An Impossibility Theorem for Node Embedding

T. Mitchell Roddenberry
Rice University, USA

Yu Zhu
Purdue University, USA

Santiago Segarra
Rice University, USA

Abstract

With the increasing popularity of graph-based methods for dimensionality reduction and representation learning, node embedding functions have become important objects of study in the literature. In this paper, we take an axiomatic approach to understanding node embedding methods. Motivated by desirable properties of node embeddings for encoding the role of a node in the structure of a network, we first state three properties for embedding dissimilarity networks. We then prove that no node embedding method can satisfy all three properties at once, reflecting fundamental difficulties inherent to the task. Having identified these difficulties, we show that mild relaxations of these axioms allow for certain node embedding methods to be admissible.

1 INTRODUCTION

Graph-structured data is pervasive in the natural and social sciences, not only as a direct model for data, but as a useful abstraction for understanding complex relational systems. This ubiquity has driven recent developments in graph representation learning (Hamilton, 2020), seeking to extract useful representations of graphs for classification and inference tasks. In particular, node embedding methods embed the nodes of a graph in a low-dimensional space, in ways that encode the structural role of each node in a graph. Indeed, the goal of node embedding methods is to represent the nodes of a graph in a space where "similar" nodes are close together, and "different" nodes are far apart (Hamilton, 2020).

Inspired by the seminal paper of Kleinberg (2002), we take an *axiomatic view* of node embedding methods.

Proceedings of the 27$^{\text{th}}$ International Conference on Artificial Intelligence and Statistics (AISTATS) 2024, Valencia, Spain. PMLR: Volume 238.

Specifically, our contributions are the following:

1) We identify three basic properties that reflect desirable qualities of node embedding methods (Properties 1 to 3).

2) We prove that no node embedding can satisfy all three properties simultaneously (Theorem 1).

3) We illustrate the necessary trade-offs implied by the impossibility result with example node embedding methods (Propositions 1 to 3).

4) We consider relaxations of these axioms (Properties 4 and 5), allowing for the construction of node embedding algorithms that satisfy the relaxed properties simultaneously (Propositions 4 and 5).

By proving that the three stated properties can be simultaneously satisfied by no node embedding method, we establish fundamental difficulties in the problem of node embedding. For the practitioner, this indicates that when choosing an algorithm or method, at least one of the three properties must be sacrificed.

2 NETWORKS AND NODE EMBEDDINGS

A typical type of data from which graphs, or networks, are constructed is a finite point cloud in a metric space. Finding the triangle inequality unnecessary, we relax the metric condition to yield a *dissimilarity network*. To construct a dissimilarity network, begin with a finite set of nodes $\mathcal{V}$. For convenience, we will typically identify $\mathcal{V}$ with the first $|\mathcal{V}|$ natural numbers, saying $\mathcal{V} = \{1, 2, \ldots, |\mathcal{V}|\}$. A dissimilarity network, then, is a finite set of nodes $\mathcal{V}$ coupled with a symmetric *dissimilarity function* $d : \mathcal{V} \times \mathcal{V} \to \mathbb{R}^{\geq 0}$, with the condition that $d(i, j) = 0$ if and only if $i = j$. Notice that metrics are dissimilarity functions, so that a finite metric space can be viewed as a dissimilarity network.

More generally, we speak of a *dissimilarity space*, which is a set $\mathcal{S}$ coupled with a dissimilarity function ρ on $\mathcal{S}$ (so that $\rho : \mathcal{S} \times \mathcal{S} \to \mathbb{R}^{\geq 0}$). This allows us to endow a dissimilarity network with features on the nodes. For a dissimilarity space $(\mathcal{S}, \rho)$, an $\mathcal{S}$-*featured* *dissimilarity network* is a dissimilarity network $(\mathcal{V}, d)$ and a *feature map* $F : \mathcal{V} \to \mathcal{S}$, denoted $(\mathcal{V}, d, F)$. The set of all such objects is denoted $\mathcal{N}(\mathcal{S})$. We will speak of *featured dissimilarity networks* when $\mathcal{S}$ is understood from context.

Graph representation learning seeks to represent networks in amenable spaces, typically Euclidean space or on a low-dimensional manifold (Hamilton, 2020). This motivates our object of study, which is the class of *node embedding functions*. Node embedding functions endow the set of nodes in a network with a *pseudodissimilarity* structure. For a dissimilarity function ρ, if we relax the identity of indiscernibles, so that $x = y$ implies $\rho(x, y) = 0$, but $\rho(x, y) = 0$ does not imply $x = y$, as well as allow for ρ to take values in $\mathbb{R}^{\geq 0} \cup \{+\infty\}$, we call ρ a *pseudodissimilarity function*. A *pseudodissimilarity space* is then a set coupled with a pseudodissimilarity function on it. Let $\mathcal{M}$ be the set of all finite pseudodissimilarity spaces. A node embedding function is a map $\xi : \mathcal{N}(\mathcal{S}) \to \mathcal{M}$ with the condition that for any featured dissimilarity network $N = (\mathcal{V}, d, F)$, the corresponding dissimilarity space has the same underlying set: $\xi(N) = (\mathcal{V}, \phi)$, for some (pseudo)dissimilarity function ϕ on $\mathcal{V}$. In the context of embedding nodes into some predefined space, such as $\mathbb{R}^n$, this is equivalent to restricting the dissimilarity function on that space to the embedded points of $\mathcal{V}$. For instance, if the embedding function maps the nodes into $\mathbb{R}^n$ with the usual metric, then ϕ is determined by the Euclidean distances between the embedded nodes. Some examples of node embedding functions include single-linkage clustering, motif-based embedding, and spectral methods, all discussed further in Section 4.

3 MAIN RESULT

A typical goal in node embedding tasks is to embed nodes in a way that *preserves* and *reflects* the original network structure. We begin by defining three properties of node embedding functions that capture these goals, namely being *self-contained*, *consistent*, and *graph-aware*. Throughout, let $(\mathcal{S}, \rho)$ be a dissimilarity space, and consider node embedding functions $\xi : \mathcal{N}(\mathcal{S}) \to \mathcal{M}$.

In many graph representation learning tasks, we seek embeddings that are invariant/equivariant to the labeling of nodes. In particular, the embedding of nodes via some node embedding function should only be dependent on their pairwise dissimilarities and features. We state this for dissimilarity networks with 2 nodes as follows.

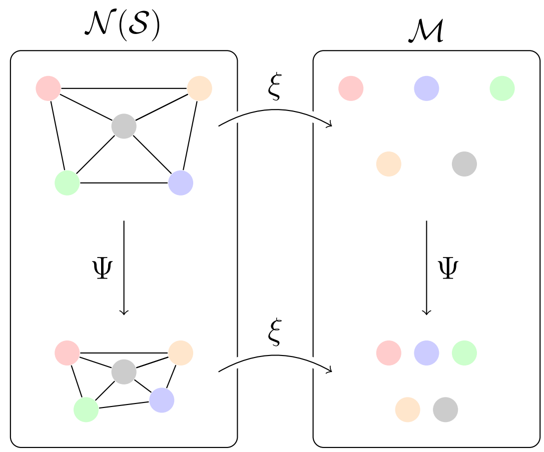

Figure 1: Consistency property of a node embedding ξ. For any two networks N_1, N_2 and a CoDNI map Ψ from N_1 to N_2, the same map applied to their corresponding embeddings is contractive.

Property 1 (Self-containedness)**.** *Let $N_2 = (\{i, j\}, d, F)$ be a featured dissimilarity network on 2 nodes, and put $\alpha = d(i, j)$, so that $\alpha > 0$. For some node embedding function ξ, put $(\{i, j\}, \phi) = \xi(N_2)$. Suppose there exists a function g such that for all such networks,*

$$\phi(i, j) = g(F(i), F(j); \alpha). \tag{1}$$

Under these conditions, we say that ξ is self-contained.

Self-containedness is perhaps the most self-evident property we propose. In particular, it requires node embedding functions to depend on the network structure alone, rather than other external information. Stated in the most basic case for networks on two nodes, it is a primitive form of permutation invariance. That is, when embedding a two-node network, the dissimilarity in the embedded space should only depend on the dissimilarity in the original network and the features of both nodes. Notably, our main theorem demonstrates that defining self-containedness only for two-node networks is enough to yield the impossibility result.

Another useful property of a node embedding function is for it to preserve local proximity between nodes: that is, if two nodes i, j are similar, their embeddings should be similar as well. This is stated as a desirable property of a node embedding, for instance, in (Shaw and Jebara, 2009; Wang et al., 2016; Ma et al., 2017). Given that networks are often used to model sparse, pairwise, and localized interactions, it is preferable for a node embedding algorithm to respect the structure of each node's neighborhood. So, if we shrink the dissimilarity between a pair of nodes, their dissimilarity in the embedding space should decrease accordingly. We state this by first defining the notion of a "contractive and dissimilarity non-increasing" map from one network to another.

Definition 1. *Consider two featured dissimilarity net-*

works $N_1 = (\mathcal{V}_1, d_1, F_1), N_2 = (\mathcal{V}_2, d_2, F_2)$ *with possibly different numbers of nodes. A map* $\Psi : \mathcal{V}_1 \to \mathcal{V}_2$ *is said to be contractive and dissimilarity non-increasing (CoDNI) if, for all* $i, j \in \mathcal{V}_1$,

$$\rho(F_1(i), F_1(j)) \geq \rho(F_2(\Psi(i)), F_2(\Psi(j))) \tag{2}$$
$$d_1(i,j) \geq d_2(\Psi(i), \Psi(j)). \tag{3}$$

Notice that the image nodes under a CoDNI map are closer in feature space and have smaller network dissimilarities than their corresponding preimages. Thus, given a CoDNI map from one network to another, a reasonable property of a node embedding is to reduce the distance between nodes in the embedding space.

Property 2 (Consistency)**.** *Consider two featured dissimilarity networks* $N_1 = (\mathcal{V}_1, d_1, F_1), N_2 = (\mathcal{V}_2, d_2, F_2)$ *with possibly different numbers of nodes. For some node embedding function* ξ*, put* $(\mathcal{V}_1, \phi_1) = \xi(N_1)$ *and* $(\mathcal{V}_2, \phi_2) = \xi(N_2)$*. We say that* ξ *is consistent if, for all* $i, j \in \mathcal{V}_1$ *and CoDNI maps* $\Psi : \mathcal{V}_1 \to \mathcal{V}_2$,

$$\phi_1(i,j) \geq \phi_2(\Psi(i), \Psi(j)). \tag{4}$$

The property of consistency is essentially a monotonicity condition: as networks are contracted via CoDNI maps, the corresponding embeddings must also be contracted. Property 2 is illustrated in Fig. 1.

Finally, we wish for the node embedding to reflect the global dissimilarity structure of the underlying network as well. Ignoring the features, one could envision an identity mapping as a good embedding that satisfies Property 2, since all nodes maintain the exact same dissimilarity with their neighbors. However, an approach such as this one fails to capture the role of a node *globally* in the network. Thus, we wish to state an axiom that reflects the sensitivity of a node embedding to the *global* structure of the network. We express this in terms of the dissimilarity between the embedding of a given pair of nodes, where we increase all other pairwise dissimilarities in the network in order to change the embedded dissimilarity between the two nodes.

Property 3 (Graph-awareness)**.** *Consider a featured dissimilarity network* $N = (\mathcal{V}, d, F)$ *on at least* 3 *nodes, and consider an arbitrary pair of nodes* $i, j \in \mathcal{V}$ *such that* $i \neq j$*. Let* $\mathcal{D}_{ij}$ *be the set of dissimilarity functions* d' *such that* $d'(i,j) = d(i,j)$*, and for all* $k, \ell \in \mathcal{V}$ *it holds that* $d'(k,\ell) \geq d(k,\ell)$*. For some node embedding function* ξ*, if for all such networks* N *there exists some* $d' \in \mathcal{D}_{ij}$ *such that the embeddings* $(\mathcal{V}, \phi) = \xi(N), (\mathcal{V}, \phi') = \xi((\mathcal{V}, d', F))$ *satisfy* $\phi(i,j) \neq \phi'(i,j)$*, then we say that* ξ *is* graph-aware*.*

A good node embedding ought to reflect the role of each node in the broad context of the network. An example of properties that are only represented on a global scale comes up when considering the *betweenness centrality* (Freeman, 1977). A node that only has a few neighbors may not have a notable role in a network when viewed locally, but if the node lies in a bottleneck between two large regions of the network, it will have high betweenness centrality. To reflect cases like this, graph-awareness requires that the embedded dissimilarity between two nodes is not only dependent on their dissimilarity in the original network, but also on the graph structure surrounding that pair of nodes. In contrast with consistency, graph-awareness is a property describing *sensitivity to global information*, as illustrated in Fig. 2.

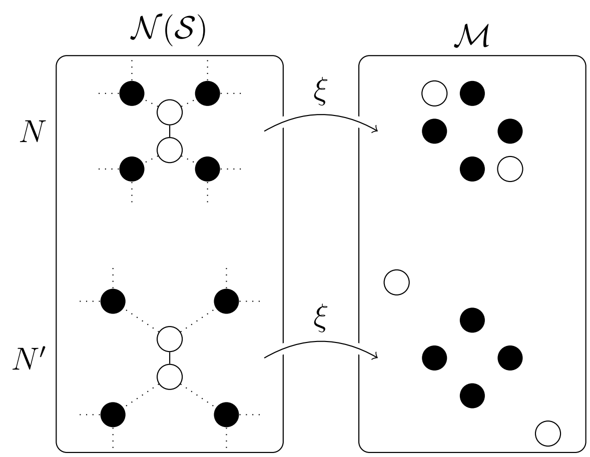

Figure 2: Graph-awareness property of a node embedding function ξ. For a given pair of nodes (labelled in white) in a network N, there is an alternative dissimilarity function that preserves their dissimilarity while expanding the dissimilarity between other pairs, yielding the network N'. The dissimilarity of the given pair under the embedding ξ is different between N and N'.

The importance of embeddings that *preserve* proximity structures (*i.e.*, are consistent) and *reflect* the global structure (*i.e.*, are graph-aware) is noted by Xu (2021), where embeddings of complex networks into simpler, low-dimensional spaces is shown to have applications in the analysis of networks in fields ranging from social sciences to biology. Indeed, the incorporation of global network structure in embedding methods was shown to yield performance improvements in multiple tasks by Cao et al. (2015), highlighting the need to enforce sensitivity to network structure beyond strictly local proximity.

With these definitions in place, we now state our main result.

Theorem 1. *There does not exist a node embedding function that is simultaneously self-contained, consistent, and graph-aware.*

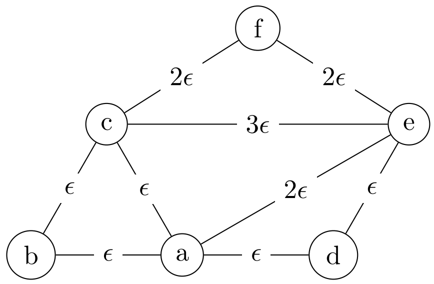

Figure 3: Single-linkage and motif-based embeddings. Pictured is a dissimilarity network on node set $\{a, b, c, d, e, f\}$, with dissimilarities labeled on the edges. For simplicity, assume that omitted dissimilarities in the figure take on some large value. Under the single-linkage embedding, the ϵ-ball about the node a consists of the nodes $\{a, b, c, d, e\}$. In contrast, the triangle-linkage embedding only yields an ϵ-ball about a consisting of $\{a, b, c\}$.

We leave the proof to Section 5. Theorem 1 reflects a fundamental difficulty inherent to the node embedding task. One wishes for a node embedding method to preserve *local* structures, while still being sensitive to *global* structure as well. However, preserving both of these in a way consistent with our axioms is impossible: intuitively, one cannot suitably capture both types of information within a single embedding. Indeed, this tension between preserving local and global information in node embedding tasks has been noted in the literature (Shaw and Jebara, 2009; Wang et al., 2016; Ma et al., 2017). Moreover, this tradeoff is even implicitly recognized in the design of certain embedding algorithms, such as node2vec (Grover and Leskovec, 2016), where the user-chosen parameters reflect the bias towards local or global exploration of the graph. Guided by the implicit tradeoffs of existing methods, Theorem 1 is the first instance in the literature of an explicit, formal recognition of the difficulty of accomplishing both local and global embeddings of networks.

Remark 1. *The relationships between $\mathcal{S}$-featured dissimilarity networks via CoDNI maps gives the collection of such networks the structure of a category Mac Lane (1971); Riehl (2017), in which case a node embedding function is best understood as a functor from this category to a category of pseudodissimilarity spaces. We discuss this perspective in Appendix B, followed by its implications for similarity networks in Appendix C.*

4 EXAMPLES

In light of Theorem 1, any node embedding method that satisfies two of the three properties of self-containedness, consistency, and graph-awareness must fail to satisfy the third. We illustrate this with a few examples of node embedding functions previously considered in the literature (Kleinberg, 2002; Carlsson and Mémoli, 2013), which we examine through the lens of our proposed axioms.

4.1 Single-Linkage Clustering

We consider the single-linkage clustering procedure, which embeds dissimilarity networks in ultrametric spaces (Johnson, 1967). Indeed, dendrograms represent ultrametric spaces, where the distance between two points is the depth at which they are merged into the same cluster.

We first discuss a useful construction for defining the embedding distance of the single-linkage procedure.

Definition 2. *For a dissimilarity network $N = (\mathcal{V}, d, F)$, and two nodes $i, j \in \mathcal{V}$, a* path *from i to j is a finite sequence $P = [x_1, x_2, \ldots, x_{L-1}, x_L]$ of nodes in $\mathcal{V}$ such that $x_1 = i, x_L = j$. The set of all paths from i to j is denoted $\mathcal{P}_{ij}$. We refer to the* ℓth *element of a path P by $P(\ell)$. The* length *of a path is the number of elements in the path, denoted $L(P)$.*

For a dissimilarity network $N = (\mathcal{V}, d, F)$, the single-linkage procedure $(\mathcal{V}, \phi) = \xi(N)$ is such that the embedding distance between two nodes $i, j \in \mathcal{V}$ is given by

$$\phi(i, j) = \min_{P \in \mathcal{P}_{ij}} \max_{1 \le \ell < L(P)} d(P(\ell), P(\ell + 1)), \qquad (5)$$

with an illustrative example given in Fig. 3. One can check that this is equivalent to the construction of the dendrogram in the standard formulation of single-linkage clustering. Notice that this embedding disregards node features. Let us examine this embedding distance with respect to each of our proposed properties.

Proposition 1. *The single-linkage procedure is self-contained and consistent, but not graph-aware.*

The proof is in Section D of the SM. Proposition 1 indicates the highly local nature of the single-linkage procedure. Indeed, although in many networks global structure is uncovered by single-linkage clustering, the distance between nodes that are initially close together is oblivious to the surrounding network structure.

Remark 2. *The single-linkage procedure for embedding nodes in an ultrametric space is a special case of a metric projection of a dissimilarity network, as described* *by Segarra et al. (2020). Similar arguments show that such metric projections satisfy self-containedness and consistency, but fail to satisfy graph-awareness.*

Remark 3. *By their definitions, the first link made in methods such as complete or average linkage clustering is identical to that of single-linkage clustering. For this reason, Proposition 1 holds for these methods as well.*

4.2 Motif-Based Embedding

Single-linkage clustering relies strictly on pairwise distances between nodes, leading to undesirable properties such as sensitivity to noise, chaining (Lance and Williams, 1967), and failure to incorporate global structure, as reflected by the fact that the single-linkage procedure is not graph-aware. To remedy this, we consider another ultrametric embedding based on triangular motifs, which we refer to as the *triangle-linkage procedure*, based on the clustering scheme discussed in (Carlsson and Mémoli, 2013, Section 6.7).

For a dissimilarity network $N = (\mathcal{V}, d, F)$, the triangle-linkage procedure, denoted $(\mathcal{V}, \phi) = \xi_T(N)$, is such that the embedding dissimilarity between any distinct nodes $i, j \in \mathcal{V}$ is $+\infty$ when $|\mathcal{V}| < 3$. Otherwise, we have

$$\begin{aligned}\phi(i,j) = \min\{\epsilon > 0 : \exists k \in \mathcal{V}, i \neq k, j \neq k, \\ \max\{d(i,j), d(i,k), d(j,k)\} \leq \epsilon\}.\end{aligned} \tag{6}$$

That is, two nodes i, j have an embedded dissimilarity of at most ϵ if there is a third node k such that the triple i, j, k forms a triangle where each side has dissimilarity less than or equal to ϵ. This is illustrated in Fig. 3. As before, we now examine this embedding distance with respect to our proposed properties.

Proposition 2. *The triangle-linkage procedure is self-contained and graph-aware, but not consistent.*

The proof is in Section E of the SM. The incorporation of higher-order network structure (namely, triangles) in the triangle-linkage procedure gains the property of graph-awareness over single-linkage, but this is at the cost of consistency, by Theorem 1. Indeed, the monotonicity property of the triangle-linkage procedure can fail when two nodes have the same image under a CoDNI map, allowing for extraneous higher-order structures to distort the embedding.

Remark 4. *Although the triangle-linkage procedure uses a fully-connected triple of nodes as a means of comparison, this definition could easily by modified to account for other motifs, such as cliques, loops, and others.*

4.3 Spectral Embedding

While the above single-linkage and triangle-linkage procedures are combinatorial methods for embedding dissimilarity networks, *spectral methods* are also a common approach (Belkin and Niyogi, 2003). In particular, one constructs a graph given pairwise distances between nodes, and weighs the edges of that graph based on some kernel. Then, spectral methods are used to embed the nodes in Euclidean space. We consider a very simple procedure inspired by this, where the leading eigenvector of the constructed adjacency matrix is used to embed nodes in $\mathbb{R}$, taking the standard Euclidean metric as the embedding distance. The use of the leading eigenvectors of the adjacency matrix was used, for instance, by Larroca et al. (2021); Marenco et al. (2022) in changepoint detection tasks.

One common kernel for converting dissimilarities to edge weights is via a Gaussian radial basis kernel. We formulate a general kernel condition that captures the essential properties of such a method. Let $\kappa : \mathbb{R}^{\geq 0} \times \mathcal{S} \times \mathcal{S} \to \mathbb{R}^{\geq 0}$ be a continuous (with respect to the real-valued argument) function mapping dissimilarities and pairs of features to the nonnegative real numbers such that for all $x \in \mathbb{R}^{\geq 0}, s_1, s_2 \in \mathcal{S}$, it holds that $\kappa(0; s_1, s_1) = 1, \kappa(x; s_1, s_2) = \kappa(x; s_2, s_1) > 0$, and κ is monotonically decreasing with respect to x and $\rho(s_1, s_2)$, with the condition that $\kappa(x; s_1, s_2) \to 0$ as $x \to \infty$. One can check that these conditions hold for common radial basis functions.

For some $\mathcal{V} = \{1, 2, \ldots, n\}$, let $N = (\mathcal{V}, d, F)$ be an $\mathcal{S}$-featured dissimilarity network, and construct an adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$ where $A_{ij} = \kappa(d(i,j); F(i), F(j))$. Let $\mathbf{u}$ be a normalized leading eigenvector of $\mathbf{A}$, with corresponding eigenvalue $\lambda > 0$, as guaranteed by the Perron-Frobenius theorem. Here, $\mathbf{u}$ is the *eigenvector centrality vector* of $\mathbf{A}$ (Bonacich, 1987). Denoting the node embedding of N as $(\mathcal{V}, \phi) = \xi_C(N)$, the embedding dissimilarity ϕ is defined for each pair of nodes $i, j \in \mathcal{V}$ as

$$\phi(i,j) = \sqrt{\lambda}|u_i - u_j|. \tag{7}$$

We refer to this embedding as the *eigenvector centrality embedding*. Its behavior with respect to Theorem 1 is characterized as follows:

Proposition 3. *The eigenvector centrality embedding procedure is self-contained and graph-aware, but not consistent.*

The proof is in Section F of the SM. This is not too surprising, since eigenvectors of matrices are reflective of the *global* matrix structure, so graph-awareness can be readily demonstrated. Moreover, the eigenvector centrality is not sensitive to permutations of nodes, so that self-containedness holds as well.

4.4 Clustering and Kleinberg's Impossibility Theorem

Kleinberg (2002) established three axioms for *clustering functions*, which take as input a set of points $\mathcal{V}$ coupled with a dissimilarity function d, and yields a partition of $\mathcal{V}$. By appropriately defining the notions of *scale-invariance, richness,* and *consistency,*[1] he proved an impossibility result for clustering functions: namely, that no clustering function satisfies these three properties simultaneously.

Our results draw inspiration from this approach (although our proof techniques are completely different), and indeed apply to clustering functions as special types of node embedding functions. To see this, let a partition of $\mathcal{V}$ be given, so that $\mathcal{V} = \bigcup_{j=1}^{L} \mathcal{V}_j$, where $\mathcal{V}_j \cap \mathcal{V}_k = \emptyset$ for $j \neq k$. For each $v \in \mathcal{V}$, there exists a unique integer j such that $v \in \mathcal{V}_j$. If two (not necessarily distinct) nodes $v, w \in \mathcal{V}$ are contained in the same partition $\mathcal{V}_j$, we say that $v \sim w$. Define a pseudodissimilarity function on $\mathcal{V}$ in the following way:

$$\phi(v, w) = \begin{cases} 0 & v \sim w \\ 1 & \text{otherwise.} \end{cases} \tag{8}$$

One can see that ϕ encodes the partition $\{\mathcal{V}_j\}_{j=1}^{L}$, where two distinct nodes are in the same partition if and only if they have dissimilarity 0, and are in differing partitions if and only if they have dissimilarity 1. One can see that there is an injection from the set of partitions of $\mathcal{V}$ and all such pseudodissimilarity functions on $\mathcal{V}$. Thus, our axioms yield a new impossibility result for clustering, as a particular type of node embedding. Moreover, the analogous axioms for embeddings of *similarity* networks also yield an impossibility result for clustering, via Theorem 3.

5 PROOF OF THEOREM 1

Let $(\mathcal{S}, \rho)$ be a dissimilarity space. Suppose, for the sake of contradiction, that $\xi : \mathcal{N}(\mathcal{S}) \to \mathcal{M}$ is a node embedding function that is self-contained, consistent, and graph-aware. For some $n \geq 3$, let $\mathcal{V} = \{1, 2, \ldots, n\}$ and some $F : \mathcal{V} \to \mathcal{S}$ be given. Pick $i, j \in \mathcal{V}$ that achieves minimum dissimilarity in $\mathcal{S}$, so that

$$i, j \in \operatorname*{arg\,min}_{i', j' \in \mathcal{V} : i' \neq j'} \rho(F(i), F(j)), \tag{9}$$

assuming without loss of generality that $i < j$. Fix some $\delta > 0$ and define a dissimilarity function d on $\mathcal{V}$ so that $d(k, \ell) = \delta$ for all $k \neq \ell$, taking value 0 otherwise. This yields a featured dissimilarity network $N = (\mathcal{V}, d, F)$.

[1]Consistency as defined by Kleinberg is distinct from Property 2, although related in studying contractive maps.

Since ξ is graph-aware, choose a dissimilarity function $d_1 \in \mathcal{D}_{ij}$ so that if $N_1 = (\mathcal{V}, d_1, F)$ and $(\mathcal{V}, \phi) = \xi(N), (\mathcal{V}, \phi_1) = \xi(N_1)$, we have $\phi(i, j) \neq \phi_1(i, j)$.

Additionally, put $N_2 = (\{i, j\}, d\big|_{\{i,j\}^2}, F\big|_{\{i,j\}})$. That is to say, N_2 is the restriction of N to the nodes i, j. Put $(\{i, j\}, \phi_2) = \xi(N_2)$, so that by self-containment we have

$$\phi_2(i, j) = g(F(i), F(j); d(i, j)) = g(F(i), F(j); \delta), \tag{10}$$

where g is the function whose existence is guaranteed by self-containedness.

Consider the map $\Psi : \{i, j\} \hookrightarrow \mathcal{V}$ defined via inclusion, so that $\Psi(i) = i, \Psi(j) = j$. Notice that Ψ is a CoDNI map from N_2 to N_1. By consistency of ξ, then, we have

$$\phi_1(i, j) \leq \phi_2(i, j) = g(F(i), F(j); \delta), \tag{11}$$

where the equality follows from (10). Similarly, consider the map $\Psi' : \mathcal{V} \to \{i, j\}$ where $\Psi'(k) = i$ for $k \leq i$ and $\Psi'(k) = j$ for $k > i$ (in particular, $\Psi'(j) = j$). For any $k, \ell \in \mathcal{V}$ such that $k \neq \ell$, we have

$$\rho(F(k), F(\ell)) \overset{(a)}{\geq} \rho(F(i), F(j)) \overset{(b)}{\geq} \rho(F(\Psi'(k)), F(\Psi'(\ell))), \tag{12}$$

where (a) follows from our choice of i, j in (9), and (b) follows from the fact that $\rho(F(\Psi'(k)), F(\Psi'(\ell)))$ is either equal to $\rho(F(i), F(j))$ or zero. Moreover, again for $k \neq \ell$,

$$d_1(k, \ell) \overset{(c)}{\geq} d(k, \ell) = \delta \geq d_2(\Psi'(k), \Psi'(\ell)), \tag{13}$$

where (c) follows from our choice of $d_1 \in \mathcal{D}_{ij}$.

Examining (12) and (13), we see that Ψ' is a CoDNI map from N_1 to N_2. By consistency of ξ, we have

$$\phi_1(i, j) \geq \phi_2(i, j) \overset{(d)}{=} g(F(i), F(j); \delta), \tag{14}$$

where (d) invokes (10).

Combining the inequalities (11) and (14), we have

$$\phi_1(i, j) = g(F(i), F(j); \delta). \tag{15}$$

The same argument can be applied to N (that is, constructing Ψ, Ψ' between N and N_2), yielding

$$\phi(i, j) = g(F(i), F(j); \delta). \tag{16}$$

Combining (15) and (16), we see that $\phi(i, j) = \phi_1(i, j)$, thus contradicting graph-awareness, as desired. ■

Remark 5. *In the proof of Theorem 1, the networks N, N_2 have dissimilarity functions that satisfy the triangle inequality, and are thus metrics. If one restricts the domain of node embedding functions to dissimilarity networks with metric dissimilarity functions, the impossibility result still holds.*

6 RELAXATIONS

Although all node embedding methods must violate at least one of the three properties we put forth due to Theorem 1, it is worthwhile to study relaxations of these properties that yield admissible node embeddings. In this section, we consider the implications of relaxing graph-awareness and consistency.

6.1 Weak Graph-Awareness

Recall the definition of graph-awareness, where the embedding dissimilarity of two nodes i, j must be sensitive to sufficient increases in the values taken by the network dissimilarity function, described by the set $\mathcal{D}_{ij}$. Here, we consider a relaxation of this property, where we allow for any perturbation of the network dissimilarity that preserves the dissimilarity between nodes i, j. We refer to this property as *weak graph-awareness*:

Property 4 (Weak graph-awareness)**.** *Consider a featured dissimilarity network $N = (\mathcal{V}, d, F)$ on at least 3 nodes, and consider an arbitrary pair of nodes $i, j \in \mathcal{V}$ such that $i \neq j$. Let $\mathcal{C}_{ij}$ be the set of distance functions d' such that $d'(i,j) = d(i,j)$. For some node embedding function ξ, if for all such networks there exists some $d' \in \mathcal{C}_{ij}$ such that the embeddings $(\mathcal{V}, \phi) = \xi(N), (\mathcal{V}, \phi') = \xi((\mathcal{V}, d', F))$ satisfy $\phi(i,j) \neq \phi'(i,j)$, then we say that ξ is* weakly graph-aware*.*

Observe that all graph-aware node embedding functions are weakly graph-aware, since $\mathcal{D}_{ij} \subseteq \mathcal{C}_{ij}$. Looking back to our analysis of the single-linkage procedure, this relaxation is sufficient for the single-linkage embedding to satisfy all three axioms.

Proposition 4. *The single-linkage procedure is self-contained, consistent, and weakly graph-aware.*

The proof is in Section G of the SM. The relaxation of graph-awareness to weak graph-awareness can be interpreted as making the node embedding function less sensitive to the global structure of the network. In doing so, we allow for arbitrary perturbations of the network structure, rather than ones that only increase dissimilarity.

6.2 Injective Consistency

In the proof of Theorem 1, a key step is forming a CoDNI map from a network of many nodes to a network of two nodes. By the pigeonhole principle, a function mapping a set to another set with smaller cardinality is not injective. This motivates a stronger requirement on the map Ψ in our definition of consistency, yielding a property that we call *injective consistency*:

Property 5 (Injective consistency)**.** *Consider two featured dissimilarity networks $N_1 = (\mathcal{V}_1, d_1, F_1)$ and $N_2 = (\mathcal{V}_2, d_2, F_2)$ with possibly different numbers of nodes. For some node embedding function ξ, let $(\mathcal{V}_1, \phi_1) = \xi(N_1), (\mathcal{V}_2, \phi_2) = \xi(N_2)$. We say that ξ is* injectively consistent *if, for all $i, j \in \mathcal{V}_1$ and injective CoDNI maps $\Psi : \mathcal{V}_1 \to \mathcal{V}_2$,*

$$\phi_1(i,j) \geq \phi_2(\Psi(i), \Psi(j)). \tag{17}$$

Observe that all node embedding functions that are consistent are also injectively consistent, since injective consistency depends on a stronger hypothesis (namely, the CoDNI map Ψ being injective). We now reconsider the triangle-linkage procedure given this relaxed notion of consistency.

Proposition 5. *The triangle-linkage procedure is self-contained, injectively consistent, and graph-aware.*

The proof is in Section H of the SM. In the same way that weak graph-awareness makes the node embedding less sensitive to changes in the global network structure, relaxing consistency to injective consistency weakens the sensitivity of node embedding functions to local changes in the network structure. This illustrates the fundamental tradeoff described by Theorem 1: a node embedding function can either be sensitive to global structure or local structure, but not both.

7 RELATED WORK

7.1 Node embedding

Node embedding has achieved significant progress in recent years. Existing works can be summarized into the following three categories (Zhang et al., 2018; Cai et al., 2018; Goyal and Ferrara, 2018).

Matrix factorization methods factorize a matrix S encoding node similarities into $X^\top X$ or $X^\top Y$ then adopt the columns of X (and possibly Y) as node embeddings. Examples include classical methods such as spectral clustering (von Luxburg, 2007), Laplacian eigenmaps (Belkin and Niyogi, 2003) and multidimensional scaling (Cox and Cox, 2008), as well as modern approaches such as HOPE (Ou et al., 2016), NetMF (Qiu et al., 2018) and GMF-FE (Zhu et al., 2023).

Word embedding algorithms leverage word2vec (Mikolov et al., 2013a,b), which was originally proposed for learning word embeddings from text. DeepWalk (Perozzi et al., 2014) first proposes to treat nodes as words and random walks as sentences, so that word2vec can be directly applied to generate node embeddings. Variants of DeepWalk consider different neighborhood sampling strategies, such as node2vec (Grover and Leskovec,

2016) employing biased second-order random walks. It has been proven that these methods *implicitly* factorize similarity matrices (Levy and Goldberg, 2014; Qiu et al., 2018), providing a link to the first category.

Neural architectures are inspired by the recent success of geometric deep learning (Bronstein et al., 2017). Many such works apply autoencoder architectures to construct meaningful node embeddings, including DNGR (Cao et al., 2016), SDNE (Wang et al., 2016), and ARGA/ARVGA (Pan et al., 2018). Moreover, instead of taking only the graph structure into consideration, approaches have also been proposed which are able to incorporate additional features on the nodes and edges (Shervashidze et al., 2011; Hamilton et al., 2017; Wang et al., 2017).

Researchers have observed tradeoffs that resemble ours when using graph neural networks (GNNs) as node embedding functions, such as the tradeoff between "over-squashing" and "over-smoothing" Arnaiz-Rodríguez et al. (2022); Rampášek et al. (2022); Di Giovanni et al. (2023). It would be an interesting line of work to understand how these very practical interests of practitioners using GNNs can be understood in light of the properties of consistency and graph-awareness.

7.2 Theoretical analysis of general node embeddings

Although various node embedding methods have been proposed, theoretical analysis regarding their properties and limitations independent of any particular algorithm is sparsely found in the literature. Srinivasan and Ribeiro (2019) studied the relationships between node embeddings in Euclidean space and so-called "structural representations". A paper by Seshadhri et al. (2020) proves that graphs generated from low-dimensional embeddings (using dot products as a similarity measure) cannot be both sparse and have high triangle density, two hallmarks of real-world networks. However, a follow-up paper (Chanpuriya et al., 2020) shows that the impossibility results in (Seshadhri et al., 2020) are a consequence of the specific model considered, and can be avoided by obtaining the embeddings from a different matrix factorization procedure.

7.3 Axiomatic approaches

In this work, we provide theoretical insights for node embedding using an axiomatic framework. A similar approach was originally taken by Kleinberg (2002) in the context of clustering, which proposes three axioms – namely scale invariance, richness, and consistency – and shows that it is impossible for a clustering function to simultaneously satisfy all of them. A group of works followed Kleinberg (2002), where suitable relaxations of the clustering axioms are formed to yield possibility and uniqueness results (Ben-David and Ackerman, 2008; Carlsson and Mémoli, 2013; Carlsson et al., 2013, 2014, 2021; Cohen-Addad et al., 2018; Meila, 2005). Beyond clustering, axiomatic approaches have been widely adopted in other areas including recommendation systems (Pennock et al., 2000; Andersen et al., 2008), computer vision (Kenney et al., 2005; Chessel et al., 2006), resource allocation (Lan et al., 2010), and social choice theory (Arrow, 1963).

8 CONCLUSION

We present the first axiomatic analysis of node embeddings for featured networks. Beyond establishing fundamental trade-offs inherent to the problem of node embedding, our impossibility result motivates a thorough look at the nature of the node embedding problem. In the context of clustering, Carlsson and Mémoli (2013) bypasses the impossibility results of Kleinberg (2002) by considering a relaxed codomain consisting of persistent clusterings, rather than simple partitions. Similarly, other works such as (Ben-David and Ackerman, 2008; Cohen-Addad et al., 2018) show how clustering informed by a particular loss function can satisfy a similar set of axioms. We envision the proposed framework motivating similar advances in the field of node embedding, ultimately resulting in better theoretical understanding and practical algorithms.

Acknowledgements

This work was partially supported by NSF under award CCF-2008555.

References

Andersen, R., Borgs, C., Chayes, J., Feige, U., Flaxman, A., Kalai, A., Mirrokni, V., and Tennenholtz, M. (2008). Trust-based recommendation systems: An axiomatic approach. In *The World Wide Web Conference*, pages 199–208.

Arnaiz-Rodríguez, A., Begga, A., Escolano, F., and Oliver, N. M. (2022). Diffwire: Inductive graph rewiring via the lovász bound. In *The First Learning on Graphs Conference*.

Arrow, K. J. (1963). *Social choice and individual values*. Yale University Press.

Belkin, M. and Niyogi, P. (2003). Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396.

Ben-David, S. and Ackerman, M. (2008). Measures of clustering quality: A working set of axioms for clustering. *Advances in Neural Information Processing Systems*, 21:121–128.

Bonacich, P. (1987). Power and centrality: A family of measures. *American journal of sociology*, 92(5):1170–1182.

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., and Vandergheynst, P. (2017). Geometric deep learning: going beyond Euclidean data. *IEEE Signal Processing Magazine*, 34(4):18–42.

Cai, H., Zheng, V. W., and Chang, K. C.-C. (2018). A comprehensive survey of graph embedding: Problems, techniques, and applications. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1616–1637.

Cao, S., Lu, W., and Xu, Q. (2015). GraRep: Learning graph representations with global structural information. In *ACM International Conference on Information and Knowledge Management*, pages 891–900.

Cao, S., Lu, W., and Xu, Q. (2016). Deep neural networks for learning graph representations. In *AAAI Conference on Artificial Intelligence*.

Carlsson, G. and Mémoli, F. (2013). Classifying clustering schemes. *Foundations of Computational Mathematics*, 13(2):221–252.

Carlsson, G., Mémoli, F., Ribeiro, A., and Segarra, S. (2013). Axiomatic construction of hierarchical clustering in asymmetric networks. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 5219–5223.

Carlsson, G., Mémoli, F., Ribeiro, A., and Segarra, S. (2014). Hierarchical quasi-clustering methods for asymmetric networks. In Xing, E. P. and Jebara, T., editors, *International Conference on Machine Learning*, volume 32 of *Proceedings of Machine Learning Research*, pages 352–360, Bejing, China. PMLR.

Carlsson, G., Mémoli, F., and Segarra, S. (2021). Robust hierarchical clustering for directed networks: An axiomatic approach. *SIAM Journal on Applied Algebra and Geometry*, 5(4):675–700.

Chanpuriya, S., Musco, C., Sotiropoulos, K., and Tsourakakis, C. (2020). Node embeddings and exact low-rank representations of complex networks. *Advances in Neural Information Processing Systems*, 33.

Chessel, A., Cao, F., and Fablet, R. (2006). Interpolating orientation fields: An axiomatic approach. In *European Conference on Computer Vision*, pages 241–254. Springer.

Cohen-Addad, V., Kanade, V., and Mallmann-Trenn, F. (2018). Clustering redemption – beyond the impossibility of Kleinberg's axioms. In *Advances in Neural Information Processing Systems*, pages 8526–8535.

Cox, M. A. and Cox, T. F. (2008). Multidimensional scaling. In *Handbook of data visualization*, pages 315–347. Springer.

Davis, C. and Kahan, W. M. (1970). The rotation of eigenvectors by a perturbation. iii. *SIAM Journal on Numerical Analysis*, 7(1):1–46.

Di Giovanni, F., Giusti, L., Barbero, F., Luise, G., Lio, P., and Bronstein, M. M. (2023). On over-squashing in message passing neural networks: The impact of width, depth, and topology. In *International Conference on Machine Learning*, pages 7865–7885. PMLR.

Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41.

Goyal, P. and Ferrara, E. (2018). Graph embedding techniques, applications, and performance: A survey. *Knowledge-Based Systems*, 151:78–94.

Grover, A. and Leskovec, J. (2016). node2vec: Scalable feature learning for networks. In *ACM International Conference on Knowledge Discovery and Data Mining*, pages 855–864.

Hamilton, W. L. (2020). Graph representation learning. *Synthesis Lectures on Artifical Intelligence and Machine Learning*, 14(3):1–159.

Hamilton, W. L., Ying, R., and Leskovec, J. (2017). Inductive representation learning on large graphs. In *Advances in Neural Information Processing Systems*, pages 1025–1035.

Johnson, S. C. (1967). Hierarchical clustering schemes. *Psychometrika*, 32(3):241–254.

Kenney, C. S., Zuliani, M., and Manjunath, B. (2005). An axiomatic approach to corner detection. In *Conference on Computer Vision and Pattern Recognition*, volume 1, pages 191–197. IEEE.

Kleinberg, J. (2002). An impossibility theorem for clustering. *Advances in Neural Information Processing Systems*, 15:463–470.

Lan, T., Kao, D., Chiang, M., and Sabharwal, A. (2010). An axiomatic theory of fairness in network resource allocation. In *IEEE Conference on Computer Communications*, pages 1343–1351.

Lance, G. N. and Williams, W. T. (1967). A general theory of classificatory sorting strategies: 1. Hierarchical systems. *The Computer Journal*, 9(4):373–380.

Larroca, F., Bermolen, P., Fiori, M., and Mateos, G. (2021). Change point detection in weighted and directed random dot product graphs. In *European Signal Processing Conference*, pages 1810–1814. IEEE.

Levy, O. and Goldberg, Y. (2014). Neural word embedding as implicit matrix factorization. In *Advances in Neural Information Processing Systems*, pages 2177–2185.

Ma, Y., Wang, S., Ren, Z., Yin, D., and Tang, J. (2017). Preserving local and global information for network embedding. *arXiv preprint*.

Mac Lane, S. (1971). *Categories for the working mathematician*. Springer Verlag.

Marenco, B., Bermolen, P., Fiori, M., Larroca, F., and Mateos, G. (2022). Online change point detection for weighted and directed random dot product graphs. *IEEE Transaction on Signal and Information Processing over Networks*, 8:144–159.

Meila, M. (2005). Comparing clusterings: An axiomatic view. In *International Conference on Machine Learning*, page 577–584.

Mikolov, T., Chen, K., Corrado, G., and Dean, J. (2013a). Efficient estimation of word representations in vector space. In *International Conference on Learning Representations*.

Mikolov, T., Sutskever, I., Chen, K., Corrado, G., and Dean, J. (2013b). Distributed representations of words and phrases and their compositionality. In *Advances in Neural Information Processing Systems*, pages 3111–3119.

Ou, M., Cui, P., Pei, J., Zhang, Z., and Zhu, W. (2016). Asymmetric transitivity preserving graph embedding. In *ACM International Conference on Knowledge Discovery and Data Mining*, pages 1105–1114.

Pan, S., Hu, R., Long, G., Jiang, J., Yao, L., and Zhang, C. (2018). Adversarially regularized graph autoencoder for graph embedding. In *International Joint Conference on Artificial Intelligence*, pages 2609–2615.

Pennock, D. M., Horvitz, E., and Giles, C. L. (2000). Social choice theory and recommender systems: Analysis of the axiomatic foundations of collaborative filtering. In *AAAI Conference on Artificial Intelligence*, pages 729–734.

Perozzi, B., Al-Rfou, R., and Skiena, S. (2014). DeepWalk: Online learning of social representations. In *ACM International Conference on Knowledge Discovery and Data Mining*, pages 701–710.

Qiu, J., Dong, Y., Ma, H., Li, J., Wang, K., and Tang, J. (2018). Network embedding as matrix factorization: Unifying DeepWalk, LINE, PTE, and node2vec. In *ACM International Conference on Web Search and Data Mining*, pages 459–467.

Rampášek, L., Galkin, M., Dwivedi, V. P., Luu, A. T., Wolf, G., and Beaini, D. (2022). Recipe for a general, powerful, scalable graph transformer. *Advances in Neural Information Processing Systems*, 35:14501–14515.

Riehl, E. (2017). *Category theory in context*. Courier Dover Publications.

Segarra, S., Roddenberry, T. M., Mémoli, F., and Ribeiro, A. (2020). Metric representations of networks: A uniqueness result. In *International Conference on Acoustics, Speech, and Signal Processing*, pages 5345–5349.

Seshadhri, C., Sharma, A., Stolman, A., and Goel, A. (2020). The impossibility of low-rank representations for triangle-rich complex networks. *Proceedings of the National Academy of Sciences*, 117(11):5631–5637.

Shaw, B. and Jebara, T. (2009). Structure preserving embedding. In *International Conference on Machine Learning*, pages 937–944.

Shervashidze, N., Schweitzer, P., Van Leeuwen, E. J., Mehlhorn, K., and Borgwardt, K. M. (2011). Weisfeiler-Lehman graph kernels. *Journal of Machine Learning Research*, 12(9).

Srinivasan, B. and Ribeiro, B. (2019). On the equivalence between positional node embeddings and structural graph representations. In *International Conference on Learning Representations*.

von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and computing*, 17(4):395–416.

Wang, D., Cui, P., and Zhu, W. (2016). Structural deep network embedding. In *ACM International Conference on Knowledge Discovery and Data Mining*, pages 1225–1234.

Wang, Q., Mao, Z., Wang, B., and Guo, L. (2017). Knowledge graph embedding: A survey of approaches and applications. *IEEE Transactions on Knowledge and Data Engineering*, 29(12):2724–2743.

Xu, M. (2021). Understanding graph embedding methods and their applications. *SIAM Review*, 63(4):825–853.

Zhang, D., Yin, J., Zhu, X., and Zhang, C. (2018). Network representation learning: A survey. *IEEE Transactions on Big Data*, 6(1):3–28.

Zhu, Y., Swami, A., and Segarra, S. (2023). Free energy node embedding via generalized skip-gram with negative sampling. *IEEE Transactions on Knowledge and Data Engineering*, 35(8):8024–8036.

Checklist

1. For all models and algorithms presented, check if you include:
   (a) A clear description of the mathematical setting, assumptions, algorithm, and/or model. **Yes**.
   (b) An analysis of the properties and complexity (time, space, sample size) of any algorithm. **Not Applicable**. We do not propose a new

algorithm but rather a way to study and categorize existing ones through relevant properties.

   (c) (Optional) Anonymized source code, with specification of all dependencies, including external libraries. **Not Applicable**.

2. For any theoretical claim, check if you include:
   (a) Statements of the full set of assumptions of all theoretical results. **Yes**.
   (b) Complete proofs of all theoretical results. **Yes**. Our main theorem is shown in Section 5 and all other results are shown in the supplementary material.
   (c) Clear explanations of any assumptions. **Yes**.

3. For all figures and tables that present empirical results, check if you include:
   (a) The code, data, and instructions needed to reproduce the main experimental results (either in the supplemental material or as a URL). **Not Applicable**. No empirical results are presented, given the theoretical nature of the paper.
   (b) All the training details (e.g., data splits, hyperparameters, how they were chosen). **Not Applicable**. No empirical results are presented, given the theoretical nature of the paper.
   (c) A clear definition of the specific measure or statistics and error bars (e.g., with respect to the random seed after running experiments multiple times). **Not Applicable**. No empirical results are presented, given the theoretical nature of the paper.
   (d) A description of the computing infrastructure used. (e.g., type of GPUs, internal cluster, or cloud provider). **Not Applicable**. No empirical results are presented, given the theoretical nature of the paper.

4. If you are using existing assets (e.g., code, data, models) or curating/releasing new assets, check if you include:
   (a) Citations of the creator If your work uses existing assets. **Not Applicable**.
   (b) The license information of the assets, if applicable. **Not Applicable**.
   (c) New assets either in the supplemental material or as a URL, if applicable. **Not Applicable**.
   (d) Information about consent from data providers/curators. **Not Applicable**.
   (e) Discussion of sensible content if applicable, e.g., personally identifiable information or offensive content. **Not Applicable**.

5. If you used crowdsourcing or conducted research with human subjects, check if you include:
   (a) The full text of instructions given to participants and screenshots. **Not Applicable**.
   (b) Descriptions of potential participant risks, with links to Institutional Review Board (IRB) approvals if applicable. **Not Applicable**.
   (c) The estimated hourly wage paid to participants and the total amount spent on participant compensation. **Not Applicable**.

An Impossibility Theorem for Node Embedding: Supplementary Materials

A Table of Notation

Notation	Description
$(\mathcal{V}, d)$	Dissimilarity network
$(\mathcal{S}, \rho)$	Dissimilarity space
$\mathcal{N}(\mathcal{S})$	$\mathcal{S}$-featured dissimilarity networks
$\mathcal{M}$	Finite pseudodissimilarity spaces
$\xi : \mathcal{N}(\mathcal{S}) \to \mathcal{M}$	Node embedding function
$\Psi : \mathcal{V}_1 \to \mathcal{V}_2$	CoDNI map
$\mathcal{D}_{ij}$ for some $(\mathcal{V}, d)$	$\{d' : d'(i,j) = d(i,j), d'(k,\ell) \geq d(k,\ell)\ \forall k,\ell \in \mathcal{V}\}$

Table 1: Table of notation.

B A Categorical Variation of Theorem 1

We frame the study of node embedding functions in terms of functors between categories, inspired by the work of Carlsson and Mémoli (2013) on clustering. In order to do so, let us first recall basic definitions in category theory. We refer the reader to the books (Mac Lane, 1971; Riehl, 2017) for further reference.

Definition 3. *A* category C *consists of:*

1. *A collection of* objects $\mathrm{ob}(\mathsf{C})$
2. *For each* $X, Y \in \mathrm{ob}(\mathsf{C})$*, a collection of* morphisms $\mathrm{Mor}_{\mathsf{C}}(X, Y)$*,*

such that

1. *For each* $X \in \mathrm{ob}(\mathsf{C})$*, there is a distinguished* identity morphism $\mathrm{id}_X \in \mathrm{Mor}_{\mathsf{C}}(X, X)$
2. *There is a* composition map*, so that for any* $X, Y, Z \in \mathrm{ob}(\mathsf{C})$*,*

$$\circ : \mathrm{Mor}_{\mathsf{C}}(X, Y) \times \mathrm{Mor}_{\mathsf{C}}(Y, Z) \to \mathrm{Mor}_{\mathsf{C}}(X, Z) \tag{18}$$

   where $\circ$ *is associative*

3. *For any* $X, Y \in \mathrm{ob}(\mathsf{C}), f \in \mathrm{Mor}_{\mathsf{C}}(X, Y), g \in \mathrm{Mor}_{\mathsf{C}}(Y, X)$*, we have* $\mathrm{id}_Y \circ f = g \circ \mathrm{id}_X$*.*

A type of morphism of particular interest is an isomorphism.

Definition 4. *Let* C *be a category, and let* $X, Y \in \mathrm{ob}(\mathsf{C}), f \in \mathrm{Mor}_{\mathsf{C}}(X, Y)$ *be given. If there exists* $g \in \mathrm{Mor}_{\mathsf{C}}(Y, X)$ *such that* $g \circ f = \mathrm{id}_X, f \circ g = \mathrm{id}_Y$*, we say that* f *is an* isomorphism*. If there exists such an isomorphism, we say that* X *and* Y *are* isomorphic *in* C*.*

The canonical example of a category is Set, where $\mathrm{ob}(\mathsf{Set})$ is the collection of sets, and $\mathrm{Mor}_{\mathsf{Set}}(X, Y)$ consists of all functions from X to Y. We also define the following relevant categories.

Definition 5. *Let $\mathcal{S}$ be a dissimilarity space. The category of $\mathcal{S}$-featured dissimilarity networks, denoted $\mathsf{N}\mathcal{S}$, is such that*

1. $\mathrm{ob}(\mathsf{N}\mathcal{S})$ *is the collection of all triples $(\mathcal{V}, d, F)$, where $\mathcal{V}$ is a finite set, d is a dissimilarity function on $\mathcal{V}$, and $F : \mathcal{V} \to \mathcal{S}$*
2. *For $N_1, N_2 \in \mathrm{ob}(\mathsf{N}\mathcal{S})$, $\mathrm{Mor}_{\mathsf{N}\mathcal{S}}(N_1, N_2)$ is the collection of all CoDNI maps from N_1 to N_2*
3. *For $N \in \mathrm{ob}(\mathsf{N}\mathcal{S})$, id_N is the canonical identity map on the nodes, and composition is also defined as expected.*

Definition 6. *The category of finite pseudodissimilarity spaces, denoted M, is such that*

1. $\mathrm{ob}(\mathsf{M})$ *is the collection of all tuples $(\mathcal{M}, \rho)$, where $\mathcal{M}$ is a finite set, and ρ is a pseudodissimilarity function on $\mathcal{M}$*
2. *For $(M_1, \rho_1), (M_2, \rho_2) \in \mathrm{ob}(\mathsf{M})$, $\mathrm{Mor}_{\mathsf{M}}(M_1, M_2)$ is the collection of all non-expansive maps from M_1 to M_2, that is, maps $f : M_1 \to M_2$ such that for all $x, y \in M_1$:*
$$\rho_1(x, y) \geq \rho_2(f(x), f(y)) \tag{19}$$
3. *For $M \in \mathrm{ob}(\mathsf{M})$, id_M is the canonical identity map on M, and composition is also defined as expected.*

One suspects that node embedding functions take objects in the category $\mathsf{N}\mathcal{S}$, and yield corresponding objects in the category M. Moreover, these maps should preserve structure in some way. The way to describe this formally is via a functor:

Definition 7. *Let C, D be categories. A* functor $F : \mathsf{C} \to \mathsf{D}$ *consists of*

1. *An* object function $F : \mathrm{ob}(\mathsf{C}) \to \mathrm{ob}(\mathsf{D})$, *whose application we denote by FX for each $X \in \mathrm{ob}(\mathsf{C})$*
2. *A* morphism function, *so for every $X, Y \in \mathrm{ob}(\mathsf{C})$, $F : \mathrm{Mor}_{\mathsf{C}}(X, Y) \to \mathrm{Mor}_{\mathsf{D}}(FX, FY)$, whose application we denote by Ff for each $f \in \mathrm{Mor}_{\mathsf{C}}(X, Y)$,*

such that

1. *For each $X \in \mathrm{ob}(\mathsf{C})$, we have $F\mathrm{id}_X = \mathrm{id}_{FX}$*
2. *For each $X, Y, Z \in \mathrm{ob}(\mathsf{C}), f \in \mathrm{Mor}_{\mathsf{C}}(X, Y), g \in \mathrm{Mor}_{\mathsf{C}}(Y, Z)$, we have $F(g \circ f) = Fg \circ Ff$.*

Two functors can be composed by composing their respective object and morphism functions. A particularly useful type of functor is the *forgetful functor*. For the category of featured dissimilarity networks $\mathsf{N}\mathcal{S}$, the forgetful functor $\alpha : \mathsf{N}\mathcal{S} \to \mathsf{Set}$ is such that for any given dissimilarity networks $N_1 = (\mathcal{V}_1, d_1, F_1), N_2 = (\mathcal{V}_2, d_2, F_2)$ in $\mathrm{ob}(\mathsf{N}\mathcal{S})$ and CoDNI map $\Psi : \mathcal{V}_1 \to \mathcal{V}_2$ in $\mathrm{Mor}_{\mathsf{N}\mathcal{S}}(N_1, N_2)$, we have

$$\begin{gathered} \alpha : \mathrm{ob}(\mathsf{N}\mathcal{S}) \to \mathrm{ob}(\mathsf{Set}) \\ (\mathcal{V}_1, d_1, F_1) \mapsto \mathcal{V}_1 \\ \\ \alpha : \mathrm{Mor}_{\mathsf{N}\mathcal{S}}(N_1, N_2) \to \mathrm{Mor}_{\mathsf{Set}}(\mathcal{V}_1, \mathcal{V}_2) \\ \Psi \mapsto \Psi. \end{gathered} \tag{20}$$

That is, the forgetful functor on $\mathsf{N}\mathcal{S}$ discards the dissimilarity and feature information of a network, and preserves morphisms as maps between sets. Similarly, we define the forgetful functor $\beta : \mathsf{M} \to \mathsf{Set}$ for the category of finite pseudodissimilarity spaces so that for any $(M_1, \rho_1), (M_2, \rho_2) \in \mathrm{ob}(\mathsf{M})$ and non-expansive map $f : M_1 \to M_2 \in \mathrm{Mor}_{\mathsf{N}\mathcal{S}}((M_1, \rho_1), (M_2, \rho_2))$, we have

$$\begin{gathered} \beta : \mathrm{ob}(\mathsf{M}) \to \mathrm{ob}(\mathsf{Set}) \\ (M_1, \rho_1) \mapsto M_1 \\ \\ \beta : \mathrm{Mor}_{\mathsf{M}}((M_1, \rho_1), (M_2, \rho_2)) \to \mathrm{Mor}_{\mathsf{Set}}(M_1, M_2) \\ f \mapsto f. \end{gathered} \tag{21}$$

Before restating our impossibility result, we define the appropriate notion of graph-awareness in this context.

Property 6 (Functorial graph-awareness)**.** *Let $\alpha : \mathsf{N}\mathcal{S} \to \mathsf{Set}$ and $\beta : \mathsf{M} \to \mathsf{Set}$ be the forgetful functors as previously described. Let $\xi : \mathsf{N}\mathcal{S} \to \mathsf{M}$ be a functor such that $\beta \circ \xi = \alpha$. Thus, for any network $N = (\mathcal{V}, d, F)$, the finite pseudodissimilarity space $(M, \rho) = \xi N$ satisfies $M = \mathcal{V}$; in light of this, we simply write $(\mathcal{V}, \rho) = \xi N$. Consider a featured dissimilarity network $N = (\mathcal{V}, d, F)$ on at least 3 nodes, and consider an arbitrary pair of distinct nodes $i, j \in \mathcal{V}$. Let $\mathcal{D}_{ij}$ be the set of dissimilarity functions d' on $\mathcal{V}$ such that $d'(i, j) = d(i, j)$, and for all $k, \ell \in \mathcal{V}$ it holds that $d'(k, \ell) \geq d(k, \ell)$. For any such d', note that $\mathrm{id}_{\mathcal{V}} \in \mathrm{Mor}_{\mathsf{N}\mathcal{S}}(N, (\mathcal{V}, d', F))$.*

If for all such networks N and pairs of distinct nodes i, j there exists a $d' \in \mathcal{D}_{ij}$ such that putting $(\mathcal{V}, \rho) = \xi N, (\mathcal{V}, \rho') = \xi(\mathcal{V}, d', F)$ yields $\rho(i, j) \neq \rho'(i, j)$, we say that the functor ξ is graph-aware.

Note that this definition of graph-awareness for functors is essentially identical to that for node embedding functions, except stated in a way that includes the condition of preserving the underlying finite set. With these definitions in place, we restate Theorem 1 in categorical language before proceeding to its proof.

Theorem 2. *There exists no graph-aware functor $\xi : \mathsf{N}\mathcal{S} \to \mathsf{M}$.*

Proof. We prove the categorical version in Theorem 2 by reducing it to the original statement in Theorem 1. That is, we show that the existence of such a functor would imply the existence of a node embedding function that is self-contained, consistent, and graph-aware, which is absurd. Let $\xi : \mathsf{N}\mathcal{S} \to \mathsf{M}$ be a functor.

We first consider the implications of the condition $\beta \circ \xi = \alpha$ stated in the redefinition of graph-awareness, which we state in Lemmas 1 to 3.

Lemma 1. *If $\beta \circ \xi = \alpha$, then the object function $\xi : \mathrm{ob}(\mathsf{N}\mathcal{S}) \to \mathrm{ob}(\mathsf{M})$ is a node embedding function.*

Proof. (Lemma 1) The statement immediately follows from the definition of a node embedding function. □

Since the condition $\beta \circ \xi = \alpha$ implies that the object function of ξ is a node embedding function, it makes sense to speak of it being self-contained and consistent.

Lemma 2. *If $\beta \circ \xi = \alpha$, then the object function $\xi : \mathrm{ob}(\mathsf{N}\mathcal{S}) \to \mathrm{ob}(\mathsf{M})$ is self-contained.*

Proof. (Lemma 2) Consider two dissimilarity networks $N_1 = (\{i_1, j_1\}, d_1, F_1), N_2 = (\{i_2, j_2\}, d_2, F_2)$, where $d_1(i_1, j_1) = d_2(i_2, j_2)$ and $F_1(i_1) = F_2(i_2), F_1(j_1) = F_2(j_2)$. Then, the bijection

$$\begin{aligned} f : \{i_1, j_1\} &\to \{i_2, j_2\} \\ i_1 &\mapsto j_1 \\ i_2 &\mapsto j_2 \end{aligned} \tag{22}$$

is an isomorphism. Thus, the morphism $\xi f \in \mathrm{Mor}_{\mathsf{M}}(\xi N_1, \xi N_2)$ is also an isomorphism. That is to say, the embeddings $(M_1, \rho_1) = \xi N_1, (M_2, \rho_2) = \xi N_2$ under ξ are isomorphic in M. By the hypothesis $\beta \circ \xi = \alpha$, we have that $\xi f = f$ and $M_1 = \{i_1, j_2\}$ and $M_2 = \{i_2, j_2\}$. One can then check that the isomorphism of these spaces implies $\rho_1(i_1, j_1) = \rho_2(f(i_1), f(i_2)) = \rho_2(i_2, j_2)$. That is, the embedding distance between nodes in a two-node network is invariant under labeling of the nodes, thus satisfying self-containedness. □

Lemma 3. *If $\beta \circ \xi = \alpha$, then the object function $\xi : \mathrm{ob}(\mathsf{N}\mathcal{S}) \to \mathrm{ob}(\mathsf{M})$ is consistent.*

Proof. (Lemma 3) Let $N_1 = (\mathcal{V}_1, d_1, F_1), N_2 = (\mathcal{V}_2, d_2, F_2) \in \mathrm{ob}(\mathcal{N})$ be given such that there exists a CoDNI map $\Psi : \mathcal{V}_1 \to \mathcal{V}_2$ in $\mathrm{Mor}_{\mathsf{N}\mathcal{S}}(N_1, N_2)$. Put $(M_1, \rho_1) = \xi N_1, (M_2, \rho_2) = \xi N_2$. By the hypothesis $\beta \circ \xi = \alpha$, we have $M_1 = \mathcal{V}_1, M_2 = \mathcal{V}_2$, and $\xi\Psi = \Psi$. Since $\Psi \in \mathrm{Mor}_{\mathsf{M}}((M_1, \rho_1), (M_2, \rho_2))$, the following holds for all $i, j \in \mathcal{V}_1$:

$$\rho_1(i, j) \geq \rho_2(\Psi(i), \Psi(j)), \tag{23}$$

so that the object function $\xi : \mathrm{ob}(\mathsf{N}\mathcal{S}) \to \mathrm{ob}(\mathsf{M})$ satisfies consistency. □

We now show that assuming graph-awareness of ξ implies that the object function is graph-aware. By the definition of a functor being graph-aware, $\beta \circ \xi = \alpha$, so that the object function of ξ is a node embedding function by Lemma 1. Thus, it makes sense to speak of the object function of ξ being graph-aware as well.

Lemma 4. *If ξ is graph-aware, then the object function $\xi : \mathrm{ob}(\mathsf{N}\mathcal{S}) \to \mathrm{ob}(\mathsf{M})$ is graph-aware.*

Proof. *(Lemma 4)* The statement immediately follows from the definition of graph-awareness for a node embedding function. □

To conclude the proof, suppose for the sake of contradiction that ξ is graph-aware. Then, by Lemmas 1 to 4, the object function $\xi : \mathrm{ob}(\mathsf{N}\mathcal{S}) \to \mathrm{ob}(\mathsf{M})$ is a node embedding function that is self-contained, consistent, and graph-aware. This is impossible, as desired. □

C Impossibility Results for Similarity Networks

A common type of network in practice is a similarity network, where the weights on each edge indicate similarity, rather than dissimilarity, between nodes. A *similarity network* is a finite set of nodes $\mathcal{V}$ coupled with a symmetric *similarity function* $e : \mathcal{V} \times \mathcal{V} \to \mathbb{R}^{>0} \cup \{+\infty\}$, with the condition that $e(i, j) = +\infty$ if and only if $i = j$. An $\mathcal{S}$*-featured similarity network* is defined much in the same way as a dissimilarity network, for some dissimilarity space $(\mathcal{S}, \rho)$.

Similarity networks behave much like dissimilarity networks. Indeed, we will now show that modifying the notion of a CoDNI map to address the interpretation of a similarity network yields a category of featured similarity networks that is isomorphic to the category of featured dissimilarity networks. Let us now define the notion of a "contractive and similarity non-decreasing" map from one similarity network to another.

Definition 8. *Consider two featured similarity networks $N_1 = (\mathcal{V}_1, e_1, F_1), N_2 = (\mathcal{V}_2, e_2, F_2)$ with possibly different numbers of nodes. A map $\Psi : \mathcal{V}_1 \to \mathcal{V}_2$ is said to be* contractive and similarity non-decreasing *(CoSND) if, for all $i, j \in \mathcal{V}_1$,*

$$\rho(F_1(i), F_1(j)) \geq \rho(F_2(\Psi(i)), F_2(\Psi(j))) \tag{24}$$
$$e_1(i, j) \leq e_2(\Psi(i), \Psi(j)). \tag{25}$$

Observe that the definition of a CoSND map is dual to that of a CoDNI map for dissimilarity networks: rather than decreasing network dissimilarities, it increases network similarities. Taking note of this symmetry, we find it convenient to define the *dual* of a (dis)similarity network.

Definition 9. *Let $N = (\mathcal{V}, d, F)$ be a featured dissimilarity network. The* dual *of N, denoted N^*, is the similarity network $N^* = (\mathcal{V}, e, F)$, where for any $i, j \in \mathcal{V}$ with $i \neq j$, e is a similarity function such that*

$$\begin{aligned} e : \mathcal{V} \times \mathcal{V} &\to \mathbb{R}^{\geq 0} \cup \{+\infty\} \\ (i, i) &\mapsto +\infty \\ (i, j) &\mapsto \frac{1}{d(i, j)}. \end{aligned} \tag{26}$$

Similarly, for a featured similarity network $M = (\mathcal{V}, e, F)$, the dual *of M, also denoted M^*, is the dissimilarity network $M^* = (\mathcal{V}, d, F)$, where for any $i, j \in \mathcal{V}$ with $i \neq j$, d is a dissimilarity function such that*

$$\begin{aligned} d : \mathcal{V} \times \mathcal{V} &\to \mathbb{R}^{\geq 0} \\ (i, i) &\mapsto 0 \\ (i, j) &\mapsto \frac{1}{e(i, j)}. \end{aligned} \tag{27}$$

Observe that the double-dual of a (dis)similarity network is itself: $(N^*)^* = N$. Taking the dual of a network transforms CoDNI maps to CoSND maps and vice versa, as stated in the following lemma.

Lemma 5. *Let $N_1 = (\mathcal{V}_1, d_1, F_1), N_2 = (\mathcal{V}_2, d_2, F_2)$ be dissimilarity networks. Suppose $\Psi : \mathcal{V}_1 \to \mathcal{V}_2$ is a CoDNI map from N_1 to N_2. Then, Ψ is also a CoSND map from N_1^* to N_2^*.*

Similarly, let $M_1 = (\mathcal{V}_1, e_1, F_1), M_2 = (\mathcal{V}_2, e_2, F_2)$ be similarity networks. Suppose $\Psi : \mathcal{V}_1 \to \mathcal{V}_2$ is a CoSND map from M_1 to M_2. Then Ψ is also a CoDNI map from M_1^* to M_2^*.

We omit the proof, as it is essentially a direct application of the definitions. With these tools in place, we are now ready to define the category of featured similarity networks.

Definition 10. *Let $\mathcal{S}$ be a dissimilarity space. The category of $\mathcal{S}$-featured similarity networks, denoted $\mathsf{S}\mathcal{S}$, is such that*

1. $\mathrm{ob}(\mathsf{S}\mathcal{S})$ *is the collection of all triples $(\mathcal{V}, e, F)$, where $\mathcal{V}$ is a finite set, e is a similarity function on $\mathcal{V}$, and $F : \mathcal{V} \to \mathcal{S}$*
2. *For $N_1, N_2 \in \mathrm{ob}(\mathsf{S}\mathcal{S})$, $\mathrm{Mor}_{\mathsf{N}\mathcal{S}}(N_1, N_2)$ is the collection of all CoSND maps from N_1 to N_2*
3. *For $N \in \mathrm{ob}(\mathsf{N}\mathcal{S})$, id_N is defined in the obvious way, and composition is also defined as expected.*

Of course, there is a natural forgetful functor $\gamma : \mathsf{S}\mathcal{S} \to \mathsf{Set}$, similar to the forgetful functor $\alpha : \mathsf{N}\mathcal{S} \to \mathsf{Set}$ as described before. As suggested by the double-dual property and Lemma 5, taking the dual network ought to respect the structure of $\mathsf{N}\mathcal{S}$ and $\mathsf{S}\mathcal{S}$. Indeed, one can explicitly construct an isomorphism between these two categories, as follows.

Lemma 6. *Let $\mathcal{S}$ be a dissimilarity space. Let $A : \mathsf{N}\mathcal{S} \to \mathsf{S}\mathcal{S}$ be a functor such that*

$$\begin{gathered} A : \mathrm{ob}(\mathsf{N}\mathcal{S}) \to \mathrm{ob}(\mathsf{S}\mathcal{S}) \\ N \mapsto N^* \\ \\ A : \mathrm{Mor}_{\mathsf{N}\mathcal{S}}(N_1, N_2) \to \mathrm{Mor}_{\mathsf{S}\mathcal{S}}(FN_1, FN_2) \\ \Psi \mapsto \Psi. \end{gathered} \tag{28}$$

Similarly, let $B : \mathsf{S}\mathcal{S} \to \mathsf{N}\mathcal{S}$ be a functor such that

$$\begin{gathered} B : \mathrm{ob}(\mathsf{S}\mathcal{S}) \to \mathrm{ob}(\mathsf{N}\mathcal{S}) \\ M \mapsto M^* \\ \\ B : \mathrm{Mor}_{\mathsf{S}\mathcal{S}}(M_1, M_2) \to \mathrm{Mor}_{\mathsf{S}\mathcal{S}}(GM_1, GM_2) \\ \Psi \mapsto \Psi. \end{gathered} \tag{29}$$

Then, A, B form an isomorphism between $\mathsf{N}\mathcal{S}$ and $\mathsf{S}\mathcal{S}$, so that $A \circ B = \mathrm{id}_{\mathsf{S}\mathcal{S}}$ and $B \circ A = \mathrm{id}_{\mathsf{N}\mathcal{S}}$. We call A, B the dual functors.

Given this machinery, we now develop the notion of graph-awareness for functors from the category of similarity networks to the category of dissimilarity spaces. This definition mirrors that for dissimilarity networks.

Property 7 (Functorial similarity graph-awareness)**.** *Let $\gamma : \mathsf{S}\mathcal{S} \to \mathsf{Set}$ and $\beta : \mathsf{M} \to \mathsf{Set}$ be the forgetful functors as previously described. Let $\eta : \mathsf{S}\mathcal{S} \to \mathsf{M}$ be a functor such that $\beta \circ \eta = \gamma$. Thus, for any similarity network $N = (\mathcal{V}, e, F)$, the finite pseudodissimilarity space $(M, \rho) = \eta N$ satisfies $M = \mathcal{V}$; in light of this, we simply write $(\mathcal{V}, \rho) = \eta N$. Consider a featured similarity network $N = (\mathcal{V}, d, F)$ on at least 3 nodes, and consider an arbitrary pair of distinct nodes $i, j \in \mathcal{V}$. Let $\mathcal{E}_{ij}$ be the set of dissimilarity functions e' on $\mathcal{V}$ such that $e'(i, j) = e(i, j)$, and for all $k, \ell \in \mathcal{V}$ it holds that $e'(k, \ell) \leq e(k, \ell)$. For any such e', note that $\mathrm{id}_{\mathcal{V}} \in \mathrm{Mor}_{\mathsf{S}\mathcal{S}}(N, (\mathcal{V}, e', F))$.*

If for all such networks N and pairs of distinct nodes i, j there exists a $e' \in \mathcal{E}_{ij}$ such that putting $(\mathcal{V}, \rho) = \eta N, (\mathcal{V}, \rho') = \eta(\mathcal{V}, e', F)$ yields $\rho(i, j) \neq \rho'(i, j)$, we say that the functor η is graph-aware.

This definition not only mirrors that for the category of dissimilarity networks, it respects the dual functor in the following sense.

Lemma 7. *Let $A : \mathsf{N}\mathcal{S} \to \mathsf{S}\mathcal{S}$ and $B : \mathsf{S}\mathcal{S} \to \mathsf{N}\mathcal{S}$ be the dual functors. If $\eta : \mathsf{S}\mathcal{S} \to \mathsf{M}$ is a graph-aware functor, then $\eta \circ B : \mathsf{N}\mathcal{S} \to \mathsf{M}$ is a graph-aware functor. Similarly, if $\xi : \mathsf{N}\mathcal{S} \to \mathsf{M}$ is a graph-aware functor, then $\xi \circ A : \mathsf{S}\mathcal{S} \to \mathsf{M}$ is a graph-aware functor.*

We omit the proof of this, as it follows quite simply from the definitions. Mirroring Theorem 2, this observation directly implies the following theorem.

Theorem 3. *Denote the category of $\mathcal{S}$-featured similarity networks by $\mathsf{S}\mathcal{S}$ with contractive and similarity non-decreasing maps as morphisms. There exists no graph-aware functor $\eta : \mathsf{S}\mathcal{S} \to \mathsf{M}$.*

The above theorem can similarly be read as "there is no node embedding function for similarity networks that satisfies self-containedness, consistency, and graph-awareness," where consistency and graph-awareness are redefined to take the modified definitions into account.

D Proof of Proposition 1

Although in light of Theorem 1 it is sufficient to verify that self-containedness and consistency hold to prove Proposition 1, we discuss all three properties for the sake of completeness.

Self-containedness. Let $N_2 = (\{i,j\}, d, F)$ be a featured dissimilarity network on 2 nodes, and let $(\{i,j\}, \phi) = \xi(N_2)$. Observe that $P = [i,j]$ is the only path from i to j (disregarding paths with repeated entries), so that by (5) we have

$$\phi(i,j) = d(i,j). \tag{30}$$

This is consistent with Property 1, where $g(F(i), F(j); \alpha) = \alpha$, as desired.

Consistency. Let two featured dissimilarity networks $N_1 = (\mathcal{V}_1, d_1, F_1), N_2 = (\mathcal{V}_2, d_2, F_2)$ be given such that there exists a CoDNI map $\Psi : \mathcal{V}_1 \to \mathcal{V}_2$. Let $(\mathcal{V}_1, \phi_1) = \xi(N_1), (\mathcal{V}_2, \phi_2) = \xi(N_2)$. Let nodes $i, j \in \mathcal{V}_1$ be given, and consider a path P from i to j (in $\mathcal{V}_1$). Construct the path $\widehat{P}$ from $\Psi(i)$ to $\Psi(j)$ by applying Ψ to each element in P. Since Ψ is CoDNI, we have that $d_1(P(\ell), P(\ell+1)) \geq d_2(\widehat{P}(\ell), \widehat{P}(\ell+1))$ for all $1 \leq \ell < L(P) = L(\widehat{P})$. It follows that

$$\max_{1\leq \ell < L(P)} d_1(P(\ell), P(\ell+1)) \geq \max_{1\leq \ell < L(\widehat{P})} d_2(\widehat{P}(\ell), \widehat{P}(\ell+1)). \tag{31}$$

Examining (5), this implies that

$$\phi_1(i,j) \geq \phi_2(\Psi(i), \Psi(j)), \tag{32}$$

which is consistent with Property 2, as desired.

Graph-awareness. Since the other properties hold, Theorem 1 implies that the single-linkage procedure is not graph-aware. To verify this, let a featured dissimilarity network $N = (\mathcal{V}, d, F)$ on at least 3 nodes be given, and take $i, j \in \mathcal{V}$ to be the pair of nodes with minimum dissimilarity. That is,

$$i, j \in \operatorname*{arg\,min}_{i',j' \in \mathcal{V} : i' \neq j'} d(i', j'). \tag{33}$$

One can check that the path $P = [i,j]$ always attains the minimum distance in (5). Moreover, for any $d' \in \mathcal{D}_{ij}$ as described in Property 3, the expression (5) increases with respect to the distance d'. Since $d(i,j) = d'(i,j)$ by definition, the path $P = [i,j]$ still attains the minimum distance in (5) for any $d' \in \mathcal{D}_{ij}$, and said distance is the same as that induced by d. Thus, the single-linkage procedure is not graph-aware. ■

E Proof of Proposition 2

As before, we verify this statement for all three properties.

Self-containedness. Since the node set of a 2-node network $N = (\{i,j\}, d, F)$ has cardinality fewer than 3 nodes, the embedding distance between its constituent nodes under the triangle-linkage procedure is always $+\infty$: this satisfies self-containedness where $g(F(i), F(j); \alpha) = +\infty$.

Graph-awareness. Let a featured dissimilarity network $N = (\mathcal{V}, d, F)$ be given such that $|\mathcal{V}| \geq 3$. Let $(\mathcal{V}, \phi) = \xi_T(N)$ be the triangle-linkage embedding of N, and pick distinct nodes $i, j \in \mathcal{V}$. Put $\delta = \phi(i,j)$, and take $\mathcal{W} \subseteq \mathcal{V}$ to be the set of all nodes $k \in \mathcal{V}$ such that $d(i,k) \leq \delta, d(j,k) \leq \delta$. Notice that $\mathcal{W}$ is nonempty. Define the dissimilarity function $d' \in \mathcal{D}_{ij}$ such that for all $k \in \mathcal{W}$, we have $d'(i,k) = d'(j,k) = 2\delta$. Letting $(\mathcal{V}, \phi') = \xi_T((\mathcal{V}, d', F))$, one can verify that $\phi'(i,j) > \delta$, so that the triangle-linkage procedure is graph-aware, as desired.

Consistency. Since the other properties hold, Theorem 1 implies that the triangle-linkage procedure is not consistent. To verify this, consider two featured dissimilarity networks, $N_1 = (\{i_1, j_1, k_1\}, d_1, F_1), N_2 = (\{i_2, j_2, k_2\}, d_2, F_2)$, where d_1, d_2 are defined so that

$$\begin{aligned} d_1(i_1, j_1) &= \delta & \qquad d_2(i_2, j_2) &= \delta/2 \\ d_1(i_1, k_1) &= \delta & \qquad d_2(i_2, k_2) &= 2\delta \\ d_1(j_1, k_1) &= \delta & \qquad d_2(j_2, k_2) &= 2\delta, \end{aligned}$$

for some $\delta > 0$. Letting $(\{i_1, j_1, k_1\}, \phi_1) = \xi_T(N_1), (\{i_2, j_2, k_2\}, \phi_2) = \xi_T(N_2)$, we see that $\phi_1(i_1, j_1) = \delta$ and $\phi_2(i_2, j_2) = 2\delta$. Define the CoDNI map $\Psi : \{i_1, j_1, k_1\} \to \{i_2, j_2, k_2\}$ so that

$$\Psi \begin{cases} i_1 \mapsto i_2 \\ j_1 \mapsto j_2 \\ k_1 \mapsto j_2. \end{cases} \tag{34}$$

Despite Ψ being a CoDNI map, we have $\phi_1(i_1, j_1) < \phi_2(\Psi(i_1), \Psi(j_1))$, so that the triangle-linkage procedure is not consistent. ■

F Proof of Proposition 3

As before, we verify this statement for all three properties.

Self-containedness. Let a dissimilarity network $N_2 = (\{i, j\}, d, F)$ on two nodes is given. Put $(\{i, j\}, \phi) = \xi_C(N_2)$. One can easily see that $\phi(i, j) = 0$, satisfying self-containedness.

Graph-awareness. Let a dissimilarity network $N = (\mathcal{V}, d, F)$ on at least three nodes be given, and identify $\mathcal{V} = \{1, 2, \ldots, |\mathcal{V}|\}$. Pick two nodes in $\mathcal{V}$, and identify them with the integers $1, 2$, without loss of generality. Put $\alpha = \kappa(d(1, 2); F(1), F(2))$, and

$$\beta = \underset{\substack{i,j \in \mathcal{V} \\ j \notin \{1,2\}}}{\arg\min}\, \kappa(d(i, j); F(i), F(j)). \tag{35}$$

Consider the following $n \times n$ adjacency matrix:

$$\mathbf{A}' = \begin{bmatrix} 0 & \alpha & \beta & \\ \alpha & 0 & \beta & \mathbf{0} \\ \beta & \beta & 0 & \\ & & & \\ & \mathbf{0} & & \mathbf{0} \\ & & & \end{bmatrix}. \tag{36}$$

One can check that the first and second entries of the leading eigenvector of $\mathbf{A}'$, which we denote $\mathbf{u}'$, are equal, yielding an embedding distance of zero for the two corresponding nodes. Alternatively, consider the following $n \times n$ matrix for the same values of α, β:

$$\mathbf{A}'' = \begin{bmatrix} 0 & \alpha & \beta & \\ \alpha & 0 & \beta/2 & \mathbf{0} \\ \beta & \beta/2 & 0 & \\ & & & \\ & \mathbf{0} & & \mathbf{0} \\ & & & \end{bmatrix}. \tag{37}$$

In this case, the first and second entries of the leading eigenvector of $\mathbf{A}''$, which we denote $\mathbf{u}''$, are not equal, yielding a nonzero embedding distance for the two corresponding nodes.

For any function $f : \mathcal{V} \times \mathcal{V} \setminus \{1, 2\} \to \mathbb{R}^{\geq 0}$ such that the range of f is bounded by β and $f(i, j) = 0$ if and only if $i = j$, there exists a dissimilarity function $\widehat{d} \in \mathcal{D}_{12}$ such that $f(i, j) = \kappa(\widehat{d}(i, j); F(i), F(j))$, due to the conditions on the function κ and the choice of β. Therefore, one can choose dissimilarity functions $\widehat{d'}, \widehat{d''} \in \mathcal{D}_{12}$ such that the adjacency matrix $\widehat{\mathbf{A}}'$ constructed from the dissimilarity network $(\mathcal{V}, \widehat{d'}, F)$ is arbitrarily close to $\mathbf{A}'$ in the Frobenius norm, and the adjacency matrix $\widehat{\mathbf{A}}''$ constructed from the dissimilarity network $(\mathcal{V}, \widehat{d''}, F)$ is arbitrarily close to $\mathbf{A}''$ in the Frobenius norm. By choosing such dissimilarity functions, one can find embeddings $(\mathcal{V}, \widehat{\phi}') = \xi_C((\mathcal{V}, \widehat{d}', F)), (\mathcal{V}, \widehat{\phi}'') = \xi_C((\mathcal{V}, \widehat{d}'', F))$ so that $\widehat{\phi}'(1,2)$ is arbitrarily close to zero, and $\widehat{\phi}''(1,2)$ is arbitrarily close to the nonzero dissimilarity obtained from the leading eigenvector of (37), due to the Davis-Kahan $\sin\theta$ Theorem (Davis and Kahan, 1970). It follows that the eigenvector centrality embedding procedure is graph-aware.

Consistency. Choose an arbitrary $s \in \mathcal{S}$, and consider a dissimilarity network $N = (\{1,2\}, d, F)$, where $F(1) = F(2) = s$ and d is such that $\kappa(d(1,2); s, s) = 0.5$. Letting $(\{1,2\}, \phi) = \xi_C(N)$, we showed previously that $\phi(1,2) = 0$. Now consider a dissimilarity network $N' = (\{1,2,3\}, d', F')$, where $F'(1) = F'(2) = F'(3) = s$, and d' is chosen such that $d'(1,2) = d(1,2), \kappa(d'(1,3); s, s) = 0.25, \kappa(d'(2,3); s, s) = 0.125$, which is possible due to the properties of κ. Letting $(\{1,2,3\}, \phi') = \xi_C(N')$, one can check that $\phi'(1,2) \neq 0$. Notice that the inclusion map $\Psi : \{1,2\} \hookrightarrow \{1,2,3\}$ is a CoDNI map from N to N', but $\phi(1,2) \leq \phi'(\Psi(1), \Psi(2))$, violating consistency. ∎

G Proof of Proposition 4

By Proposition 1, the single-linkage procedure is self-contained and consistent, leaving weak graph-awareness to be shown.

Let a featured dissimilarity network $N = (\mathcal{V}, d, F)$ on at least 3 nodes be given, and let $(\mathcal{V}, \phi) = \xi(N)$ be the single-linkage node embedding. Take nodes $i, j \in \mathcal{V}$ such that $i \neq j$, and put $\alpha = \phi(i,j)$, noting that $\alpha \neq 0$. Pick a third node $k \in \mathcal{V}$, and choose any dissimilarity function $d' : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}^{\geq 0}$ such that $d'(i,j) = d(i,j), d'(i,k) = d'(j,k) = \alpha/2$. Note that $d' \in \mathcal{C}_{ij}$. Let $(\mathcal{V}, \phi') = \xi((\mathcal{V}, d', F))$, and observe that $\phi'(i,j) \leq \alpha/2$. Thus, $\phi'(i,j) \neq \phi(i,j)$, as desired. ∎

H Proof of Proposition 5

By Proposition 2, the triangle-linkage procedure is self-contained and graph-aware, leaving injective consistency to be shown.

Let two featured dissimilarity networks $N_1 = (\mathcal{V}_1, d_1, F_1), N_2 = (\mathcal{V}_2, d_2, F_2)$ be given, and suppose there exists an injective CoDNI map $\Psi : \mathcal{V}_1 \to \mathcal{V}_2$. Put $(\mathcal{V}_1, \phi_1) = \xi_T(N_1), (\mathcal{V}_2, \phi_2) = \xi_T(N_2)$. For any distinct nodes $i, j \in \mathcal{V}_1$, put $\delta = \phi_1(i,j)$, so that there exists $k \in \mathcal{V}_1$ such that $k \neq i, k \neq j$, and $\max\{d_1(i,j), d_1(i,k), d_1(j,k)\} = \delta$. Since Ψ is an injective CoDNI map, $\Psi(i), \Psi(j), \Psi(k)$ are distinct elements of $\mathcal{V}_2$, and

$$\max\{d_2(\Psi(i), \Psi(j)), d_2(\Psi(i), \Psi(k)), d_2(\Psi(j), \Psi(k))\} \leq \delta, \tag{38}$$

so that $\phi_2(\Psi(i), \Psi(j)) \leq \delta$. That is to say, the triangle-linkage procedure is injectively consistent, as desired. ∎