What relations are reliably embeddable in Euclidean space?

Robi Bhattacharjee   RCBHATTA@ENG.UCSD.EDU
Department of Computer Science
University of California, San Diego
San Diego, CA 92093, USA

Sanjoy Dasgupta   DASGUPTA@ENG.UCSD.EDU
Department of Computer Science
University of California, San Diego
San Diego, CA 92093, USA

Editor:

Abstract

We consider the problem of embedding a relation, represented as a directed graph, into Euclidean space. For three types of embeddings motivated by the recent literature on knowledge graphs, we obtain characterizations of which relations they are able to capture, as well as bounds on the minimal dimensionality and precision needed.

Keywords: graph embeddings; knowledge graphs.

1. Introduction

The problem of *embedding graphs in Euclidean space* has arisen in a variety of contexts over the past few decades. Most recently, it has been used for making symbolic knowledge available to neural nets, to help with basic reasoning tasks (Nickel et al., 2016). This knowledge consists of relations expressed in tuples, like (`Tokyo`, `is-capital-of`, `Japan`). Alternatively, each relation (like `is-capital-of`) can be thought of as a directed graph whose nodes correspond to *entities* (such as cities and countries).

A wide array of methods have been proposed for embedding such relations in vector spaces (Paccanaro and Hinton, 2001; Kemp et al., 2006; Sutskever et al., 2009; Bordes et al., 2011; Nickel et al., 2011; Bordes et al., 2013; Socher et al., 2013; Nickel and Kiela, 2017). For instance, *translational embeddings* (Bordes et al., 2013) map each entity x to a vector $\phi(x) \in \mathbb{R}^d$ and each relation r to a vector $\Psi(r) \in \mathbb{R}^d$. The intention is that for any entities x, y and any relation r,

$$\text{relation } (x, r, y) \text{ holds} \iff \phi(x) + \Psi(r) \approx \phi(y).$$

This is motivated in part by the success of *word embeddings* (Mikolov et al., 2013a), which embed words in Euclidean space so that words with similar co-occurrence statistics lie close to one another. It has been observed that these embeddings happen to obey linear relationships of the type above for certain relations and entities, making it possible, for instance, to use them for simple analogical reasoning (Mikolov et al., 2013b). Rather than relying upon these haphazard coincidences, it makes sense to explicitly embed relations of interest so that this property is assured.

An alternative scheme, *structured embeddings* (Bordes et al., 2011), again assigns each entity x a vector $\phi(x) \in \mathbb{R}^d$, but assigns each relation r a pair of $d \times d$ matrices, L_r and R_r, so that

$$\text{relation } (x, r, y) \text{ holds} \iff L_r \phi(x) \approx R_r \phi(y).$$

Notice that this is more general than translational embeddings because L_r can capture any affine transformation by adding a constant-valued feature to ϕ.

Another example of an embedding method is *bilinear embedding* (Nickel et al., 2011), in which each entity x gets a vector $\phi(x) \in \mathbb{R}^d$ and each relation r gets a matrix A_r, so that

$$\text{relation } (x, r, y) \text{ holds} \iff \phi(x)^T A_r \phi(y) \geq \text{ some threshold.}$$

These three embedding methods—translational, structured, and bilinear—broadly represent the various schemes that have been proposed in the recent machine learning literature, and many other suggestions are variants of these. For instance, *linear relational embedding* (Paccanaro and Hinton, 2001) assigns each entity x a vector $\phi(x)$ and each relation r a matrix M_r so that $(x, r, y) \iff \phi(y) \approx M_r\phi(x)$: a special case of structured embedding.

Typically the parameters of the embeddings (the mapping ϕ as well as the vectors and matrices for each relation) are fit to a given list of relation triples, using some suitable loss function. They can then be used for simple reasoning tasks, such as link prediction.

In this paper, we take a formal approach to this whole enterprise.

1. What kinds of relations can be embedded using these methods? Can arbitrary relations be accurately represented?

2. What dimensionality is needed for these embeddings?

3. What precision is needed for these embeddings? This question turns out to play a central role.

In particular, we will think of a relation as being reliably embeddable if it admits an embedding that does not require too much precision or too high a dimension. We wish to gauge what kinds of relations have this property.

In order to answer these questions, it is enough to look at a single relation at a time. We therefore look at the problem of *embedding a given directed graph in Euclidean space*.

1.1 Related work

There is a substantial literature on embedding *undirected* graphs into Euclidean space. A key result is the following.

Theorem 1 (Maehara (1984)) *For any undirected graph $G = (V, E)$, there is a mapping $\phi : V \to \mathbb{R}^d$ such that $\{u, v\} \in E \iff \|\phi(u) - \phi(v)\| \leq 1$. Here $d \leq |V|$.*

We will call this an *undirected distance embedding* to avoid confusion with embeddings of *directed* graphs, our main focus. The *sphericity* of an undirected graph G is defined as the smallest dimension d for which such an embedding exists; computing it was proved NP-hard by Kang and Muller (2012). The same paper showed an even more troubling result, that

2

embeddings achieving this minimum dimension sometimes require precision (number of bits per coordinate) *exponential* in $|V|$. This has been a key consideration in our formulation of robustness.

An embedding of an undirected graph can also be based on dot products rather than Euclidean distance. We call these *undirected similarity embeddings*. The following is known.

Theorem 2 (Reiterman et al. (1989)) *For any undirected graph $G = (V, E)$, there is a mapping $\phi : V \to \mathbb{R}^d$ such that $\{u, v\} \in E \iff \phi(u) \cdot \phi(v) \geq 1$. Here $d \leq |V|$.*

The minimum dimension d needed for an undirected similarity embedding is at most the sphericity of G, but can be much smaller. A complete binary tree on n nodes, for instance, has sphericity $\Omega(\log n)$ but can be embedded in $\mathbb{R}^3$ using a dot-product embedding (Reiterman et al., 1989).

The present paper is about embeddings of *directed* graphs, and many of the results we obtain are qualitatively different from the undirected case. We also diverge from earlier theory work by giving *precision* a central role in the analysis, via a suitable notion of robustness.

Another body of work, popular in theoretical computer science, has looked at *embeddings of distance metrics* into Euclidean space (Linial et al., 1995). Here a metric on finitely many points is specified by an undirected graph with edge lengths, where the distance between two nodes is the length of the shortest path between them. The idea is to find an embedding of the nodes into Euclidean space that preserves all distances. For many graphs, an embedding of this type is not possible: for constant-degree expander graphs, for instance, a multiplicative distortion of $\Omega(\log n)$ is inevitable, where n is the number of nodes. The problem we are considering differs in two critical respects: first, we only need to preserve immediate neighborhoods, and second, we are dealing with directed graphs.

The machine learning literature has proposed many methods for embedding, such as those mentioned above, along with empirical evaluation. There has also been work on embeddings into non-Euclidean spaces: complex-valued (Trouillon et al., 2016) or hyperbolic (Nickel and Kiela, 2017). In this paper, we focus on the Euclidean case.

1.2 Overview of results

Let $G = (V, E)$ be a directed graph representing a relation we wish to embed. Here V is the set of entities, and an edge (u, v) means that the relation holds from u to v.

We begin by considering a formalization of *translational embeddings*. We find that only a limited class of relations can be embedded this way: a directed cycle does not have a translational embedding (Theorem 4), but any directed acyclic graph does (Theorem 5).

Next, we consider more powerful classes of embeddings: abstractions of the structured and bilinear embeddings mentioned above, that we call *distance embeddings* and *similarity embeddings*, respectively. We find, first, that all directed graphs admit both types of embeddings (Theorem 10). Moreover, the minimum dimension achievable in the two cases differs by at most 1 (Theorem 18), and is closely related to the sign rank of the adjacency matrix of the graph (Theorem 26). We present several examples of embeddings for canonical types of graphs: paths, cycles, trees, and so on.

We also explicitly focus on the precision of embeddings, which has not been a feature of the earlier theory work on undirected graphs. In particular, we introduce a notion of δ-robustness, where larger values of δ correspond to more robust embeddings. We relate this directly to precision by showing that any graph that admits a δ-robust embedding also has distance and similarity embeddings into the $O((1/\delta^2)\log n)$-dimensional Boolean hypercube (Theorem 23). In this way, the δ parameter translates directly into an upper bound on the number of bits needed. We look at the robustness achievable on different families of graphs. We find, for instance, that for any graph of maximum degree D, robustness $\delta \geq 1/D$ can be attained (Theorem 10). On the other hand, random dense graphs are not robustly embeddable (Corollary 29).

Our analysis of embeddings focuses on two parameters: dimension and robustness. We show that the former is NP-hard to minimize (Appendix B), while the latter can be maximized efficiently by semidefinite programming (Section 6). Thus robustness is a promising optimization criterion for designing embeddings.

2. Translational embeddings

Definition 3 *A* translational embedding *of a directed graph $G = (V, E)$ is given by a mapping $\phi : V \to \mathbb{R}^d$, a unit vector $z \in \mathbb{R}^d$, and thresholds $\{t_u \geq 0 : u \in V\}$, such that for all $u \neq v$,*

$$(u, v) \in E \iff \|\phi(v) - (\phi(u) + z)\| \leq t_u.$$

If all the thresholds t_u are identical, then we call it a uniform *translational embedding.*

Note that (i) the requirement that z be a unit vector is without loss of generality, and (ii) we avoid checking self-edges in order to sidestep various complications. Paraphrasing, the definition imposes an ordinal constraint: if $(u, v) \in E$ but $(u, w) \notin E$, then $\phi(v)$ must lie closer to $\phi(u) + z$ than does $\phi(w)$.

For instance, let P_n denote the directed path $1 \to 2 \to \cdots \to n$. A uniform translational embedding in $\mathbb{R}$ is given by $\phi(k) = k$, $z = 1$, and any $0 < t < 1$.

As another example, consider the directed complete bipartite graph containing all edges from node set V_1 to complementary node set V_2. A uniform translational embedding to $\mathbb{R}$ is again available: map

$$\phi(u) = \begin{cases} 0 & \text{for } u \in V_1 \\ 1 & \text{for } u \in V_2 \end{cases}$$

with $z = 1$ and any $0 < t < 1$.

It is of interest to determine what kinds of graphs can be embedded translationally. We begin with a negative result.

Theorem 4 *C_n, the directed cycle on n nodes, does not admit a translational embedding for any $n \geq 3$.*

Proof Assume that $u_1 \to u_2 \to \cdots \to u_n \to u_1$ has a translational embedding $(\phi, z, \{t_u\})$; we will arrive at a contradiction.

First, for any edge $x \to y$, the conditions for $(x, y) \in E$ and $(y, x) \notin E$ are, respectively,

$$\|\phi(y) - (\phi(x) + z)\|^2 \leq t_x^2$$
$$\|\phi(x) - (\phi(y) + z)\|^2 > t_y^2$$

which can be rewritten

$$\|(\phi(y) - \phi(x)) - z\|^2 \le t_x^2$$
$$\|(\phi(y) - \phi(x)) + z\|^2 > t_y^2.$$

The left-hand sides have $\|\phi(y) - \phi(x)\|^2$ and $\|z\|^2$ in common. Subtracting, we get

$$z \cdot (\phi(y) - \phi(x)) > \frac{1}{4}(t_y^2 - t_x^2).$$

Now we can apply this to the n edges of the cycle to yield the system of inequalities

$$z \cdot (\phi(u_2) - \phi(u_1)) > \frac{1}{4}(t_{u_2}^2 - t_{u_1}^2)$$
$$z \cdot (\phi(u_3) - \phi(u_2)) > \frac{1}{4}(t_{u_3}^2 - t_{u_2}^2)$$
$$\vdots$$
$$z \cdot (\phi(u_1) - \phi(u_n)) > \frac{1}{4}(t_{u_1}^2 - t_{u_n}^2)$$

The left-hand sides add up to zero, as do the right-hand sides, a contradiction. ∎

On the other hand, any directed *acyclic* graph can be translationally embedded.

Theorem 5 *Suppose directed graph $G = (V, E)$ is acyclic. Then G admits a uniform translational embedding.*

Proof By topologically ordering G, assume without loss of generality that $V = \{1, 2, \ldots, n\}$ and that all edges $(i, j) \in E$ have $i < j$. Let $G' = (V, E')$ denote the undirected version of G, with an edge $\{i, j\} \in E'$ for every $(i, j) \in E$. By applying a result of Frankl and Maehara (1988), we obtain an embedding $\psi : V \to \mathbb{R}^d$ of G' with the following characteristics:

- $\|\psi(i)\|^2 = \Delta$, where $\Delta \ge 1$ is at most the maximum degree of G'.

- If $\{i, j\} \in E'$ then $\|\psi(i) - \psi(j)\|^2 = 2(\Delta - 1)$.

- If $\{i, j\} \notin E'$ then $\|\psi(i) - \psi(j)\|^2 = 2\Delta$.

We then define a uniform translational embedding of G into $\mathbb{R}^{d+1}$ as follows:

$$\phi(i) = (i\delta, \psi(i)),$$

where $\delta = 1/(n-1)$. Take $z = e_1$, the first coordinate direction, and threshold $t = \sqrt{2\Delta - 1}$.

To see that this works, pick any $i < j$. First off, if $(i, j) \in E$, then $\{i, j\} \in E'$ and

$$\|\phi(j) - (\phi(i) + z)\|^2 = (j\delta - i\delta - 1)^2 + \|\psi(i) - \psi(j)\|^2$$
$$\le (1 - \delta)^2 + 2(\Delta - 1) \ \le \ t^2.$$

On the other hand, if $(i, j) \notin E$, then $\{i, j\} \notin E'$, and we have

$$\|\phi(j) - (\phi(i) + z)\|^2 \ > \ \|\psi(i) - \psi(j)\|^2 \ = \ 2\Delta \ > \ t^2.$$

Finally, we confirm that the embedding does not suggest a back edge from j to i:

$$\|\phi(i) - (\phi(j) + z)\|^2 = (i\delta - j\delta - 1)^2 + \|\psi(i) - \psi(j)\|^2$$
$$\geq (1 + \delta)^2 + 2(\Delta - 1) \; > \; t^2.$$

∎

Open problem 1 *What characterization can be given for the* minimum *dimension of a translational embedding of a dag?*

3. Distance embeddings

Definition 6 *A* distance embedding *of a directed graph $G = (V, E)$ is given by a pair of mappings $\phi_{in}, \phi_{out} : V \to \mathbb{R}^d$, and a threshold t, such that for all pairs of nodes u, v,*

$$(u, v) \in E \iff \|\phi_{out}(u) - \phi_{in}(v)\| \leq t.$$

We will sometimes be interested in distance embeddings into the unit sphere, *where all $\phi_{in}(u)$ and $\phi_{out}(v)$ have length one.*

This formalism captures several types of embedding that have been proposed in the machine learning literature. Recall, for instance, the notion of a *structured embedding* (Bordes et al., 2011), given by $\phi : V \to \mathbb{R}^d$ and $d \times d$ matrices L and R, where $(u, v) \in E \iff L\phi(u) \approx R\phi(v)$. This can be converted into a distance embedding by taking $\phi_{out}(u) = L\phi(u)$ and $\phi_{in}(u) = R\phi(u)$. Conversely, if a graph has distance embedding $\phi_{in}, \phi_{out} : V \to \mathbb{R}^d$, then it has a structured embedding $(\phi : V \to \mathbb{R}^{2d}, L, R)$, where $\phi(u)$ is the concatenation of $\phi_{in}(u)$ and $\phi_{out}(u)$ and matrices L and R retrieve the bottom and top d coordinates, respectively, of a $2d$-dimensional vector.

In the above formulation of distance embedding, there is a single threshold, t, that applies for all points. An alternative would be to allow a different threshold t_u for each node u, so that

$$(u, v) \in E \iff \|\phi_{out}(u) - \phi_{in}(v)\| \leq t_u.$$

This is easily simulated under our current definition, by adding an extra dimension. Given an embedding $\phi_{in}, \phi_{out} : V \to \mathbb{R}^d$ with varying thresholds t_u, we can define $\tilde{\phi}_{in} : \tilde{\phi}_{out} : V \to \mathbb{R}^{d+1}$ by $\tilde{\phi}_{in}(u) = (\phi_{in}(u), 0)$ and $\tilde{\phi}_{out}(u) = (\phi_{out}(u), \sqrt{t^2 - t_u^2})$, where $t = \max_u t_u$. Then

$$\|\phi_{out}(u) - \phi_{in}(v)\| \leq t_u \iff \|\tilde{\phi}_{out}(u) - \tilde{\phi}_{in}(v)\| \leq t.$$

We will shortly see that every directed graph has a distance embedding. It is of interest, then, to characterize the minimum achievable dimension.

Definition 7 *Let $d_{dist}(G)$ be the smallest dimension d of any distance embedding of G. Let $d_{dist}^\circ(G)$ be the smallest dimension of any distance embedding into the unit sphere.*

A useful observation is that d_{dist} and d_{dist}° do not differ by much.

Theorem 8 *For any directed graph G, we have $d_{dist}(G) \leq d_{dist}^{\circ}(G) \leq d_{dist}(G) + 1$.*

Proof The first inequality is trivial. We give an informal sketch of the second, since the details appear in the proof of Theorem 22. A distance embedding ϕ of G in $\mathbb{R}^d$ can be mapped to an embedding ϕ' in a small neighborhood of the unit sphere $S^d \subset \mathbb{R}^{d+1}$. To see this, notice that scaling down ϕ (and t) by a constant factor maintains the embedding property. If they are sufficiently downscaled that the set of embedded points lies within a d-dimensional ball of very small radius, then this ball can be placed close to the surface of the unit sphere in $\mathbb{R}^{d+1}$, and the points can be projected to the surface of the sphere while inducing an arbitrarily small multiplicative distortion in pairwise distances. ∎

As described in the introduction, earlier work has brought out troubling pathologies in the precision required for embedding an undirected graph: achieving the minimum possible dimension could require the vectors to be specified using a number of bits that is exponential in $|V|$ (Kang and Muller, 2012). For this reason, we keep careful track of precision. Our key tool in doing so is a notion of *robustness*, which we will later relate to both precision and dimension.

Definition 9 *Suppose a distance embedding of a directed graph $G = (V, E)$ is given by $(\phi_{in}, \phi_{out}, t)$. We say the embedding is δ-robust, for $\delta > 0$, if*

- $(u, v) \in E \implies \|\phi_{out}(u) - \phi_{in}(v)\|^2 \leq t^2.$

- $(u, v) \notin E \implies \|\phi_{out}(u) - \phi_{in}(v)\|^2 \geq t^2(1 + \delta).$

We now show that all directed graphs have distance embeddings.

Theorem 10 *Let $G = (V, E)$ be any directed graph. Let A be its $|V| \times |V|$ adjacency matrix: that is, A_{uv} is 1 if $(u, v) \in E$ and 0 otherwise. Let k denote the rank of A and σ_1 its largest singular value. Then G has a distance embedding into the unit sphere in $\mathbb{R}^k$ that is $(1/\sigma_1)$-robust.*

Proof For convenience, label the vertices $1, 2, \ldots, n$. Take the singular value decomposition of A so that $A = U^T \Sigma V$ where U and V are $n \times n$ orthogonal matrices, and Σ is a diagonal matrix with entries $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n$. If rank $k < n$, then $\sigma_{k+1} = \cdots = \sigma_n = 0$.

Writing $A = (\Sigma^{1/2}U)^T(\Sigma^{1/2}V)$, take $\phi_{out}(i) \in \mathbb{R}^k$ to be the first k coordinates of the ith column of $\Sigma^{1/2}U$ (the remaining coordinates are zero), and $\phi_{in}(i) \in \mathbb{R}^k$ to be the first k coordinates of the ith column of $\Sigma^{1/2}V$. Then $A_{ij} = \phi_{out}(i) \cdot \phi_{in}(j)$. These vectors all have length at most $\sqrt{\sigma_1}$; normalize them to unit length, to get $\widehat{\phi}_{out}, \widehat{\phi}_{in} : V \to S^{k-1}$. Then

- $(i, j) \in E \implies \widehat{\phi}_{out}(i) \cdot \widehat{\phi}_{in}(j) \geq 1/\sigma_1$ and $\|\widehat{\phi}_{out}(i) - \widehat{\phi}_{in}(j)\|^2 \leq 2(1 - 1/\sigma_1).$

- $(i, j) \notin E \implies \widehat{\phi}_{out}(i) \cdot \widehat{\phi}_{in}(j) = 0$ and $\|\widehat{\phi}_{out}(i) - \widehat{\phi}_{in}(j)\|^2 = 2.$

Setting $t = \sqrt{2(1 - 1/\sigma_1)}$, we see the embedding is δ-robust for $\delta \geq 1/(1 - 1/\sigma_1) - 1 \geq 1/\sigma_1$. ∎

As a consequence, any graph of constant degree is robustly embeddable. The proof of the following corollary is deferred to the appendix.

Corollary 11 *Suppose all nodes in G have indegree $\leq \Delta_-$ and outdegree $\leq \Delta_+$. Then G has a distance embedding that is $\sqrt{1/(\Delta_+ \Delta_-)}$-robust.*

3.1 Differences from undirected embeddings

At first glance, directed embeddings may not seem all that different from undirected embeddings considering standard transformations between the two types of graphs, as defined below. To show that the two cases are quite different, we now exhibit examples in which a directed graph can be embedded in much lower dimension than its undirected counterpart.

Definition 12   *1. For an undirected graph G, let $\overleftrightarrow{G}$ be the directed graph which has edges $(u, v), (v, u) \in E(\overleftrightarrow{G})$ for every $\{u, v\} \in E(G)$.*

   2. For a directed graph G, let $\overline{G}$ be the undirected graph with 2 vertices v_{out}, v_{in} for every $v \in V(G)$ and with edge $\{u_{out}, v_{in}\} \in E(\overline{G})$ for every $(u, v) \in E(G)$.

In both cases, it is easily seen that the minimum dimension required to embed the directed version of the graph, ignoring self-edges, is at most the dimension needed for the undirected version (that is, the sphericity). But the converse is not true even approximately.

Theorem 13 *Let $G = K_{n,n}$ be the undirected complete bipartite graph. Then G has sphericity $\geq n$ whereas $d_{dist}(\overleftrightarrow{G}) \leq 1$.*

The proof appears in the appendix. It is well-known that the sphericity of the complete bipartite graph $K_{n,n}$ is at least n (Maehara, 1986); a basic intuition is that embedding undirected bipartite graphs can be difficult because they have large independent sets with many common neighbors. However, for directed graphs this problem is mitigated by the flexibility that comes from having two embeddings, ϕ_{in} and ϕ_{out}.

Our next example follows the same principle.

Theorem 14 *Let $G = \overleftrightarrow{K_n}$ be a directed graph with every possible edge (including self loops). Then $d_{dist}(G) = 0$ while $\overline{G}$ has sphericity $\geq n$.*

3.2 Robustness yields low dimensionality

We now show that any graph with a δ-robust embedding can be embedded in dimension $O((1/\delta^2) \log n)$.

Theorem 15 *If G has a δ-robust distance embedding (in any dimension), then it also has a $\frac{\delta}{2}$-robust embedding in $O(\frac{1}{\delta^2} \log n)$ dimensions.*

Proof   This is a consequence of a lemma of Johnson and Lindenstrauss (1984). Let $\phi_{out}, \phi_{in} : V \to \mathbb{R}^d$, with threshold t, be a δ-robust embedding of G. The JL lemma states that for any $\epsilon > 0$, there exists a map $f : \mathbb{R}^d \to \mathbb{R}^m$, with $m = O((\log n)/\epsilon^2)$, so that

$$(1 - \epsilon)\|\phi_{out}(u) - \phi_{in}(v)\|^2 \leq \|f(\phi_{out}(u)) - f(\phi_{in}(v))\|^2 \leq (1 + \epsilon)\|\phi_{out}(u) - \phi_{in}(v)\|^2,$$

for all $u, v \in V$. To ensure that the new embedding is $(\delta/2)$-robust, it suffices to take $\epsilon = \delta/8$. ∎

Later, we will see that a graph with a δ-robust embedding is in fact robustly embeddable in the $O((1/\delta^2) \log n)$-dimensional *Hamming cube*. In this way, robustness implies the existence of a low-dimensional embedding that requires only one bit of precision per coordinate.

4. Similarity embeddings

Definition 16 *A similarity embedding of a directed graph $G = (V, E)$ is given by a pair of mappings $\phi_L, \phi_R : V \to \mathbb{R}^d$ and a threshold t, such that*

$$(u, v) \in E \iff \phi_L(u) \cdot \phi_R(v) \geq t.$$

We will often be interested in embeddings into the unit sphere, where the $\phi_L(u)$ and $\phi_R(u)$ have unit norm. We use L, R notation as opposed to $\{\text{in}, \text{out}\}$ to help distinguish similarity embeddings from distance embeddings.

This is closely related to the notion of bilinear embedding (Nickel et al., 2011), which assigns each node u to a vector $\phi(u) \in \mathbb{R}^d$ so that $(u, v) \in E \iff \phi(u)^T A \phi(v) \geq t$, for some $d \times d$ matrix A. To obtain a similarity embedding, take $\phi_L(u) = \phi(u)$ and $\phi_R(u) = A\phi(u)$. Conversely, given a similarity embedding $\phi_L, \phi_R : V \to \mathbb{R}^d$, we can construct a bilinear embedding by setting $\phi(u)$ to the $2d$-dimensional concatenation of $\phi_L(u)$ and $\phi_R(u)$, and taking A to be $\begin{pmatrix} 0 & I \\ 0 & 0 \end{pmatrix}$.

The distance embedding constructed in Theorem 10 also functions as a similarity embedding. Thus, such embeddings exist for every graph.

Definition 17 *For directed graph G, let $d_{sim}(G)$ denote the smallest dimension into which a similarity embedding can be given.*

We now see that the dimensions $d_{dist}, d_{dist}^\circ, d_{sim}$ are almost identical.

Theorem 18 $d_{dist}^\circ(G) - 1 \leq d_{sim}(G) \leq d_{dist}^\circ(G) \leq d_{dist}(G) + 1$ *for any directed graph G.*

Proof The inequality $d_{sim}(G) \leq d_{dist}^\circ(G)$ is immediate: any distance embedding into the unit sphere automatically meets the requirements of a similarity embedding. The final inequality is from Theorem 8. It thus remains to show that $d_{dist}^\circ(G) \leq d_{sim}(G) + 1$.

Let $\phi_L, \phi_R : V \to \mathbb{R}^d$ be a similarity embedding of G with threshold t. Indexing vertices as $1, 2, \ldots, n$, let M be an $n \times n$ matrix with $M_{ij} = \phi_L(i) \cdot \phi_R(j)$.

If J is the all-ones matrix, then $M - tJ$ is matrix of rank at most $d + 1$ such that

$$(M - tJ)_{ij} \geq 0 \iff (i, j) \in E.$$

We will extract a distance embedding into the unit sphere from this matrix.

Express $M - tJ$ as $U^T W$ for $U, W \in \mathbb{R}^{(d+1) \times n}$. Next, normalize the columns of U and W to unit length, to get $\widehat{U}$ and $\widehat{W}$. The key idea is that the pairwise dot products between

these unit vectors still satisfy the above criterion. In short, taking $\phi_{out}(i)$ to be the ith column of $\widehat{U}$ and $\phi_{in}(i)$ to be the ith column of $\widehat{W}$, we get a distance embedding of G into the unit sphere in $\mathbb{R}^{d+1}$:

$$(i, j) \in E \iff \phi_{out}(i) \cdot \phi_{in}(j) \geq 0 \iff \|\phi_{out}(i) - \phi_{in}(j)\|^2 \leq 2.$$

∎

Combined with Theorem 8, this means that $|d_{dist}(G) - d_{sim}(G)| \leq 1$. In contrast, for undirected graphs, the minimum dimension needed by a dot-product embedding could be significantly less than for a distance embedding (Reiterman et al., 1989).

4.1 Robust similarity embeddings

Measuring the robustness of a similarity embedding is a bit different than for distance embeddings. For instance, the threshold for a similarity embedding need not even be positive, and thus a term of form $t(1 + \delta)$ is not meaningful. We use an additive rather than multiplicative notion of robustness.

Definition 19 *We say a similarity embedding given by (ϕ_L, ϕ_R, t) is δ-robust, for $\delta > 0$, if*

$$(u, v) \notin E \implies \phi_L(u) \cdot \phi_R(v) \leq t - \delta \max_{w \in V} \max(\|\phi_L(w)\|^2, \|\phi_R(w)\|^2).$$

The term $\max_{w \in V} \|\phi(w)\|^2$ ensures that rescaling a similarity embedding does not change its robustness parameter.

Theorem 10 produces a distance embedding in the unit sphere, which is therefore also a similarity embedding. The following is an immediate corollary.

Corollary 20 *Let $G = (V, E)$ be any directed graph. If its adjacency matrix has rank k and largest singular value σ_1, then G has a $(1/\sigma_1)$-robust similarity embedding into the unit sphere in $\mathbb{R}^k$.*

4.2 Relationship between similarity-robustness and distance-robustness

The two presented definitions of robustness are closely linked. Most immediately, robust similarity embeddings necessarily imply the existence of robust distance embeddings.

Theorem 21 *Let G be a graph that has a δ-robust similarity embedding. Then G has distance-robustness at least $\delta/2$.*

An approximate converse is also true.

Theorem 22 *Let $G = (V, E)$ be a directed graph with a δ-robust distance embedding $(\phi_{in}, \phi_{out}, t)$ into $\mathbb{R}^d$. Let B be the largest length of the vectors ϕ_{in}, ϕ_{out}, and define the scaled diameter of the embedding as*

$$\Delta = \max\left(\frac{B}{t}, 1\right).$$

Then G has a similarity embedding into $\mathbb{R}^{d+1}$ with robustness $\delta^2/(18\Delta^4)$.

4.3 Embeddings into the Hamming cube

We now show that any graph that has a δ-robust similarity embedding (into any dimension) can be embedded robustly into the $O((1/\delta^2)\log n)$-dimensional Hamming cube. Thus this notion of robustness translates directly into a bound on the number of *bits* of precision needed for embedding.

Theorem 23 *Suppose directed graph $G = (V, E)$ has a δ-robust similarity embedding into the unit sphere. Then it has an $O(\delta)$-robust distance embedding into $\{0,1\}^k$, where $k = O((1/\delta^2)\log n)$, that is simultaneously an $O(\delta)$-robust similarity embedding.*

A proof is found in the appendix.

Notice that by combining Theorems 10 and 23, we see that any directed graph whose indegrees and outdegrees are bounded by Δ has both distance and similarity embeddings into the $O((1/\Delta^2)\log n)$-dimensional Hamming cube that are $O(1/\Delta)$-robust.

A partial converse is immediate: any distance or similarity embedding into $\{0,1\}^k$ is necessarily at least $(1/k)$-robust. Thus, robustness can serve as an approximate proxy for dimension.

On the other hand, it is unclear whether embeddability in low-dimensional Euclidean space necessarily implies the existence of a robust embedding.

Open problem 2 *Does the existence of a low-dimensional embedding imply that there also exists a robust embedding?*

5. Lower bounds

5.1 Sign rank

Our previous construction from the proof of Theorem 18 with $M - tJ$ yields a matrix in which positive elements correspond to edges in G, and negative elements correspond to non-edges. This reveals a natural relationship between finding similarity embeddings and finding low rank sign matrices of an adjacency graph.

Definition 24 *A sign matrix M is a matrix whose entries are either $+$ or $-$. The sign rank of M is the minimum rank of any real-valued matrix whose entries agree in sign with M. We use the convention that 0 is treated as neither $+$ nor $-$, and consequently the minimum rank matrix must have all non-zero elements.*

There is a natural bijection between directed graphs on n nodes and sign matrices of size $n \times n$, with a matrix entry of $+$ corresponding to an edge and $-$ corresponding to a non-edge. Let $M(G)$ denote this sign matrix for directed graph G.

Definition 25 *The sign rank of a graph G, denoted $d_{sign}(G)$, is taken to be the sign rank of $M(G)$.*

Using the same construction we used in Theorem 18 we find that $d_{sign}(G)$ is closely linked to our other notions of dimension.

Theorem 26 *For any graph G, we have $d_{sim}(G) \leq d_{sign}(G) \leq d_{sim}(G) + 1$.*

5.2 Random graphs

In this section, we show that with high probability, random dense directed graphs have large embedding dimension, and thus low robustness. This is because there are far more directed graphs on n nodes than there are $n \times n$ sign matrices of low sign rank.

Theorem 27 *(Alon et al. (2016)) For $r \leq n/2$, the number of $n \times n$ sign matrices of sign rank at most r does not exceed $2^{O(rn \log n)}$.*

Theorem 28 *Let G be a random directed graph over n vertices such that each edge is chosen independently with constant probability $0 < p < 1$. Then for large enough n, with high probability,*

$$d_{sign}(G) = \Omega\left(\frac{nH(p)}{\log n}\right),$$

where $H(p) = -p \log p - (1-p) \log(1-p)$.

Proof Consider a random graph G drawn by selecting each edge independently with probability p. Fix any $\epsilon > 0$ and consider the typical set T_ϵ induced by these random graphs, as per the asymptotic equipartition property (Cover and Thomas, 2006). It follows that for sufficiently large n,

1. $G \in T_\epsilon$ with probability $\geq 1 - \epsilon$.

2. For any $G \in T_\epsilon$, we have $\Pr(G) \leq 2^{-n^2(H(p)-\epsilon)}$.

By Theorem 27, for any r, the maximum number of elements $G \in T_\epsilon$ for which $M(G)$ has sign rank $\leq r$ is $2^{O(rn \log n)}$. Consequently, the random graph G has sign rank at least r with probability at least

$$P(d_{sign}(G) \geq r) \geq 1 - \epsilon - 2^{O(rn \log n)} 2^{-(n^2)(H(p)-\epsilon)}.$$

Selecting $r = \frac{CnH(p)}{\log n}$ for a sufficiently small constant C finishes the proof. ∎

Applying Theorems 15, 21 then shows that random graphs have low robustness.

Corollary 29 *Let G be a random directed graph over n vertices such that each edge is chosen with constant probability p. Then, with high probability, G has distance robustness and similarity robustness at most $O(\frac{\log n}{\sqrt{n}})$.*

6. Algorithms for robust embedding

We show in the appendix that computing $d_{dist}(G)$ and $d_{sim}(G)$ are both NP-hard problems. On the other hand, computing the robustness of a graph G is more tractable.

We present a semidefinite programming approach to finding the distance-robustness and similarity-robustness of G. This can be used (see Theorems 15 and 23) to construct low-dimensional robust embeddings of G.

6.1 Distance embeddings

Given a graph $G = (V, E)$ with $V = \{v_1, v_2, \ldots, v_n\}$, we find a robustness-maximizing distance embedding with a semidefinite program. For convenience, we will let x_i denote $\phi_{out}(v_i)$ and y_i denote $\phi_{in}(v_i)$. We also include a scalar variable δ which represents the robustness, and assume (without loss of generality) that our threshold $t = 1$. Then, the following semidefinite program suffices.

$$\begin{aligned} \underset{x,y,t,\delta}{\text{maximize}} \quad & \delta \\ \text{subject to} \quad & \langle x_i, x_i \rangle + \langle y_j, y_j \rangle - 2\langle x_i, y_j \rangle \leq 1, \ (v_i, v_j) \in E \\ & \langle x_i, x_i \rangle + \langle y_j, y_j \rangle - 2\langle x_i, y_j \rangle \geq 1 + \delta, \ (v_i, v_j) \notin E \end{aligned}$$

6.2 Similarity embeddings

This similarity case is analogous, but has the detail that we restrict ourselves to unit vectors. This is still guaranteed to find the optimal robustness since any similarity embedding can be converted into a spherical embedding with the same robustness (albeit higher dimension).

$$\begin{aligned} \underset{x,y,t,\delta}{\text{maximize}} \quad & \delta \\ \text{subject to} \quad & \langle x_i, y_j \rangle \geq t, \ (v_i, v_j) \in E \\ & \langle x_i, y_j \rangle \leq t - \delta, \ (v_i, v_j) \notin E \\ & \langle x_i, x_i \rangle = \langle y_i, y_i \rangle = 1, \ 1 \leq i \leq n \end{aligned}$$

Acknowledgments

References

N. Alon, S. Moran, and A. Yehudayoff. Sign rank versus VC dimension. In *Conference on Learning Theory*, 2016.

A. Bordes, J. Weston, R. Collobert, and Y. Bengio. Learning structured embeddings of knowledge bases. In *Twenty-Fifth AAAI Conference on Artificial Intelligence*, 2011.

A. Bordes, N. Usunier, A. Garcia-Duran, J. Weston, and O. Yakhnenko. Translating embeddings for modeling multi-relational data. In *Advances in Neural Information Processing Systems*, 2013.

M.S. Charikar. Similarity estimation techniques from rounding algorithms. In *Thirty-Fourth Annual ACM Symposium on Theory of Computing*, 2002.

T. Cover and J. Thomas. *Elements of Information Theory*. 2006.

P. Frankl and H. Maehara. The Johnson-Lindenstrauss lemma and the sphericity of some graphs. *Journal of Combinatorial Theory Series B*, 44(3):355–362, 1988.

M.X. Goemans and D.P. Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM*, 42 (6):1115–1145, 1995.

W.B. Johnson and J. Lindenstrauss. Extensions of Lipschitz maps into a Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

R.J. Kang and T. Muller. Sphere and dot product representations of graphs. *Discrete and Computational Geometry*, 47(3):548–568, 2012.

C. Kemp, J.B. Tenenbaum, T.L. Griffiths, T. Yamada, and N. Ueda. Learning systems of concepts with an infinite relational model. In *Twenty-First National Conference on Artificial Intelligence*, 2006.

N. Linial, E. London, and Y. Rabinovich. The geometry of graphs and some of its algorithmic applications. *Combinatorica*, 15(2):215–245, 1995.

H. Maehara. Space graphs and sphericity. *Discrete Applied Mathematics*, 7(1):55–64, 1984.

H. Maehara. Sphericity exceeds cubicity for almost all complete bipartite graphs. *Journal of Combinatorial Theory Series B*, 40:231–235, 1986.

T. Mikolov, I. Sutskever, K. Chen, G. Corrado, and J. Dean. Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, 2013a.

T. Mikolov, W.-T. Yih, and G. Zweig. Linguistic regularities in continuous space word representations. In *Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2013b.

M. Nickel and D. Kiela. Poincare embeddings for learning hierarchical representations. In *Advances in Neural Information Processing Systems*, 2017.

M. Nickel, V. Tresp, and H.-P. Kriegel. A three-way model for collective learning on multi-relational data. In *International Conference on Machine Learning*, 2011.

M. Nickel, K.P. Murphy, V. Tresp, and E. Gabrilovich. A review of relational machine learning for knowledge graphs. *Proceedings of the IEEE*, 104(1):11–33, 2016.

A. Paccanaro and G.E. Hinton. Learning hierarchical structures with linear relational embedding. In *Advances in Neural Information Processing Systems*, 2001.

J. Reiterman, V. Rodl, and E. Sinajova. Geometrical embeddings of graphs. *Discrete Mathematics*, 74(3):291–319, 1989.

R. Socher, D. Chen, C.D. Manning, and A. Ng. Reasoning with neural tensor networks for knowledge base completion. In *Advances in Neural Information Processing Systems*, 2013.

I. Sutskever, J.B. Tenenbaum, and R. Salakhutdinov. Modeling relational data using Bayesian clustered tensor factorization. In *Advances in Neural Information Processing Systems*, 2009.

T. Trouillon, J. Welbl, S. Reidel, E. Gaussier, and G. Bouchard. Complex embeddings for simple link prediction. In *International Conference on Machine Learning*, 2016.

Appendix A. Proofs to selected theorems

A.1 Proof of Corollary 11

Proof This follows immediately from Theorem 10 because the largest singular value of the adjacency matrix will be at most $\sqrt{\Delta_+ \Delta_-}$. This is doubtless a well-known fact, but for completeness we give a brief explanation here.

The top singular value σ_1 is the square root of the top eigenvalue of $A^T A$, call it λ. Let v be the corresponding eigenvector. Since $A^T A$ has no negative entries, we may assume $v \geq 0$ (flipping every entry of v to its absolute value can only increase $v^T A^T A v$). If v_i is the largest entry of v,

$$\lambda v_i = (A^T A v)_i = \sum_{j=1}^{n} (A^T A)_{ij} v_j \leq v_i \sum_j (A^T A)_{ij} = v_i \left(\sum_\ell A_{\ell i} \left(\sum_j A_{\ell j} \right) \right) \leq v_i \Delta_- \Delta_+.$$

Thus $\lambda \leq \Delta_+ \Delta_-$ and $\sigma_1 \leq \sqrt{\Delta_+ \Delta_-}$. ∎

A.2 Proof of Theorem 13

Proof It is known that G has sphericity $\geq n$ (Maehara, 1986). To embed $\overleftrightarrow{G}$, let A, B be its partition into two independent sets. Then for any $a \in A$, $\phi_{out}(a) = -1, \phi_{in}(a) = 1$. For any $b \in B$, $\phi_{out}(b) = 1, \phi_{in}(b) = -1$. This embedding ϕ, with $t = 0$, is a distance embedding into $\mathbb{R}$. ∎

A.3 Proof of Theorem 21

Proof Let ϕ be a δ-robust similarity embedding of G. Rescale the embedding so that $\phi_{L,R}(v)$ all have norm at most 1. Then it follows that for some $t \in [-1, 1]$,

1. $(u, v) \in E$ if and only if $\langle \phi_L(u), \phi_R(v) \rangle \geq t$.

2. $(u, v) \notin E$, if and only if $\langle \phi_L(u), \phi_R(v) \rangle \leq t - \delta$.

Next, we convert this embedding into a spherical embedding as follows. Let e, f be unit vectors that are orthogonal to each other and to all vectors in our embedding. We append suitable multiples of e to each $\phi_L(v)$ vector and of f to each $\phi_R(v)$, so that the resulting vectors all have unit length. This operation preserves dot products and thus gives a spherical embedding ϕ' with robustness δ.

Since ϕ' is spherical, it is also a distance embedding with $\phi'_{L,R} = \phi'_{out,in}$ where

1. $(u, v) \in E$ if and only if $\|\phi'_{out}(u) - \phi'_{in}(v)\|^2 \leq 2 - 2t$.

2. $(u, v) \notin E$ if and only if $\|\phi'_{out}(u) - \phi'_{in}(v)\|^2 \geq 2 - 2t + 2\delta$.

From here, we can lower-bound the distance robustness of ϕ' by

$$\frac{2 - 2t + 2\delta}{2 - 2t} - 1 \;\geq\; \frac{\delta}{2}.$$

$\blacksquare$

A.4 Proof of Theorem 22

Proof Recall our earlier idea, in the proof sketch for Theorem 8, of placing distance embeddings on the unit sphere by mapping them onto a small neighborhood of the sphere. We will now look at a particular realization of this method.

The distance embedding given by $\phi_{in}, \phi_{out} : V \to \mathbb{R}^d$ can be scaled so that $t = 1$. We then get the following, for all $u, v \in V$.

(a) $(u, v) \notin E$ if and only if $\|\phi_{out}(u) - \phi_{in}(v)\|^2 \geq 1 + \delta$.

(b) $(u, v) \in E$ if and only if $\|\phi_{out}(u) - \phi_{in}(v)\|^2 \leq 1$.

(c) All $\|\phi_{in}(u)\|, \|\phi_{out}(u)\|$ are $\leq \Delta$.

Let e be a unit vector orthogonal to all embedded vectors (i.e. a new dimension). For some constant $c > 0$ whose value we will later set, let $\phi'_L, \phi'_R : V \to \mathbb{R}^{d+1}$ be defined by

$$\phi'_R(v) = \frac{e + c\,\phi_{in}(v)}{\sqrt{1 + c^2 \|\phi_{in}(v)\|^2}}$$

$$\phi'_L(v) = \frac{e + c\,\phi_{out}(v)}{\sqrt{1 + c^2 \|\phi_{out}(v)\|^2}}$$

Notice that these vectors have unit length. It follows that for any u, v,

$$\langle \phi'_L(u), \phi'_R(v) \rangle = \frac{1 + c^2 \langle \phi_{out}(u), \phi_{in}(v) \rangle}{\sqrt{(1 + c^2 \|\phi_{out}(u)\|^2)(1 + c^2 \|\phi_{in}(v)\|^2)}}$$
$$= \frac{(1 + c^2 \|\phi_{out}(u)\|^2) + (1 + c^2 \|\phi_{in}(v)\|^2) - c^2 \|\phi_{out}(u) - \phi_{in}(v)\|^2}{2\sqrt{(1 + c^2 \|\phi_{out}(u)\|^2)(1 + c^2 \|\phi_{in}(v)\|^2)}}$$

We now bound this quantity in the cases that $(u, v) \in E$ and $(u, v) \notin E$. In doing so, we will show that ϕ' is a similarity embedding, into the unit sphere, of the desired robustness.

Suppose $(u, v) \in E$. Using the inequality $A + B \geq 2\sqrt{AB}$ for $A, B \geq 0$ as well as property (b), we have

$$\langle \phi'_L(u), \phi'_R(v) \rangle \geq 1 - \frac{c^2}{2} \cdot \frac{\|\phi_{out}(u) - \phi_{in}(v)\|^2}{\sqrt{(1 + c^2 \|\phi_{out}(u)\|^2)(1 + c^2 \|\phi_{in}(v)\|^2)}}$$

$$\geq 1 - \frac{c^2}{2}.$$

On the other hand, if $(u, v) \notin E$, then by properties (a) and (c),

$$\langle \phi'_L(u), \phi'_R(v) \rangle \leq \frac{1}{2} \left(\sqrt{\frac{1 + c^2 \|\phi_{out}(u)\|^2}{1 + c^2 \|\phi_{in}(v)\|^2}} + \sqrt{\frac{1 + c^2 \|\phi_{in}(v)\|^2}{1 + c^2 \|\phi_{out}(u)\|^2}} \right) - \frac{c^2}{2} \cdot \frac{1 + \delta}{1 + c^2 \Delta^2}.$$

We can simplify the first term using the inequalities $(1 + x) + 1/(1 + x) \leq 2 + 2x^2$ and $\sqrt{1 + x} \leq 1 + x/2$ for $x \geq 0$. Again using property (c), we get

$$\langle \phi'_L(u), \phi'_R(v) \rangle \leq 1 + \frac{c^4 \Delta^4}{4} - \frac{c^2}{2} \cdot \frac{1 + \delta}{1 + c^2 \Delta^2}.$$

Set $c = \sqrt{\delta/(3\Delta^4)}$. Then $c^2 \Delta^2 \leq \delta/3$ and we get

$$\langle \phi'_L(u), \phi'_R(v) \rangle \leq 1 - \frac{c^2}{2} \left(1 + \frac{\delta}{2} - \frac{c^2 \Delta^4}{2} \right) = 1 - \frac{c^2}{2} \left(1 + \frac{\delta}{3} \right).$$

The robustness of a similarity embedding is measured additively, and follows by taking the difference of the expressions for the cases when $(u, v) \in E$ and $(u, v) \notin E$. ∎

A.5 Proof of Theorem 23

Proof Write $n = |V|$. Suppose the $2n$ vectors $\phi_L(u)$, $\phi_R(u)$ lie on S^{d-1}, the unit sphere in $\mathbb{R}^d$, and constitute a δ-robust similarity embedding: for some threshold t, and any u, v,

- $(u, v) \in E \implies \phi_L(u) \cdot \phi_R(v) \geq t + \delta$, and

- $(u, v) \notin E \implies \phi_L(u) \cdot \phi_R(v) \leq t$.

We will embed these vectors into the Hamming cube using the random halfspace method of Goemans and Williamson (1995) and Charikar (2002). Specifically, pick k vectors $r_1, \ldots, r_k$ uniformly at random from S^{d-1}, and define the embedding $h : \mathbb{R}^d \to \{0, 1\}^k$ by $h(x) = (h_1(x), \ldots, h_k(x))$, where the ith hash function $h_i : \mathbb{R}^d \to \{0, 1\}$ is

$$h_i(x) = \begin{cases} 1 & \text{if } r_i \cdot x \geq 0 \\ 0 & \text{if } r_i \cdot x < 0 \end{cases}$$

Now, for any vectors x, y,

$$\Pr(h_i(x) \neq h_i(y)) = \Pr((r_i \cdot x)(r_i \cdot y) \leq 0) = \frac{\theta}{\pi},$$

where θ is the angle between x and y. Thus for nodes u, v in G,

$$\Pr(h_i(\phi_L(u)) \neq h_i(\phi_R(v))) = \arccos(\phi_L(u) \cdot \phi_R(v))/\pi$$
$$\begin{cases} \leq \arccos(t + \delta)/\pi & \text{if } (u, v) \in E \\ \geq \arccos(t)/\pi & \text{if } (u, v) \notin E \end{cases} \tag{1}$$

The difference between the two options is:

$$\frac{1}{\pi}\left(\arccos(t) - \arccos(t+\delta)\right) = \frac{1}{\pi}\int_t^{t+\delta} \frac{dz}{\sqrt{1-z^2}} \geq \frac{\delta}{\pi}.$$

Write $h_L(u) = h(\phi_L(u))$ and $h_R(u) = h(\phi_R(u))$. Letting $d(\cdot)$ denote Hamming distance in $\{0,1\}^k$, we have that the expected value of $d(h_L(u), h_R(v))$ is k times the quantity in equation (1). A simple Chernoff-Hoeffding bound, unioned over all pairs u, v, then suffices to show that if $k = O((1/\delta^2)\log n)$, then with probability at least $1 - 1/n$,

- $(u,v) \in E \implies d(h_L(u), h_R(v)) \leq k(\arccos(t) - 2\delta/3)/\pi$, and

- $(u,v) \notin E \implies d(h_L(u), h_R(v)) \geq k(\arccos(t) - \delta/3)/\pi$.

Thus h_L, h_R constitute an $O(\delta)$-robust distance embedding. To see that this is also an $O(\delta)$-robust similarity embedding, notice that all the embedded vectors $h_L(u)$ and $h_R(u)$ have expected squared Euclidean norm $k/2$, and given the setting of k, these norms will be tightly concentrated, within multiplicative factor $1 \pm O(\delta)$, about their expected values. ∎

Appendix B. NP hardness results

In this section, we show that it is NP-hard to find distance or similarity embeddings of minimum dimension. We do so by adapting the results of Kang and Muller (2012) on undirected embeddings to the directed case. First, we briefly review some definitions from their paper. Readers interested in further details should consult their very clear presentation.

Definition 30 *(Kang and Muller (2012))*

1. *An* oriented k-hyperplane arrangement *$\mathcal{H} = \{h_1, h_2, \ldots, h_n\}$ is a set of hyperplanes in $\mathbb{R}^k$ each of which is given an orientation, so that all points in $\mathbb{R}^k$ are either on the positive side of h_i, denoted h_i^+, or the negative side, denoted h_i^-, or on h_i itself.*

2. *The* sign vector *of a point $p \in \mathbb{R}^k$ with respect to $\mathcal{H}$ is the vector $\sigma(p) \in \{+, -, 0\}^n$ such that*
$$\sigma(p)_i = \begin{cases} + & \text{if } p \in h_i^+ \\ - & \text{if } p \in h_i^- \\ 0 & \text{if } p \in h_i \end{cases}$$

3. *The* combinatorial description *of $\mathcal{H}$ is defined to be the set of all sign vectors, $\mathcal{D}(\mathcal{H}) = \{\sigma(p) : p \in \mathbb{R}^k\}$.*

4. *Consider any set $S \subset \{-, +\}^n$ containing $(-, -, -, \ldots, -)$ and $(+, +, \ldots, +)$. We say S is k-realizable if there exists an oriented k-hyperplane arrangement $\mathcal{H}$ with $S \subset \mathcal{D}(\mathcal{H})$.*

5. *k-REALIZABILITY denotes the algorithmic problem of deciding, given a set $S \subset \{-, +\}^n$ as input, whether S is k-realizable.*

Theorem 31 *(Kang and Muller (2012)) k-REALIZABILITY is NP-hard for all $k > 1$.*

B.1 Distance embeddings

The main idea is to reduce from k-REALIZABILITY. Given a set $S \subset \{-,+\}^n$, we will construct a graph $G(S)$ in polynomial time, such that S is k-realizable if and only if $G(S)$ has a k-dimensional distance embedding. For convenience, we start by presenting the construction from Kang and Muller (2012).

Definition 32 *(Kang and Muller (2012)) For any $S \subset \{-,+\}^n$, define $G_U(S) = (V, E)$ to be the undirected graph with vertices*

$$V = \{a_1, a_2, \ldots a_n\} \cup \{b_1, b_2, \ldots b_n\} \cup \{c_\sigma : \sigma \in S\}$$

and edges

- $\{c_\sigma, c_\pi\} \in E$ *for all* $\sigma, \pi \in S$

- $\{a_i, a_j\}, \{b_i, b_j\} \in E$ *for all* $i \neq j$

- $\{a_i, c_\sigma\} \in E$ *if and only if* $\sigma_i = +$

- $\{b_i, c_\sigma\} \in E$ *if and only if* $\sigma_i = -$.

Theorem 33 *(Kang and Muller (2012)) S is k-realizable if and only if $G_U(S)$ has a k-dimensional undirected distance embedding.*

Our directed construction is very similar to $G_U(S)$.

Definition 34 *For $S \subset \{-,+\}^n$, let $G_D(S) = (V, E)$ be the directed graph with vertices*

$$V = \{a_1, a_2, \ldots a_n\} \cup \{b_1, b_2, \ldots b_n\} \cup \{c_\sigma : \sigma \in S\}$$

and edges

- $(a_i, c_\sigma) \in E$ *if and only if* $\sigma_i = +$

- $(b_i, c_\sigma) \in E$ *if and only if* $\sigma_i = -$.

Theorem 35 *S is k-realizable if and only if $G_D(S)$ has a distance embedding of dimension at most k.*

Proof

$\Rightarrow$ Suppose $G_D(S)$ has a distance embedding ϕ with dimension k. Let h_i be the hyperplane that is the perpendicular bisector of $\phi_{out}(a_i)$ and $\phi_{out}(b_i)$; orient it so that $\phi_{out}(a_i)$ lies on the positive side and $\phi_{out}(b_i)$ on the negative side. Letting $\mathcal{H} = \{h_1, h_2, \ldots, h_n\}$, we claim that $\phi_{in}(c_\sigma)$ has sign vector exactly σ with respect to $\mathcal{H}$. Thus $S \subset \mathcal{D}(\mathcal{H})$.

To prove our claim, consider any $\sigma \in S$, $1 \leq i \leq n$. c_σ has an edge from exactly one of a_i and b_i and consequently $\phi_{in}(c_\sigma)$ is closer to the corresponding $\phi_{out}(a_i)$ or $\phi_{out}(b_i)$. Thus $\phi_{in}(c_\sigma)$ is on the σ_i side of h_i. Combining this over all i, we see that $\phi_{in}(c_\sigma)$ has sign vector σ as desired.

$\Leftarrow$ Suppose S is k-realizable. Then by Theorem 33, $G_U(S)$ has a k-dimensional undirected distance embedding ϕ (in the sense of Theorem 1). We construct a k-dimensional distance embedding ϕ' of $G_D(S)$ from ϕ as follows. $\phi'_{out}(a_i) = \phi(a_i)$, $\phi'_{out}(b_i) = \phi(b_i)$, $\phi'_{in}(c_\sigma) = \phi(c_\sigma)$, and the remaining vectors are assigned so that they are each at distance > 1 from all the other vectors. The only possible edges in such an embedding are edges from $\{a_1, a_2, \dots a_n, b_1, b_2, \dots b_n\} \to \{c_\sigma : \sigma \in S\}$. Since the corresponding edges are in $G_U(S)$, it follows that ϕ' is a valid embedding of $G_D(S)$ as desired.

■

Since the construction of $G_D(S)$ from S takes polynomial time, the hardness of k-REALIZABILITY implies the following.

Corollary 36 *Computing $d_{dist}(G)$ for a directed graph G is NP-hard.*

B.2 Similarity embeddings

This section is very similar to the previous one.

Definition 37 *(Kang and Muller (2012)) For $S \subset \{-, +\}^n$, define $G_U(S) = (V, E)$ to be the undirected graph with vertices*

$$V = \{a_1, a_2, \dots a_n\} \cup \{c_\sigma : \sigma \in S\}$$

and edges

- $\{a_i, a_j\} \in E$ *for all $i \neq j$*

- $\{a_i, c_\sigma\} \in E$ *if and only if $\sigma_i = +$.*

Theorem 38 *(Kang and Muller (2012)) S is k-realizable if and only if $G_U(S)$ has a k-dimensional undirected similarity embedding (in the sense of Theorem 2).*

Definition 39 *For $S \subset \{-, +\}^n$, define $G_D(S) = (V, E)$ as the directed graph with vertices*

$$V = \{a_1, a_2, \dots a_n\} \cup \{c_\sigma : \sigma \in S\}$$

and edges $E = \{(a_i, c_\sigma) : \sigma_i = +\}$.

Theorem 40 *S is k-realizable if and only if $d_{sim}(G_D(S)) \leq k$.*

Proof

$\Rightarrow$ Suppose $G_D(S)$ has a similarity embedding ϕ_L, ϕ_R with dimension k and threshold t. Let

$$h_i = \{v : \langle v, \phi_L(a_i) \rangle = t\},$$

and orient it so that $h_i^+ = \{v : \langle v, \phi_L(a_i) \rangle > t\}$. Let $\mathcal{H} = \{h_1, h_2, \dots, h_n\}$. We claim, as in the corresponding case of Theorem 35, that $\phi_R(c_\sigma)$ has sign vector exactly σ with respect to $\mathcal{H}$. This clearly suffices as it shows that $\sigma \in \mathcal{D}(\mathcal{H})$ for all $\sigma \in S$.

$\Leftarrow$ Suppose S is k-realizable. Then by Theorem 38, $G_U(S)$ has a k-dimensional undirected similarity embedding ϕ with threshold $t = 1$. Define $\phi_L(a_i) = \phi(a_i)$, $\phi_R(c_\sigma) = \phi(c_\sigma)$, and set all remaining ϕ_L, ϕ_R vectors to zero. It is then easily checked that ϕ_L, ϕ_R is a similarity embedding of $G_D(S)$ with threshold 1.

$\blacksquare$

Corollary 41 *Computing $d_{sim}(G)$ for a directed graph G is NP-hard.*