



**HAL**  
open science

# Whole-Graph Representation Learning For the Classification of Signed Networks

Noé Cecillon, Vincent Labatut, Richard Dufour, Nejat Arınik

► **To cite this version:**

Noé Cecillon, Vincent Labatut, Richard Dufour, Nejat Arınik. Whole-Graph Representation Learning For the Classification of Signed Networks. IEEE Access, In press, 10.1109/ACCESS.2024.3472474 . hal-04712854

**HAL Id: hal-04712854**

**<https://hal.science/hal-04712854v1>**

Submitted on 27 Sep 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

# Whole-Graph Representation Learning For the Classification of Signed Networks

Noé Cécillon<sup>1</sup>, Vincent Labatut<sup>1</sup>, Richard Dufour<sup>2</sup>, and Nejat Arinik<sup>3</sup>

<sup>1</sup>LIA UPR 4128, Avignon Université, Avignon, F-84911, France  
(e-mail: {firstname.lastname}@univ-avignon.fr)

<sup>2</sup>LS2N UMR 6004, Nantes Université, Nantes, F-44322, France  
(e-mail: richard.dufour@univ-nantes.fr)

<sup>3</sup>CRIL UMR 8188, Université d’Artois, Lens, F-62307, France  
(e-mail: arinik@cril.fr)

September 27, 2024

## Abstract

Graphs are ubiquitous for modeling complex systems involving structured data and relationships. Consequently, graph representation learning, which aims to automatically learn low-dimensional representations of graphs, has drawn a lot of attention in recent years. The overwhelming majority of existing methods handle *unsigned* graphs. However, *signed* graphs appear in an increasing number of application domains to model systems involving two types of opposed relationships. Several authors took an interest in signed graphs and proposed methods for providing *vertex*-level representations, but only one exists for *whole-graph* representations, and it can handle only *fully connected* graphs. In this article, we tackle this issue by proposing two approaches to learning *whole-graph* representations of general *signed* graphs. The first is a SG2V, a signed generalization of the whole-graph embedding method Graph2vec that relies on a modification of the Weisfeiler–Lehman relabelling procedure. The second one is WSGCN, a whole-graph generalization of the signed vertex embedding method SGCN that relies on the introduction of master nodes into the GCN. We propose several variants of both these approaches. A bottleneck in the development of whole-graph-oriented methods is the lack of data. We constitute a benchmark composed of three collections of signed graphs with corresponding ground truths. We assess our methods on this benchmark, and our results show that the signed whole-graph methods learn better representations for this task. Overall, the baseline obtains an  $F$ -measure score of 58.57, when SG2V and WSGCN reach 73.01 and 81.20, respectively. Our source code and benchmark are publicly available online.

**Keywords**— Whole-Graph Embedding, Signed Graphs, Graph Classification, Graph Neural Networks

# 1 Introduction

Graph representation learning is a general task consisting of automatically learning a data-driven, low-dimensional and fixed-size vector representation of graphs, or parts of graphs such as vertices, edges, and subgraphs, that preserves the information conveyed by their structure [23]. The main benefit of such representations is that one can feed them to standard machine learning tools, and thus process graphs indirectly with general and efficient methods, instead of designing task-specific methods required to handle graphs directly. Due to this advantage over traditional feature engineering methods, the development of graph representation learning approaches has been the object of many publications in the last few years, as attested by the many recent surveys on the topic [30, 36, 57]. Most existing methods focus on vertices or edges, which is very useful in several tasks such as community detection or link prediction. However, the representation of graphs *as a whole* is also crucial, especially in graph classification [41, 50]. This task typically involves a collection of graphs, each one belonging to a specific class to be predicted.

In addition, the vast majority of existing representation learning methods are meant to handle *unsigned* graphs. This is understandable, as the number of available unsigned graph datasets is much larger than that for signed graphs. Nevertheless, signed graphs appear in many application domains such as Sociology [43], Neurosciences [49], International Relations [15], Business Science [28], Finance [39], Political Science [16], and Computer Science [1]; and their processing requires appropriate tools. Such graphs were originally introduced in Psychology, to represent the attitudes of people toward other people or objects [25]. More generally, they can be used to model any system involving two types of semantically opposed relationships (like/dislike, similar/different, etc.). In general, this duality makes it impossible to directly apply standard unsigned methods. Several methods have been proposed recently to handle signed graph representation learning at the level of *vertices* [11, 29, 55]. However, to the best of our knowledge, only one approach [49] allows learning *whole-graph* representations of signed graphs, and it is limited to *fully connected* graphs (i.e., every vertex is connected to all other vertices). As a consequence, it cannot handle most of the real-world signed networks from the literature, which are typically sparse (e.g., [6, 18]).

In this paper, we tackle this issue by proposing two approaches to learning the whole-graph representation of signed networks. The first is a signed generalization of the unsigned whole-graph embedding method *Graph2vec* [42]. The second one is a whole-graph generalization of the signed vertex embedding method Signed Graph Convolutional Network (SGCN) [11]. Our contributions are threefold:

- The first is methodological and concerns the two proposed representation learning approaches, for which we define several variants. All are able to handle not only fully connected graphs, but also sparse ones.
- The second is resource-oriented, as we constitute and share a benchmark annotated for signed graph classification, and constituted of three distinct collections. This is a first, as all similar datasets only focus on unsigned graphs.

- The third contribution is experimental, as we apply our proposed methods to our benchmark to assess their performance. Our results show that the proposed approaches perform better than the baseline.

The rest of this article is organized as follows. In Section 2, we introduce the main concepts and notations used later and review existing methods for graph representation learning. Next, we present our signed graph classification benchmark, in Section 3. Then, in Section 4, we describe the methods that we propose to handle signed whole-graph representation learning. We present and compare our results in Section 5. Finally, we summarize our main findings in Section 6, and discuss possible perspectives.

## 2 Background

This section describes the main notions used in the rest of the article. We first introduce the principal concepts and notations related to signed graphs in Section 2.1). We then review existing methods for graph representation learning in Section 2.2.

### 2.1 Signed Graphs

Formally, a signed graph is a triple  $G = (V, E, s)$  composed of a set of vertices  $V$ , a set of edges  $E \subseteq V^2$  between them, and a function  $s : E \rightarrow \{-, +\}$  that associates a sign to each edge. We denote  $E^-$  and  $E^+$  as the subsets of negative and positive edges, respectively. Consequently,  $E = E^- \cup E^+$ . As is customary in graph theory [5], we denote  $n$  as the *order* of the graph (i.e., its number of vertices), while  $m$  is its *size* (i.e., its number of edges).

In this work, we focus on undirected unweighted signed graphs. The *unsigned* neighborhood  $N(u)$  of a vertex  $u$  ignores edge signs and includes all vertices attached to this vertex:  $N(u) = \{v \in V : (u, v) \in E\}$ . On the contrary, the positive  $N^+(u)$  and negative  $N^-(u)$  neighborhoods focus only on one edge sign:  $N^\pm(u) = \{v \in V : (u, v) \in E^\pm\}$ . We similarly define the unsigned, negative and positive degrees as the cardinalities of the corresponding neighborhoods, i.e.,  $k(u) = |N(u)|$ ,  $k^-(u) = |N^-(u)|$ , and  $k^+(u) = |N^+(u)|$ . The sign of a path or cycle corresponds to the product of its constituting edge signs. Consequently, this sign is negative if the path or cycle contains an odd number of negative edges, and positive otherwise. The positive (resp. negative) *reachable set* of a vertex  $u$  is the subset of vertices that are connected to  $u$  through positive (resp. negative) shortest paths.

*Structural Balance* (SB) is a fundamental property of signed graphs [24, 25]. In its strict definition, a graph is said to be structurally balanced when all its cycles are positive [7]. Equivalently, for a structurally balanced graph, it is possible to find a bisection of  $V$  such that all positive edges are internal, i.e., they connect vertices from the same cluster, whereas all negative edges are external, i.e., they lie in between clusters. Fig. 1.a) illustrates this situation: the graph contains two clusters  $\{v_1, v_2, v_3\}$  and  $\{v_4, \dots, v_7\}$ ; all positive edges are inside these clusters; all negative edges are between them. In real-world networks, though, graphs are rarely *perfectly* balanced, and no bisection exists that respects the SB definition. In this case, one may want to measure the *amount* of imbalance in the graph. This is typically done by computing the

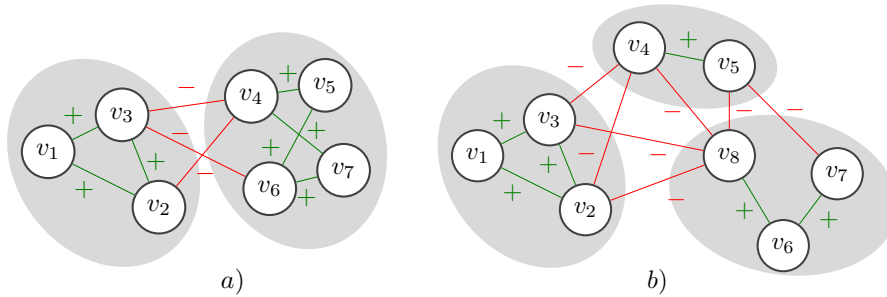


Figure 1: Examples of perfectly balanced graphs according to a) Structural Balance; and b) Generalized Balance ( $k = 3$ ).

*Frustration* measure (a.k.a. *Line Index* or *Imbalance*), which requires solving a combinatorial optimization problem [13]. Let us consider an arbitrary bisection of  $V$ . The positive edges located in between clusters and the negative edges located inside them are said to be *frustrated*, as they do not respect SB. For instance, in Fig. 1.a), if edges  $(v_1, v_2)$  and  $(v_2, v_4)$  were negative and positive, respectively, they would be frustrated. The Frustration of this bisection is the number of such edges. The Frustration of the graph is the minimal Frustration over all possible bisections. Put differently, the graph Frustration is the minimal number of edges whose sign must be switched to reach perfect SB.

The notion of SB was generalized to allow partitions composed of more than two antagonistic clusters [10]. A graph that can be split into  $k$  such clusters is said to be  $k$ -balanced. Equivalently, this *Generalized Balance* (GB) requires that a graph have no cycles with *exactly one* negative edge. Fig. 1.b) illustrates the notion of GB: the presented graph contains three clusters  $\{v_1, v_2, v_3\}$ ,  $\{v_4, v_5\}$ , and  $\{v_6, \dots, v_8\}$ ; all positive edges are inside the clusters; all negative edges are between them. Note that Frustration remains a valid imbalance measure for this generalization of the SB. The combinatorial problem consisting of finding the partition of  $V$  that minimizes Frustration is called the *Correlation Clustering* (CC) problem [4]. Finally, it is important to stress that the ubiquity of SB among real-world systems, which is widely advertised in the literature, is nevertheless disputed among specialists of signed graphs [12]. Therefore, it is not guaranteed that building a graph representation method that takes SB into account will perform better than one that ignores it. Assessing this point is one of the goals of our experiments.

## 2.2 Graph Representation Learning

Graph representation learning is a very popular research topic, and the literature provides a profusion of methods allowing us to automatically train models to represent various *types* of graphs (directed, weighted, signed, multiplex, etc.), and various *parts* of graphs (vertices, edges, subgraphs, whole graphs, etc.). However, to the best of our knowledge, none of them can handle *signed* graphs as a *whole*, except Hierarchical Signed Graph Representation Learning (HS-GRL) [49]. Moreover, this method, which we discuss later in this section, suffers from a limitation as it can only process *fully connected* graphs. Consequently, to position our work, in the following, we summarize the main approaches de-

signed to deal with one of two situations that are the closest to our own: on the one hand, representing *whole* but *unsigned* graphs, and on the other hand, representing the *vertices* of *signed* graphs. The methods used as a basis for our work are described in further detail later, in Section 4.

### 2.2.1 Whole Graphs

Whole-graph embeddings can be obtained through handcrafted approaches such as feature engineering [46], matrix factorization [40], or graph kernels [33]. However, they often suffer from generalization problems [42]. Representation learning methods are data-driven and allow solving this issue by adapting the graph representation to the targeted dataset.

A popular approach involves leveraging models from the field of Natural Language Processing (NLP), such as in Graph2vec [42] and Graph Classification via Graph Structure Learning (GC-GSL) [27]. Graph2vec, one of the earliest methods for whole-graph representation learning, is an unsupervised and task-agnostic approach that leverages the Doc2vec model [34] from NLP and applies it to graphs. Instead of treating a text as a collection of paragraphs, Graph2vec considers a graph as a collection of subgraphs, which are then used to train a SkipGram model (cf. Section 4.3 for more details).

Another family of approaches relies on graph autoencoders to learn the representation in an unsupervised way, e.g., Permutation-Invariant Graph-level Autoencoder (PIGAE) [53], or the Denoising Autoencoder-based (DAE) method from [22]. Such neural networks are composed of two parts. First, the encoder receives a raw representation of the graph, which is compressed to remove redundant information and superfluous variability and get a fixed-sized and compact representation. Second, the decoder is in charge of reconstructing the original input based on the compressed representation. The autoencoder is trained by minimizing the reconstruction error.

The literature contains another family of approaches, that adapt Convolutional Neural Networks (CNN) from the field of image processing to handle graphs, resulting in supervised methods able to learn whole-graph representations for specific classification tasks, e.g., Patchy-San Convolutional Network (PSCN) [45] and NgramCNN [37]. PSCN adapts the notion of convolution to the context of graphs, which allows applying the same principle as for image processing. The graph is represented by a collection of subgraphs, on which PSCN performs convolutions. These are then aggregated to create higher-level representations of the graph.

Finally, another strategy is to modify vertex-oriented Graph Neural Networks (GNN) to produce whole-graph representations, e.g., Message Passing Neural Network (MPNN) [21] or Virtual Column Network (VCN) [47]. This is conducted by the addition of a so-called master node, which is connected to all the other vertices. At the end of the training, the vector associated with this master node can be used as a representation of the graph.

### 2.2.2 Signed Graphs

Although much less common than for *unsigned* graphs, many *vertex* representation learning methods have been developed for *signed* graphs. Historically, the first type of such method relied on random walks, e.g., Signed Network Embed-

ding (SNE) [55] or Signed Directed Embeddings (SIDE) [29]. The general idea is to sample the graph using random walks and feed them to a standard neural network, which learns a representation that preserves both graph structure and edge signs. Put differently, the representations of two well-connected vertices tend to be close in the embedding space, whereas those of two vertices connected by a negative edge tend to be distant.

Signed Network Embedding (SiNE) [51] relies on a deep learning framework. Unlike the previous methods, it does not use random walks to decompose the network, but simpler subgraphs as it extracts all the open triads. The neural network is trained to learn a representation of a vertex that is similar to its direct positive neighbors, and dissimilar to its direct negative neighbors. We describe SiNE in further detail in Section 4.1.

More recently, a group of methods leveraged GNN (Graph Neural Networks) to learn a representation of vertices in signed graphs. The first one, Signed Graph Convolutional Network (SGCN) [11], directly generalizes Graph Convolutional Networks (GCNs) [32] to signed graphs, by proposing a dual representation and a dual message passing rule to take positive and negative paths into account (cf. Section 4.3 for more details). Methods such as Signed Graph Attention Networks (SiGAT) [26] or Signed Network Embedding via Graph Attention (SNEA) [35] introduce attention in the process, to give more importance to relevant neighbors during the message passing step.

Hierarchical Signed Graph Representation Learning (HSGRL) [49] extends SNEA by including an additional pooling module. Based on an information-based centrality metric, it selects a fixed number of vertices and uses their individual representations to build an overall representation of the *whole* signed graph. According to our review of the literature, HSGRL is the *only* method able to do so. However, it is designed only for *fully connected* graphs: this restricts its application to a very specific type of graphs and constitutes a serious limitation of this method. Indeed, the signed networks used in the literature to model and study real-world systems are typically sparse, see for instance [6, 18].

### 3 Signed Graph Datasets

To assess the proposed methods, we constitute a benchmark composed of three datasets of signed networks. This is not a trivial work, as most publicly available signed networks are *individual* graphs that are used for single graph problems such as vertex classification or link prediction. To perform graph classification, not only do we need *collections* of signed graphs, but these collections must be *annotated* for classification (i.e., each graph must be explicitly associated with a class). Our three datasets come from various sources and differ in their structures and sizes. The first one (Section 3.1) is based on an existing collection of real-world *unsigned* networks originally designed for a binary classification task, which we extend to obtain *signed* networks. The second one (Section 3.2) is an existing collection of artificially generated signed graphs, and the third one (Section 3.3) is an existing collection of real-world vote networks. These last two collections were not initially designed for graph classification: we leverage their metadata to repurpose them and define proper classification tasks. In principle, we could also include the *fully connected* signed networks used to assess the performance of the HSGRL method in [49]. Unfortunately, they are not publicly

available, therefore, we focus on the three datasets that we constructed. Global statistics describing these datasets are provided in Table 1, and described later (Section 3.4). The datasets themselves are all publicly available online<sup>1</sup>.

### 3.1 SpaceOrigin Conversations

The *SpaceOrigin* collection (SO) was originally proposed in [46]. Papegnies *et al.* extract a collection of conversational networks from a corpus of chat conversations taking place between players of the online video game SpaceOrigin. Each network is built around a message of interest, called the *targeted message*, and aims at modeling its conversational context. Its vertices represent players, and its weighted edges reflect the intensity of their verbal interactions. Each network integrates the messages present in a so-called *context period*, which contains a fixed number of messages occurring right before and after the targeted message. Temporal integration is performed by sliding a fixed-sized window over the context period, and incrementing edge weights based on the co-occurrence of speakers in this window. Papegnies *et al.* tackle the task of automatic moderation, which they formulate as a binary classification problem consisting of determining whether the targeted message is *Abusive* or *Non-abusive*. The available ground truth is based on manual annotation. For more details on the graph extraction process and the task itself, see [46].

The networks produced in [46] are unsigned, though. To obtain signed networks instead, we change some parts of the extraction process. When sliding the window over the context period, we leverage a sentiment analyzer<sup>2</sup> to determine the polarity of the players’ interactions, based on their exchanged textual content. The resulting weight change can thus be either negative (hostile interaction) or positive (neutral or friendly). Consequently, the total weight obtained when integrating over the whole context period can also be negative or positive. We call the resulting dataset *Signed SpaceOrigin* (SSO), and it contains 2,545 conversational graphs.

### 3.2 Correlation Clustering Instances

This dataset is proposed in [3], originally as a means to study the space of optimal solutions to the *Correlation Clustering* problem [4] (CC), described in Section 2.1. Arımk *et al.* want to study the multiplicity and diversity of the optimal solutions to CC. For this purpose, they define a random model and generate a collection of artificial graphs with planted partitions, applying various levels of noise to control the difficulty of the problem. They use an exact method to identify all possible optimal solutions for each graph in this collection and study how certain graph characteristics relate to the number of solutions. Due to the NP-hard nature of CC, they focus on relatively small graphs, with a maximal order (number of vertices) of  $n = 50$ . They produce a total of 24,660 unweighted signed graphs, including 22,560 completely connected graphs (i.e., every pair of vertices is connected), while the remaining 2,100 graphs are not completely connected, with a density ranging from 0.25 to 0.75.

To use this dataset in the present work, we define a classification problem by associating each graph of the collection with a label. This problem, named

---

<sup>1</sup><https://doi.org/10.5281/zenodo.13851362>

<sup>2</sup><https://github.com/TheophileBlard/french-sentiment-analysis-with-bert>



Table 1: Statistics describing our three datasets. Notations  $\pm$  and  $[\ ]$  respectively denote the standard deviation, and minimum & maximum.

<b>Data</b>	<b>Number of Graphs</b>	<b>Average Number of Vertices</b>		<b>Average density</b>
SSO	2,545	47.74 $\pm$ 20.34	[2; 214]	0.48 $\pm$ 0.16 [0.10; 1.00]
CCS	24,660	27.31 $\pm$ 7.44	[16; 50]	0.95 $\pm$ 0.17 [0.19; 1.00]
EPF	6,000	67.34 $\pm$ 59.21	[20; 274]	0.70 $\pm$ 0.19 [0.07; 1.00]

<b>Data</b>	<b>Nbr. of Classes</b>	<b>Gini Index</b>	<b>Average SB Frustration</b>		<b>Average GB Frustration</b>	
SSO	2	0.74	0.30 $\pm$ 0.04	[0.01; 0.48]	0.25 $\pm$ 0.04	[0.01; 0.46]
CCS	2	0.66	0.37 $\pm$ 0.05	[0.03; 0.51]	0.33 $\pm$ 0.04	[0.01; 0.49]
EPF	3	0.44	0.28 $\pm$ 0.04	[0.01; 0.46]	0.22 $\pm$ 0.03	[0.00; 0.45]

<b>Data</b>	<b>Average Number of Negative Edges</b>		<b>Average Number of Positive Edges</b>	
SSO	166.1 $\pm$ 22.96	[1; 1,692]	245.9 $\pm$ 30.41	[1; 2,323]
CCS	220.6 $\pm$ 130.45	[25; 833]	131.0 $\pm$ 80.11	[24; 392]
EPF	333.9 $\pm$ 877.20	[0; 15,933]	2,552.2 $\pm$ 5,761.46	[0; 33,153]

<b>Data</b>	<b>Average proportion of Positive Edges</b>
SSO	59.61 $\pm$ 24.73 [0.00; 100.00]
CCS	37.54 $\pm$ 11.99 [20.10; 79.01]
EPF	78.30 $\pm$ 22.22 [0.00; 100.00]

*Correlation Clustering Solutions* (CCS), consists of predicting whether there are a single vs. several optimal CC solutions for the graph of interest.

### 3.3 European Parliament Roll-Calls

The last dataset is based on a collection of signed graphs extracted in [2] from a description of the voting activity at the European Parliament (EP). The raw data corresponds to roll-call votes cast individually by Members of the EP (MEPs) during plenary sessions, in the course of the 7th term (2009–2014). Such votes can take one of three values: FOR (MEP supporting the proposition), AGAINST (MEP opposing the proposition) or ABSTENTION (MEP not taking a stand despite being present). MEPs can also be absent, and consequently, not take part at all in a roll-call. Each network extracted by Arnik *et al.* corresponds to a specific roll-call, using vertices to model MEPs and edges to represent an agreement between them: a positive sign represents an identical vote, and a negative one represents a disagreement. The goal of Arnik *et al.* is to study the polarization of the EP, and more specifically, its voting patterns, and how these are affected by various criteria, such as the topic of the voted proposition. For this purpose, they first identify factions of similarly voting MEPs in each roll-call network. Next, they compare the resulting vertex partitions to identify the types of situations that result in a comparable voting pattern.

The raw data contains 6,595 roll calls, from which the Arnik *et al.* extract

many more networks by leveraging the metadata associated with the voting activity. Since each MEP belongs to a member state and a European political group, they extract not only overall networks containing all MEPs, but also state- and group-specific networks, containing only the MEPs from a given member state or political group, respectively. In total, they produce 244,015 networks. Some of them are too small or too sparse (almost empty) to be interesting in a classification context, though. We constitute our dataset by first filtering out these unusable instances and then randomly sampling 6,000 networks.

As for the previous dataset, the initial collection of networks considered here was not originally used to perform any prediction tasks. We leverage the clusters of networks exhibiting similar voting patterns identified by Arnik *et al.*, and define a classification task consisting of predicting the number of factions identified in a network. We call the resulting dataset *European Parliament Factions* (EPF).

### 3.4 Brief Comparison

Table 1 provides a few descriptive statistics for all three datasets, to help interpret the classification results in Section 5. For each dataset, the top part shows the number of graphs, the average number of vertices by graph, and the average density. Graph density is the proportion of edges present in the graph, relative to a fully connected graph. CCS is by far the largest dataset in terms of the number of graphs, however, these are smaller. Moreover, the order of graphs, (i.e., their number of vertices) is not as uniform in both SSO and EPF, covering two orders of magnitude.

The middle part of the table shows the number of classes in each dataset, the Gini Index, and the average Frustration. The datasets contain roughly the same number of classes. The Gini Index is used here to characterize class imbalance. The Frustration measure, which we compute for both types of considered balance (SB vs. GB) is expressed as a proportion over the total number of edges in the graph, to have comparable values. By definition, the Frustration obtained for the strict version of structural balance is greater or equal to that of the generalized version: equal if the optimal partition contains two clusters, and greater if it contains three or more clusters.

Finally, the bottom part of the table exhibits the average numbers of positive and negative edges by graph, and the average proportion of positive edges. The number of edges is quite variable in all three datasets. In EPF, a few graphs have no positive or negative edges at all. Overall, the graphs tend to be denser than *unsigned* real-world networks [17].

## 4 Representation Methods

We now describe the three families of methods that we propose to handle signed whole-graph representation learning. The first can be considered a baseline and relies on the aggregation of signed *vertex* embeddings (Section 4.1). The second is an adaptation of an *unsigned* whole-graph embedding method to *signed* graphs (Section 4.2). The third is based on a Graph Convolutional Network able

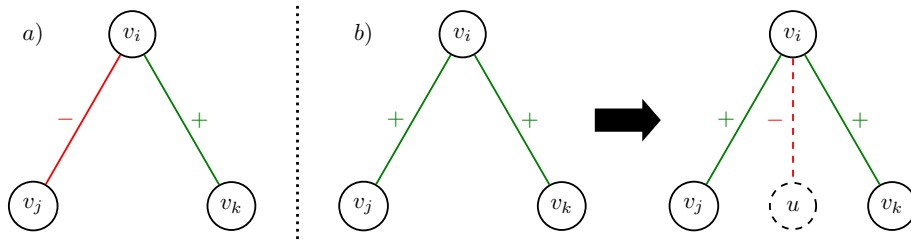


Figure 2: Triads used in SiNE, as inputs of the dual neural network. Mixed triplets (a) are used directly, positive ones (b) require a transformation, and negative ones (not represented) are not used.

to learn signed *vertex* representations, which we adapt to handle *whole* signed graphs (Section 4.3).

#### 4.1 Aggregated Signed Network Embedding

Signed Network Embedding (SiNE) is a deep learning framework for vertex embedding in signed graphs, proposed by Wang *et al.* [51]. Following [9], it is based on the assumption that the representation of a vertex should be similar to its positive neighbors, and dissimilar to its negative ones. To model signed networks based on this principle, SiNE proceeds at a very local level by focusing on a very specific type of subgraph. It extracts the set of all *open triads* (i.e., three vertices connected by two edges) present in the graph, focusing on those containing one positive and one negative edge, as illustrated by Figure 2.a. This set of triads represents the graph and is fed to a deep learning framework composed of two neural networks sharing certain weights. Following the principle mentioned earlier, this model is trained to minimize the similarity between the representation of the vertex located at the center of the triad and its negative neighbors, while maximizing its similarity with its positive ones. One neural network within the framework is dedicated to the positive neighbors, while the other network handles the negative ones.

A limitation of this approach is that it ignores vertices whose neighbors are all negative or all positive, such as the first triad in Figure 2.b. To handle the latter case, SiNE introduces a dummy vertex, denoted by  $u$  in the figure, that is connected through a negative edge to the central vertex of the positive triad,  $v_i$ . This allows creating as many dummy triads containing one positive and one negative edge, which can be processed by the framework. A similar principle could be used to handle negative triads, by adding positive dummy edges. However, Wang *et al.* stress that this is not justified by the signed network literature, and prefer to discard negative triads [51].

At the end of the training process, SiNE learns a compact representation for each vertex in the input graph, based only on its direct neighborhood. To get a graph-level representation, we simply aggregate the representations of all the vertices in the graph. For the sake of completeness, we consider two approaches: averaging these representations and summing them. We select SiNE as a baseline because it reaches top performances on several tasks and datasets in the literature [11, 29, 48]. Of course, these are vertex-level tasks,

and not graph-level tasks as in the present article.

According to Wang *et al.* [51], the time complexity of SiNE is  $O(RnSTd^2)$ , where  $R$  is the number of epochs,  $n$  the number of vertices in the input graph,  $S$  the number of triplets used to describe each vertex,  $T$  the number of layers in the neural network, and  $d$  the dimension of the representation (or of the layers, whichever is larger). In our case, we do not work with a single graph, but rather a whole collection of graphs  $\mathcal{G}$ . Consequently, we must include an additional multiplicative factor corresponding to the number of graphs in the collection. Moreover, we must replace  $n$  by  $N$ , the number of vertices in the largest graph in  $\mathcal{G}$ . As a result, the total expression is  $O(|\mathcal{G}|RNSTd^2)$ . Variables  $R$ ,  $S$ ,  $T$  and  $d$  are user-controlled parameters, whereas  $\mathcal{G}$  and  $N$  depend on the data.

## 4.2 Signed Graph2vec

Graph2vec [42] is an embedding method designed to learn representations of whole *unsigned* graphs, and is based on an analogy with the Doc2vec approach defined for text [34]. Its principle is to consider graphs (documents, in the analogy) as collections of subgraphs (words). The procedure enumerates rooted subgraphs around all vertices of the considered graphs. Each one represents the neighborhood of a vertex (the so-called root) in a certain order.

These rooted subgraphs are named using labels obtained with the relabeling procedure of the Weisfeiler–Lehman isomorphism test [52] (WL for short). Starting with the degree as the initial vertex label, this procedure goes through two phases to iteratively update these labels. First, each vertex is described by a tuple consisting of its previous label, and a sorted multiset containing those of its neighbors. Second, each unique tuple is replaced by a new label, to be used in the next iteration. Two identical tuples are replaced by the same label, but two different ones get distinct labels. This phase allows for a compact representation of the vertices. At the end of the process, each rooted subgraph is represented by its root’s label. More formally, the label update rule is

$$\ell_t(u) = f(\ell_{t-1}(u), \{\ell_{t-1}(v) : v \in N(u)\}), \quad (1)$$

where  $\ell_t(u)$  is the label of the subgraph rooted in  $u$  at iteration  $t$ ,  $N(u)$  is the neighborhood of  $u$ , and  $f$  is an injective function used to replace the tuples by new labels. Note that the set of the neighbors’ labels is ordered. Consider  $v_1$  in the left graph of Figure 3, for instance. Assuming degree is used to initialize the vertex labels, the composite label produced by the above rule is  $3, 122$ : the initial label of  $v_1$  is  $3$ , and those of its neighbors are  $1, 2, 2$  when ranked in increasing order. This composite label is then fetched to  $f$ , which returns the compressed label. Here, it could be, for example  $4$ , which has not been used yet at this stage of the process.

The number of iterations corresponds to the desired order of the neighborhood covered by the rooted subgraph (i.e., how far the subgraph spreads). At the end of this process, two isomorphic rooted subgraphs should get the same label. The obtained labels are then used to train the standard Doc2vec Skip-Gram model. Graph2vec has proven its effectiveness in many tasks involving the classification of *unsigned* graphs [19, 44, 59].

This method is not able to take advantage of the additional information present in signed graphs (i.e., edge signs), though. For this purpose, we define

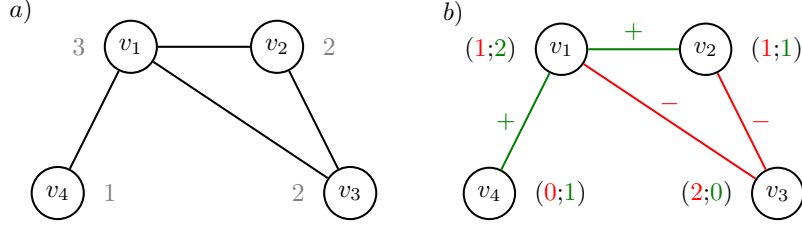


Figure 3: Unsigned (a) and signed (b) graphs used to illustrate the relabelling rules of G2V and SG2V (cf. text). The numeric values are the vertex degrees.

Signed Graph2vec, an adaptation of Graph2vec that relies on two variants of the WL relabeling procedure able to handle edge signs. The first one, denoted by **SG2Vn** (n for neutral), is straightforward and does not assume that the network has any form of structural balance. Regarding label initialization, instead of using the degree, we use both positive and negative degrees. For each vertex  $u$ , we first define a tuple  $(k^+(u); k^-(u))$ , and then replace each unique pair with a unique label using  $f$ . For the rest of the iterations, we proceed as in Graph2vec, except that we append the sign of the concerned edge in front of each neighbor when building the labels:

$$\ell_t(u) = f\left(\ell_{t-1}(u), \{[s(u, v), \ell_{t-1}(v)] : v \in N(u)\}\right), \quad (2)$$

where  $s(u, v)$  is the sign of edge  $(u, v)$ , and  $[\ ]$  denotes string concatenation. This allows distinguishing vertices holding the same label but connected to the root with edges of opposed signs. In our example from Figure 3 (right), applying  $f$  to the pairs of negative and positive degrees could produce the initial labels 1, 2, 3 and 4 (in this order), as all these pairs are different. According to the above rule, the composite label of  $v_1$  at the first iteration would be  $1, +2-3+4$ .

The second relabeling method, denoted by **SG2Vsb**, is the one proposed (for a different purpose) by Zhang *et al.* in [58], and it assumes that the network is structurally balanced. Each vertex is represented by two labels, based on its positive and negative reachable sets, respectively (cf. Section 2.1). The authors do not explain how they perform their initialization, so we use the positive (resp. negative) degree for the positive (resp. negative) label. The method requires one update rule for each label:

$$\ell_t^+(u) = f(\ell_{t-1}^+(u), \{\ell_{t-1}^+(v) : v \in N^+(u)\}, \{\ell_{t-1}^-(v) : v \in N^-(u)\}) \quad (3)$$

$$\ell_t^-(u) = f(\ell_{t-1}^-(u), \{\ell_{t-1}^-(v) : v \in N^+(u)\}, \{\ell_{t-1}^+(v) : v \in N^-(u)\}). \quad (4)$$

In our example from Figure 3 (right), the positive composite label of  $v_1$  at the first iteration is  $2, 11, 2$ , because it has an initial positive label of 2, two positive neighbors with an initial positive label of 1, and one negative neighbor with an initial negative label of 2. Symmetrically, the negative label of  $v_1$  is  $1, 01, 0$ . At the end of the process,  $f$  is applied to tuples formed by the positive and negative labels of each vertex, resulting in the final rooted subgraph labels.

Based on the algorithm described in [42], the time complexity of Graph2Vec is  $O(|\mathcal{G}|RNT^2d)$ , where  $|\mathcal{G}|$  is the size of the collection,  $R$  the number of epochs,  $N$  the number of vertices of the largest graph in  $\mathcal{G}$ ,  $T$  the order of the neigh-

borhood covered by the rooted subgraphs, and  $d$  the dimension of the representation. The complexity is identical for our signed generalization, since our modifications do not affect the computational cost of the original algorithm. Like for SiNE, time complexity is linear in the collection size  $\mathcal{G}$  and in the graph order  $N$ , but it also depends on some user-controlled parameters ( $R$ ,  $T$ , and  $d$ ).

### 4.3 Whole-graph Signed GCN

Graph Convolutional Networks (GCNs) are a family of neural networks that adapt traditional Convolutional Neural Networks (CNNs) so that they can process graph data instead of standard tabular data [54]. Shallow embedding methods such as Graph2vec may miss complex patterns in the graphs, and deep learning methods like CNNs are likely to solve this limitation [56]. GCNs generalize the convolution operation by considering graph neighborhoods instead of linear or grid neighborhoods as in standard CNNs used in NLP and image processing.

Most GCNs are designed to produce *vertex*-level representations. The general principle is as follows. Each vertex is initially represented by a vector, that can be generated randomly or based on some vertex features. In a convolution layer, the representation of a given vertex is combined with that of its neighbors. The result is then fed to a generally non-linear function (e.g., multilayer perceptron) to get the updated vertex representation. The information is propagated through multiple layers to incorporate information from multi-hop neighbors. The maximal number of hops corresponds to the number of convolution layers in the network. GCNs achieve state-of-the-art performances on many tasks [11] such as vertex classification, edge prediction, and community detection.

Standard GCNs are only able to handle *unsigned* networks, though. Leveraging the information conveyed by edge signs mainly requires adapting the message-passing rules used when computing the graph-based convolution. A few methods allow doing so, and in this work, we focus on *Signed Graph Convolutional Networks* (SGCN) [11], for two reasons. First, SGCN obtains strong performances on many tasks, such as node classification and link sign prediction, and is used as a basis for multiple other methods [31, 38, 49]. Second, its implementation is conveniently available online<sup>3</sup> and can be modified to fit our needs. Similarly to **SG2Vsb** (Section 4.2), this method relies on a dual hidden representation of a vertex, corresponding to its positive vs. negative reach sets, and it uses balance theory to aggregate and propagate vertex representation across layers. Formally, the hidden representations  $\mathbf{h}$  are updated as follows:

$$\mathbf{h}_t^+(u) = \sigma \left( \mathbf{W}_t^+ \left[ \sum_{v \in N^+} \frac{\mathbf{h}_{t-1}^+(v)}{k^+(u)}, \sum_{v \in N^-} \frac{\mathbf{h}_{t-1}^-(v)}{k^-(u)}, \mathbf{h}_{t-1}^+(u) \right] \right) \quad (5)$$

$$\mathbf{h}_t^-(u) = \sigma \left( \mathbf{W}_t^- \left[ \sum_{v \in N^+} \frac{\mathbf{h}_{t-1}^-(v)}{k^+(u)}, \sum_{v \in N^-} \frac{\mathbf{h}_{t-1}^+(v)}{k^-(u)}, \mathbf{h}_{t-1}^-(u) \right] \right), \quad (6)$$

where  $\sigma$  is a non-linear activation function,  $[\ ]$  denotes the concatenation, and the  $\mathbf{W}$  matrices are learnable weights. Ultimately, the dual hidden representations are concatenated to obtain a single vertex representation.

<sup>3</sup><https://github.com/benedekrozemberczki/SGCN>

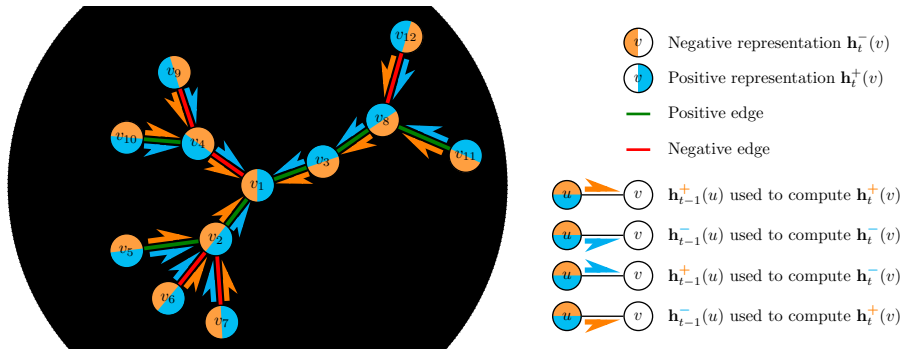


Figure 4: Update rules of SGCN, applied to an example graph. Only the edges belonging to a shortest path between the vertex of interest  $v_1$  and some other vertex are shown.

Figure 4 illustrates how these update rules work, focusing on vertex  $v_1$ . The dual representation of the vertices is shown using two colors: orange (negative) and cyan (positive). The update rules are depicted by pairs of arrows. Each arrow is attached either to the positive or negative representation of the *source* vertex, and shows how it is used to update the representation of the target vertex. The arrow color indicates if this update affects the negative (orange) or positive (cyan) representation of the *target* vertex (see the figure legend). Coming back to  $v_1$ , the first convolution layer combines the positive representations of its positive neighbors  $v_2$  and  $v_3$ , as well as the negative representation of its negative neighbor  $v_4$ , to build  $\mathbf{h}_1^+(v_1)$ . Symmetrically,  $\mathbf{h}_1^-(v_1)$  is based on the negative representations of  $v_2$  and  $v_3$  and the positive representation of  $v_4$ . The second layer performs an update that takes into account the representations of the second-order neighbors, in a way that respects structural balance. For instance, since  $v_9$  is connected to  $v_1$  by a positive path, its positive representation is (indirectly) used to compute  $\mathbf{h}_2^+(v_1)$ , and its negative representation to compute  $\mathbf{h}_2^-(v_1)$ . On the contrary,  $v_{10}$  is connected to  $v_1$  by a negative path, so its positive and negative representations affect  $\mathbf{h}_2^-(v_1)$  and  $\mathbf{h}_2^+(v_1)$ , respectively.

SGCN effectively learns representations of vertices in signed graphs. However, our objective is to handle *whole* graphs. A solution proposed for unsigned graphs in the literature [21] consists of adding a *master node* (also called a *virtual* or *super* node), which is connected to all other vertices in the graph. One then uses the representation of this master node as the representation of the whole graph. The intuition is that, as the master node is connected to all parts of the graph, its representation aggregates all its information.

It is not straightforward to adapt this approach to signed graphs, though: connecting a master node to the rest of the graph is not trivial, as there are two different types of edges. We propose five interconnection schemes, which we call collectively *Whole-graph SGCN* (WSGCN). They are represented in Fig. 5: circular vertices and solid lines correspond to the original graph, whereas the master nodes are shown as rounded rectangles denoted by  $MN$ , and their connections to the original vertices are shown as dotted lines. The first three interconnection schemes do not respect structural balance and can be considered baselines: WSGCN+ (Fig. 5.a) and WSGCN- (Fig. 5.b) consist of connecting a single



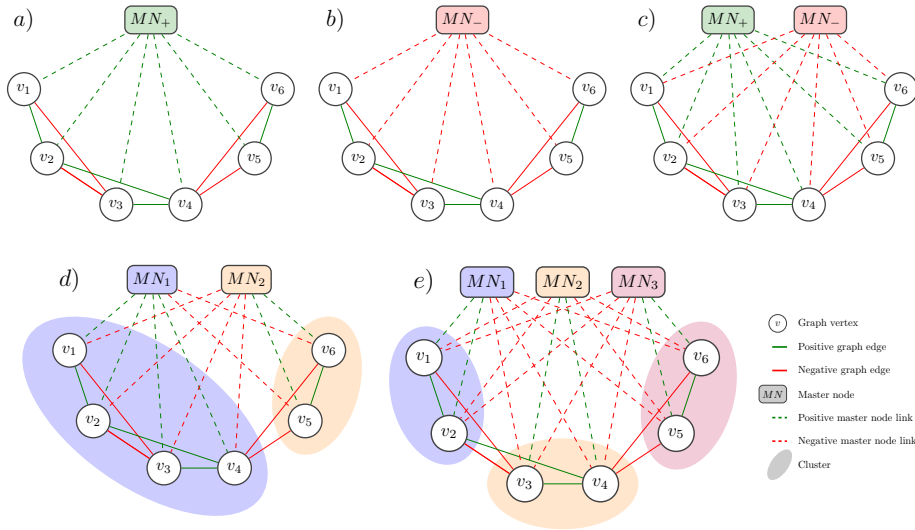


Figure 5: Examples of the 5 proposed interconnection schemes: WSGCN+ (a), WSGCN- (b), WSGCN± (c), WSGCNsb (d) and WSGCNgb (e). *MN* stands for Master Node. Green and red edges represent positive and negative connections, respectively. Each colored ellipse is a cluster.

master node to the rest of the graph using only positive and negative edges, respectively. With WSGCN± (Fig. 5.c), we use two distinct master nodes (one positive and one negative), which allows us to combine both previous schemes at once. The whole-graph representation is obtained by summing both master node representations. The fourth scheme, WSGCNsb (Fig. 5.d), is based on a (strict) structural balance. Using the *signnet* library<sup>4</sup>, we detect the optimal graph bisection. Then, we connect one distinct master node to each cluster by enforcing structural balance, i.e., positive edges within the cluster and negative ones within the other cluster. Like before, the whole-graph embedding is obtained by summing the representations of both master nodes. Finally, the fifth scheme, WSGCNgb (Fig. 5.e), relies on the *generalized* structural balance. As before, we find the optimal graph partition, but this time there may be more than two clusters. We add one master node for each cluster and sum their representations to get the whole-graph embedding.

SGCN generalizes a vanilla GCN architecture by using a dual representation of the vertices. This does not affect the time complexity of such architecture, which is  $O(RT(md + nd^2))$  according to the literature [8, 54], where  $R$  is the number of epochs,  $T$  the number of layers in the neural network,  $m$  the size of the graph (i.e., its number of edges),  $n$  its order (i.e., number of vertices), and  $d$  the dimension of the representation. The master nodes that we introduce in our whole-graph adaptation of SGCN do not imply any change in this complexity either. However, we iteratively apply our WSGCN method to all graphs in a collection  $|\mathcal{G}|$ , so its total time complexity is  $O(|\mathcal{G}|RT(Md + Nd^2))$ , with  $M$  the size of the largest graph, and  $N$  its order. Like for the other methods described in this section, some of these variables correspond to user-controlled parameters

<sup>4</sup><https://github.com/schochastics/signnet>



Table 2: Results in terms of macro  $F$ -measure obtained with SiNE. Each row focuses on a task, whereas columns represent the function used to aggregate vertex representations.

<b>Task</b>	<b>Sum</b>	<b>Average</b>
SSO	<b>55.42</b>	50.21
CCS	<b>50.48</b>	48.62
EPF	<b>69.81</b>	67.58

( $R$ ,  $T$ , and  $d$ ), whereas others depend on the data ( $\mathcal{G}$ ,  $M$ , and  $N$ ).

## 5 Results

Our experimental protocol consists of assessing all the methods described in Section 4 on the datasets and tasks presented in Section 3. We produce the representations, then train an SVM to perform the classification, using a 10-fold cross-validation. Our code is available online<sup>5</sup>. We conduct our experiments on an Nvidia RTX 2080 Ti GPU.

We discuss each family of methods separately: SiNE (Section 5.1), Graph2vec (Section 5.2), and Signed Graph Convolutional Networks (Section 5.3); before comparing them (Section 5.4). All the performance values are expressed in terms of the macro-averaged  $F$ -measure, i.e., by computing the  $F$ -measure for each class separately, then averaging them to get the overall performance. This allows giving the same importance to all classes, even in imbalanced datasets. For the sake of completeness, performance expressed in terms of Precision and Recall is provided in the Appendix.

### 5.1 Aggregated Signed Network Embedding

In this section, we present the results obtained with SiNE, the signed vertex embedding method that we consider our baseline. The performance scores are shown in Table 2 for all our datasets, in terms of macro  $F$ -measure. Each column focuses on one of the two functions used to aggregate vertex representations: sum and average. The corresponding tables describing performance in terms of Precision (Table 5) and Recall (Table 6) are available in the Appendix. Overall, the Precision tends to be slightly higher than the Recall for the SSO and CCS datasets, whereas the opposite is true for the EPF dataset. However, both metrics follow the trends described below for the  $F$ -measure.

On SSO and CCS, the performances are close to a random classifier, which would get an expected 50%  $F$ -measure. SiNE performs better on EPF, with the best result reaching 69.81%. We assume that for this task, the local information available is often sufficiently discriminative. Moreover, as the graphs in this dataset tend to be larger, SiNE can extract a greater number of training triads, contributing to an improved model performance. Furthermore, the choice of the aggregation function has a notable impact on performance, with the sum operation yielding better results across all tasks.

<sup>5</sup><https://github.com/CompNet/SWGE>

Table 3: Results in terms of macro  $F$ -measure obtained with **Graph2vec** and our two proposed signed adaptations. Each column focuses on a specific number of iterations used when extracting rooted subgraphs.

<b>Task</b>	<b>Method</b>	<b>1 it.</b>	<b>2 it.</b>	<b>3 it.</b>	<b>4 it.</b>	<b>5 it.</b>
SSO	G2V	75.09	71.32	72.62	73.96	73.77
	SG2Vn	74.85	71.44	72.15	72.88	72.37
	SG2Vsb	67.29	72.01	74.88	76.98	<b>77.44</b>
CCS	G2V	48.43	48.12	45.97	51.85	52.07
	SG2Vn	49.70	49.25	48.37	<b>52.57</b>	52.12
	SG2Vsb	49.84	51.81	49.70	51.37	51.62
EPF	G2V	45.63	49.76	52.31	60.43	63.14
	SG2Vn	75.63	81.44	84.18	86.16	88.68
	SG2Vsb	80.83	82.90	86.31	87.99	<b>89.98</b>

## 5.2 Signed Graph2vec

In this section, we present the results obtained with our two proposed variants of the *Signed Graph2vec* method: **SG2Vn** (which does not enforce SB) and **SG2Vsb** (which does). As a reference, we also include in our study **Graph2vec** (noted **G2V**), which simply ignores all edge signs. The performances are shown in Table 3, in terms of macro  $F$ -measure, whereas scores expressed in terms of Precision (Table 7) and Recall (Table 8) are provided in the Appendix. The columns show the performance for an increasing number of iterations, ranging from 1 to 5. As a reminder, this parameter controls the order of the rooted subgraphs extracted to describe the graph and, therefore, the range that is taken into account when characterizing vertex neighborhoods. The Precision and the Recall scores are very comparable, which is why we now focus on the  $F$ -measure. In the following, we discuss each dataset separately.

### 5.2.1 Signed Space Origin

The top part of Table 3 presents the performances of the three **Graph2vec** variants on the *Signed SpaceOrigin* dataset. It appears that they are affected diversely by the number of iterations. On the one hand, **G2V** and **SG2Vn** get their best score with a single iteration, and increasing iterations tend to reduce the performance. On the other hand, **SG2Vsb** starts low but increases consistently with the number of iterations. In the end, it reaches an  $F$ -measure of 77.44% and outperforms both other variants. The average diameter is 5.47 in this dataset, which means five iterations correspond to almost complete graph coverage.

From these observations, we can assume that a lot of information is already conveyed by the unsigned graph structures, in this dataset. This is consistent with the results obtained in the article that published the original data, as its authors already had some success performing a similar task with the unsigned version of the graphs [46]. Nevertheless, the signs bring some additional discriminative power, which **SG2Vsb** leverages to improve the classification performance. By comparison, this is not the case with **SG2Vn**. The relatively low level of SB Frustration in this dataset (0.30) may favor methods relying on SB.

### 5.2.2 Correlation Clustering Solutions

The results obtained for this dataset are shown in the middle part of Table 3. The performance is clearly lower, for all three variants, and even similar to the score expected from a random classifier (50), with a maximum  $F$ -measure of 52.57% obtained by **SG2Vn**. As explained in Section 3.2, the dataset contains two subsets: some graphs are completely connected, and the rest are not. Assuming that one type of graph might be more difficult to handle than the other, we try training separately on these subsets. However, we do not see any significant difference between the obtained results, which are similar to those already presented in Table 3.

We assume that either this classification task is too hard, in the sense that the information available in the graphs is not sufficient to perform the prediction, or that none of the three Graph2vec variants manage to capture the relevant information. The value predicted in this task is directly related to the distribution of edge signs (cf. Section 3.2), so we know with certainty that the information conveyed by signs is essential. The fact that **G2V** has similar performance to its signed counterparts hints at the second assumption (methods are unable to capture relevant information). In addition, the higher level of SB Frustration (0.37) may hinder the performance of **SG2Vsb**, compared to **SG2Vn**.

Increasing the number of iterations eventually improves the performance, but the effect is not as strong and stable as for the previous dataset. The graphs are more compact in this dataset, with an average diameter of only 3.6, which may partially explain this observation. Indeed, a few iterations are enough to retrieve all available information, and increasing their number does not bring any new neighbors.

### 5.2.3 European Parliament Factions

The results for this dataset are shown in the bottom part of Table 3. It appears that both our signed adaptations perform drastically better than the original unsigned method. The difference in  $F$ -measure is the largest of the three datasets: 63.14 (**G2V**) vs. 88.68 (**SG2Vn**) and 89.98 (**SG2Vsb**). Edge signs thus appear to be even more important for this task than they were for the *Signed SpaceOrigin* dataset. The SB-based variant **SG2Vsb** is slightly above **SG2Vn**, which seems to indicate that this type of structure may be relevant to this classification task. The average SB Frustration is 0.28 for this dataset.

This dataset contains larger graphs than the others, with an average of 67.34 vertices (vs. 47.74 and 27.31 previously) and a mean diameter of 4.12. This may explain the very strong effect of the number of iterations on the performance, even for **G2V**, the unsigned variant. It seems that even a small increase (proportionally to the order of the network) in the part of the graph covered when extracting rooted subgraphs, is enough to greatly improve the quality of the classification.

## 5.3 Whole-graph Signed GCN

In this section, we present the results obtained with the five variants that we proposed for the WSGCN method. Each one relies on the addition of one or several master nodes, through different interconnection schemes. The first three

Table 4: Results in terms of macro  $F$ -measure obtained with the original **SGCN** and our five proposed **WSGCN** interconnection schemes. Each column focuses on a specific number of convolution layers.

<b>Task</b>	<b>Method</b>	<b>1 lay.</b>	<b>2 lay.</b>	<b>3 lay.</b>	<b>4 lay.</b>	<b>5 lay.</b>
SSO	SGCN	66.48	67.21	68.06	68.87	69.54
	WSGCN+	65.12	66.78	68.42	68.98	69.29
	WSGCN-	54.19	54.89	55.56	55.59	55.89
	WSGCN±	49.28	48.95	49.01	49.25	49.39
	WSGCNsb	52.69	54.99	55.49	55.08	55.67
	WSGCNgb	66.59	68.85	71.21	72.28	<b>73.69</b>
CCS	SGCN	70.29	70.65	71.27	71.43	71.89
	WSGCN+	70.50	70.86	71.13	71.27	71.35
	WSGCN-	70.14	70.55	70.84	71.02	71.18
	WSGCN±	70.76	71.02	71.00	71.20	71.37
	WSGCNsb	69.59	70.12	70.86	71.21	71.46
	WSGCNgb	71.75	72.20	72.98	73.24	<b>73.49</b>
EPF	SGCN	90.16	90.87	91.63	92.04	92.65
	WSGCN+	88.56	89.30	90.49	91.11	91.87
	WSGCN-	88.61	89.30	90.49	91.09	91.80
	WSGCN±	86.32	88.45	90.01	90.78	91.56
	WSGCNsb	91.11	92.09	93.31	94.17	94.99
	WSGCNgb	92.36	93.65	95.29	96.04	<b>96.43</b>

(**WSGCN+**, **WSGCN-**, **WSGCN±**) ignore any type of structural balance, whereas the others enforce, respectively, SB (**WSGCNsb**) and GB (**WSGCNgb**). We considered two methods to extract a graph representation: using only the last layer vs. the sum of all layers. Preliminary experiments showed that the former performs better, so we only focus on this approach in the following discussion.

We also include **SGCN** as a reference in our study, i.e., the original method without any master node. To get a graph-level representation, we proceed like for **SiNE**, and sum the representations of all vertices. We alternatively experimented with averaging them, but got lower performance, which is why we focus only on the sum, here. The results are shown in Table 4, in terms of macro  $F$ -measure. As before, the performance in terms of Precision (Table 9) and Recall (Table 10) is provided in the Appendix. The columns show the results for an increasing number of convolution layers, ranging from 1 to 5. When comparing Precision and Recall, we observe very similar scores, as with the previous methods, which is why we focus on the  $F$ -measure results. In the following, we discuss each dataset separately.

### 5.3.1 Signed Space Origin

The top part of Table 4 shows the results on the *Signed SpaceOrigin* dataset. Increasing the number of layers in the convolutional network results in better performances for all variants, but to different extents. For instance, when going from 1 to 5 layers, the performance gain is +7,1  $F$ -measure point for **WSGCNgb** but only +0.11 for **WSGCN±**. With an average diameter of 5.47 in this dataset, using 5 layers allows for almost complete graph coverage.

All three variants that ignore any form of balance (WSGCN+, WSGCN-, and WSGCN±) are below the unsigned baseline (SGCN), which indicates that using signs improperly is counterproductive as it decreases the classification performance. WSGCN+ largely outperforms both other variants, probably because there are many more positive than negative edges in this dataset (cf. Table 1).

Among the variants that take balance into account, WSGCNgb consistently outperforms WSGCNsb: 55.67 vs. 73.69 with five layers. This shows that the type of balance selected when learning the representation must match the structural properties of the considered graphs. Interestingly, vanilla SGCN is the second-best method, which illustrates the methodological importance of the interconnection scheme when using a master node approach. It is on par with WSGCNgb when using a single layer because they are equivalent for this specific parameter value, however, the difference quickly grows with the number of layers.

### 5.3.2 Correlation Clustering Solutions

The middle part of Table 4 shows the  $F$ -measure scores for the *Correlation Clustering Solutions*. On this dataset, all methods yield quite similar results, except for WSGCNgb, which once again obtains the best performances. In particular, this method is able to capture more information at the whole-graph level than SGCN at the vertex level. Increasing the number of layers still improves the results, but the effect is much weaker than for the previous dataset. The graphs are smaller there, which may explain this, as more layers do not bring more information after a certain point. These results also show that this task is not as challenging as assumed when discussing *Graph2vec* results, since it is possible to get scores much higher than the expected performance of a random classifier. There is still room for improvement, though, as we are far from a perfect classification.

### 5.3.3 European Parliament Factions

The bottom part of Table 4 shows the results obtained for the EPF graphs. The behavior on this dataset is similar enough to that on *Signed SpaceOrigin*: the three variants that ignore balance (WSGCN+, WSGCN-, and WSGCN±) are below the original SGCN method (the latter being the worst, again), whereas WSGCNgb gets the best results, peaking at 96.43% when using 5 layers. There are three differences, though: First, the overall performance is much better, with a minimal  $F$ -measure of 88.45. This is in line with the behavior exhibited by the Graph2vec variants on the same dataset, and could be explained by the low level of Frustration (around 0.20). Second, WSGCN+ and WSGCN- perform very similarly. Third, WSGCNsb is above SGCN, which could mean that many graphs have a 2-cluster structure in this dataset.

## 5.4 Comparison and Concluding Remarks

We now compare and analyze the results of the three families of methods. Our baseline, which relies on SiNE, is consistently the least efficient method for all tasks. The gap in  $F$ -measure reaches up to 27.23 points with the best method, on the *Signed SpaceOrigin* dataset. On the one hand, this could be explained in part by the order and size of the graphs that constitute our benchmark: these

are relatively small, whereas SiNE was designed to handle large graphs, with hundreds or thousands of vertices. On the other hand, one could assume that working with smaller graphs should be an advantage when aggregating vertex representations to produce a whole-graph representation, as the most important vertices are likely to have a stronger effect on the produced representation. Nevertheless, our results indicate that applying a method designed to handle whole graphs directly leads to much better classification results than simply aggregating multiple vertex representations.

Signed Graph2vec, through its **SG2Vsb** variant, obtains the best performance on the *Signed SpaceOrigin* dataset, while Whole-graph SGCN, through its **WSGCNgb** variant, largely dominates on both other datasets (and is close on SPO). Regarding the use of signs to learn whole-graph representations, we identify three main results. First, the best signed methods systematically dominate their unsigned counterparts, often by a large margin. This shows the interest of leveraging this information to produce a relevant representation of signed graphs. Second, among the signed methods, those based on some form of SB, be it strict (**SG2Vsb** and **WSGCNsb**) or generalized (**WSGCNgb**), obtain better results than signed methods that ignore this property (**SG2Vn**, **WSGCN+**, **WSGCN-**, **WSGCN#**). Moreover, the latter generally gets a performance comparable to unsigned methods. This confirms that using signs is not sufficient: the notion of structural balance should be integrated in the design of the representation learning method, so that this property is preserved in the representation space. Third, the generalized version of structural balance seems to work better than the strict version. This is probably because GB is the most general of the two definitions, and subsumes SB. In other words, if the graph exhibits strict structural balance, it is captured by the GB-based representations, whereas SB-based ones cannot handle generalized balance.

Increasing the number of iterations in G2V and SG2V, or that of layers in SGCN and WSGCN, has a positive impact on performances for all datasets, and for almost all variants. This effect is generally stronger when the appropriate type of balance is leveraged to aggregate the representation of direct and indirect neighbors, though, which confirms the importance of this concept when dealing with signed graphs. For the WSGCN variants, it also shows that the method does not suffer from oversmoothing on the considered datasets. However, at some point, using more iterations or layers does not bring any significant performance gain.

Graph Convolutional Networks are the best-performing method, overall. However, this comes at a cost: they are also the most expensive in terms of computational runtime. Learning the representation of a graph with WSGCN variants takes more than 25 seconds on average, over the three datasets. As a comparison, SG2V variants take an average of 0.15 seconds per graph, and the SiNE baseline takes 0.65 seconds. SG2V is more than 100 times faster than WSGCN, but its performance is 8 F-measure points below, overall.

The time complexity of all three methods is linear in  $|\mathcal{G}|$ , the number of graphs in the collection, so they are likely to scale well on larger datasets. Their complexity is also linear in the number of vertices  $N$ , therefore, they can reasonably be expected to handle larger graphs as well. The complexity of SiNE and SG2V is independent of the number of edges  $M$ , whereas for WSGCN, it depends linearly on this graph size. This can be an issue if dealing with much denser graphs, but, as already explained in Section 2.2, real-world networks are

typically sparse. The complexity of both SG2V and WSGCN depends on  $T$ , a variable that controls the extent of the vertex neighborhood considered when building the graph representation. As a consequence, dealing with larger graphs may require increasing  $T$  to produce relevant representations. On this point, SWGCN has an advantage over SG2V, as its complexity only depends linearly on  $T$ , whereas this dependence is quadratic for SG2V.

## 6 Conclusion

In this paper, we tackle the problem of learning signed whole-graph representations, and use them for the classification of signed networks. In the absence of any appropriate method in the literature, we generalize two existing models: 1) we adapt Graph2vec to signed graphs by generalizing the Weisfeiler–Lehman relabeling procedure in two different ways; 2) we adapt Signed Graph Convolutional Networks to whole graphs by introducing master nodes and five interconnection schemes. Their implementation is shared online<sup>5</sup>. We build a benchmark of signed graphs annotated for classification, which is publicly available online<sup>1</sup>. It is constituted of three datasets, each one corresponding to a real-world application of graph classification: detecting abusive behavior in online conversations; estimating the multiplicity of optimal solutions to a combinatorial problem; and predicting the polarity at the European Parliament. We use this benchmark to assess our methods and compare them with a baseline relying on SiNE, a standard signed vertex embedding method. Our results show that a signed whole-graph embedding manages to learn better representations for the classification of signed networks. Our proposed method, WSGCNgb, which relies on the notion of generalized structural balance, obtains the best results, overall.

Our work could be extended in several ways. First, we want to expand our benchmark by including new collections containing larger graphs. This could be done by complementing existing unsigned graph datasets, as we did for the *Signed Space Origin* collection. Repositories of correlation matrices constitute another promising source, as these can be considered signed adjacency matrices. Regarding the methods, we want to apply the master node approach to other types of signed GNN, such as Graph Attention Networks (GATs). It would also be interesting to generalize our methods so that they can use edge weights, which are available in two of the original datasets used in this article. Another promising perspective is to experiment with other master node interconnection schemes, based on other variants of Structural Balance. The concept of *Relaxed Balance* [14, 20], in particular, is very interesting, as it allows inter-cluster (resp. intra-cluster) edges to be positive (resp. negative).

## References

- [1] N. Arımk. *Multiplicity in the Partitioning of Signed Graphs*. Phd thesis, Avignon Université, 2021. URL <https://tel.archives-ouvertes.fr/tel-03384624>.
- [2] N. Arımk, R. Figueiredo, and V. Labatut. Multiple partitioning of mul-



- tiplex signed networks: Application to european parliament votes. *Social Networks*, 60:83–102, 2020. doi:[10.1016/j.socnet.2019.02.001](https://doi.org/10.1016/j.socnet.2019.02.001).
- [3] N. Arımk, R. Figueiredo, and V. Labatut. Multiplicity and diversity: analysing the optimal solution space of the correlation clustering problem on complete signed graphs. *Journal of Complex Networks*, 8(6), 2021. doi:[10.1093/comnet/cnaa025](https://doi.org/10.1093/comnet/cnaa025).
- [4] N. Bansal, A. Blum, and S. Chawla. Correlation clustering. In *43rd Annual IEEE Symposium on Foundations of Computer Science*, pages 238–247, 2002. doi:[10.1109/SFCS.2002.1181947](https://doi.org/10.1109/SFCS.2002.1181947).
- [5] B. Bollobas. *Modern Graph Theory*. Springer, 1998. doi:[10.1007/978-1-4612-0619-4](https://doi.org/10.1007/978-1-4612-0619-4).
- [6] J. Brito-Montes, E. Canto-Lugo, and R. Huerta-Quintanilla. Modularity, balance, and frustration in student social networks: The role of negative relationships in communities. *PLoS ONE*, 17(12):e0278647, 2022. doi:[10.1371/journal.pone.0278647](https://doi.org/10.1371/journal.pone.0278647).
- [7] D. Cartwright and F. Harary. Structural balance: A generalization of Heider’s theory. *Psychological Review*, 63:277–293, 1956. doi:[10.1037/h0046049](https://doi.org/10.1037/h0046049).
- [8] M. Chen, Z. Wei, B. Ding, Y. Li, Y. Yuan, X. Du, and J.-R. Wen. Scalable graph neural networks via bidirectional propagation. In *34th International Conference on Neural Information Processing Systems*, pages 14556–14566, 2020. URL [https://papers.nips.cc/paper\\_files/paper/2020/hash/a7789ef88d599b8df86bbee632b2994d-Ab](https://papers.nips.cc/paper_files/paper/2020/hash/a7789ef88d599b8df86bbee632b2994d-Ab)
- [9] M. Cygan, M. Pilipczuk, M. Pilipczuk, and J.O. Wojtaszczyk. Sitting closer to friends than enemies, revisited. In *Mathematical Foundations of Computer Science*, pages 296–307, 2012. doi:[https://doi.org/10.1007/978-3-642-32589-2\\_28](https://doi.org/10.1007/978-3-642-32589-2_28).
- [10] J. A. Davis. Clustering and structural balance in graphs. *Human Relations*, 20(2):181–187, 1967. doi:[10.1177/001872676702000207](https://doi.org/10.1177/001872676702000207).
- [11] T. Derr, Y. Ma, and J. Tang. Signed graph convolutional network. In *18th IEEE International Conference on Data Mining*, page 929–934, 2018. doi:[10.1109/ICDM.2018.00113](https://doi.org/10.1109/ICDM.2018.00113).
- [12] P. Doreian. Reflections on studying signed networks. *Journal of Interdisciplinary Methodologies and Issues in Science*, 2:2.1–2.14, 2017. doi:[10.18713/JIMIS-170117-2-1](https://doi.org/10.18713/JIMIS-170117-2-1).
- [13] P. Doreian and A. Mrvar. A partitioning approach to structural balance. *Social Networks*, 18(2):149–168, 1996. doi:[10.1016/0378-8733\(95\)00259-6](https://doi.org/10.1016/0378-8733(95)00259-6).
- [14] P. Doreian and A. Mrvar. Partitioning signed social networks. *Social Networks*, 31(1):1–11, 2009. doi:[10.1016/j.socnet.2008.08.001](https://doi.org/10.1016/j.socnet.2008.08.001).
- [15] P. Doreian and A. Mrvar. Structural balance and signed international relations. *Journal of Social Structure*, 16:1–49, 2015. URL <https://www.cmu.edu/joss/content/articles/volume16/DoreianMrvar.pdf>.



- [16] P. Doreian and A. Mrvar. Signed networks for the us supreme court overturning its prior decisions. *Connections*, 39(1), 2019. doi:[10.21307/connections-2019-00](https://doi.org/10.21307/connections-2019-00).
- [17] P. Esmailian and M. Jalili. Community detection in signed networks: the role of negative ties in different scales. *Scientific Reports*, 5:14339, 2015. doi:[10.1038/srep14339](https://doi.org/10.1038/srep14339).
- [18] G. Facchetti, G. Iacono, and C. Altafini. Exploring the low-energy landscape of large-scale signed social networks. *Physical Review E*, 86(3):036116, 2012. doi:[10.1103/PhysRevE.86.036116](https://doi.org/10.1103/PhysRevE.86.036116).
- [19] C. Fang, Z. Liu, Y. Shi, J. Huang, and Q. Shi. Functional code clone detection with syntax and semantics fusion learning. In *International Symposium on Software Testing and Analysis*, page 516–527, 2020. doi:[10.1145/3395363.3397362](https://doi.org/10.1145/3395363.3397362).
- [20] R. Figueiredo and G. Moura. Mixed integer programming formulations for clustering problems related to structural balance. *Social Networks*, 35(4):639–651, 2013. doi:[10.1016/j.socnet.2013.09.002](https://doi.org/10.1016/j.socnet.2013.09.002).
- [21] J. Gilmer, S. S. Schoenholz, P. F. Riley, O. Vinyals, and G. E. Dahl. Neural message passing for quantum chemistry. In *34th International Conference on Machine Learning*, page 1263–1272, 2017. URL <https://proceedings.mlr.press/v70/gilmer17a.html>.
- [22] L. Gutiérrez-Gómez and J.-C. Delvenne. Unsupervised network embeddings with node identity awareness. *Applied Network Science*, 4:82, 2019. doi:[10.1007/s41109-019-0197-1](https://doi.org/10.1007/s41109-019-0197-1).
- [23] W. L. Hamilton. *Graph Representation Learning*, volume 46 of *Synthesis Lectures on Artificial Intelligence and Machine Learning*. Morgan & Claypool, San Rafael, USA, 2020. doi:[10.2200/s01045ed1v01y202009aim046](https://doi.org/10.2200/s01045ed1v01y202009aim046).
- [24] F. Harary. On the notion of balance of a signed graph. *Michigan Mathematical Journal*, 2(2):143–146, 1953. doi:[10.1307/mmj/1028989917](https://doi.org/10.1307/mmj/1028989917).
- [25] F. Heider. Attitudes and cognitive organization. *Journal of Psychology*, 21(1):107–112, 1946. doi:[10.1080/00223980.1946.9917275](https://doi.org/10.1080/00223980.1946.9917275).
- [26] J. Huang, H. Shen, L. Hou, and X. Cheng. Signed graph attention networks. In *International Conference on Artificial Neural Networks*, 2019.
- [27] T. Huynh, T. T. Thi Ho, and B. Le. Graph classification via graph structure learning. In *Asian Conference on Intelligent Information and Database Systems*, volume 13758 of *Lecture Notes in Computer Science*, pages 269–281. Springer, 2022. doi:[10.1007/978-3-031-21967-2\\_22](https://doi.org/10.1007/978-3-031-21967-2_22).
- [28] P. Jensen. Network-based predictions of retail store commercial categories and optimal locations. *Physical Review E*, 74(3):035101, 2006. doi:[10.1103/PhysRevE.74.035101](https://doi.org/10.1103/PhysRevE.74.035101).
- [29] J. Kim, H. Park, J.-E. Lee, and U. Kang. SIDE: Representation learning in signed directed networks. In *World Wide Web Conference*, pages 509–518, 2018. doi:[10.1145/3178876.3186117](https://doi.org/10.1145/3178876.3186117).

- [30] J.-Y. Kim and S.-B. Cho. A systematic analysis and guidelines of graph neural networks for practical applications. *Expert Systems with Applications*, 184:115466, 2021. doi:[10.1016/j.eswa.2021.115466](https://doi.org/10.1016/j.eswa.2021.115466).
- [31] M.-J. Kim, Y.-C. Lee, and S.-W. Kim. TrustSGCN: Learning trustworthiness on edge signs for effective signed graph convolutional networks. In *International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 2451–2455, 2023. doi:[10.1145/3539618.3592075](https://doi.org/10.1145/3539618.3592075).
- [32] T. N. Kipf and M. Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, 2017.
- [33] N. M. Kriege, F. D. Johansson, and C. Morris. A survey on graph kernels. *Applied Network Science*, 5:6, 2020. doi:[10.1007/s41109-019-0195-3](https://doi.org/10.1007/s41109-019-0195-3).
- [34] Q. Le and T. Mikolov. Distributed representations of sentences and documents. In *31st International Conference on International Conference on Machine Learning*, volume 32, page II–1188–II–1196, 2014. URL <http://proceedings.mlr.press/v32/le14.html>.
- [35] Y. Li, Y. Tian, J. Zhang, and Y. Chang. Learning signed network embedding via graph attention. In *34th AAAI Conference on Artificial Intelligence*, 2020.
- [36] Y. Liu, M. Jin, S. Pan, C. Zhou, Y. Zheng, F. Xia, and P. Yu. Graph self-supervised learning: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 35(6):5879–5900, 2022. doi:[10.1109/tkde.2022.3172903](https://doi.org/10.1109/tkde.2022.3172903).
- [37] Z. Luo, L. Liu, J. Yin, Y. Li, and Z. Wu. Deep learning of graphs with ngram convolutional neural networks. *IEEE Transactions on Knowledge and Data Engineering*, 29(10):2125–2139, 2017. doi:[10.1109/tkde.2017.2720734](https://doi.org/10.1109/tkde.2017.2720734).
- [38] M. Ma, S. Na, and H. Wang. AEGCN: An autoencoder-constrained graph convolutional network. *Neurocomputing*, 432:21–31, 2021. doi:[10.1016/j.neucom.2020.12.061](https://doi.org/10.1016/j.neucom.2020.12.061).
- [39] M. MacMahon and D. Garlaschelli. Community detection for correlation matrices. *Physical Review X*, 5(2):021006, 2015. doi:[10.1103/PhysRevX.5.021006](https://doi.org/10.1103/PhysRevX.5.021006).
- [40] L. Maddalena, I. Manipur, M. Manzo, and M. R. Guarracino. On whole-graph embedding techniques. In *International Symposium on Mathematical and Computational Biology*, pages 115–131. Springer, 2020. doi:[10.1007/978-3-030-73241-7\\_8](https://doi.org/10.1007/978-3-030-73241-7_8).
- [41] C. Morris. Graph neural networks: Graph classification. In *Graph Neural Networks: Foundations, Frontiers, and Applications*, chapter 9, pages 179–193. Springer, 2022. doi:[10.1007/978-981-16-6054-2\\_9](https://doi.org/10.1007/978-981-16-6054-2_9).
- [42] A. Narayanan, M. Chandramohan, R. Venkatesan, L. Chen, Y. Liu, and S. Jaiswal. graph2vec: Learning distributed

- representations of graphs. In *13th International Workshop on Mining and Learning with Graph*, 2017. URL [http://www.mlgworkshop.org/2017/paper/MLG2017\\_paper\\_21.pdf](http://www.mlgworkshop.org/2017/paper/MLG2017_paper_21.pdf).
- [43] J. W. Neal, Z. P. Neal, and C. E. Durbin. Inferring signed networks from preschoolers’ observed parallel and social play. *Social Networks*, 71:80–86, 2022. doi:10.1016/j.socnet.2022.07.002.
- [44] Q.-D. Ngo, H.-T. Nguyen, H.-A. Tran, and D.-H. Nguyen. IoT botnet detection based on the integration of static and dynamic vector features. In *8th International Conference on Communications and Electronics*, pages 540–545, 2021. doi:10.1109/ICCE48956.2021.9352145.
- [45] M. Niepert, M. Ahmed, and K. Kutzkov. Learning convolutional neural networks for graphs. In *33rd International Conference on International Conference on Machine Learning*, pages 2014–2023, 2016. URL <http://proceedings.mlr.press/v48/niepert16.pdf>.
- [46] E. Papegnies, V. Labatut, R. Dufour, and G. Linares. Conversational networks for automatic online moderation. *IEEE Transactions on Computational Social Systems*, 6(1):38–55, 2019. doi:10.1109/TCSS.2018.2887240.
- [47] T. Pham, T. Tran, H. Dam, and S. Venkatesh. Graph classification via deep learning with virtual nodes. *arXiv*, cs.LG:1708.04357, 2017. URL <https://arxiv.org/abs/1708.04357>.
- [48] X. Shen and F.-L. Chung. Deep network embedding for graph representation learning in signed networks. *IEEE Transactions on Cybernetics*, 50:1556–1568, 2020. doi:10.1109/TCYB.2018.2871503.
- [49] H. Tang, G. Ma, L. Guo, X. Fu, H. Huang, and L. Zhan. Contrastive brain network learning via hierarchical signed graph pooling model. *IEEE Transactions on Neural Networks and Learning Systems*, in press, 2022. doi:10.1109/TNNLS.2022.3220220.
- [50] K. Tsuda and H. Saigo. Graph classification. In *Managing and Mining Graph Data*, volume 40 of *Advances in Database Systems*, pages 337–363. Springer, 2010. doi:10.1007/978-1-4419-6045-0\_11.
- [51] S. Wang, J. Tang, C. Aggarwal, Y. Chang, and H. Liu. Signed network embedding in social media. In *17th SIAM International Conference on Data Mining*, pages 327–335, 2017. doi:10.1137/1.9781611974973.37.
- [52] B. Y. Weisfeiler and A. A. Leman. A reduction of a graph to a canonical form and an algebra arising during this reduction. *Nauchno-Technicheskaya Informatsia*, 2(9):12–16, 1968. URL [https://www.iti.zcu.cz/wl2018/pdf/wl\\_paper\\_translation.pdf](https://www.iti.zcu.cz/wl2018/pdf/wl_paper_translation.pdf).
- [53] R. Winter, F. Noe, and D.-A. Clevert. Permutation-invariant variational autoencoder for graph-level representation learning. In *Conference on Neural Information Processing Systems*, 2021. URL <https://proceedings.neurips.cc/paper/2021/hash/4f3d7d38d24b740c95da2b03dc3a2333-Abstract>

- [54] Z. Wu, S. Pan, F. Chen, G. Long, C. Zhang, and P. S. Yu. A comprehensive survey on graph neural networks. *IEEE Transactions on Neural Networks and Learning Systems*, 32(1):4–24, 2021. doi:[10.1109/tnnls.2020.2978386](https://doi.org/10.1109/tnnls.2020.2978386).
- [55] S. Yuan, X. Wu, and Y. Xiang. SNE: Signed network embedding. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, volume 10235 of *Lecture Notes in Computer Science*, pages 183–195, 2017. doi:[10.1007/978-3-319-57529-2\\_15](https://doi.org/10.1007/978-3-319-57529-2_15).
- [56] S. Zhang, H. Tong, J. Xu, and R. Maciejewski. Graph convolutional networks: Algorithms, applications and open challenges. In *International Conference on Computational Social Networks*, volume 11280 of *Lecture Notes in Computer Science*, pages 79–91, 2018. doi:[10.1007/978-3-030-04648-4\\_7](https://doi.org/10.1007/978-3-030-04648-4_7).
- [57] S. Zhang, H. Tong, J. Xu, and R. Maciejewski. Graph convolutional networks: a comprehensive review. *Computational Social Networks*, 6:11, 2019. doi:[10.1186/s40649-019-0069-y](https://doi.org/10.1186/s40649-019-0069-y).
- [58] Z. Zhang, J. Liu, X. Zheng, Y. Wang, P. Han, Y. Wang, K. Zhao, and Z. Zhang. RSGNN: A model-agnostic approach for enhancing the robustness of signed graph neural networks. In *ACM Web Conference*, page 60–70, 2023. doi:[10.1145/3543507.3583221](https://doi.org/10.1145/3543507.3583221).
- [59] X. Zhou, Y. Zhang, Z. Li, X. Wang, J. Zhao, and Z. Zhang. Large-scale cellular traffic prediction based on graph convolutional networks with transfer learning. *Neural Computing and Applications*, 34:5549–5559, 2022. doi:[10.1007/s00521-021-06708-x](https://doi.org/10.1007/s00521-021-06708-x).

## A Additional Results

This appendix provides additional results regarding the experiments conducted in Section 5. Tables 5 and 6 show the performance of the SiNE baseline, similarly to Table 2, but in terms of Precision and Recall instead of  $F$ -measure, respectively. Likewise, Tables 7 and 8 show the performance of the G2V-based methods, similarly to Table 3, but in terms of Precision and Recall. Finally, Tables 9 and 10 show the performance of the SGCN-based methods, similarly to Table 4, but in terms of Precision and Recall. The pairs of matching Precision and Recall scores are very similar. As a result, the  $F$ -measure scores discussed in the main text are also very similar, since the  $F$ -measure is the harmonic mean of Precision and Recall.

Table 5: Results in terms of macro Precision obtained with SiNE. Each row focuses on a task, whereas columns represent the function used to aggregate vertex representations.

<b>Task</b>	<b>Sum</b>	<b>Average</b>
SSO	56.12	51.01
CCS	50.88	49.11
EPF	69.07	66.74

Table 6: Results in terms of macro Recall obtained with SiNE. Each row focuses on a task, whereas columns represent the function used to aggregate vertex representations.

<b>Task</b>	<b>Sum</b>	<b>Average</b>
SSO	54.74	49.44
CCS	50.09	48.14
EPF	70.57	68.45

Table 7: Results in terms of macro Precision obtained with Graph2vec and our two proposed signed adaptations. Each column focuses on a specific number of iterations used when extracting rooted subgraphs.

<b>Task</b>	<b>Method</b>	<b>1 it.</b>	<b>2 it.</b>	<b>3 it.</b>	<b>4 it.</b>	<b>5 it.</b>
SSO	G2V	76.31	71.49	73.01	73.94	74.03
	SG2Vn	76.02	72.20	72.89	73.49	72.98
	SG2Vsb	66.68	71.58	74.79	76.50	77.03
CCS	G2V	48.67	47.97	46.21	52.87	51.79
	SG2Vn	50.33	49.88	48.53	53.20	51.84
	SG2Vsb	50.31	52.13	49.67	51.49	52.02
EPF	G2V	47.68	51.24	54.53	61.85	64.78
	SG2Vn	74.75	81.05	84.29	86.35	88.13
	SG2Vsb	79.66	82.85	86.46	87.64	90.44

Table 8: Results in terms of macro Recall obtained with Graph2vec and our two proposed signed adaptations. Each column focuses on a specific number of iterations used when extracting rooted subgraphs.

<b>Task</b>	<b>Method</b>	<b>1 it.</b>	<b>2 it.</b>	<b>3 it.</b>	<b>4 it.</b>	<b>5 it.</b>
SSO	G2V	73.90	71.15	72.23	73.98	73.51
	SG2Vn	73.71	70.69	71.42	72.28	71.77
	SG2Vsb	67.91	72.44	74.97	77.46	77.85
CCS	G2V	48.19	46.40	45.73	50.76	52.35
	SG2Vn	49.08	48.63	48.21	51.95	52.40
	SG2Vsb	49.37	51.49	49.73	51.25	51.22
EPF	G2V	43.74	48.36	50.26	59.07	61.58
	SG2Vn	76.53	81.83	84.07	85.97	89.23
	SG2Vsb	82.03	82.95	86.16	88.34	89.52

Table 9: Results in terms of macro Precision obtained with the original **SGCN** and our five proposed **WSGCN** interconnection schemes. Each column focuses on a specific number of convolution layers.

<b>Task</b>	<b>Method</b>	<b>1 lay.</b>	<b>2 lay.</b>	<b>3 lay.</b>	<b>4 lay.</b>	<b>5 lay.</b>
SSO	SGCN	65.88	66.76	68.02	68.45	69.40
	WSGCN+	65.24	66.98	68.57	69.68	70.00
	WSGCN-	54.42	55.11	55.78	55.48	56.03
	WSGCN±	49.84	48.99	49.25	49.18	49.58
	WSGCNsb	53.16	55.84	57.10	56.42	56.59
	WSGCNgb	68.32	68.59	72.35	72.46	73.51
CCS	SGCN	71.04	71.21	71.86	72.01	72.07
	WSGCN+	71.25	71.49	71.67	72.23	71.77
	WSGCN-	70.42	70.41	70.76	71.04	71.40
	WSGCN±	70.99	71.45	71.20	71.43	71.33
	WSGCNsb	69.84	70.33	71.25	71.49	71.94
	WSGCNgb	71.43	72.35	73.56	74.03	74.18
EPF	SGCN	90.46	91.20	91.89	92.08	92.58
	WSGCN+	89.40	89.96	90.48	91.24	92.23
	WSGCN-	88.95	89.39	90.66	91.54	92.03
	WSGCN±	86.11	88.23	89.86	90.67	91.60
	WSGCNsb	91.32	92.20	93.56	94.67	95.45
	WSGCNgb	92.66	93.81	95.45	96.16	96.51

Table 10: Results in terms of macro Recall obtained with the original **SGCN** and our five proposed **WSGCN** interconnection schemes. Each column focuses on a specific number of convolution layers.

<b>Task</b>	<b>Method</b>	<b>1 lay.</b>	<b>2 lay.</b>	<b>3 lay.</b>	<b>4 lay.</b>	<b>5 lay.</b>
SSO	SGCN	67.09	67.66	68.10	69.29	69.68
	WSGCN+	65.00	66.58	68.27	68.29	68.59
	WSGCN-	53.96	54.67	55.34	55.70	55.75
	WSGCN±	48.73	48.91	48.77	49.32	49.20
	WSGCNsb	52.22	54.16	53.96	53.80	54.77
	WSGCNgb	64.94	69.11	70.10	72.10	73.87
CCS	SGCN	69.55	70.09	70.68	70.85	71.71
	WSGCN+	69.76	70.24	70.59	70.33	70.93
	WSGCN-	69.86	70.69	70.92	71.00	70.96
	WSGCN±	70.53	70.59	70.80	70.97	71.41
	WSGCNsb	69.34	69.91	70.47	70.93	70.98
	WSGCNgb	72.07	72.05	72.40	72.46	72.81
EPF	SGCN	89.86	90.54	91.37	92.00	92.72
	WSGCN+	87.73	88.64	90.50	90.98	91.51
	WSGCN-	88.27	89.21	90.32	90.64	91.57
	WSGCN±	86.53	88.67	90.16	90.89	91.52
	WSGCNsb	90.90	91.98	93.06	93.67	94.53
	WSGCNgb	92.06	93.49	95.13	95.92	96.35