a good solution. The elbow method is a common heuristic in mathematical optimization to choose a point where increasing the number of variables is no longer justified due to the non-increase of the attached information.

Figure 4.19 shows the above mentioned cutoff selection. Sorting the number of loans according to descending order, we find an elbow corresponding to 22 eligible Field Partners.

Once we selected the most active Field Partners, we started to build the temporal network for Field Partners interactions. Since we wanted to study the similarities of Field Partners business models in time, we thought that the best representation of the interaction could be building a temporal network in the following way. The n nodes are the 22 Field Partners, and we put a link between them if they have posted a loan in the same sector of investment at the same time. We selected a monthly granularity for the analysis as a balance between a low granularity (day), which causes an increase in noise in the data, and a high one (year), which nevertheless causes a loss of information at a low level.

In the end, we obtained an undirected temporal network that allows us to focus on Field Partners interactions, specifically on their similarity on investing in the same sector.
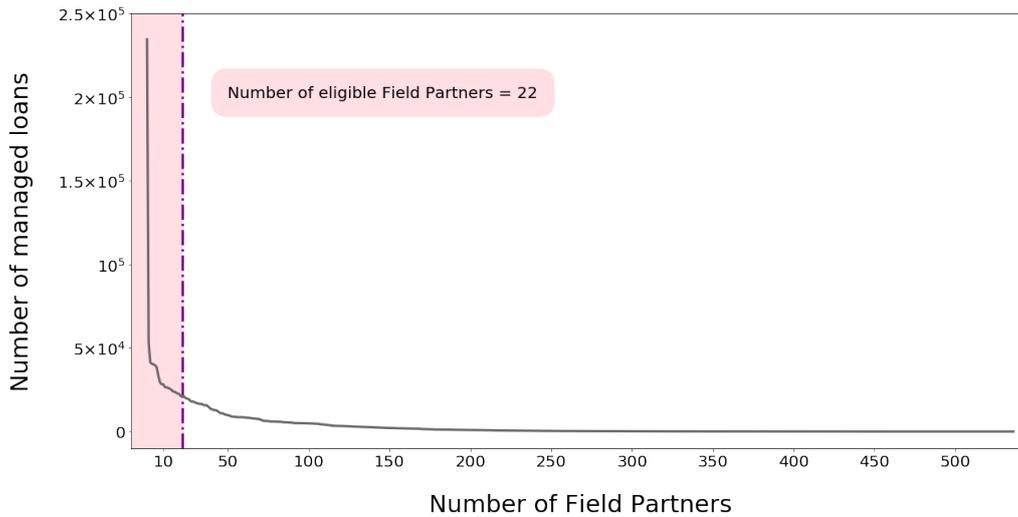
Figure 4.19: Cutoff selection on the Field Partners. Sorting the number of loans according to descending order, we find an elbow corresponding to 22 eligible Field Partners. The dashed line indicates the elbow, while the pink area highlights the eligible Field Partners.

4.5.2 Tensor decomposition of the Field Partners temporal network

As we explained in Section 3.1.4.iv, we operated a transformation on the above described temporal network. We can indeed think of a temporal network as a collection of T static adjacency matrices, one for each timestamp $t \in T$ - in our case, we have taken into account 157 months [2].

We have thus obtained a 3-rank tensor whose dimensions are $n \times n \times t$ representing our real time-varying system - see Figure 3.5. As explained in Section 3.1.4.iv, the tensor representation of the system helps us analyse the community-activity structures of the temporal network (we will refer to them also as "mesoscale structures"). These structures can shed light on the similarities of the Field Partner activities, taking the temporal evolution into account.

4.5.2.i Tensor rank: the selection

The step we are going to face now is the definition of the number of mesoscale structures. In Chapter 3 we used the core consistency metric (see Section 3.1.4.iv, [74]). In this case, we decided to use another method for computa-

---

[2]Our data range almost 15 years, minus the months in which we do not have an interaction between any of the 22 Field Partner under analysis.

tional reasons. The new technique we decided to adopt exploits the approximation error that the tensor decomposition algorithm makes in the attempt to reproduce the tensor through the three factors A, B, C (see Figure 3.5). In more detail, we know that each time we decompose a tensor in factors, it happens with an approximation. We thus calculated this approximation error for each possible rank for 10 different decomposition, for 10 different possible ranks. We then selected the optimal rank using the elbow method heuristic (see Section 4.5.1).

We decided to keep a low value for the rank to have a low dimensional representation of the Field Partner system. A low dimensionality of the problem makes reading the results more intuitive, and it might give us insights into the Field Partners activity.

Figure 4.20 show the selection of the best rank for our tensor. It represents the approximation error in tensor decomposition with respect to the tensor rank. The error decreases as the rank increases, with an elbow corresponding to the optimal rank $r = 3$.
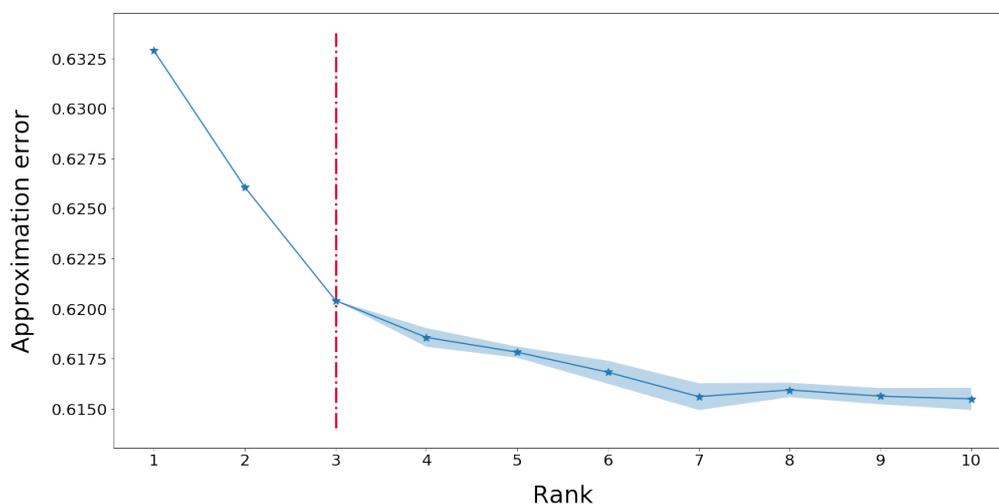
Figure 4.20: Approximation error in tensor decomposition with respect to the rank. The red dashed line indicates the elbow corresponding to the optimal rank $r = 3$. The blue line and the shaded area represent respectively the average and the variance among the samples we used for the analysis.

4.5.2.ii Mesoscale structures in the Field Partners tensor

Once we selected the optimal rank, we proceeded with the tensor decomposition. As we discusses in Section 3.1.4.iv, we expressed the 3-rank tensor

with an approximated decomposition (see Equation 3.5). As in the case of the
temporal network embedding discussed in the previous Chapter, this technique
let us group our links at a given time into mesoscale structures. We now have
each link at a given time in our temporal network corresponding to a 1-rank
tensor. We can thus assign a mesoscale structure to each link.

   We can now assign each given link to a specific sector of investment and
a specific timestamp since we linked Field Partners who operated in the same
sector at the same time. Figure 4.21 shows the different distribution in time
and sector of the 3 different mesoscale structures we found. From a sector
distributions point of view, we do not detect any substantial difference between
the mesoscale structures, except for the number of elements that compose
them (a). Looking instead at the timestamp distributions, we can see how
our model can capture complementary and defined groups in each mesoscale
structure (b).

   At this level of analysis, we can only state that the representation of the
temporal network through the tensor decomposition has found mesoscale struc-
tures corresponding to well-defined time intervals. At this point, it is worth
asking what this can mean. Are there any non-trivial patterns in Field Partners
interactions? Or do we observe an arbitrary temporal subdivision inexorably
linked to an absence or scarcity of signal in the data? To answer these ques-
tions, we need to resort to a reference null model. Like in the previous Chapter,
this can help us understand what kind of signal we have in the data and what
part of it our model can capture.

4.5.3   Null models in tensor decomposition

We want to spend a few words on the importance of reference null models.
We know that the interactions described in temporal networks give rise to
temporal and/or structural components. They can be more or less defined,
more or less interacting, and more or less easy to understand and represent.
The null models help us to outline the salient features by disentangling the
various components.

   Each problem can refer to different and *ad hoc* null models. In this frame-
work, there is no principle solution. Our question, whether there is an intrinsic
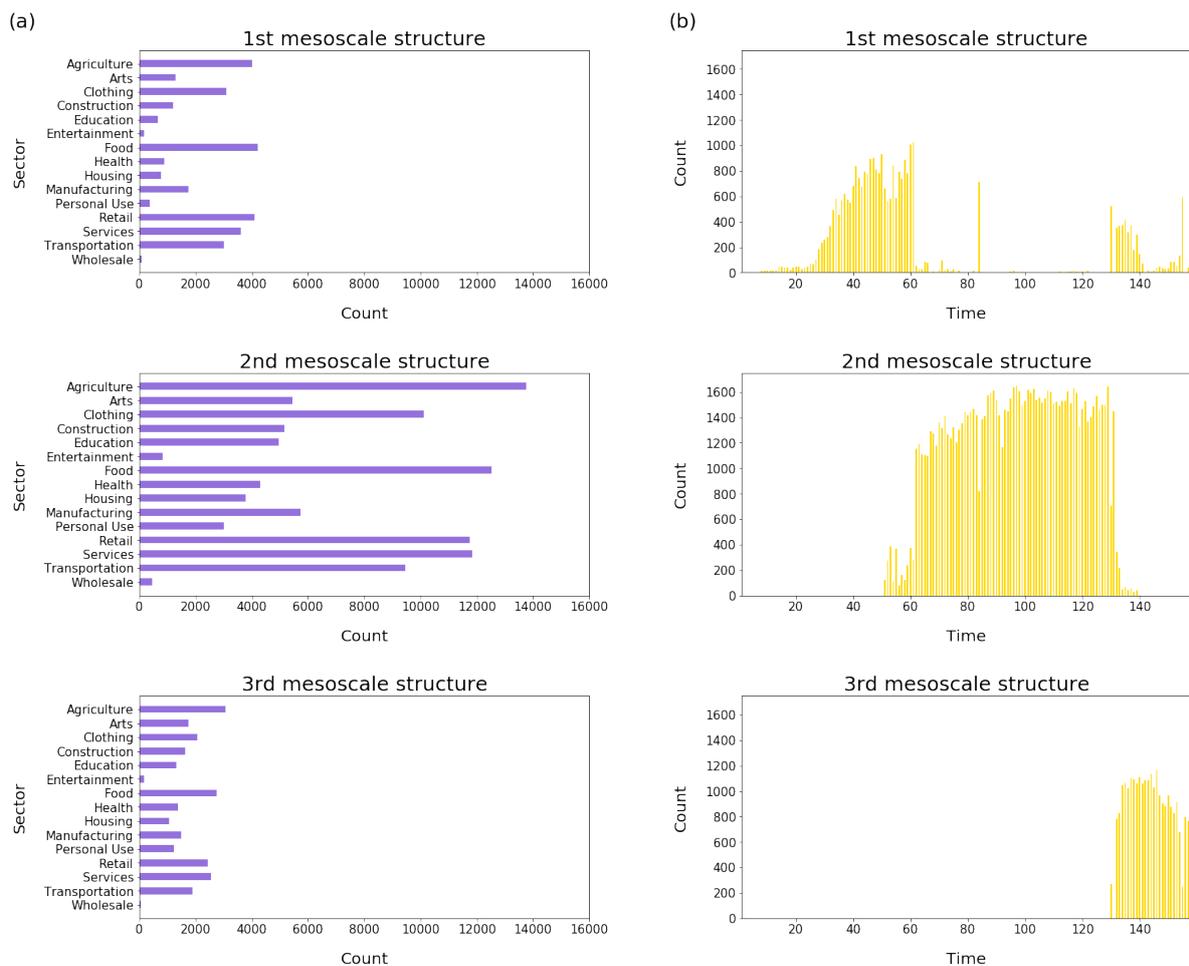
Figure 4.21: Distribution of the sectors of investment (a) and the time intervals (b) in the mesoscale structures obtained through tensor decomposition.

structure in our data, can help us to outline an effective one. In Figure 4.21 in the mesoscale structures, we have seen that there is a defined timestamps distribution (b), which does not, however, correspond to very different sector distributions (a). Therefore, we can ask ourselves whether the spatial correlations will also fall by destroying any structural correlations. In this case, it would mean that we have upstream a lack of signal in the data that causes the tensor decomposition to create randomly defined time intervals. However, if the temporal correlations also fall, we can assume higher-order structural correlations that we cannot see by simply looking at the sector distributions in the mesoscale structures. With that in mind, we thought an optimal null model would be one where we would shuffle the nodes for each timestamp in the time network. In this way, we would have for each timestamp the same nodes but connected randomly, thus destroying any structural correlation.

We thus built a temporal network that would preserve the timeline of the original one but would destroy the possible structural correlation by shuffling, for each timestamp, the original nodes. We will refer to this shuffle tensor as the *null-model tensor*.

This null model is purposely simple: this will allow us to understand what structures exist in our data. A first clue about the quality of the signal coming from our data is to understand how the approximation error varies on the decomposition of the null-model tensor. If we observe a trend similar to that observed in Figure 4.20, think that there may not be an intrinsic a priori structure. Therefore, we have repeated the same experiment described in Section 4.5.2.i and shown in Figure 4.20 also for the null-model tensor. The result is shown in Figure 4.22.
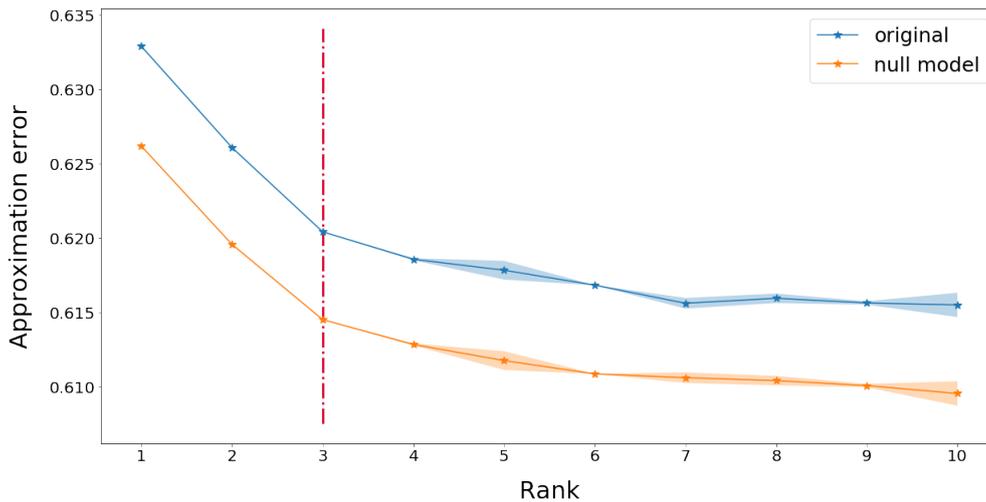
Figure 4.22: Comparison of the approximation error in the original (blue line) and in the null-model (orange line) tensor decomposition with respect to the rank. The red dashed line indicates the elbow corresponding to the optimal rank $r = 3$. The lines and the shaded areas around them represent respectively the average and the variance among the samples we used for the analysis.

As we can observe, the trend found in Figure 4.22 is not very different from that in Figure 4.20 . The approximation error method used to find the optimal rank also fixes that for the null-model tensor at $r = 3$, the same result obtained for the original tensor. It might be a signal of a lack of intrinsic structure of the data.

To have a confirmation, we proceeded with the analysis of the mesoscale structures of the null-model tensor. Figure 4.23 shows the results. Once again,

we do not observe any peculiar distribution of sectors in any of the mesoscale structures. On the other hand, we find a well-defined and complementary distribution of the timestamps for each mesoscale structure.
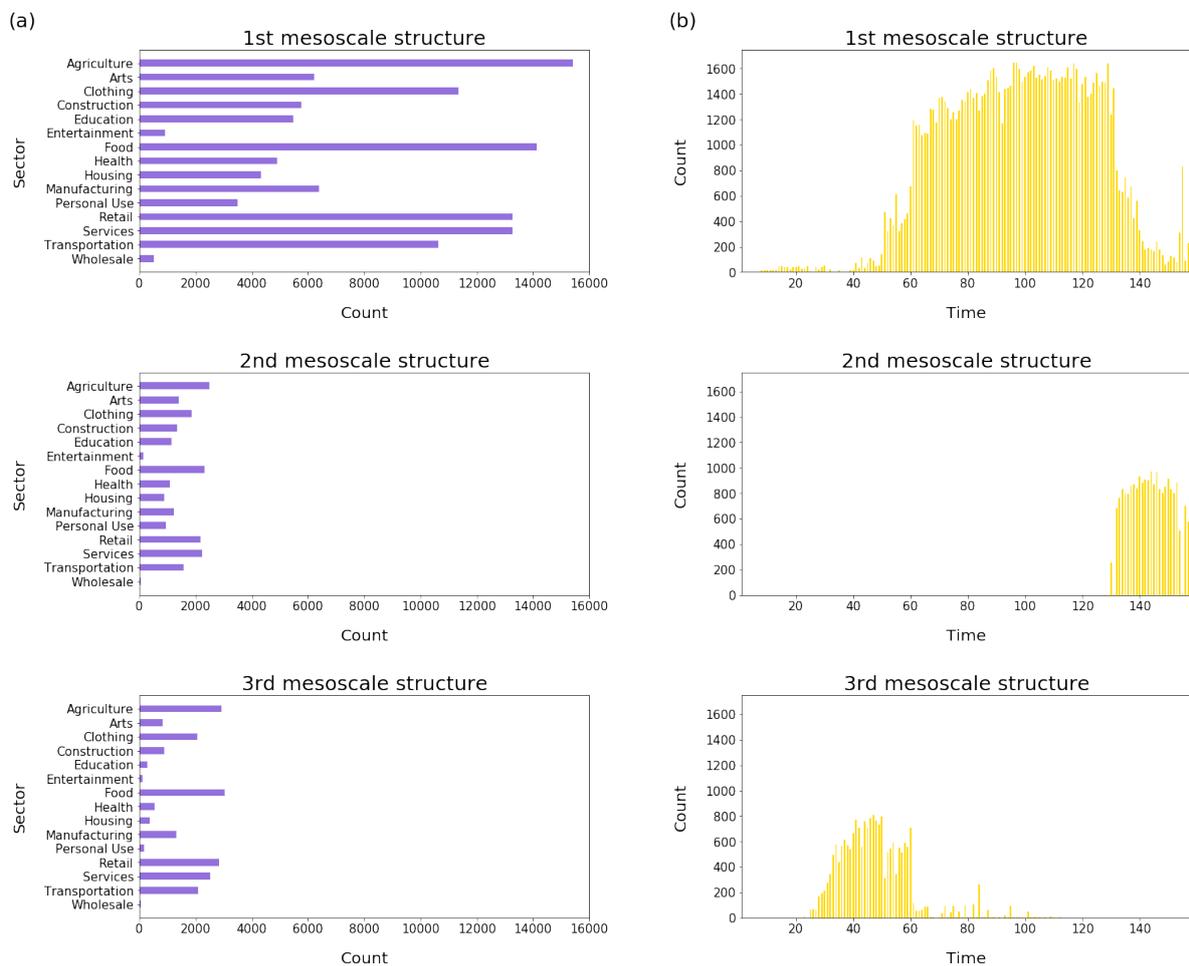
Figure 4.23: Distribution of the sectors of investment (a) and the time intervals (b) in the mesoscale structures obtained through tensor decomposition. The analysis has been done on the null-model tensor.

The results for the null-model tensor are analogous to the previous ones. We can therefore say that the model arbitrarily decomposes temporal interactions lacking an actual structure at the origin. It suggests that there is no such clear signal in the data or that it is even absent.

However, we can say that the tensor decomposition, in this case, has helped us to understand that we probably cannot find an explanation for what we found in Figure 4.17 by looking at the simple interactions between Field Partners in the different sectors. It again confirms how a correct representation of the problem can help us explain some non-trivial trends and patterns that we observe in the exploratory phase. Unfortunately, in this specific case, the already

discussed data limitations (see Section 4.3.2) can lead to signal weakness that even a good representation of the problem can struggle to express and capture.

4.6 Final Remarks

With this mix of analysis, exploration, and modelling of interactions on Kiva, we introduced the crowdfunding platform's history, evolution, and main characteristics as an example of a specific type of social interaction, the event-type. We can look at this Chapter in two ways. It is an application of the models and techniques seen in the previous Chapter, and it represents the result of convergent analyses both from the more mathematical-computational sphere and from those of the social sciences.

Following the silver thread of this Thesis on the low dimensional representation of the social interactions, we have proposed a variety of analyzes on different levels. In this final part of the Chapter, we want to highlight how we first introduced dimensionality reduction by setting the Field Partners as the focus of our analyses. Second, we devised a representation of the Field Partners interactions using tensor decomposition, which, confirming its effectiveness, helped delineate the scarceness of structure of the information in the data.

5

Conclusions and perspectives

HE study of social interactions is crucial to understand the complexity of human relationships, society, and socio-technical systems. Social interactions have been central to socio-economic studies for a long time. With the advancement of digital technologies, studies on social interactions have gained access to large basins of digital data on human interactions. In this context, disciplines such as computational social sciences and social network analysis have quickly progressed thanks to the significant advance in data collection and analysis. Representing, analysing and understanding the structure and dynamics of social interactions have posed a challenge in this research area. Networks have thus become a common language, a bridge established across many disciplines such as statistical physics, applied mathematics, social sciences and complex systems sciences, aiming both at answering established research questions with new tools and at tackling new and more ambitious challenges. Networks are indeed a natural representation of social interactions: network datasets naturalistically generated by digital platforms are challenging in terms of sheer size and heterogeneity, as they encode many different aspects such as temporal and geographical features and user-generated contents, both structured and unstructured. Dealing with such high-dimensional system encoding social interactions is challenging in terms of computation and analysis. Many methods of network dimensionality reduction and low-dimensional repre-

sentation of graphs have been developed to tackle this issue. There are many possibilities for the design of low-dimensional representations of nodes, edges, or the whole graph that preserve given relations of the original entities. Network embeddings have been individuated as an important solution: by reducing the dimensionality of the network, they facilitate making sense of the original data, esposing the relevant structures, and enable us to conceive analyses that would otherwise be impossible to be carried out. Many machine learning algorithms have the potential to solve problems such as link prediction, nodes classification or clustering tasks, but most of them are designed to accept as an input vector-based object. Network embeddings are useful to represent networks interactions and dynamics in terms of vector-based object, learning and preserving feature representations of nodes and links, and are a suitable input for otherwise impracticable machine learning tasks such as classification and prediction.

In the first part of this Thesis, we focus on the development and the application of new embedding techniques for the representation of time-resolved interaction networks. Temporal networks have received great attention from the network science community, because allow to understand the understanding of how the characteristics and dynamics of social interactions change over time. In particular, temporal networks allow the study of dynamical processes, such as spreading processes. Spreading processes appear in diverse natural and technological systems, such as the spread of infectious diseases and the dissemination of information. They are interconnected with human interactions because complex networked structures, especially changing over time, may affect the behaviour of spreading processes. Given the importance of spreading processes in real-world scenarios, understanding the characteristics of social interactions that impact epidemic processes have crucial significance.

We present a temporal network embedding technique whose originality lies in the embedding of interaction events, not of nodes, Focusing on the event, i.e. on the interaction between nodes at a given time, our method preserves information on the causal structure of the network; hence it is in principle suitable to study dynamical processes on temporal networks such as epidemic spreading. We apply our embedding method on empirical temporal network data on human proximity in space in numerous real-world environments, be-

cause these settings are optimal to study epidemic processes. The embedded representation retains important temporal and structural features of the data, that allow us to understand how these aspect of the interactions plays a central role in the dynamic of spreading processes.

Our method can obtain a representation for any temporal network, but the focus on event embedding makes it particularly suitable and extensible for representing temporal networks on top of which to study a causal process such as an spreading process.

In the second part of the Thesis, we analyse an online crowdfunding system, for which is available a large-scale interaction dataset. The data contains information on the interactions and the transactions between different actors operating on this system worldwide for more than 15 years. Specifically, we can observe how the users (lenders) invest in specific sectors by financing other users (borrowers). There are also intermediate actors, the so called Field Partners, who manage the borrowers' loans in these transactions. Field Partners act as brokers between borrowers and lenders: activities on the system are almost entirely in the hands of the Field Partners. We provide a general descriptive analysis of the principal characteristics of this crowdfunding system and its anatomy. Alongside this descriptive analysis, we focus on the temporal dimension of the network, a complex challenge because the representation of this system as a network presents various actors. Since that the role of Field Partners is crucial in the transactions, to understand the whole system is necessary to understand how Field Partners interactions work. Our research interest mainly focuses on the transformation of system structure over time and how Field Partners transactions shaped it, and we opt for a representation of the system that considers both community structure and temporal evolution. A tensor representation of this system can help us capture the correlation between the latent structure of the system and its temporal patterns. In order to detect these underlying structures and to deal with a challenging dataset, we operate a a tensor factorisation, obtaining a compact representation capable of learning and representing community-activity structures of the tensor. Analysing these underlying structures paves the way to characterise and better outline the Field Partners' activity on the system, and the evolution of the system in general.

The results leave us with some open questions. Can we design other meth-

ods to study data on social interactions? Can we find different data sources worth exploring with our dimensionality reduction methods? In this Thesis we introduced different social interactions systems and dimensionality reduction methods that are only a small part of the big, continuously evolving framework of low dimensional social network representation. Contact networks and crowd-funding systems are just an example of interaction system data: mobility data on proximity interactions between people, social media data that describe interactions among users, financial transactions are others high-resolution data on social interactions that we can explore. Moreover, we deal in this Thesis with nodes and interactions of temporal networks, without considering any meta-data that can describe them. In real-world social interactions, people may have roles or specific demographic characteristics; interactions are exchanges whose features depend on the people and the setting in which they occur. A possible direction could be implementing a dimensional reduction method that can take into account the metadata of users and events that describe a social interaction. Another possible follow-up work is including a wider range of dynamical processes in our analysis. We could consider other epidemic processes, such SIS or SIR, more complex than the SI process simulated in this Thesis. Given the peculiarity of focusing on event embedding, suitable for studying different spreading processes, it would be interesting to adapt our embedding technique to study specific spreading processes, misinformation, and fake news spreading in this context. We could even expand our horizon to other important dynamical processes, such as consensus formation on networks and synchronisation. Both dynamical processes are central in social networks research, and studying the emergence of consensus in social systems or their synchronisation may shed light on the dynamics of social interactions.

By focusing on social networks as a representation for social interactions that will evolve and change over time, we develop different dimensional reduction methodologies adapted to different settings and objectives to have a compact but informative representation. The network representation and the dimensional reduction techniques we proposed here showed us new insights and perspectives on the different systems we dealt with, and in general, on the characteristics and structures of social interactions.

The contribution of this Thesis consists in the exploration and application to

real-world systems of methods for dimensionality reduction of time-resolved so-
cial interaction data. This work lies in the broader context of low-dimensional
graph data representations, which has witnessed the development of a vast
spectrum of methods spanning embeddings, low-rank tensor representations,
graph convolutions, and more. To date, no single approach is generally suitable
to deal with all data sources, and complex tradeoffs exist between performance
in the context of a prediction task and interpretability of the learned representa-
tions. We dare hope to have contributed to exploring the rich research domain
at the intersection of network science, machine learning, and their applications
to decoding and understanding digital data on social behaviours.

6

Appendix - weg2vec: event embedding in temporal networks

HIS Appendix reports the article we published on Scientific Reports [57], as a complementary resource to understand in more detail what we have treated in Chapter 3. You may find additional information here: `https://www.nature.com/articles/s41598-020-63221-2`.

SCIENTIFIC REPORTS

natureresearch

Check for updates

OPEN

weg2vec: Event embedding for temporal networks

Maddalena Torricelli[1,2], Márton Karsai[3,4,1] & Laetitia Gauvin[1] ✉

Network embedding techniques are powerful to capture structural regularities in networks and to identify similarities between their local fabrics. However, conventional network embedding models are developed for static structures, commonly consider nodes only and they are seriously challenged when the network is varying in time. Temporal networks may provide an advantage in the description of real systems, but they code more complex information, which could be effectively represented only by a handful of methods so far. Here, we propose a new method of event embedding of temporal networks, called *weg2vec*, which builds on temporal and structural similarities of events to learn a low dimensional representation of a temporal network. This projection successfully captures latent structures and similarities between events involving different nodes at different times and provides ways to predict the final outcome of spreading processes unfolding on the temporal structure.

An interacting group of people, the collectively active neurons in the brain, or the transportation system of a city are only a few examples of complex systems, which are all intrinsically dynamical[1]. They can be commonly interpreted as a set of entities, which interact over time and form a network structure coding the architecture of the system in hand[2,3]. This duality of the structure and temporal nature of interactions can be effectively captured by temporal networks, proposing a new and precise representation of complex systems as compared to earlier strategies[4,5]. On the finest level, temporal networks consist of time-varying events between interacting nodes, and as a whole they appear as systems with high complexity and dimensionality. Events in real temporal structures, however, are not random but correlated with each other and arguably driven by several microscopic mechanisms leading to several generative characters of the network. Emerging properties like the heterogeneous number or strength of interactions[6,7], community structure[8–10], degree correlations, bursty temporal patterns[11,12], or temporal motifs of causally correlated events[13,14] are all arguably induced by such mechanisms. In turn, temporal events and their correlations largely influence dynamical processes as well[15], like they determine the speed and final outcome of information or epidemic spreading[16–19]. The recognition of these impacts has put temporal networks in the focus point of several investigations recently, which yet struggle to propose efficient representations to capture the complex temporal/structural information coded in them, while reducing their dimensionality to ease their analysis.

Correlated patterns in the structure and dynamics of networks usually can be described by certain higher-order representations[20]. For static networks, *line graphs* propose an efficient description[21,22], which in their simplest form identify static links as nodes and connect them if they are adjacent, i.e., share an ending node in common. Other technique is based on simplicial complexes[23] considering homology of the network topology to capture higher order structures. At the same time, recent *network embedding* methods propose inventive ways to obtain a reduced dimensional representation of static structures. Their common goal is to map complex networks to a low-dimensional space, while conserving certain similarities of nodes reflected by some distance metrics in the embedding. Most common methods use random walk sampling[24,25] or graph convolution[26,27] to capture the local structural context of network nodes. In case of temporal networks, the recently proposed *event graph* representation[28,29] defines a higher-order description by identifying relations between events, which are adjacent, i.e. not simultaneous and share at least one ending node. Adjacent events then are connected by links with direction respecting the arrow of time, and with weight defined as the absolute time difference between the connected events. This type of description is very useful as it codes all time-respecting paths in a temporal network at once, while proposing an information lossless representation of the temporal structure as a static weighted directed acyclic graph. Recently a few *dynamical network embedding* methods[30–33] have been developed to consider dynamical changes in the structure in the learned network representations. At the base of many methods there is the

[1]ISI Foundation, Turin, Italy. [2]Alma Mater Studiorum University of Bologna, Bologna, Italy. [3]Department of Network and Data Science, Central European University, H-1051, Budapest, Hungary. [4]Univ de Lyon, ENS de Lyon, INRIA, CNRS, UMR 5668, IXXI, 69364, Lyon, France. ✉e-mail: gauvinlaetitia@gmail.com
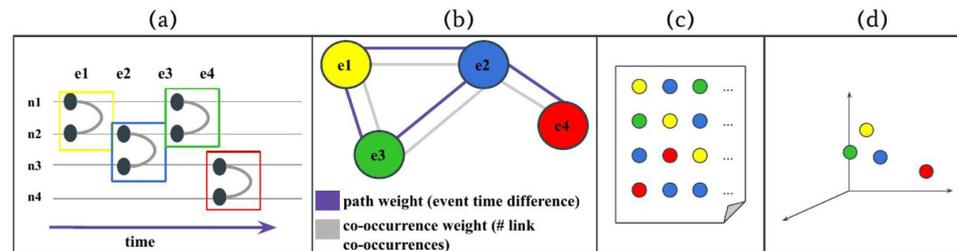
Figure 1. Schematic presentation of the methodological pipeline of the presented temporal network embedding method, which takes a temporal network to (**a**) project it into a weighted event graph; (**b**) to sample a set of environments for each event; (**c**) and uses it as input for a Skip-Gram model; (**d**) to obtain an event embedding of the original network.

modification of the standard representation of the temporal network, whether it is in the form of a list of events, a tensor[9] or a supra-adjacency matrix[34]. All of these methods like DyANE[34], Online Node2Vec[32], STWalk[30], or the one proposed by Singer *et al.*[31] commonly aim to solve a node embedding problem by locally sampling the temporal-structural neighbourhood of nodes to create contexts, which they feed to a Skip-Gram learning architecture borrowed from the text representation literature[35]. As a solution, they build a sequence of correlated/ updated embeddings of network snapshots, which consider short term history of the network backward in time. However, these methods have some common limitations; first of all, it can be hard to manage a high number of hyper parameters for the control of the sampling random walk process and the embedding itself. At the same time, the embedding of nodes may miss to reflect the dynamical changes of temporal interactions. Finally, taking into account only past and present interactions in the embedding can crucially limit the performance of the prediction, while the consideration of future events can significantly improve this task.

Here we propose a new temporal network embedding method, that we call *weg2vec* (weighted event graph to vector), which aims to tackle all these shortcomings. This is an event embedding method, which represents an entire temporal network in the same reduced dimensional abstract space. It is based on combined event contexts built by sampling locally a higher-order static representation of temporal networks, which in turn code the complex patterns characterising the structure and dynamics of real world networks. This is an unsupervised representation learning technique, which can consider the past and future context of an event simultaneously. It is sampling without using dynamical processes, thus it can be controlled by a handful of hyper parameters. It identifies similarity between different events/nodes, which may be active at different times, but influence a similar set of nodes in the future. To demonstrate the power of this representation, we showcase its utility through the prediction of the final outcome of modelled spreading processes on several real world temporal networks. This prediction task performs significantly better when it builds on our representation as compared to other dynamical network embeddings.

As follows, first we will present the pipeline to build our embedding method. We will show the characteristics of our representation, measuring its stability and its ability to capture temporal and structural information from the network. We will show then the results obtained in estimating an epidemic spreading outcome, as a showcase of the potentials of our embedding method in the analysis of dynamic processes. Finally, we will compare our results to similar computations performed with two other embedding methods. In the final Sections, we present the discussion of the results and the analysis of the methods.

Results

An embedding method of temporal networks may take a list of temporal interactions as input, and provide a lower dimensional representation, in which vectors corresponding to similar nodes or events in the original structure ideally point close to each other in the embedding. In our pipeline we solve this problem in three consecutive methodological steps. First, we turn the original temporal network into a higher-order representation, which captures pairs of adjacent and consecutive events, which may be in causal relationship. Second, we use this representation to generate environments for each event, sampled from their adjacent neighbours. Finally we obtain an embedding of events by training a Skip-Gram model on the obtained environments. These steps are schematically drawn in Fig. 1) and introduced next in the coming sections.

To demonstrate our method we used four different datasets all obtained from the SocioPatterns project[36]. These open datasets record the time evolving physical proxy interactions of a large number of people in different settings like in *conference*, *high school*, *hospital*, or *primary school*. The data comes as a long sequence of network snapshots recording simultaneous interactions between any participants in every 20 seconds. While for demonstration most of the results in the paper are shown only for the conference and primary school settings, a detailed data description together with the analysis for the other networks are presented in the Supplementary Information.

Temporal networks as weighted event graphs.   Let us consider a temporal network

$$G_T = (N, E_T, T),$$ (1)

where E_T denotes a set of events (temporal edges) among nodes in n at times $t \in T$. Specifically, we define an event $e = (i, j, t)$ as an interaction between two nodes $(i, j) \in N \times N$ at a given timestamp $t \in T$. The time aggregation of interactions in G_T over t maps the underlying structure into a static graph $G = (N, E)$ defined over the same set of nodes N, which are connected if they interacted at least once. For simplicity here we assume that events are undirected and no self events/links exist, i.e. for any event (i, j, t) or link $(i, j), i \neq j$.

We define two events $e_1 = (i, j, t_1)$ and $e_2 = (k, l, t_2)$ to be *adjacent* ($e_1 \rightarrow e_2$) if they share at least one node ($\{i, j\} \cap \{k, l\} \neq \varnothing$) and they are time consecutive ($t_1 < t_2$). Note that adjacency in a static network G can be similarly defined between links $l_1 = (i, j)$ and $l_2 = (k, l)$, which share at least one node ($\{i, j\} \cap \{k, l\} \neq \varnothing$). More restrictively, we called them δt-*adjacent* ($e_1 \xrightarrow{\delta t} e_2$) if they are adjacent and $|t_2 - t_1| \leq \delta t$, thus follow each other within a given period of time ($0 \leq \delta t \leq \tau$) where τ corresponds to the total period of time of the interactions. Adjacency introduces a directed relation between events, with orientation respecting their order in time. Using this notion we can formally define a static directed network representation $D = (E_T, E_D)$ of any temporal network, where original events in E_T are defined as nodes and they are connected by directed links $e_D \in E_D$ if they are adjacent $e_D = e_1 \rightarrow e_2$. The obtained network is a directed acyclic graph called the *event graph*, defined earlier in[28,29]. It can be interpreted as a temporal line graph, which provides a higher-order representation to map out simultaneously all time respecting paths of the original temporal network without any loss of information. Note, that to simplify our representation, for a given event if it has multiple future adjacent events with the same pair of nodes, we only consider the earliest one.

Event graphs can be easily enriched with various types of link weights reflecting some temporal or structural information coded in the original structure. Here, to better capture the strength of potential causal relationships, first we consider a weight defined as $w_{path} = \frac{1}{(1 + |t_2 - t_1|)}$, which is a measure inversely proportional to the absolute time difference between adjacent events at t_1 and t_2. This definition of the weight allows us to include the temporality of interactions such as long decay in social activities. At the same time we define a second weight for adjacent events (links of the event graph), based on the total number of co-occurring events on the underlying adjacent links in the static graph. More precisely, the $w_{co}(e_1, e_2)$ co-occurrence weight counts the number of δt-adjacent events in G_T appearing on a given pair of adjacent links $l_1(i, j)$ and $l_2(k, l)$ in the static graph G. Note that adjacent events connected in d, which corresponds to the same links in the underlying network G, will have the same w_{co} values. Datasets analysed in this paper are defined as sequences of snapshots aggregating temporal interactions in consecutive time windows of size Δt. In these systems we compute w_{co} for adjacent links as the number of co-occurrence of corresponding events within a single snapshot. This definition may slightly underestimate the real co-occurrence (if $\delta t \leq \Delta t$), but provides the best plausible solution due to the un-ambiguity of timings of events within a single snapshot.

Neighbourhood sampling strategy.   In the same spirit of recent node embedding techniques[25,35] based on the Skip-Gram model, we propose an event embedding method, which samples neighbourhoods for events from the weighted event graph representation to map them to a lower dimensional space. To assign an environment to an event e_k, we sample its local neighbourhood set N_k, which consists of the set of its first in- (past) and out- (future) neighbours (also called its *predecessors* and *successors* from now on). The sampling is done according to probabilities determined by the two types of weights of the links that connect the actual event to its neighbours.

In order to consider not only the past but the future of an event in its environment, during sampling we use a combined set of its predecessor and successor events. Further, to balance the contribution of causality and temporal co-occurrence, we introduce a neighbourhood sampling strategy such that the probability $p(e_l)$ of picking an event e_l from the combined neighbourhood set N_k of the central event e_k is given by:

$$p(e_l) = \alpha F(w_{path}(e_k, e_l)) + (1 - \alpha) F(w_{co}(e_k, e_l))$$ (2)

where α is a coefficient between 0 and 1 scaling the contribution of the two types of weights and F is a normalised weighted function defined as:

$$F(w) = \frac{w(e_k, e_l)}{\sum_{n \in N_k} w(e_k, e_n)}.$$ (3)

Using such probabilities computed for each neighbour, we sample nb number of environments randomly for each event for an effective training of the model explained next. Each environment contains s events, we call the number s, the length of the environment.

Embedding of empirical temporal networks.   Once the environments have been created, they are passed as inputs to the Skip-Gram model, with parameters fixed to different values according to the analysis to conduct. The result is a d-dimensional vector space in which events are represented by Cartesian coordinates. As an illustration, Fig. 2 shows a 3-dimensional embedded representation of two of the empirical networks we used for the analysis, recorded in a conference and a primary school settings. Our aim with this setting was to investigate the performance of low-dimensional embedding on the one hand, and on the other taking into account equally the effects of causal temporal paths and co-occurrences by setting $\alpha = 0.5$. The environment parameters
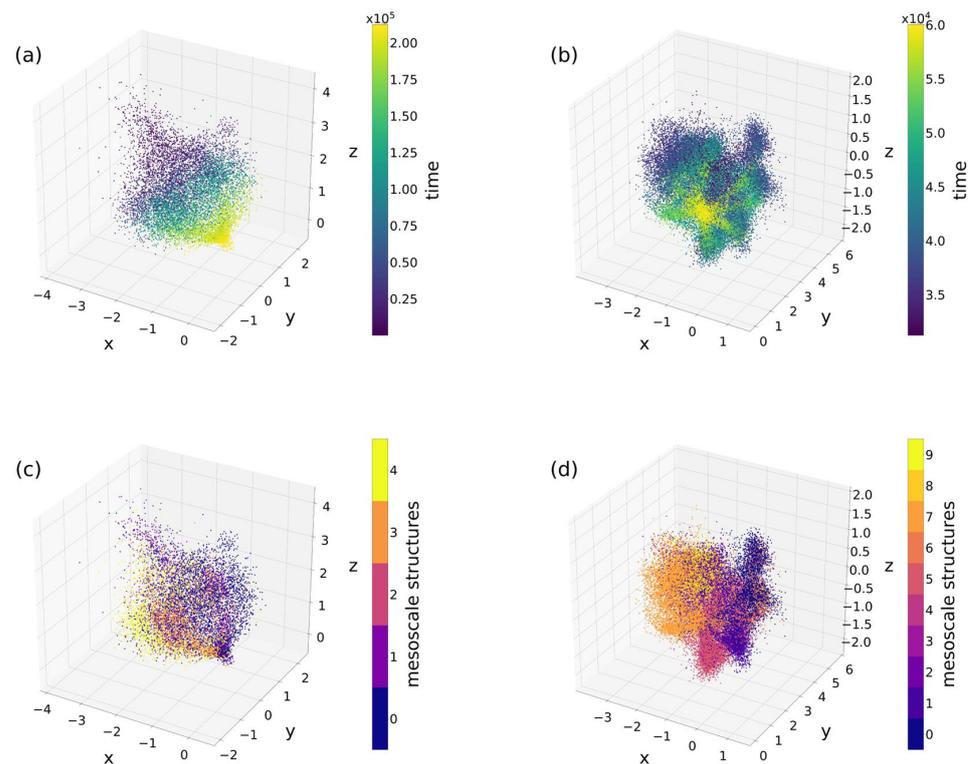
Figure 2. 3-dimensional embeddings of the conference and primary school networks. *x, y* and *z* axes indicate event coordinates, while colour in panels (a and b) shows the time at which the event occurs and in panels (c and d) mesoscale structure membership found using a tensor factorisation method (see Section Tensor Factorisation for mesoscale structure extraction) respectively set to find 5 and 10 of these structures. Hyper parameters were set to $\alpha = 0.5, nb = 10$ and $s = 10$.

nb and *s* have been set both to 10 as an example. In Fig. 2(a,b) each event is represented as a point in the embedded space with colour indicating the time at which they occurred in the original temporal network. Interestingly, the gradient change of colours indicates that these embeddings capture in large part the time ordering of the events. At the same time, Fig. 2(c,d) shows the same 3-dimensional embedded representations, but with colours representing the membership to mesoscale structures detected by tensor factorisation methods applied on the original temporal network[9] (see Section on Tensor Factorisation for mesoscale structure extraction). Evidently, as colours are not distributed randomly but similar colours are somewhat grouped together in space, it suggests that our embedding is able to capture some of these mesoscale structures as well.

We propose an additional microscopic scale analyses of our embedding method in the Supplementary Information. There we studied pairs of events and we were interested in the relation between their time difference measured in the temporal network and their euclidean distance observed in the embedding. Indeed, we found clear correlation between these quantities and demonstrated that the euclidean distances among linked events in the temporal network are significantly smaller than the distances measured between randomly selected events pairs. These observations demonstrate that our method simultaneously captures structural and temporal vicinity of events.

Effects of the dimension and of the neighbourhood sampling.   One of the most important hyper parameter of our method is the number of dimensions of the embedding vector space. Lower than optimal number of dimensions may lead to neglected but otherwise relevant latent correlations in the temporal structure, while overestimation of this number may give us a highly redundant embedded space. We test here the consistency and robustness of our embedding technique in terms of this parameter. We argue that as we increase the number of dimensions, once it reaches and overpasses an optimal number, it starts coding increasingly redundant information in the embedding. As a consequence, further dimensions would not alter the relative positions of embedded events and the pairwise euclidean distances among them would stabilise. To check this assumption, we use an entropy measure capturing the fluctuations of pairwise euclidean distances over several realisations with the same number of embedding dimensions. More precisely, for selected event pairs, we measured the probability distribution of their pairwise euclidean distances over 10 realisations (see Section Entropy Computation) and used it to compute an entropy score. Averaging these scores over all the event pairs provided us an indicator of the stability of the embedding.
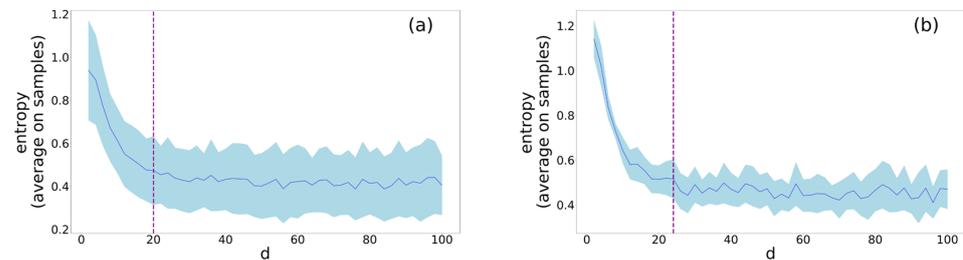
Figure 3. Entropy values with respect to d number of embedding dimensions for the conference (**a**) and the primary school (**b**) networks at $\alpha = 0.5$. The dash line represents the value ($d = 20$ and 24 respectively for (**a** and **b**)) of the optimal embedding dimension in which stability is reached. The blue line and the shaded area represent respectively the average and the variance among the samples we used for the analysis.

r^2					
Data		Original	Snapshot	Timeline	Link
Conference	($d = 20$)	0.79 ± 0.01	0.53 ± 0.04	0.66 ± 0.03	0.57 ± 0.01
Hospital	($d = 14$)	0.53 ± 0.03	0.11 ± 0.02	0.35 ± 0.06	0.50 ± 0.04
High School	($d = 26$)	0.56 ± 0.02	0.23 ± 0.01	0.53 ± 0.02	0.76 ± 0.04
Primary School	($d = 24$)	0.68 ± 0.02	0.12 ± 0.01	0.31 ± 0.02	0.55 ± 0.02

Table 1. R-squared values, r^2 obtained between the simulated and predicted epidemics outcomes using embedding of the real empirical temporal networks and of the randomised model. We set the environment parameters s and nb both to 10. The optimal embedding dimension were chosen as found in Effects of the dimension and of the neighbourhood sampling.

We show results in Fig. 3 for two empirical networks using a balanced embedding with $\alpha = 0.5$ in both cases. As we expected, the entropy decreases as we increase the number of dimensions. This is due to stabilisation of the distribution of pairwise euclidean distances, which in turn gives us a hint on the optimal dimension at which the local neighbourhoods (as defined by the sampling) are well captured by the embedding. To select this optimal dimension, we identified the lower bound at which entropy starts to fluctuate around a constant value (see Methods 4 and Supplementary Information). Note that to identify the optimal number of dimensions we have taken into account other algorithms[37] as explained in the Supplementary Information. However, our final choice fell on the entropy based method we introduced above as it maintained a good trade-off between the compactness (i.e. low dimensionality) and stability of the embedding. Other tested methods suggested an unrealistically high number of dimensions to be optimal, probably due to their incompatibility with the actual setting, as they were developed for word embedding problems.

Spreading process prediction with embedding events. Beyond the demonstrated capacities of our model to capture the temporal ordering and the underlying mesoscopic structures, it may provide further useful information about the embedded events. As a temporal network embedding, it positions events in proximity with similar neighbourhoods. In other words, it can help to identify similar events maybe involving different nodes at different times, but influencing a similar set of other nodes via overlapping temporal paths. As a consequence, this information can be used to predict the outcome of information diffusion processes on temporal networks.

To explore this problem, we model a Susceptible-Infected (SI) process, which is the simplest schematic model of epidemic or information spreading (see Section Spreading Process in Methods). Defined on networks, this model assumes that each node can be in one of two mutually exclusive states (susceptible (S) or infected (I)) at a given time. While initially each node is susceptible, infection can spread from a selected infected seed node/event via temporal interactions and can reach all other nodes via connected valid temporal paths. To obtain the expected outcome of SI process on a temporal network we took each event as the seed and simulated the spreading on the empirical temporal network to measure the final epidemic size in each case. Note that our aim to investigate the final outcome of an epidemic differs from the one pursued with DyANE[34]. In our case, we are not interested in the status of the node time by time, but in the final outcome of the epidemic originated by a specific event. To test the versatility of our embedding method we trained a model using the embedded coordinates of events for epidemic size predictions and compared results directly to the corresponding simulated outcomes. We used linear regression to approximate the correspondence between the embedding coordinates of each event and the size of epidemic initiated from them. As the goodness of the prediction we simply computed the r^2 scores between the predicted and simulated epidemic sizes. Note, that we tested more complicated non-linear models but obtained lower performance in prediction (not shown here). We report our results in Table 1, where we fixed the environment parameters s and nb both to 10 and chose the optimal embedding dimension for each real network detected as we explained in Section Effects of the dimension and of the neighbourhood sampling.

Correlation RRM		Local temporal	Weight- structural	Higher-order temporal	Structural
Original		✓	✓	✓	✓
Snapshot	$P[\mathbf{f}(\mathbf{t}), p_T(\Gamma)]$	×	✓	×	✓
Timeline	$P[L, \mathbf{f}(p_L(\Theta))]$	✓	×	×	✓
Link	$P[\mathbf{f}(L), p_L(\Theta)]$	✓	×	×	×

Table 2. Summary of preserved and eliminated structural and temporal correlations (Local temporal, Weight-structural, Higher-order temporal, and Structural) in the Original and different random reference models (Snapshot, Timeline and Link shuffling) of temporal networks. For further explanation see text.

Real temporal networks are interwoven by several temporal and structural correlations. First there are *local temporal correlations* induced by intrinsic or environment effects emerging between events on the same link leading to bursty behaviour, event trains or circadian activity patterns. Another type is *higher-order temporal correlations* leading to the emergence of causal temporal motifs. *Structural correlations* are responsible for the emerging assortative patterns, communities, or any non-random connection pattern in a social network, while *weight-structural correlations* induce the non-random distribution of strong and weak ties inside and between communities. In Table 2 we summarise which RRM preserves and eliminates which type of correlations.

To get a better sense about the effects of these correlations, we used three types of randomised reference models (RRM)[38] to eliminate combinations of temporal and structural correlations, and to identify which of them are determinant for the prediction task. These RRMs were

- *Snapshot shuffling*, which randomises the timestamps of events in order to eliminate any temporal correlation between them inducing burstiness, causal motifs, or group activation etc. This model is assigned as $P[f(t), p_T(\Gamma)]$ using the notation convention introduced in[38].
- *Timeline shuffling* takes the complete timeline of events between a connected pair of nodes in the temporal network and swap it with the timeline of another randomly selected connected pair of nodes. This shuffling method, noted as $P[L, f(p_L(\Theta))]$ in[38], eliminates all correlations between the underlying structure and timelines while also vanish any casual correlations between events on adjacent links.
- *Link shuffling* method (noted as $P[f(L), p_L(\Theta)]$ in[38]) randomises links of the underlying aggregated static network first to obtain a Bernoulli random structure, and then reassign randomly the original timelines of events to the new links without replacement. In this way, it destroys any structural and structural-temporal correlations in the network, while keeping local temporal correlations like burstiness unaltered.

For a summary of present and eliminated correlations in the different RRMs see Table 2.

The prediction results for the original and the RRM networks are summarised in Table 1, where we depict the observed average r^2 values with their standard deviation computed over the embedding realisations. These results suggest that in general the randomised model embeddings perform worse in predicting the final epidemic size with respect to the original network embeddings. In one way this is straightforward as some correlations have been eliminated from RRMs, which might be determinant for the prediction task. On the other hand, RRMs also appear with a less complex structure and limited irrelevant dependencies and noise, which in turn may help the prediction. It is the snapshot shuffling method, which leads consistently to a significant drop in performance, suggesting that temporal correlations (local or higher-order) are very important for the spreading process and that our embedding can capture these dependencies successfully. Timeline shuffling, which destroys weight-structural and higher-order temporal correlations but conserve the dynamics on links and the underlying network seems to perform better as compared to the snapshot shuffling method but yet worse than the original network. This suggests that while structural correlations can be captured by the embedding, local temporal correlations might be better predictors than weight-structural correlations. Interestingly, the link shuffling method, which conserves only local timeline dynamics on links, performs the best among the RRMs, sometimes even better than the original dataset. Consequently, indeed local temporal event dynamics is the most important feature of the temporal network, while removing structural correlations the system becomes homogeneous and easier to predict (for a supporting analysis see Supplementary Information). Although these general conclusions seem to be consistent over the several analysed datasets, results computed in different settings may have some fluctuations. As explained in the Supplementary Information, this can be partially explained by the variance of the epidemic size distribution reflecting the fluctuation of epidemic size started from different times and events. In most of the cases smaller variance of epidemic size correlates with higher predictive performance except for the link shuffling method, as explained in the Supplementary Information.

As a general conclusion we showed that the embedding successfully captures a combination of temporal and structural features of the network. On the other hand, the fact that temporal and structural features can be entangled has an impact on embedding performance but not on what the embedding is able to learn about them. For instance, the model can under-perform in a community-rich network where nodes of the same community have totally uncorrelated activities, while can provide precise predictions in shuffled datasets if structural and temporal correlations code redundant information.

In Methods in Section Parameter dependencies we also analyse the impact of changing the environment parameters nb and s, the dimension of the embedding d and the hyper parameter α on the predictability of the epidemic spreading outcome (i.e. with respect to the r^2 score). The tuning of the parameters for the best prediction gives a hint about the sensitivity of the prediction task on local properties of the networks and the sampling parameters.
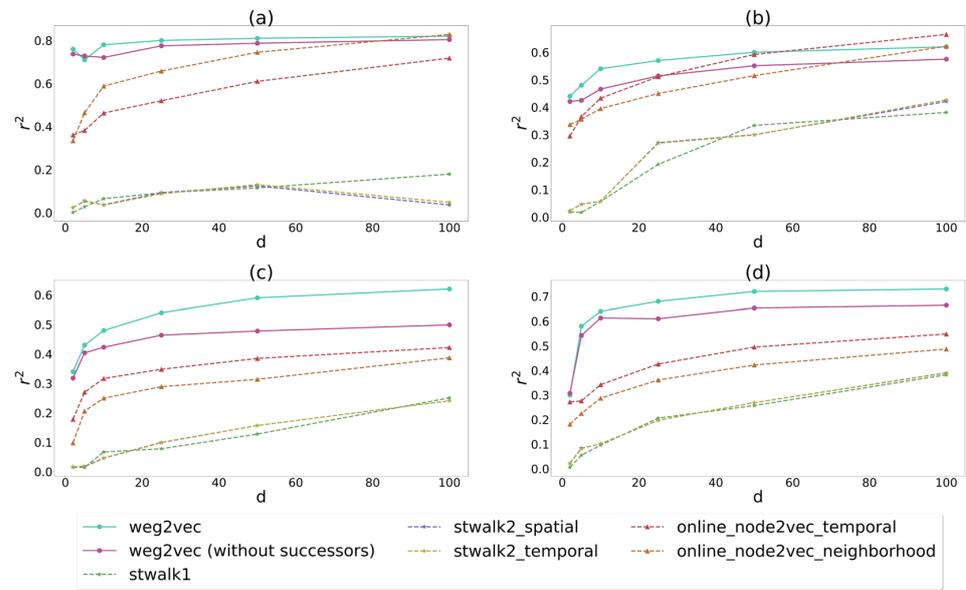
Figure 4. Comparison of STWalk, Online-Node2vec and our embedding methods in predicting spreading outcomes on empirical networks in different settings as (**a**) conference, (**b**) hospital, (**c**) high school, and (**d**) primary school. Results shown are r^2 scored obtained from linear regression on coordinates in embedding spaces with various dimensions computed for each method and empirical temporal networks.

Comparison with other methods. There are a few other recently proposed dynamical network embedding methods, which can be used for the prediction of spreading outcome. Here we consider two of the most promising ones, the STWalk[30,39], and the Online-Node2vec embedding methods[40,41] to compare their predictive performances to weg2vec. Both methods are thought to build node embeddings for dynamic graphs using the Skip-Gram model, which introduces a significant difference to our event embedding method.

STWalk is designed to learn trajectory representations of nodes in temporal graphs by operating with two graph representations, a graph at a given time step and a graph from past time steps. It performs random walks respectively called space-walk and time-walk, to sample environments to input for the Skip-Gram embedding. The authors propose two variants of STWalk, different in the way the environment is built. In STWalk1, space-walk and time-walk are performed as part of a single step on a combined graph, while in STWalk2, space- and time-walks are done separately.

Online-Node2vec is a node embedding method updating coordinates each time a new event appears in a temporal network. It also applies random walks to generate environments possibly using two strategies, the Temporal Walk algorithm and the Temporal Neighbourhood algorithm. In the Temporal Walk algorithm[42] a temporal path based centrality metric is used to capture similarity between nodes by projecting nodes on the same temporal path close to each other in the embedding. In the Temporal Neighbourhood algorithm[43], node similarity is inferred via a fingerprinting method, which projects nodes with similar neighbourhoods close to each other.

To compare the performance of the different methods, we test all of them on the four empirical networks introduced earlier. The environment parameter nb and s have been set to 10 and 10 for all cases to give them the same amount of information to learn and for a fair comparison of outcome. Further, we fix the balance parameter α to 0.5. We then compute the average r^2 scores of simulated spreading outcomes as we vary the number of embedding dimensions. Since STWalk and Online-Node2vec use only the past and the present as basis for the nodes environment, we run the simulation for our methods using only the predecessors for each event as well (see Section Neighbourhood sampling strategy). Finally, as previously, we estimate the epidemic size by using the coordinates of the actual embedding in a linear regression model (see Section Spreading Process).

According to the results in Fig. 4, our method outperforms all the other methods on any of the networks for a broad range of dimensions. The performance improves if we also consider the successors and not only the predecessors in building the environment, as expected. The exception is the hospital network, where our method gets lower scores with respect to Online-Node2vec for dimensions 50 or larger. In general, the difference in the scores can be explained due to the advantage of event embedding instead of node embedding. Indeed if we are looking at epidemic spreading mediated by temporal interactions, it becomes more natural to work with events. In the case of STWalk, the lower scores can be partly explained by the selection of the environments that are allowed to included higher-order correlations among nodes. This more complex information coded in the environments can appear less relevant or noisy for the learning task here. In case of Online-Node2vec, the relative under-performance can be due to the fact that information of the temporal and neighbourhood information are considered separately instead. Missing to join these two aspects can lead to limited information and prediction capacities.

Discussion

Embedding of networks has recently drawn a lot of attention as it both provides lower dimensional representations of networks and proves to be efficient to resolve task such as link prediction, node classification or anomaly detection. Here we proposed an embedding technique of temporal networks, a domain still little explored as most low dimensional representations of networks have been introduced for static networks. Moreover, instead of generating node embedding, we are creating link embedding, which we show to be much more efficient when we try to solve a task linked to spreading process compared to node embedding techniques.

The embedding method we introduced has the advantage to be very simple, it relies on the sampling of neighbourhood on a higher order static representation of the temporal networks and the use of the Skip-Gram model, largely developed for word embedding. The neighbourhood sampling was built so that it takes into account both the notions of causal temporal paths and co-occurrences, meaning that events that are on the same temporal paths and that tend to co-occur are projected close in space. We have shown that an embedding based on this neighbourhood sampling is particularly efficient to provide compact representations of temporal networks that retain essential features of the networks such as the time ordering and the organisation of networks in mesoscale structures. Here, we also provide a way to choose a relevant dimension for the embedding. Along with this, we show that the learned representations retain enough information of the original network to get a relevant estimate of the outcome of spreading process. Interestingly, this observation remains true even if the sampling strategy uses only information from the past for each event. This means that the technique introduced can be also used as an online method taking into account temporal events on the run. Moreover, tuning the neighbourhood sampling, i.e. playing on the trade-off between including causal temporal paths and co-occurrences, to get the best performance for the prediction of the outcome of the spreading process can be used to detect the relevant properties of the original network for the spreading process.

For future works, it would worth exploring other sampling strategies that decouples the purely structural properties, i.e. the presence of the communities in the aggregated network, from the temporal properties. Another important follow-up of this work would be the application of this embedding technique to solve questions such as the detection of key events in misinformation spreading.

Methods

Entropy Computation. To measure the entropy over the distributions of the euclidean distances between pairs of event coordinates in each of the temporal network embedding, we build 10 embeddings for 50 different dimensions (from 2 to 100 at step 2), setting $\alpha = 0.5$. The hyper parameters nb and s were set to 10 and 10 respectively. For each embedding, we then divide our dataset in 10 consecutive samples of 1000 consecutive events each, both to avoid to compute all the pairwise distances (which would be very costly computationally) and at the same time to have a representative set of events.

We compute the euclidean distance of each pair of events in the samples for each of the 10 different embeddings. We then bin the distances into $k = 10$ bins ranging from the global maximum and the global minimum values over all the possible dimension and realisations of the embedding for the same network, and measure the entropy over these sets of distances as

$$H = \sum_k p_k \log p_k,$$

(4)

where p_k represents the probability associated with the k_{th} bin. In Fig. 3 (see also Supplementary Information) the blue curve represents the entropy values with respect to the number of embedding dimensions, averaged on the 10 samples as described above and the shaded surrounding area shows the variance among the 10 samples. The vertical dash line corresponds to the dimension at which the embedding stabilises. To determine this point we looked for the best fit of a horizontal line on the average entropy curve and took the value of the first interception of the curve with its fit.

Spreading Process. The simplest epidemic models are based on the assumption that the population can be divided into compartments, each representing a phase of the disease[44–48]. The one we used for our analysis, the Susceptible-Infected model, which foresees that once a healthy node (in state S) is exposed to the infection, it will become infected (state I) with a given rate β and will never return to the original healthy state. In our specific case, the SI model has been implemented in such a way that $\beta = 1$, which defines a a deterministic process from the spreading point of view.

Taking each event as the starting of the epidemic, we simulate the epidemic spreading on the temporal network and we assign the final epidemic size to each seed event. We build a dataset containing embedding coordinates of each event, the associated epidemic size that will be the target to be predicted, the square of each coordinate and the euclidean distance from each event in the network and the first event in time, that will be used as regressors. In other terms, we assume a linear relationship between the epidemic size (y) and the embedding coordinates ($\vec{x}$) defined by the so-called *regression equation*

$$y = \beta_0 + \sum_{i=1}^{r} \beta_i x_i$$

(5)

where $x_1, \ldots, x_r$ are the predictors (the embedding coordinates, their square and the euclidean distance) and $\beta_0, \ldots, \beta_r$ are the regression coefficients. For training we operate with a 10-fold cross-validation, i.e. we first randomly partitioned the original sample into 10 equal sized sub-samples and retain a single sub-sample as the validation data for testing the model, while using the remaining 9 sub-samples for the actual training. We repeat this
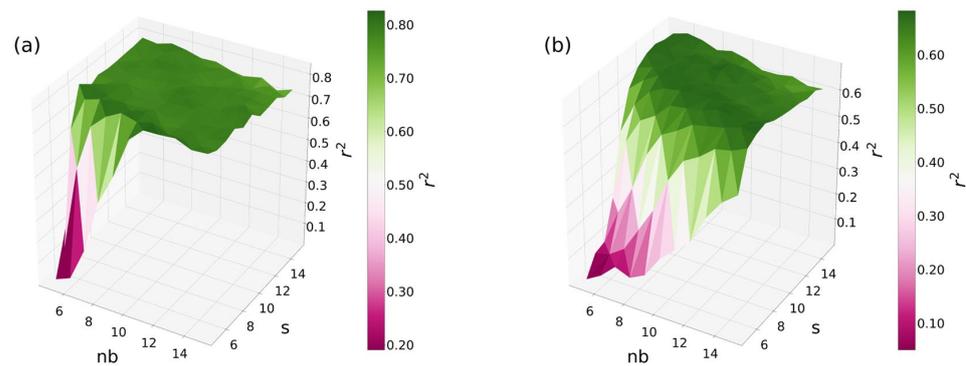
Figure 5. R-squared values, r^2, dependency on the *nb* number (x-axis) and *s* size (y-axis) of sampled environments. Figure (**a**) shows results for the conference network and Figure (**b**) for the primary school network. Colours and z-axis code the obtained average r^2 score values for given *nb* and *s* parameter pairs computed over 10 realisations. α was fixed to 0.5; we set $d = 20$ for Figure (**a**) and $d = 24$ for Figure (**b**) - see Fig. 3.

process 10 times to train the embedding to best learn the coordinates of each event in the network. We use the coefficient of determination, denoted as r^2, to understand which amount of variation in *y* can be explained by the dependence on $\overrightarrow{x}$ using the particular regression model. Larger r^2 indicates a better fit and means that the model can better explain the variation of the output with different inputs.

Tensor Factorisation for mesoscale structure extraction. A temporal network can be fully described by a time-dependent adjacency matrix, where its entries are either one or zero depending on the presence or absence of interactions of pair of nodes at a given time. This matrix can be seen as a three-way tensor, whose size is $n \times n \times t$ (*n* indicates the number of nodes and *t* the number of snapshots in the temporal network). This tensor can then be factorised into a sum of *r* rank-one tensors, i.e. number of mesoscale structures we search into the temporal network. Using this technique we can group events into mesoscale structures. Indeed with the decomposition, we now have a value for each link for each time for each rank-1 tensor, we assign a mesoscale structure to each event based on these values. Basically the mesoscale structure assigned to each event is the one for which the corresponding link at the corresponding time has the highest value. To find the optimal number of mesoscale structures, we used the core consistency metric[49]. It is based on scrutinising the *appropriateness* of the structural model based on the data and the estimated parameters of gradually augmented models. A model is called *appropriate* if adding other combinations of the same components does not improve the fit considerably. In practice, we operate different tensor decompositions for different value of the rank (ranging from 2 to 20, for all the networks) in order to estimate the best value for it.

Parameter dependencies. Here, we explain the way we investigate how hyper parameters of the environment sampling may impact the prediction score on different real networks. Figure 5 shows the r^2 scores computed for the conference and primary school networks with respect to the length *s* and number *nb* of environments sampled for each event. For these computations we fixed $\alpha = 0.5$ and the embedding dimensions to their optimal values. According to Fig. 5 on the conference network, except for very small number or length of contexts we see an emerging plateau of r^2 values. This means that the environment size compensates for the number of environments (or vice versa) when we measure the embedding performance. In other terms, increasing the length of the environment has the same effect than increasing the number of environments on the r^2 score. For the primary school network we observe a similar but somewhat weaker compensation effect (Fig. 5), while we observed the same behaviour even for the hospital and the high school networks (see Supplementary Information). These results suggest that these two parameters are highly redundant, thus we can effectively reduce our parameter set by fixing both to a large enough value.

In order to investigate the influence of the embedding dimension and the α sampling balance parameter, next we fix the environment parameters to $s = 10$ and $nb = 10$ based on our evaluation above. As shown in Fig. 6 (and in Supplementary Information for other networks) both increasing the number of embedding dimensions and α lead to better performances in predicting the spreading outcome. As the function of the number of dimensions, each case reaches a plateau in accordance with our earlier results presented in Section Effects of the dimension and of the neighbourhood sampling. On the other hand we observe somewhat stronger dependencies on α. While for the conference and the hospital networks the more one increases α the better the prediction gets, for the primary school and the high school networks the score reaches a plateau and become less sensitive to the change of α (see Fig. 6 and Supplementary Information). If we consider lower values of α, the similarity we capture between the event between adjacent events is mainly based on the co-occurrences, which are more relevant in school networks where participants might be active at the same time (e.g. in breaks between classes). This argument only
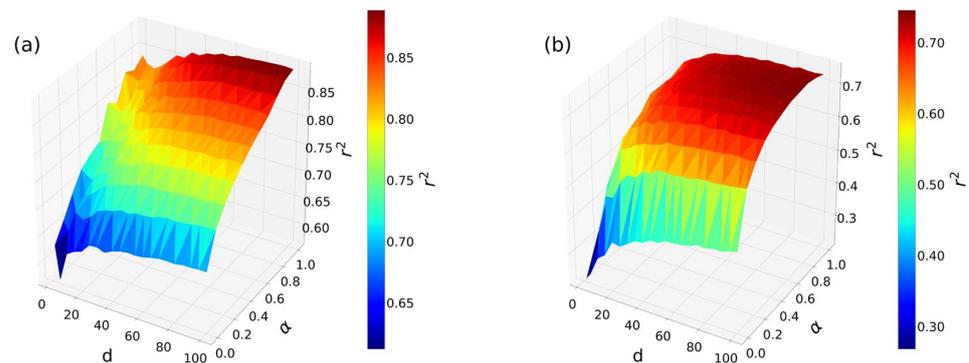
Figure 6. R-squared values, r^2, as the function of the d number of embedding dimensions (x-axis) and α sampling balance (y-axis) parameters. Figure (**a**) shows results for the conference network and Figure (**b**) for the primary school network. Colours and z-axis code the obtained average r^2 score values for a given d and α parameter pairs computed over 10 realisations. Other hyper parameters were fixed to $nb = 10$ and $s = 10$.

moderately applies to a conference or hospital where simultaneous interactions typically happen in smaller groups or not at all. Higher values of α imbalance the sampling to contain more information about temporal paths, which actually indirectly codes co-occurrence frequencies as well. This gives the advantage to the model to learn both type of similarities and to predict the epidemic outcomes with higher precision.

References

1. Bar-Yam, Y. *Dynamics of complex systems*, vol. 213 (Addison-Wesley Reading, MA, 1997).
2. Newman, M. E. The structure and function of complex networks. *SIAM review* **45**, 167–256 (2003).
3. Albert, R. & Barabási, A.-L. Statistical mechanics of complex networks. *Reviews of modern physics* **74**, 47 (2002).
4. Holme, P. & Saramäki, J. Temporal networks. *Physics reports* **519**, 97–125 (2012).
5. Masuda, N. & Lambiotte, R. *A Guide to Temporal Networks* (World Scientific, 2016).
6. Perra, N., Gonçalves, B., Pastor-Satorras, R. & Vespignani, A. Activity driven modeling of time varying networks. *Scientific reports* **2**, 469 (2012).
7. Karsai, M., Perra, N. & Vespignani, A. Time varying networks and the weakness of strong ties. *Scientific reports* **4**, 4001 (2014).
8. Laurent, G., Saramäki, J. & Karsai, M. From calls to communities: a model for time-varying social networks. *The European Physical Journal B* **88**, 301 (2015).
9. Gauvin, L., Panisson, A. & Cattuto, C. Detecting the community structure and activity patterns of temporal networks: a non-negative tensor factorization approach. *PloS one* **9**, e86028 (2014).
10. Peixoto, T. P. & Rosvall, M. Modelling sequences and temporal networks with dynamic community structures. *Nature communications* **8**, 582 (2017).
11. Barabási, A.-L. The origin of bursts and heavy tails in human dynamics. *Nature* **435**, 207–211 (2005).
12. Karsai, M., Jo, H.-H. & Kaski, K. *Bursty human dynamics* (Springer, 2018).
13. Kovanen, L., Karsai, M., Kaski, K., Kertész, J. & Saramäki, J. Temporal motifs in time-dependent networks. *Journal of Statistical Mechanics: Theory and Experiment* P11005 (2011).
14. Paranjape, A., Benson, A. R. & Leskovec, J. Motifs in temporal networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, 601–610 (ACM, 2017).
15. Masuda, N. & Holme, P. *Temporal Network Epidemiology* (Springer, 2017).
16. Karsai, M. *et al.* Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E* **83**, 025102 (2011).
17. Gauvin, L., Panisson, A., Cattuto, C. & Barrat, A. Activity clocks: spreading dynamics on temporal networks of human contact. *Scientific reports* **3**, 3099 (2013).
18. Delvenne, J.-C., Lambiotte, R. & Rocha, L. E. Diffusion on networked systems is a question of time or structure. *Nature communications* **6**, 7366 (2015).
19. Scholtes, I. *et al.* Causality-driven slow-down and speed-up of diffusion in non-markovian temporal networks. *Nature communications* **5**, 5024 (2014).
20. Benson, A. R., Gleich, D. F. & Leskovec, J. Higher-order organization of complex networks. *Science* **353**, 163–166 (2016).
21. Evans, T. & Lambiotte, R. Line graphs, link partitions, and overlapping communities. *Physical Review E* **80**, 016105 (2009).
22. Ahn, Y.-Y., Bagrow, J. P. & Lehmann, S. Link communities reveal multiscale complexity in networks. *nature* **466**, 761 (2010).
23. Petri, G., Scolamiero, M., Donato, I. & Vaccarino, F. Topological strata of weighted complex networks. *PloS one* **8**, e66506 (2013).
24. Perozzi, B., Al-Rfou, R. & Skiena, S. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 701–710 (ACM, 2014).
25. Grover, A. & Leskovec, J. node2vec: Scalable feature learning for networks. In *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*, 855–864 (ACM, 2016).
26. Hamilton, W., Ying, Z. & Leskovec, J. Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, 1024–1034 (2017).
27. Kipf, T. N. & Welling, M. Semi-supervised classification with graph convolutional networks. *arXiv preprint arXiv:1609.02907* (2016).
28. Kivelä, M., Cambe, J., Saramäki, J. & Karsai, M. Mapping temporal-network percolation to weighted, static event graphs. *Scientific reports* **8**, 12357 (2018).
29. Mellor, A. The temporal event graph. *Journal of Complex Networks* (2017).

30. Pandhre, S., Mittal, H., Gupta, M. & Balasubramanian, V. N. Stwalk: learning trajectory representations in temporal graphs. In *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data*, 210–219 (ACM, 2018).
31. Singer, U., Guy, I. & Radinsky, K. Node embedding over temporal graphs. *arXiv preprint arXiv:1903.08889* (2019).
32. Béres, F., Pálovics, R., Kelen, D., Szabó, D. & Benczúr, A. Node embeddings in dynamic graphs. *7th International Conference on Complex Networks and Their Applications, Cambridge* 178–180 (2018).
33. Kumar, S., Zhang, X. & Leskovec, J. Learning dynamic embeddings from temporal interactions. *arXiv preprint arXiv:1812.02289* (2018).
34. Sato, K., Oka, M., Barrat, A. & Cattuto, C. DyANE: Dynamics-aware node embedding for temporal networks. *arXiv e-prints* arXiv:1909.05976 (2019).
35. Mikolov, T., Chen, K., Corrado, G. & Dean, J. Efficient estimation of word representations in vector space. *arXiv preprint arXiv:1301.3781* (2013).
36. Cattuto, C. *et al.* Dynamics of person-to-person interactions from distributed rfid sensor networks. *PloS one* **5**, e11596 (2010).
37. Yin, Z. Understand functionality and dimensionality of vector embeddings: the distributional hypothesis, the pairwise inner product loss and its bias-variance trade-off. *arXiv preprint arXiv:1803.00502* (2018).
38. Gauvin, L. *et al.* Randomized reference models for temporal networks. *arXiv preprint arXiv:1806.04032* (2018).
39. Supriya Pandhre, M. G., Mittal, H. & Balasubramanian, V. N. Stwalk, https://github.com/supriya-pandhre/STWalk (Last Access: April 2019).
40. Ferenc Béres, D. M. K. D. S., Pálovics, R. & Benczúr, A. A. Node embeddings in dynamic graphs. In *Book of Abstracts of the 7th International Conference on Complex Networks and Their Applications*, 165–167 (2018).
41. Ferenc Béres, D. M. K. D. S., Pálovics, R. & Benczúr, A. A. Online node2vec, https://github.com/mislam5285/online-node2vec (Last Access: April 2019).
42. Béres, F., Pálovics, R., Oláh, A. & Benczúr, A. A. Temporal walk based centrality metric for graph streams. *Applied network science* **3**, 32 (2018).
43. Fogaras, D. & Rácz, B. Scaling link-based similarity search. In *Proceedings of the 14th international conference on World Wide Web*, 641–650 (ACM, 2005).
44. Anderson, R. M. & May, R. M. *Infectious diseases of humans: dynamics and control* (Oxford university press, 1992).
45. Bailey, N. T. *et al.* *The mathematical theory of infectious diseases and its applications* (Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975).
46. Daley, D., Gani, J. & Yakowitz, S. An epidemic with individual infectivities and susceptibilities. *Mathematical and computer modelling* **32**, 155–167 (2000).
47. Diekmann, O. & Heesterbeek, J. A. P. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation*, vol. 5 (John Wiley & Sons, 2000).
48. Barrat, A., Barthelemy, M. & Vespignani, A. *Dynamical processes on complex networks* (Cambridge university press, 2008).
49. Bro, R. & Kiers, H. A. A new efficient method for determining the number of components in parafac models. *Journal of Chemometrics: A Journal of the Chemometrics Society* **17**, 274–286 (2003).

Acknowledgements

Author contributions

L.G. and M.K. designed research; L.G., M.T. and M.K. performed research; M.T. analyzed data; and L.G., M.T. and M.K. wrote the paper.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-63221-2.

Correspondence and requests for materials should be addressed to L.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Figures

Tables

Bibliography

[1] David Lazer et al. "Social science. Computational social science." In: *Science (New York, NY)* 323.5915 (2009), pp. 721–723.

[2] R Michael Alvarez. *Computational social science.* Cambridge University Press, 2016.

[3] G Nigel Gilbert. *Computational social science.* Vol. 21. Sage, 2010.

[4] Royce M Kimmons. "Understanding collaboration in Wikipedia". In: *First Monday* (2011).

[5] Garren Gaut et al. "Improving Government Response to Citizen Requests Online". In: *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies.* 2018, pp. 1–10.

[6] Ciro Cattuto. "Data Science for Social Good: Opportunities and Challenges". In: *INFORMATIK 2019: 50 Jahre Gesellschaft für Informatik–Informatik für Gesellschaft* (2019).

[7] Samuel Carton et al. "Identifying police officers at risk of adverse events". In: *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining.* 2016, pp. 67–76.

[8] Jennifer Helsby et al. "Early intervention systems: Predicting adverse interactions between police and the public". In: *Criminal justice policy review* 29.2 (2018), pp. 190–209.

[9]   Alexandra Olteanu et al. "Social data: Biases, methodological pit-
       falls, and ethical boundaries". In: *Frontiers in Big Data* 2 (2019),
       p. 13.

[10] Ferdinand Toennies. "Gemeinschaft und Gesellschaft, Leipzig". In:
       Literature/Literaturverzeichnis 217 (1887).

[11] David Knoke and Song Yang. *Social network analysis.* Sage Publi-
       cations, 2019.

[12] Stanley Wasserman and Katherine Faust. *Social network analysis:
       Methods and applications.* Vol. 8. Cambridge university press, 1994.

[13] Francesco Martino and Andrea Spoto. "Social Network Analysis:
       A brief theoretical review and further perspectives in the study of
       Information Technology." In: *PsychNology J.* 4.1 (2006), pp. 53–86.

[14] John Scott. "Social network analysis". In: *Sociology* 22.1 (1988),
       pp. 109–127.

[15] Olivier Serrat. "Social network analysis". In: *Knowledge solutions.*
       Springer, 2017, pp. 39–43.

[16] Jacob Levy Moreno. "Sociometry and the cultural order". In: *So-
       ciometry* 6.3 (1943), pp. 299–344.

[17] Fritz Heider. "Attitudes and cognitive organization". In: *The Jour-
       nal of psychology* 21.1 (1946), pp. 107–112.

[18] Dénes König. "Theorie der endlichen und unendlichen Graphen,
       Akad". In: *Verlagsgeselschaft, Leipzig* (1936).

[19] Dorwin Cartwright and Frank Harary. "Structural balance: a gen-
       eralization of Heider's theory." In: *Psychological review* 63.5 (1956),
       p. 277.

[20] Frank Harary, Robert Zane Norman, and Dorwin Cartwright. *Struc-
       tural models: An introduction to the theory of directed graphs.* Wiley,
       1965.

[21] Mark S Granovetter. "The strength of weak ties". In: *American
       journal of sociology* 78.6 (1973), pp. 1360–1380.

[22] Elton Mayo. *The social problems of an industrial civilization, including, as an appendix: The political problem of industrial civilization.* Division of Research, Graduate School of Business Administration, Harvard . . ., 1945.

[23] William Lloyd Warner and Paul Sanborn Lunt. "The social life of a modern community." In: (1941).

[24] Charles F Manski. "Economic analysis of social interactions". In: *Journal of economic perspectives* 14.3 (2000), pp. 115–136.

[25] Aaron V Cicourel. "Cognitive sociology: Language and meaning in social interaction." In: (1974).

[26] Anthony Giddens et al. *Introduction to sociology.* WW Norton New York, NY, 1996.

[27] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. "Statistical physics of social dynamics". In: *Reviews of modern physics* 81.2 (2009), p. 591.

[28] M Weber. "Wirtschaft und gesellschaft (Trad. Economia e società) Milano". In: *Comunità* (1961).

[29] Leonhard Euler. "Solutio problematis ad geometriam situs pertinentis". In: *Commentarii academiae scientiarum Petropolitanae* (1741), pp. 128–140.

[30] Paul Erdős and Alfréd Rényi. "On the evolution of random graphs". In: ().

[31] Duncan J Watts and Steven H Strogatz. "Collective dynamics of 'small-world'networks". In: *nature* 393.6684 (1998), pp. 440–442.

[32] Albert-László Barabási and Réka Albert. "Emergence of scaling in random networks". In: *science* 286.5439 (1999), pp. 509–512.

[33] Réka Albert and Albert-László Barabási. "Statistical mechanics of complex networks". In: *Reviews of modern physics* 74.1 (2002), p. 47.

[34] Petter Holme and Jari Saramäki. "Temporal networks". In: *Physics reports* 519.3 (2012), pp. 97–125.

[35] Márton Karsai, Nicola Perra, and Alessandro Vespignani. "Time varying networks and the weakness of strong ties". In: *Scientific reports* 4 (2014), p. 4001.

[36] Jean-Pierre Eckmann, Elisha Moses, and Danilo Sergi. "Entropy of dialogues creates coherent structures in e-mail traffic". In: *Proceedings of the National Academy of Sciences of the United States of America* 101.40 (2004), pp. 14333–14337.

[37] Nathan Eagle and Alex Sandy Pentland. "Reality mining: sensing complex social systems". In: *Personal and ubiquitous computing* 10.4 (2006), pp. 255–268.

[38] Chuang Liu and Zi-Ke Zhang. "Information spreading on dynamic social networks". In: *Communications in Nonlinear Science and Numerical Simulation* 19.4 (2014), pp. 896–904.

[39] Soroush Vosoughi, Deb Roy, and Sinan Aral. "The spread of true and false news online". In: *Science* 359.6380 (2018), pp. 1146–1151.

[40] Dashun Wang et al. "Information spreading in context". In: *Proceedings of the 20th international conference on World wide web*. ACM. 2011, pp. 735–744.

[41] Tomas Mikolov et al. "Efficient estimation of word representations in vector space". In: *arXiv preprint arXiv:1301.3781* (2013).

[42] Peng Cui et al. "A survey on network embedding". In: *IEEE Transactions on Knowledge and Data Engineering* (2018).

[43] Palash Goyal et al. "Embedding networks with edge attributes". In: *Proceedings of the 29th on Hypertext and Social Media*. ACM. 2018, pp. 38–42.

[44] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. "Deepwalk: Online learning of social representations". In: *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM. 2014, pp. 701–710.

[45] Aditya Grover and Jure Leskovec. "node2vec: Scalable feature learning for networks". In: *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining.* ACM. 2016, pp. 855–864.

[46] Will Hamilton, Zhitao Ying, and Jure Leskovec. "Inductive representation learning on large graphs". In: *Advances in Neural Information Processing Systems.* 2017, pp. 1024–1034.

[47] Thomas N Kipf and Max Welling. "Semi-supervised classification with graph convolutional networks". In: *arXiv preprint arXiv:1609.02907* (2016).

[48] Vinay Setty and Katja Hose. "Event2Vec: Neural Embeddings for News Events". In: *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval.* ACM. 2018, pp. 1013–1016.

[49] Yuan Zuo et al. "Embedding Temporal Network via Neighborhood Formation". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining.* ACM. 2018, pp. 2857–2866.

[50] Palash Goyal and Emilio Ferrara. "Graph embedding techniques, applications, and performance: A survey". In: *Knowledge-Based Systems* 151 (2018), pp. 78–94.

[51] Supriya Pandhre et al. "Stwalk: learning trajectory representations in temporal graphs". In: *Proceedings of the ACM India Joint International Conference on Data Science and Management of Data.* ACM. 2018, pp. 210–219.

[52] Uriel Singer, Ido Guy, and Kira Radinsky. "Node Embedding over Temporal Graphs". In: *arXiv preprint arXiv:1903.08889* (2019).

[53] Ferenc Béres et al. "Node Embeddings in Dynamic Graphs". In: *7th International Conference on Complex Networks and Their Applications, Cambridge* (Dec. 2018), pp. 178–180.

[54]   Srijan Kumar, Xikun Zhang, and Jure Leskovec. "Learning Dynamic Embeddings from Temporal Interactions". In: *arXiv preprint arXiv:1812.02289* (2018).

[55]   Laetitia Gauvin, André Panisson, and Ciro Cattuto. "Detecting the community structure and activity patterns of temporal networks: a non-negative tensor factorization approach". In: *PloS one* 9.1 (2014), e86028.

[56]   Koya Sato et al. "Predicting partially observed processes on temporal networks by Dynamics-Aware Node Embeddings (DyANE)". In: *EPJ Data Science* 10.1 (2021), pp. 1–20.

[57]   Maddalena Torricelli, Márton Karsai, and Laetitia Gauvin. "weg2vec: Event embedding for temporal networks". In: *Scientific reports* 10.1 (2020), pp. 1–11.

[58]   Mikko Kivelä et al. "Mapping temporal-network percolation to weighted, static event graphs". In: *Scientific reports* 8.1 (2018), p. 12357.

[59]   Andrew Mellor. "The temporal event graph". In: *Journal of Complex Networks* (2017).

[60]   Ciro Cattuto et al. "Dynamics of person-to-person interactions from distributed RFID sensor networks". In: *PloS one* 5.7 (2010), e11596.

[61]   Ciro Cattuto et al. "Semantic grounding of tag relatedness in social bookmarking systems". In: *International semantic web conference*. Springer. 2008, pp. 615–631.

[62]   Philippe Vanhems et al. "Estimating potential infection transmission routes in hospital wards using wearable proximity sensors". In: *PloS one* 8.9 (2013), e73970.

[63]   Viboud Cécile et al. "High-Resolution Measurements of Face-to-Face Contact Patterns in a Primary School". In: (2011).

[64]   Julie Fournet and Alain Barrat. "Contact patterns among high school students". In: *PloS one* 9.9 (2014), e107878.

[65] Frank Harary and Robert Z Norman. "Some properties of line digraphs". In: *Rendiconti del Circolo Matematico di Palermo* 9.2 (1960), pp. 161–168.

[66] David Harris and Sarah L Harris. *Digital design and computer architecture*. Morgan Kaufmann, 2010.

[67] Simone Montangero, Montangero, and Evenson. *Introduction to Tensor Network Methods*. Springer, 2018.

[68] Frank L Hitchcock. "The expression of a tensor or a polyadic as a sum of products". In: *Journal of Mathematics and Physics* 6.1-4 (1927), pp. 164–189.

[69] Andrzej Cichocki et al. *Nonnegative matrix and tensor factorizations: applications to exploratory multi-way data analysis and blind source separation*. John Wiley & Sons, 2009.

[70] Amnon Shashua and Tamir Hazan. "Non-negative tensor factorization with applications to statistics and computer vision". In: *Proceedings of the 22nd international conference on Machine learning*. 2005, pp. 792–799.

[71] Daniel M Dunlavy, Tamara G Kolda, and Evrim Acar. "Temporal link prediction using matrix and tensor factorizations". In: *ACM Transactions on Knowledge Discovery from Data (TKDD)* 5.2 (2011), pp. 1–27.

[72] Maximilian Nickel, Volker Tresp, and Hans-Peter Kriegel. "A three-way model for collective learning on multi-relational data". In: *Icml*. 2011.

[73] Tamara G Kolda and Brett W Bader. "Tensor decompositions and applications". In: *SIAM review* 51.3 (2009), pp. 455–500.

[74] Rasmus Bro and Henk AL Kiers. "A new efficient method for determining the number of components in PARAFAC models". In: *Journal of Chemometrics: A Journal of the Chemometrics Society* 17.5 (2003), pp. 274–286.

[75] Renaud Lambiotte, Lionel Tabourier, and Jean-Charles Delvenne. "Burstiness and spreading on temporal networks". In: *The European Physical Journal B* 86.7 (2013), p. 320.

[76] M Karsai et al. "Small but slow world: How network topology and burstiness slow down spreading". In: *Physical Review E* 83.2 (2011), p. 025102.

[77] Alexei Vazquez et al. "Impact of non-Poissonian activity patterns on spreading processes". In: *Physical review letters* 98.15 (2007), p. 158702.

[78] Timothy L Wiemken and Robert R Kelley. "Machine learning in epidemiology and health outcomes research". In: *Annual review of public health* 41 (2020), pp. 21–36.

[79] William Ogilvy Kermack and Anderson G McKendrick. "A contribution to the mathematical theory of epidemics". In: *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character* 115.772 (1927), pp. 700–721.

[80] David G Kendall et al. "Deterministic and Stochastic Epidemics in Closed Populations". In: *Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 4: Contributions to Biology and Problems of Health*. The Regents of the University of California. 1956.

[81] Roy M Anderson and Robert M May. *Infectious diseases of humans: dynamics and control*. Oxford university press, 1992.

[82] Norman TJ Bailey et al. *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.

[83] DJ Daley, Joseph Gani, and S Yakowitz. "An epidemic with individual infectivities and susceptibilities". In: *Mathematical and computer modelling* 32.1-2 (2000), pp. 155–167.

[84] Odo Diekmann and Johan Andre Peter Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis and interpretation.* Vol. 5. John Wiley & Sons, 2000.

[85] Alain Barrat, Marc Barthelemy, and Alessandro Vespignani. *Dynamical processes on complex networks.* Cambridge university press, 2008.

[86] Laetitia Gauvin et al. "Randomized reference models for temporal networks". In: *arXiv preprint arXiv:1806.04032* (2018).

[87] Manish Gupta Supriya Pandhre Himangi Mittal and Vineeth N Balasubramanian. *STWalk:* `https://github.com/supriya-pandhre/STWalk`. Last Access: April 2019.

[88] Ferenc Béres et al. "Temporal walk based centrality metric for graph streams". In: *Applied network science* 3.1 (2018), p. 32.

[89] Dániel Fogaras and Balázs Rácz. "Scaling link-based similarity search". In: *Proceedings of the 14th international conference on World Wide Web.* ACM. 2005, pp. 641–650.

[90] Katherine Ann Krumme and Sergio Herrero. "Lending behavior and community structure in an online peer-to-peer economic network". In: *2009 International Conference on Computational Science and Engineering.* Vol. 4. IEEE. 2009, pp. 613–618.

[91] Dean Karlan and Jonathan Zinman. "Microcredit in theory and practice: Using randomized credit scoring for impact evaluation". In: *Science* 332.6035 (2011), pp. 1278–1284.

[92] Ghazal Zulfiqar. "Financializing the poor:'dead capital', women's gold and microfinance in Pakistan". In: *Economy and Society* 46.3-4 (2017), pp. 476–498.

[93] Bin Wang, Diego Escobari, and Xiaopeng Wang. "Social Networks in Online Peer-to-Peer Lending: The Case of Event-Type Ties as Pipes and Prisms". In: *Proceedings of the 52nd Hawaii International Conference on System Sciences.* 2019.

[94]   Raymond P Fisk et al. "Crowd-funding: transforming customers
       into investors through innovative service platforms". In: *Journal of
       service management* (2011).

[95]   Ajay Agrawal, Christian Catalini, and Avi Goldfarb. "Some simple
       economics of crowdfunding". In: *Innovation policy and the economy*
       14.1 (2014), pp. 63–97.

[96]   Bill Clinton. *Giving: How each of us can change the world.* Random
       House, 2007.

[97]   Scott E Hartley. "Kiva. org: Crowd-sourced microfinance and coop-
       eration in group lending". In: *Available at SSRN 1572182* (2010).

[98]   Roy W Chen. "Social Identity and Cooperation." PhD thesis. 2012.

[99]   Yang Liu et al. "" I loan because..." understanding motivations for
       pro-social lending". In: *Proceedings of the fifth ACM international
       conference on Web search and data mining.* 2012, pp. 503–512.

[100] Jonathan Meer and Oren Rigbi. "The effects of transactions costs
       and social distance: Evidence from a field experiment". In: *The BE
       Journal of Economic Analysis & Policy* 14.1 (2013), pp. 271–296.

[101] Todd W Moss et al. "Funding the story of hybrid ventures: Crowd-
       funder lending preferences and linguistic hybridity". In: *Journal of
       Business Venturing* 33.5 (2018), pp. 643–659.

[102] Melina Moleskis and Miguel Ángel Canela. "Crowdfunding success:
       The case of Kiva. org". In: (2016).

[103] Reem Hussein Nakity and Faisal Rana. "IMPACT OF MICROFI-
       NANCE IN ALLEVIATING POVERTY IN MENA COUNTRIES".
       In: *PalArch's Journal of Archaeology of Egypt/Egyptology* 18.13
       (2021), pp. 286–293.

[104] Megan Moodie. "Microfinance and the gender of risk: The case of
       Kiva. org". In: *Signs: Journal of Women in Culture and Society* 38.2
       (2013), pp. 279–302.

[105]   Henrik Wesemann and Joakim Wincent. "A whole new world: Counterintuitive crowdfunding insights for female founders". In: *Journal of Business Venturing Insights* 15 (2021), e00235.

[106]   Christina Jenq, Jessica Pan, and Walter Theseira. "Beauty, weight, and skin color in charitable giving". In: *Journal of Economic Behavior & Organization* 119 (2015), pp. 234–253.

[107]   Diyi Yang and Robert E Kraut. "Persuading teammates to give: Systematic versus heuristic cues for soliciting loans". In: *Proceedings of the ACM on Human-Computer Interaction* 1.CSCW (2017), pp. 1–21.

[108]   Thai T Pham and Yuanyuan Shen. "A deep causal inference approach to measuring the effects of forming group loans in online nonprofit microfinance platform". In: *arXiv preprint arXiv:1706.02795* (2017).

[109]   Jiaao Chen and Diyi Yang. "Weakly-supervised hierarchical models for predicting persuasive strategies in good-faith textual requests". In: *arXiv preprint arXiv:2101.06351* (2021).

[110]   Soumajyoti Sarkar. "Centralized Borrower and Lender Matching under Uncertainty for P2P Lending". In: *Companion Proceedings of the Web Conference 2021*. 2021, pp. 269–273.

[111]   Nasim Sonboli et al. "Fairness and Transparency in Recommendation: The Users' Perspective". In: *arXiv preprint arXiv:2103.08786* (2021).

[112]   Wei Ai et al. "Recommending teams promotes prosocial lending in online microfinance". In: *Proceedings of the National Academy of Sciences* 113.52 (2016), pp. 14944–14948.

[113]   Jaegul Choo et al. "To gather together for a better world: Understanding and leveraging communities in micro-lending recommendation". In: *Proceedings of the 23rd international conference on World wide web*. 2014, pp. 249–260.

[114]   Robin Burke, Nasim Sonboli, and Aldo Ordonez-Gauger. "Balanced neighborhoods for multi-sided fairness in recommendation". In: *Conference on Fairness, Accountability and Transparency*. PMLR. 2018, pp. 202–214.

[115]   Abraham Adam, Roger Davidson, and Maha El Choubassi. "Analysis of Kiva Lending and Team Network Structure". In: (2014).

[116]   Jaegul Choo et al. "Understanding and promoting micro-finance activities in kiva. org". In: *Proceedings of the 7th ACM international conference on Web search and data mining*. 2014, pp. 583–592.

[117]   Sara L McKinnon et al. "Kiva. org, person-to-person lending, and the conditions of intercultural contact". In: *Howard Journal of Communications* 24.4 (2013), pp. 327–347.

[118]   Roy Chen et al. "Does team competition increase pro-social lending? Evidence from online microfinance". In: *Games and Economic Behavior* 101 (2017), pp. 311–333.

[119]   Sarah Hendriks. "The role of financial inclusion in driving women's economic empowerment". In: *Development in Practice* 29.8 (2019), pp. 1029–1038.

[120]   Paul Kofman and Clare Payne. "Digital Financial Inclusion of Women: An Ethical Appraisal". In: *Handbook on Ethics in Finance* (2020), pp. 1–25.

[121]   Claude Lopez and Oscar Contreras. "Gender Equality Discussion within the G20". In: *Milken Institute W* 20 (2020).

[122]   Peter Hedström and Petri Ylikoski. "Causal mechanisms in the social sciences". In: *Annual review of sociology* 36 (2010), pp. 49–67.

[123]   Ole-Jørgen Skog. "The long waves of alcohol consumption: a social network perspective on cultural change". In: *Social networks* 8.1 (1986), pp. 1–32.

[124]   Mark Granovetter. *Getting a job: A study of contacts and careers*. University of Chicago press, 2018.

[125] Robert L Thorndike. "Who belongs in the family?" In: *Psychometrika* 18.4 (1953), pp. 267–276.

[126] Ville Satopaa et al. "Finding a" kneedle" in a haystack: Detecting knee points in system behavior". In: *2011 31st international conference on distributed computing systems workshops.* IEEE. 2011, pp. 166–171.