52nd IEEE Conference on Decision and Control
December 10-13, 2013. Florence, Italy

A Distance Metric Between Directed Weighted Graphs

Yunwen Xu, Srinivasa M. Salapaka, Carolyn L. Beck

Abstract— **Directed weighted graphs are increasingly used to model complex systems and interactions, such as networks of interconnected physical or biological subsystems. The analysis of these graphs often requires some form of *dissimilarity*, or distance measure to compare graphs. In this paper, we extend connectivity-based dissimilarity measures previously used to compare unweighted undirected graphs of the same dimensions to: (1) directed weighted graphs of the same dimensions and (2) directed weighted graphs of different dimensions. To our knowledge, this is the first approach proposed for comparing two graphs containing different numbers of nodes. We derive the conditions under which this dissimilarity measure is a pseudo-metric. This derivation provides new insights on our algorithms (previously proposed) for the graph aggregation optimization problem.**

I. INTRODUCTION

In studies of neuroscience, social networks, Internet performance and multi-agent systems, directed weighted graph models are frequently used to capture complex dynamical interactions [1]–[5]. Due to the typically large dimensions of the resulting graphs, approaches for simplifying these graph structures that maintain the underlying *ensemble* or *dominant* interactions are desired. That is, a simple *representative graph* that is *similar* to a given graph under some notion of similarity is pursued. In this paper, we focus on developing a meaningful similarity metric for comparing two graphs, for example, an original given graph and a simplified or reduced version of the same graph.

Early development of similarity/distance metrics between graphs was mostly based on graph isomorphisms, including graph isomorphism identification [6]; the edit distance given by the minimum number of adding/deleting edge operations required to obtain one graph from another [7]; the distance characterized by the maximum common subgraph or the minimum common supergraph [8]. Similarity notions that reflect neighborhood similarity have become popular in the fields of Internet analysis and social networking, and many statistical iterative updating methods have been developed [3], [9]. Another large class of graph similarity notions are defined through graph node matching or subgraph embedding [10], [11]. These methods either focus on graphs with the same number of nodes (graph isomorphisms), or deal with unweighted, typically undirected graphs (distances based on editing or sub/sup-graphs).

More recently in control and information theory, there have been specific forms of graph distances studied, such as the variation-of-information distance for comparing clustering (aggregation) results on the same dataset [12], the Kullback–Leibler (K-L) divergence rate between two Markov chains [13], and an information-based metric between two probability distributions [14]. Most of these approaches have been developed for undirected unweighted graphs and are not applicable to weighted and/or directed graphs.

In previous work, we have proposed a constructive approach for evaluating the similarity between two general directed weighted graphs of different dimensions (e.g., an original graph and an associated simplified graph) [15]. Our similarity measure uses the notion of neighborhood similarity, in particular, we say two nodes are similar if the corresponding rows in their edge weight matrices are *close* under some distance measure. Since the two graphs are of different dimensions, one-to-one node matching is not possible. To address this incompatibility, we construct a set of *composite graphs*, which provide a connection between objects in different dimensional spaces.

In this paper, we provide a proof that this constructive similarity definition is a pseudo-metric under reasonable and practical assumptions. Computing this similarity measure requires the solution of an optimal partition problem between the two node sets, which is NP-hard. We then discuss the connection between this distance metric and our previous efforts on simplifying graphs by aggregating similar nodes, and we formulate the two-fold goal of *minimizing and measuring* the distance between the original and aggregated graphs as a single optimization problem that can be approximately solved by the method proposed in [15].

II. PRELIMINARIES

A. Definitions and Notation

We review the notation and definitions that will be used throughout the paper.

– For a finite set S, we denote cardinality by $|S|$.
– The vector with n elements of value 1 is denoted by $\mathbf{1}$, and e_i denotes the vector in $\mathbb{R}^n$ with all 0 elements except for the i^{th} element, which is 1. For a matrix $X \in \mathbb{R}^{m\times n}$, its i^{th} row is denoted by $x(i) \in \mathbb{R}^{1\times n}$. For a vector $x \in \mathbb{R}^{n\times 1}$, the vector 1−norm is given by
$$\|x\|_1 = \sum_{i=1}^{n} |x_i|.$$
– We denote a weighted directed graph by a triple, $\mathcal{G}(V, E, W)$, in which $V, E \subset V\times V$ and $W \in \mathbb{R}_+^{|V|\times|V|}$ denote the set of *nodes*, *edges* and the *edge weight matrix*, respectively. We assume that $|W_{ij}| < B < \infty$ for some $B > 0$, for all (i, j) pairs. We define the node weights to be $\{p(v_i)\}$, satisfying $p(v_i) \geq 0$ with $\sum_i p(v_i) = 1$. We define the *outgoing vector* of the i^{th} node, $w(i)$, by the weights on its outgoing edges, i.e., $w(i) \triangleq [W_{i1}, W_{i2}, \cdots, W_{i|V|}]$. Here $W_{ij} = 0$ if and

This work was partially supported by the NSF grant CMMI – 1100257.
Yunwen Xu, S. Salapaka and C. L. Beck are from University of Illinois at Urbana-Champaign, Urbana, Illinois. Email: [xu27, salapaka, beck3]@illinois.edu

only if there is no edge from the i^{th} node to the j^{th} node.

- A graph $\mathcal{G}_x(V_x, E_x, X)$ is *isomorphic* to another graph $\mathcal{G}_y(V_y, E_y, Y)$, if $|V_x| = |V_y|$ and there exists an edge-weight-preserving bijective mapping $\psi : V_x \to V_y$, such that the directed edge $(i, j) \in E_x$, if and only if $(\psi(i), \psi(j)) \in E_y$, and the edge weights satisfy $X_{i,j} = Y_{\psi(i),\psi(j)}$ for all ordered pairs (i, j); we denote this isomorphism by $\mathcal{G}_x \simeq \mathcal{G}_y$.
- Let $\mathcal{N} = \{1, 2, \ldots, n\}$ and $\mathcal{M} = \{1, 2, \ldots, m\}$ be two index sets with $|\mathcal{N}| = n$ and $|\mathcal{M}| = m$, assuming $n \geq m \geq 2$. A *(hard) partition function* $\phi : \mathcal{N} \to \mathcal{M}$ is a map from $\mathcal{N}$ *onto* $\mathcal{M}$, such that $\phi^{-1}(\mathcal{M})$ is a partition of $\mathcal{N}$. That is, for any $1 \leq j \neq k \leq m$, $\phi^{-1}(j) \subset \mathcal{N}$ is non-empty, $\phi^{-1}(j) \bigcap \phi^{-1}(k) = \emptyset$ and $\bigcup_{j=1}^{m} \phi^{-1}(j) = \mathcal{N}$.
- Each partition function $\phi : \mathcal{N} \to \mathcal{M}$ defines an *aggregation matrix* $Q(\phi) \in \{0, 1\}^{n \times m}$ as

$$Q(\phi)_{i,j} = \begin{cases} 1 & \text{if } \phi(\mathrm{i}) = \mathrm{j}, \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

  Therefore, $Q(\phi)\mathbf{1} = \mathbf{1}$, and the k^{th} column of $Q(\phi)$ equals $\sum_{i \in \phi^{-1}(k)} e_i$. For a vector $w \in \mathbb{R}^{1 \times n}$, $wQ(\phi)$ provides the aggregation of w specified by partition ϕ. Remark: when $m = n$, a partition function ϕ defines a node *relabeling, or permutation*, and the associated aggregation matrix $Q(\phi)$ is a permutation matrix. In particular, if $\mathcal{G}_y$ is obtained by relabeling the nodes of $\mathcal{G}_x$, then $\mathcal{G}_x \simeq \mathcal{G}_y$.

B. Comparing graphs of the same dimensions

As an illustrative starting point, we consider the comparison of two graphs with the same number of nodes. In graph theory, a popular *distance measure* between two *connected undirected unweighted* graphs $\mathcal{G}_x(V_x, E_x, X)$ and $\mathcal{G}_y(V_y, E_y, Y)$, when $|V_x| = |V_x| = n$, is defined through *graph matching* [10], which provides a one-to-one assignment between two node sets of the same cardinality. Specifically, every *bijective* (relabeling) mapping $\psi : V_x \to V_y$ induces a ψ*-distance*, given by

$$d_\psi(\mathcal{G}_x, \mathcal{G}_y) = \sum_{i,j \in V_x} |d_{\mathcal{G}_x}(i, j) - d_{\mathcal{G}_y}(\psi(i), \psi(j))|, \tag{2}$$

in which $d_{\mathcal{G}_x}(i, j)$ and $d_{\mathcal{G}_y}(\psi(i), \psi(j))$ are the lengths of the *shortest paths* in $\mathcal{G}_x$ and $\mathcal{G}_y$ that connect the i^{th} to the j^{th} nodes, and the $\psi(i)^{th}$ to the $\psi(j)^{th}$ nodes.

Taking $d_{\mathcal{G}_x}(i, j)$ as the shortest path length is suitable when $\mathcal{G}_x$ is undirected and unweighted. For *directed weighted* graphs of the same size, we extend this notion and redefine

$$d_{\mathcal{G}_x}(i, j) := X_{i,j}, \tag{3}$$

which is the edge weight from v_i to v_j in $\mathcal{G}_x$, thus reflecting *functional influence*. Note that from the definition for the weight matrix, X, the functional influence is 0 if and only if v_i and v_j are not connected by a directed edge. Therefore, the ψ*-distance* between two directed weighted graphs (6) is given by

$$d_\psi(\mathcal{G}_x, \mathcal{G}_y) = \sum_{i,j \in V_x} |X_{i,j} - Y_{\psi(i),\psi(j)}|. \tag{4}$$

We note that every bijective mapping ψ defines an *invertible* permutation matrix $Q(\psi) \in \mathbb{R}^{n \times n}$ as in (1), with $Q(\psi)^{-1} = Q(\psi)^T$. We now introduce a matrix $Z \in \mathbb{R}^{n \times n}$ satisfying

$$Y = ZQ(\psi) \Leftrightarrow Z = YQ^T.$$

Therefore, $Y_{ij} = \sum_{k=1}^{n} Z_{ik} Q_{kj} = z(i) e_{\psi^{-1}(j)} = Z_{i,\psi^{-1}(j)}$ for all i, j. Equivalently, $Y_{i,\psi(k)} = Z_{i,k}$, for all i, k. We can then rewrite the ψ*-distance* definition (4) as

$$d_\psi(\mathcal{G}_x, \mathcal{G}_y) = \sum_{i=1}^{n} \|x(i) - z(\psi(i))\|_1. \tag{5}$$

Note that the norm $\|x(i) - z(\psi(i))\|_1$ requires only the number of columns of X and Z matrices to be the same (i.e., there are no restrictions on the number of rows).

The distance between two graphs $\mathcal{G}_x$ and $\mathcal{G}_y$ is then defined as the *minimum* ψ*-distance* achieved by the optimal matching, that is,

$$\nu(\mathcal{G}_x, \mathcal{G}_y) \stackrel{\triangle}{=} \min_{\psi} d_\psi(\mathcal{G}_x, \mathcal{G}_y). \tag{6}$$

As shown in [10], for *undirected*, *unweighted* graphs the definition for $\nu(\cdot, \cdot)$ satisfies positive definiteness, symmetry and the triangle inequality, and therefore defines a metric. However, it is worth noting that, computing this distance requires extensive computation. For example, to compute the distance between two graphs of size n requires minimizing over a set of $n!$ bijective mappings $\psi : V_x \to V_y$.

The distance metric defined above, along with the graph implicitly defined by the weight matrix Z, provide a basis for generalizing (6) to graphs with different numbers of nodes.

III. A General Graph Distance Metric

A. Comparing graphs of different dimensions

The graph node matching approach for comparing two graphs, as defined by (5) and (6), is only applicable to graphs with the same numbers of nodes. When two graphs have differing dimensions, or numbers of nodes, we propose a generalization of the matching approach. Namely, one node in the smaller graph may be assigned to a group of nodes in the larger graph, with overlap between different groups disallowed.

We now propose a definition for the *distance* between two graphs, $\mathcal{G}_x(V_x, E_x, X)$ and $\mathcal{G}_y(V_y, E_y, Y)$, with different numbers of nodes. Since the outgoing vectors $x(i) \in \mathbb{R}^{1 \times n}$ of $\mathcal{G}_x$ and $y(j) \in \mathbb{R}^{1 \times m}$ of $\mathcal{G}_y$ are of different lengths, we compare them using a set of *composite graphs* derived from $\mathcal{G}_x$ and $\mathcal{G}_y$. Moreover, we generalize the l_1 distance measure of (5) to the more general case and evaluate the dissimilarity of two outgoing vectors by the p-norm of their difference.

Definition 1 (Composite graph set) [15] Given two graphs, $\mathcal{G}_x(V_x, E_x, X)$ with n nodes and $\mathcal{G}_y(V_y, E_y, Y)$ with m nodes ($m \leq n$), the *composite graph set* associated with $\mathcal{G}_x$ and $\mathcal{G}_y$ is defined as $\mathbf{C}_{xy} \stackrel{\triangle}{=} \{\mathcal{C}_{xy}^{\phi}(V_z, E_z, Z)\}$, such that each *composite graph* $\mathcal{C}_{xy}^{\phi}(V_z, E_z, Z) \in \mathbf{C}_{xy}$ along with a partition $\phi : V_x \to V_y$ satisfy the following conditions:

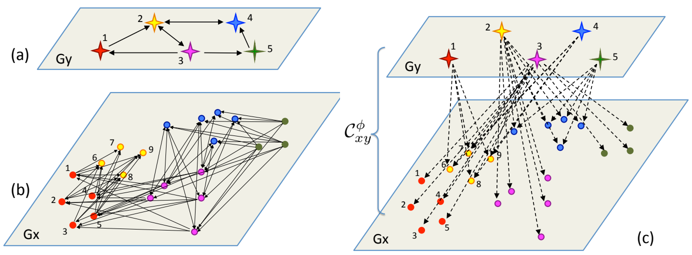

Fig. 1. The construction of one composite graph $\mathcal{C}_{xy}^{\phi}(Z) \in \mathbf{C}_{xy}$ (shown in (c)) from two graphs, $\mathcal{G}_x(V_x, E_x, X)$ (shown in (b)) and $\mathcal{G}_y(V_y, E_y, Y)$ (shown in (a)). The edges of $\mathcal{G}_x$ shown in (b) are assumed to all have unit weight; for example, the outgoing vector of the first node is $x(1) = [0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 1,\ 1,\ 1, 0, \ldots, 0]$. The "aggregated" graph $\mathcal{G}_y$ shown in (a) has aggregated outgoing vectors, e.g., $y(1) = [0,\ 4,\ 0,\ 0,\ 0]$ and $y(2) = [0,\ 0,\ 5,\ 5,\ 0]$. By Definition 1, $\mathcal{C}_{xy}^{\phi}$ contains all nodes (stars and dots) from both $\mathcal{G}_x$ and $\mathcal{G}_y$, with edges that initiate from V_y and terminate at V_x (dashed arrows). The partition function ϕ is defined from the mapping from V_x to V_y (indicated by colors). For example, for the $\mathcal{C}_{xy}^{\phi}(Z)$ shown, $\phi^{-1}(1) = \{1,\ 2,\ 3,\ 4,\ 5\}, \phi^{-1}(2) = \{6,\ 7,\ 8,\ 9\}$, are indicated by red and yellow nodes/stars, and the weighting matrix Z satisfies (7) in Definition 1 (iii), that is, $Z_{16}+Z_{17}+Z_{18}+Z_{19} = 4 = Y_{12}$ and $Z_{11}+Z_{12}+Z_{13}+Z_{14}+Z_{15} = 0 = Y_{11}$. Then we have either $\hat{z}(1) = [0, \ldots, 0,\ 1,\ 1,\ 1,\ 1]$ or $\tilde{z}(1) = [0, \ldots, 0,\ 0.5,\ 1.5,\ 1,\ 1]$ as a valid outgoing vector for $z(1)$; for this partition setting $z(1) = \hat{z}(1)$ gives a smaller value $\rho_{\phi,Z}(\mathcal{G}_x, \mathcal{G}_y)$.

(i) The node set $V_z = V_x \cup V_y$ is the union of all nodes in $\mathcal{G}_x$ and $\mathcal{G}_y$. Moreover, for notational simplicity, V_z is indexed in an order such that the first m nodes are from $\mathcal{G}_y$ and the remaining n nodes are from $\mathcal{G}_x$.

(ii) The edges in $\mathcal{C}_{xy}^{\phi}$ start from nodes in $\mathcal{G}_y$ and end at nodes in $\mathcal{G}_x$, or equivalently, $E_z \subset V_y \times V_x$. Therefore, although $\mathcal{C}_{xy}^{\phi}$ has $n+m$ nodes, we represent its weighting matrix by $Z = \left[z(1)^T\, z(2)^T \cdots z(m)^T\right]^T \in \mathbb{R}^{m \times n}$, with the outgoing vector $z(j) = [Z_{j1},\ Z_{j2}, \cdots, Z_{jn}] \in \mathbb{R}^{1 \times n}$.

(iii) The partition function $\phi : V_x \to V_y$ provides an edge weight aggregation relation between $\mathcal{C}_{xy}^{\phi}$ and $\mathcal{G}_y$:

$$Y_{jk} = \sum_{i \in \phi^{-1}(k)} Z_{ji}, \qquad j, k = 1, \cdots, m. \tag{7}$$

Using the aggregation matrix $Q(\phi)$ defined in (1), (7) can be compactly written as $Y = ZQ(\phi)$.

By construction, the outgoing vectors of any $\mathcal{C}_{xy}^{\phi} \in \mathbf{C}_{xy}$ are of the same length, n, as the outgoing vectors of $\mathcal{G}_x$ (see Figure 1 for illustration). We now adapt the concept of *ψ-distance* from (5) to obtain the *ϕ-dissimilarity* between $\mathcal{G}_x$ and $\mathcal{G}_y$,

$$\begin{aligned}\rho_{\phi,Z}(\mathcal{G}_x, \mathcal{G}_y) &\triangleq d_{\phi}(\mathcal{G}_x, \mathcal{C}_{xy}^{\phi}(V_z, E_z, Z)) \\ &= \sum_{i=1}^{n} p(v_i)\, d(x(i), z(\phi(i))),\end{aligned} \tag{8}$$

where $\{p(v_i)\}$ are node weights, and $d(\cdot,\cdot)$ takes the form of a p-norm for any $p \in [1, \infty]$. If not specified, all nodes are assumed to be equally weighted, i.e., $p(v_i) = \frac{1}{n}$ for all i. The *dissimilarity* between $\mathcal{G}_x$ and $\mathcal{G}_y$ is defined as the minimum achievable *ϕ-dissimilarity* over all partitions ϕ and weighting matrices $Z \in \mathbb{R}_{+}^{m \times n}$. Equivalently, this is the minimum over all $\mathcal{C}_{xy}^{\phi}$ in the composite graph set $\mathbf{C}_{xy}$, given by:

$$\nu(\mathcal{G}_x, \mathcal{G}_y) \triangleq \min_{\mathcal{C}_{xy}^{\phi} \in \mathbf{C}_{xy}} \rho_{\phi,Z}(\mathcal{G}_x, \mathcal{G}_y). \tag{9}$$

Similar to the findings discussed following (6), computing $\nu(\mathcal{G}_x, \mathcal{G}_y)$ requires the solution of a combinatorial optimization problem, which is NP-hard [16].

B. The distance metric $\nu(\mathcal{G}_x, \mathcal{G}_y)$

We first define graph *dilation* and *projection*.

- Given a graph $\mathcal{G}(V, E, W)$ with node weights $\{p(v_k)\}$, a *dilation* graph $\bar{\mathcal{G}}(\bar{V}, \bar{E}, \bar{W})$ of $\mathcal{G}$ is obtained by *splitting* one node, $v_k \in V$, into several nodes, $\bar{v}_{k_1}, \bar{v}_{k_2}, \ldots, \bar{v}_{k_{n_k}}$, while maintaining all given edge connections. Specifically, after splitting, the (nonnegative) node and edge weights satisfy $\sum_{i=1}^{n_k} p(\bar{v}_{k_i}) = p(v_k)$, and for $1 \le i \ne l \le n_k$, $j \ne k$, $\bar{W}_{\bar{v}_{k_i}, v_j} = W_{k,j}$, $\bar{W}_{v_j, \bar{v}_{k_i}} = \frac{p(\bar{v}_{k_i})}{p(v_k)} W_{j,k}$ and $\bar{W}_{\bar{v}_{k_i}, \bar{v}_{k_l}} = 0$. Note that $\bar{W}$ will contain n_k identical rows.
- Alternatively, if $\mathcal{G}$ contains repeated nodes, $\{v_{k_i}\}$, as indicated by identical outgoing edge vectors (or identical rows in the weight matrix W), and the node weights are $\{p(v_{k_i})\}$, then a *projection* graph $\underline{\mathcal{G}}(\underline{V}, \underline{E}, \underline{W})$ is obtained by *collapsing* or *merging* repeated nodes in $\mathcal{G}$, giving $\underline{v}_k$. Specifically, we collapse the duplicated rows (maintaining one copy) in W and get W_1. The collapse step defines a $\phi : V \to \underline{V}$. The weight matrix for the projection graph is given by $\underline{W} = W_1 Q(\phi)$, where $Q(\phi)$ is the associated aggregation matrix. The node weights are also aggregated as $p(\underline{v}_k) = \sum_i p(v_{k_i})$.

We now develop some useful properties of the graph dissimilarity (or distance) function defined in (9), and prove that this $\nu(\mathcal{G}_x, \mathcal{G}_y)$ is a pseudo-metric when d is any p-norm.

Proposition 1: The dissimilarity function $\nu(\mathcal{G}_x, \mathcal{G}_y)$ defined in (9) is nonnegative for any two graphs $\mathcal{G}_x$ and $\mathcal{G}_y$.

Proof: This is a straightforward result of (8) and (9). □

Proposition 2: The dissimilarity between isomorphic graphs is zero, that is, if $\mathcal{G}_x \simeq \mathcal{G}_y$, then $\nu(\mathcal{G}_x, \mathcal{G}_y) = 0$.

Proof: If $\mathcal{G}_x \simeq \mathcal{G}_y$, by definition $|V_x| = |V_y| = n$, and there exists a permutation $\tilde{\psi} : V_x \to V_y$, such that $(i,j) \in E_x$ if and only if $(\tilde{\psi}(i), \tilde{\psi}(j)) \in E_y$, and $X_{i,j} = Y_{\tilde{\psi}(i),\tilde{\psi}(j)}$. Let $\tilde{Q}(\tilde{\psi})$ be the aggregation matrix induced from $\tilde{\psi}$ according to (1). We have $\tilde{Q}(\tilde{\psi})^T \tilde{Q}(\tilde{\psi}) = I_n$ since $\tilde{\psi}$ is one-to-one. Select the composite graph $\mathcal{C}_{xy}^{\tilde{\psi}}(\tilde{Z})$ specified by $\tilde{\psi}$ and $\tilde{Z} \triangleq Y\tilde{Q}(\tilde{\psi})^T$. Then conditions (i), (ii) and the edge weight constraints (7) in (iii) of *Definition 1* are satisfied, so $\rho_{\tilde{\psi},\tilde{Z}}(\mathcal{G}_x, \mathcal{G}_y) = \sum_i p(v_i) d(x(i), \tilde{z}(\tilde{\psi}(i))) = 0$. To verify the equality $\nu(\mathcal{G}_x, \mathcal{G}_y) = 0$, we note that $y(i) = \tilde{z}(i)\tilde{Q}(\tilde{\psi}) = [\tilde{Z}_{i,1},\ \tilde{Z}_{i,2}, \ldots, \tilde{Z}_{i,n}][e_{\tilde{\psi}^{-1}(1)},\ e_{\tilde{\psi}^{-1}(2)}, \ldots, e_{\tilde{\psi}^{-1}(n)}] = [\tilde{Z}_{i,\tilde{\psi}^{-1}(1)},\ \tilde{Z}_{i,\tilde{\psi}^{-1}(2)}, \ldots, \tilde{Z}_{i,\tilde{\psi}^{-1}(n)}] = [Y_{i,1},\ Y_{i,2}, \ldots, Y_{i,n}]$, which implies $\tilde{Z}_{i,\tilde{\psi}^{-1}(k)} = Y_{i,k}$ for all i, k. Then $x(i) = \tilde{z}(\tilde{\psi}(i))$ for all i, since $X_{i,j} = Y_{\tilde{\psi}(i),\tilde{\psi}(j)} = \tilde{Z}_{\tilde{\psi}(i),j}$, for all i, j. □

We demonstrate *Proposition 2* via a small example.

Example 1: Consider graphs $\mathcal{G}_x$ and $\mathcal{G}_y$ depicted in Figure 2, with $V_x = \{1, 2, 3, 4\}$, $V_y = \{1', 2', 3', 4'\}$, and suppose all nodes haves the same weight, with edge weight matrices

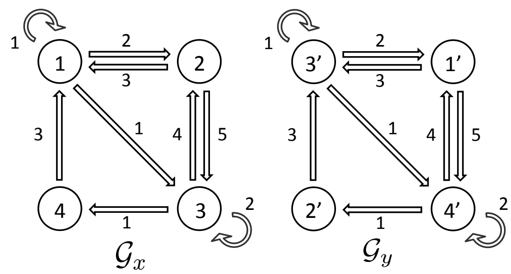

Fig. 2. $\rho(\mathcal{G}_x, \mathcal{G}_y) = 0$ if $\mathcal{G}_y \simeq \mathcal{G}_x$.

given by

$$X = \begin{bmatrix} 1 & 2 & 1 & 0 \\ 3 & 0 & 5 & 0 \\ 0 & 4 & 2 & 1 \\ 3 & 0 & 0 & 0 \end{bmatrix}, \text{ and } Y = \begin{bmatrix} 0 & 0 & 3 & 5 \\ 0 & 0 & 3 & 0 \\ 2 & 0 & 1 & 1 \\ 4 & 1 & 0 & 2 \end{bmatrix}.$$

It is easy to see that $\mathcal{G}_x \simeq \mathcal{G}_y$ through node relabeling: the permutation is given by $\tilde{\psi} : V_x \mapsto V_y$ as $\tilde{\psi}(1) = 3', \tilde{\psi}(2) = 1', \tilde{\psi}(3) = 4', \tilde{\psi}(4) = 2$. Therefore,

$$\tilde{Q}(\tilde{\psi}) = [e_2,\, e_4,\, e_1,\, e_3]\,,\; \tilde{Z} = Y\tilde{Q}(\tilde{\psi})^T = \begin{bmatrix} 3 & 0 & 5 & 0 \\ 3 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 4 & 2 & 1 \end{bmatrix}.$$

Then, $\rho_{\tilde{\psi},\tilde{Z}}(\mathcal{G}_x, \mathcal{G}_y) = \frac{1}{4}\{\|x(1) - z(3)\|_p + \|x(2) - z(1)\|_p + \|x(3) - z(4)\|_p + \|x(4) - z(2)\|_p\} = 0 \geq \nu(\mathcal{G}_x, \mathcal{G}_y) \geq 0$, and thus $\nu(\mathcal{G}_x, \mathcal{G}_y) = 0$.

Proposition 3: Given a graph $\mathcal{G}(V, E, W)$, if $\bar{\mathcal{G}}(\bar{V}, \bar{E}, \bar{W})$ and $\underline{\mathcal{G}}(\underline{V}, \underline{E}, \underline{W})$ are a *dilation* and a *projection* of $\mathcal{G}$, respectively, then $\nu(\mathcal{G}, \bar{\mathcal{G}}) = \nu(\mathcal{G}, \underline{\mathcal{G}}) = 0$.

Proof: We show $\nu(\mathcal{G}, \bar{\mathcal{G}}) = 0$ in the dilation case; the projection case is similar. Assume $|V| = n$. Since multiple node repetitions can be viewed as a sequence of single-node-repetitions, it is enough to consider the single-node-repetition case. We assume $\bar{\mathcal{G}}$ is obtained from $\mathcal{G}$ by a single-node-repetition, say $v_i \to \{\bar{v}_{i_1}, \ldots, \bar{v}_{i_k}\}$, then $|\bar{V}| = n + k - 1$. From the definition of dilation, the node weights satisfy the relation $p(v_i) = \sum_{s=1}^{k} p(\bar{v}_{i_s})$ and $p(v_j) = p(\bar{v}_j)$ for $j \neq i$.

Now we show $\nu(\mathcal{G}, \bar{\mathcal{G}}) = 0$ by defining a composite graph $\mathcal{C}^{\phi}(Z)$ with a partition function $\phi : \bar{V} \to V$ and weighting matrix $Z \in \mathbb{R}^{n \times (n+k-1)}$ that satisfies *Definition 1*, such that $\rho_{\phi,Z}(\mathcal{G}, \bar{\mathcal{G}}) = 0$. Let $\phi^{-1}(v_j) = \bar{v}_j$, if $j \neq i$, and $\phi^{-1}(v_i) = \{\bar{v}_{i_1}, \ldots, \bar{v}_{i_k}\}$, and select the rows of Z as $z(j) := \bar{w}(j)$, where $j \neq i$, and $z(i) := \bar{w}(i_1)$ (note that the $\bar{w}(i_s)$'s are identical rows for $1 \leq s \leq k$). With this choice, $\rho_{\phi,Z}(\mathcal{G}, \bar{\mathcal{G}}) = \sum_{j \neq i} p(\bar{v}_j) d(\bar{w}(j), z(j)) + \sum_{s=1}^{k} p(\bar{v}_{i_k}) d(\bar{w}(i_k), z(i)) = 0$, and hence $\nu(\mathcal{G}, \bar{\mathcal{G}}) = 0$. □

The following example illustrates the dilation/projection result.

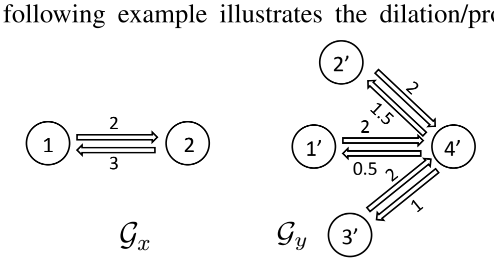

Fig. 3. Case when $\mathcal{G}_y$ is a dilation of $\mathcal{G}_x$.

Example 2: Consider graphs $\mathcal{G}_x$ and $\mathcal{G}_y$ depicted in Figure 3, with $V_x = \{1, 2\}$, $V_y = \{1', 2', 3', 4'\}$. Let $p(v_1) = p$, $p(v_2) = 1 - p$ and the weighting matrices be

$$X = \begin{bmatrix} 0 & 2 \\ 3 & 0 \end{bmatrix}, \text{ and } Y = \begin{bmatrix} 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 2 \\ 0.5 & 1.5 & 1 & 0 \end{bmatrix}.$$

Note that Y contains duplicated rows, which indicates $\{1', 2', 3'\}$ can be collapsed without inducing a positive dissimilarity. We choose $\phi : \{1', 2', 3', 4'\} \to \{1, 2\}$ as $\phi^{-1}(1) = \{1', 2', 3'\}$ and $\phi^{-1}(2) = \{4'\}$, then the corresponding $Q(\phi) = [e_1 + e_2 + e_3,\, e_4] \in \{0,1\}^{4\times 2}$. Now we construct $Z \in \mathbb{R}^{2\times 4}$ by deleting duplicated rows in Y, that is $z(1) := y(1), z(2) := y(2)$, so $ZQ(\phi) = Y$ is satisfied, and $\mathcal{C}^{\phi}_{yx}(Z) \in \mathbf{C}_{yx}$. Thus, $\rho_{\phi,Z}(\mathcal{G}_x, \mathcal{G}_y) = p(v_{1'})\|y(1) - z(1)\|_p + p(v_{2'})\|y(2) - z(1)\|_p + p(v_{3'})\|y(3) - z(1)\|_p + p(v_{4'})\|y(4) - z(2)\|_p = p(v_1)\|y(1) - z(1)\|_p + p(v_2)\|y(4) - z(2)\|_p = 0$, and this gives $\nu(\mathcal{G}_x, \mathcal{G}_y) = 0$. Alternatively, in this example $\mathcal{G}_x$ can be viewed as a projection of $\mathcal{G}_y$.

Lemma *1:* For weighted directed graphs $\mathcal{G}_x(V_x, E_x, X)$ with n nodes and $\mathcal{G}_y(V_y, E_y, Y)$ with m nodes, $\nu(\mathcal{G}_x, \mathcal{G}_y) = 0$ if and only if $\mathcal{G}_x \simeq \mathcal{G}_{y'}$ where $\mathcal{G}_{y'}$ is either $\mathcal{G}_y$ or a dilation/projection of $\mathcal{G}_y$.

Proof: The "if" part has been shown in *Proposition 2* and *Proposition 3*. For the "only if" part, assume that $n > m$. Let $\nu(\mathcal{G}_x, \mathcal{G}_y) = \min_{\mathcal{C}^{\phi}_{xy} \in \mathbf{C}_{xy}} \rho_{\phi,Z}(\mathcal{G}_x, \mathcal{G}_y) = 0$, then there exist $\tilde{\phi} : V_x \to V_y$ and $\tilde{Z}$ that define a composite graph $\mathcal{C}^{\tilde{\phi}}_{xy}$ that achieves $\rho_{\tilde{\phi},\tilde{Z}}(\mathcal{G}_x, \mathcal{G}_y) = \sum_{i=1}^{n} p(v_i) d(x(i), \tilde{z}(\tilde{\phi}(i))) = 0$. Since $p(v_i) \geq \delta > 0$, the zero dissimilarity is only achieved when $d(x(i), \tilde{z}(\tilde{\phi}(i))) = \|x(i) - \tilde{z}(\tilde{\phi}(i))\|_p = 0$ for all i. This implies $\tilde{z}(k) = x(j)$ for all $j \in \tilde{\phi}^{-1}(k)$, and $k = 1, \ldots, m$ In other words, all outgoing vectors $x(i)$ are identical (and equal to $\tilde{z}(\tilde{\phi}(i))$ within each cell of the partition. This can happen in two cases: (i) the set $\tilde{\phi}^{-1}(k)$ only contains a single element for all k, that is, all partitions are one-to-one, which implies $\mathcal{G}_x$ and $\mathcal{G}_y$ are permutations of each other, thus $\mathcal{G}_x \simeq \mathcal{G}_y$; (ii) some partitions consist of nodes with repeated outgoing vectors; this is case when the larger graph is a dilation of the smaller graph. In both cases, using the notation $\mathcal{G}_{y'}$, we have $\mathcal{G}_{y'} \simeq \mathcal{G}_x$. □

Lemma *2:* Let $\mathcal{G}_x(V_x, E_x, X)$, $\mathcal{G}_y(V_y, E_y, Y)$ and $\mathcal{G}_w(V_w, E_w, W)$ be three weighted directed graphs, then the following inequality holds for the dissimilarity function defined in (9), with $d(\cdot,\cdot)$ being the p-norm for $1 \leq p \leq \infty$:

$$\nu(\mathcal{G}_x, \mathcal{G}_y) \leq \nu(\mathcal{G}_x, \mathcal{G}_w) + \nu(\mathcal{G}_w, \mathcal{G}_y). \tag{10}$$

Proof: Assume $|V_x| = n$, $|V_w| = q$, $|V_y| = m$ and $n \geq q \geq m$, and let $\{p(x_i)\}$, $\{p(w_i)\}$ and $\{p(y_i)\}$ be the corresponding node weights. Let $\mathcal{C}^{\phi_1}_{xw}(Z_1) = \arg\min_{\mathcal{C}^{\phi}_{xw} \in \mathbf{C}_{xw}} \rho(\mathcal{G}_x, \mathcal{G}_w)$ and $\mathcal{C}^{\phi_2}_{wy}(Z_2) = \arg\min_{\mathcal{C}^{\phi}_{wy} \in \mathbf{C}_{wy}} \rho(\mathcal{G}_w, \mathcal{G}_y)$, that is, $\rho(\mathcal{G}_x, \mathcal{G}_w) = \rho_{\phi_1,Z_1}(\mathcal{G}_x, \mathcal{G}_w)$ and $\rho(\mathcal{G}_w, \mathcal{G}_y) = \rho_{\phi_2,Z_2}(\mathcal{G}_w, \mathcal{G}_y)$.

We define a partition function $\hat{\phi} : V_x \to V_z$ by $\hat{\phi}(i) = \phi_2(\phi_1(i))$ for $1 \leq i \leq n$, and a weight matrix $\hat{Z} \in \mathbb{R}^{m \times n}$ as

$$\hat{Z}_{k,l} = \gamma_{k,l}[Z_1]_{i,l} \tag{11}$$

for $1 \leq k \leq m, 1 \leq l \leq n$ and $\phi_2(i) = k$, $\gamma_{k,l} \triangleq \frac{[Z_2]_{k\phi_1(l)}}{W_{i\phi_1(l)}}$, ($\gamma_{k,l} \triangleq 0$ when both numerator and denominator are 0). The pair $\hat{\phi}$ and $\hat{Z}$ satisfies the constraint

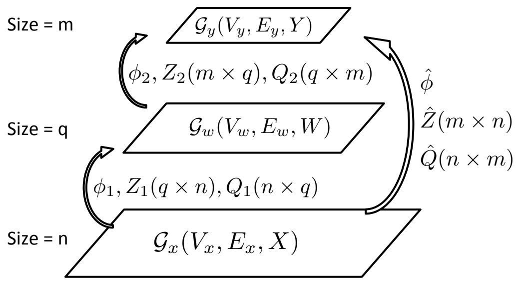

Fig. 4. $\nu(\mathcal{G}_x,\mathcal{G}_y) \le \rho(\mathcal{G}_x,\mathcal{G}_w) + \rho(\mathcal{G}_w,\mathcal{G}_y)$.

(7), since $\sum_{l\in\hat{\phi}^{-1}(j)} \hat{Z}_{k,l} = \sum_{r\in\phi_2^{-1}(j)} \left[\sum_{l\in\phi_1^{-1}(r)} \hat{Z}_{kl}\right] = \sum_{r\in\phi_2^{-1}(j)} [Z_2]_{k,r} = Y_{k,j}$, and they define a valid (although not necessarily optimal) compositie graph $\mathcal{C}^{\hat{\phi}}(\hat{Z})$. Then we have

$$
\begin{aligned}
&\nu(\mathcal{G}_x,\mathcal{G}_y) \le \rho_{\hat{\phi},\hat{Z}}(\mathcal{G}_x,\mathcal{G}_y) \qquad (12)\\
=\;& \sum_{i=1}^{n} p(x_i) d(x(i), \hat{z}(\hat{\phi}(i)))\\
\overset{(a)}{\le}\;& \sum_{i=1}^{n} p(x_i) \underbrace{\left[d(x(i), z_1(\phi_1(i))) + d(z_1(\phi_1(i)), \hat{z}(\hat{\phi}(i)))\right]}_{(I)}.
\end{aligned}
$$

in which we apply the triangle inequality for p-norms in (a). Also by assumption,

$$
\begin{aligned}
&\nu(\mathcal{G}_x,\mathcal{G}_w) + \nu(\mathcal{G}_w,\mathcal{G}_y) \qquad (13)\\
=&\sum_{i=1}^{n} p(x_i) d(x(i), z_1(\phi_1(i))) + \sum_{j=1}^{q} p(w_j) d(w(j), z_2(\phi_2(j)))\\
\overset{(b)}{=}&\sum_{i=1}^{n} p(x_i) d(x(i), z_1(\phi_1(i)))\\
&+\sum_{j=1}^{q}\left[\sum_{i\in\phi_1^{-1}(j)} p(x_i) d(w(\phi_1(i)), z_2(\phi_2\phi_1(i)))\right]\\
=&\sum_{i=1}^{n} \underbrace{p(x_i)\left[d(x(i), z_1(\phi_1(i))) + d(w(\phi_1(i)), z_2(\phi_2\phi_1(i)))\right]}_{(II)}.
\end{aligned}
$$

Equality (b) follows from the fact that the summation over the index j (representing the nodes in $\mathcal{G}_w$) can be replaced by the summation over the index i (representing the nodes in $\mathcal{G}_x$) by using the partition function $\phi_1 : V_x \to V_w$.

Let $d(\cdot,\cdot)$ be the p-norm for $1 \le p < \infty$. We also denote $j := \phi_1(i)$, $k = \phi_2(j)$ to simplify the indices. To prove (10), we need to show that the summation (I) is less than summation (II) and therefore it is sufficient to show the individual terms in (I) are less than the corresponding terms in (II). More specifically, it is sufficient to show the following inequality

$$\|z_1(j) - \hat{z}(k)\|_p^p \le \|w(j) - z_2(k)\|_p^p \qquad (14)$$

holds for each $1 \le j \le q$. Then an upper bound of $\nu(\mathcal{G}_x,\mathcal{G}_y)$ as given in (12) will be smaller than (13), which is the right hand side of (10). For a finite p, the right hand side of (14) is given by

$$
\begin{aligned}
RHS &= \sum_{h=1}^{q} |W_{j,h} - [Z_2]_{k,h}|^p\\
&\overset{(c)}{=} \sum_{h=1}^{q} \left|\sum_{l\in\phi_1^{-1}(h)} ([Z_1]_{j,l} - \hat{Z}_{k,l})\right|^p\\
&= \sum_{h=1}^{q} \left|\sum_{l\in\phi_1^{-1}(h)} (1-\gamma_{k,l})([Z_1]_{j,l})\right|^p\\
&\overset{(d)}{\ge} \sum_{h=1}^{q} \left[\sum_{l\in\phi_1^{-1}(h)} |(1-\gamma_{k,l})[Z_1]_{j,l}|^p\right]\\
&= \sum_{h=1}^{q} \left[\sum_{l\in\phi_1^{-1}(h)} |[Z_1]_{j,l} - \hat{Z}_{k,l}|^p\right] = LHS
\end{aligned}
$$

(c) uses the definition for $\hat{Z}$ in (11), and (d) holds since for all $l \in \phi_1^{-1}(h)$, $\gamma_{k,l}$ is a constant and $Z_1 \ge 0$, and thus all terms $(1-\gamma_{k,l})([Z_1]_{j,l})$ are of the same sign. The inequality follows since $|a+b|^p \ge |a|^p + |b|^p$, for $a, b \in \mathbb{R}, ab > 0$ and $1 \le p < \infty$.

When $p = \infty$, the right hand side of (14) is given by

$$
\begin{aligned}
RHS &= \max_{1\le h\le q} |W_{j,h} - [Z_2]_{k,h}|\\
&= \max_{1\le h\le q} \left|\sum_{l\in\phi_1^{-1}(h)} (1-\gamma_{k,l})[Z_1]_{j,l}\right|\\
&\overset{(e)}{\ge} \max_{1\le h\le q} \left\{\max_{l\in\phi_1^{-1}(h)} |(1-\gamma_{k,l})[Z_1]_{j,l}|\right\}\\
&= \max_{1\le h\le q} \left\{\max_{l\in\phi_1^{-1}(h)} |[Z_1]_{j,l} - \hat{Z}_{k,l}|\right\} = LHS.
\end{aligned}
$$

The inequality in (e) holds for the same reason as that in (d). We have proved (14) for $1 \le p \le \infty$, which implies $\nu(\mathcal{G}_x,\mathcal{G}_y)$ satisfies the triangle inequality when d is a p-norm.

Remark: The case for $p = \infty$ can also be seen by the equivalence of p-norms, for $1 \le p \le \infty$. On the other hand, the proof given is for case $|V_x| \ge |V_w| \ge |V_y|$, but can be easily adapted to cases when $|V_w| > |V_x|$ or $|V_w| < |V_y|$ by imposing dilation/projection. For example, when $|V_w| \ge |V_x| \ge |V_y|$, we can first dilate $\mathcal{G}_x$ to the same size as $\mathcal{G}_w$ such that $\nu(\mathcal{G}_x,\mathcal{G}_{x'}) = 0$, then $\mathcal{G}_{x'}, \mathcal{G}_w$ and $\mathcal{G}_y$ are in the same case in our proof. □

We have shown that $\nu(\mathcal{G}_x,\mathcal{G}_y) \ge 0$ for all $\mathcal{G}_x$, $\mathcal{G}_y$, and 0 is achieved if and only if $\mathcal{G}_x \simeq \mathcal{G}_{y'}$. Note that, $\nu(\mathcal{G}_x,\mathcal{G}_y)$ in (9) is defined for $|V_x| \ge |V_y|$. We extend the definition of $\nu(\mathcal{G}_x,\mathcal{G}_y)$ for $|V_x| < |V_y|$ by imposing the symmetry condition, that is $\nu(\mathcal{G}_x,\mathcal{G}_y) \triangleq \nu(\mathcal{G}_y,\mathcal{G}_x)$. We impose the symmetry $\nu(\mathcal{G}_x,\mathcal{G}_y) = \nu(\mathcal{G}_y,\mathcal{G}_x)$ by define the distance function from holds by definition: the partition function ϕ is defined on the larger node set and maps to the smaller node set. We also have demonstrated the triangle inequality holds. Therefore, we have the following result.

Theorem *1:* If $d(\cdot,\cdot)$ in (8) is p-norm $(1 \le p \le \infty)$, The dissimilarity function defined in (9) is a pseudo-metric between two directed weighted graphs.

IV. Discussion: Applications to Graph Simplification

A. Extension: when $d(\cdot,\cdot)$ is not a p-norm

We have shown that when $d(\cdot,\cdot)$ is a p-norm ($1 \leq p \leq \infty$) on $\mathbb{R}^n$, the dissimilarity function $\nu(\mathcal{G}_x, \mathcal{G}_y)$ defined in (9) satisfies the conditions for being a pseudo-metric. The flexibility provided by using p-norms fulfills a range of practical needs. For instance, in an Internet graph, the l_1 distance between two nodes quantifies the total number of uncommon hyperlinks directed to two web pages [17]; in transportation networks, the l_2 (or the Euclidean) distance measures geological distance. In these examples, the dissimilarity/distance $\nu(\mathcal{G}_x, \mathcal{G}_y)$ defined in our framework provides a metric. However, satisfying the conditions of a metric is not always necessary (nor possible) for practical problems. Nevertheless, for problems with the goal of minimizing distance/maximizing dissimilarity, alternative meaningful measures of similarity can be used as utility or cost functions.

One example is the *squared* Euclidean distance, instead of the l_2 distance, which provides three advantages: first, the squared Euclidean distance has a nice differential property, which is often preferred in algorithm design; second, the squared distance yields a larger penalty for large errors than the l_2 distance; and third, in many applications, the squared Euclidean distance is a natural choice for representing energy based costs. When $\mathcal{G}_y$ closely approximates $\mathcal{G}_x$, these distance measures yield a similar error. Another example arises in state-order reduction of Markov chains. Although the K-L divergence does not satisfy the symmetry property and is not a metric, it provides useful interpretations for problems related to probability distributions. The usage of the K-L divergence rate as a measure of distance between two Markov chains has been widely accepted [13], [18]. Therefore, although taking $d(\cdot,\cdot)$ as the K-L divergence in (5) does not provide a metric, some notion of similarity can still be inferred. In ongoing work we are investigating the case $0 < p < 1$, and other distance functions, for use in $d(\cdot,\cdot)$.

B. Connections: graph model aggregation

Note that computing the distance $\nu(\mathcal{G}_x, \mathcal{G}_y)$ for given $\mathcal{G}_x$ and $\mathcal{G}_y$ requires optimizing over a combinatorial number of partition functions. This problem becomes even more computationally intense when searching for a graph $\mathcal{G}_{y^\star}$ that minimizes $\nu(\mathcal{G}_x, \mathcal{G}_y)$ for a given $\mathcal{G}_x$. Most existing algorithms propose iterative solution candidates $\mathcal{G}_y$ and compute $\nu(\mathcal{G}_x, \mathcal{G}_y)$ at each iteration. Our formulation of composite graphs makes it possible to combine the optimization problems of computing $\nu(\mathcal{G}_x, \mathcal{G}_y)$ and minimizing over possible $\mathcal{G}_y$. This perspective gives a reinterpretation of the graph aggregation method proposed in [15], where a single optimization problem

$$\phi^\star, Z^\star = \arg\min_{\phi, Z} \rho_{\phi,Z}(\mathcal{G}_x, \mathcal{G}_y), \text{ for } \mathcal{C}^{\phi}_{xy} \in \bigcup_{\mathcal{G}_y : |V_y| = m} \mathbf{C}_{xy}. \quad (15)$$

is posed and solved, approximately. This single optimization leads to an optimal composite graph $\mathcal{C}^{\phi^\star}_{xy}(Z^\star) \in \{\bigcup_{\mathcal{G}_y : |V_y| = m} \mathbf{C}_{xy}\}$, for all $\mathcal{G}_y$. Moreover, the optimal aggregated graph $\mathcal{G}^\star_y$ may be determined from $\phi^\star$ and $Z^\star$. A detailed solution is presented in [15].

V. Conclusion

We have proposed a distance metric for comparing two general directed weighted graphs with (possibly) different dimensions, for the purpose of graph simplification via node aggregation. This distance is based on evaluating node similarities between the graphs' outgoing vectors. In particular, a composite graph set is introduced for matching a node in a smaller graph to a group of nodes in a larger graph. Computing and minimizing this distance (as the objective of aggregation) is NP-hard, thus heuristics addressing combinatorial graph aggregation problems, such as [15], are relevant.

References

[1] S. Kim, D. Putrino, S. Ghosh, and E. N. Brown, "A Granger causality measure for point process models of ensemble neural spiking activity," *PLoS Comput Biol*, vol. 7, no. 3, p. e1001110, Mar 2011.

[2] L. Yen, F. Fouss, C. Decaestecker, P. Francq, and M. Saerens, "Graph nodes clustering with the sigmoid commute-time kernel: A comparative study," *Data Knowledge Engineering*, vol. 68, no. 3, pp. 338–361, Mar. 2009.

[3] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM (JACM)*, vol. 46, no. 5, pp. 604–632, 1999.

[4] M. E. Mehran Mesbahi, *Graph Theoretic Methods in Multiagent Networks*. Princeton University Press, 2010.

[5] D. Dimarogonas, E. Frazzoli, and K. Johansson, "Distributed event-triggered control for multi-agent systems," *IEEE Transactions on Automatic Control*, vol. 57, no. 5, pp. 1291–1297, 2012.

[6] D. G. Corneil and C. C. Gotlieb, "An efficient algorithm for graph isomorphism," *Journal of the ACM*, vol. 17, no. 1, pp. 51–64, 1970.

[7] B. T. Messmer and H. Bunke, "A new algorithm for error-tolerant subgraph isomorphism detection," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 20, no. 5, pp. 493–504, 1998.

[8] M.-L. Fernández and G. Valiente, "A graph distance metric combining maximum common subgraph and minimum common supergraph," *Pattern Recognition Letters*, vol. 22, no. 6, pp. 753–758, 2001.

[9] V. D. Blondel, A. Gajardo, M. Heymans, P. Senellart, and P. Van Dooren, "A measure of similarity between graph vertices: Applications to synonym extraction and web searching," *SIAM review*, vol. 46, no. 4, pp. 647–666, 2004.

[10] G. Chartrand, G. Kubicki, and M. Schultz, "Graph similarity and distance in graphs," *Aequationes Mathematicae*, vol. 55, no. 1, pp. 129–145, 1998.

[11] Y. Keselman, A. Shokoufandeh, M. F. Demirci, and S. Dickinson, "Many-to-many graph matching via metric embedding," in *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition.*, vol. 1, 2003, pp. I–850.

[12] M. Meilă, "Comparing clusterings - an information based distance," *Journal of Multivariate Analysis*, vol. 98, no. 5, pp. 873–895, 2007.

[13] Z. Rached, F. Alajaji, and L. Campbell, "The Kullback-Leibler divergence rate between Markov sources," *IEEE Trans. on Information Theory*, vol. 50, no. 5, pp. 917–921, May 2004.

[14] M. Vidyasagar, "A metric between probability distributions on finite sets of different cardinalities and applications to order reduction," *IEEE Trans. on Automatic Control*, vol. 57, no. 10, pp. 2464–2477, Oct. 2012.

[15] Y. Xu, S. M. Salapaka, and C. L. Beck, "On reduction of graphs and markov chain models," in *50th IEEE Conference on Decision and Control and European Control Conference, 2011*. IEEE, 2011, pp. 2317–2322.

[16] D. Aloise, A. Deshpande, P. Hansen, and P. Popat, "NP-hardness of Euclidean sum-of-squares clustering," *Machine Learning*, vol. 75, no. 2, pp. 245–248, 2009.

[17] I. Dhillon, Y. Guan, and B. Kulis, "Weighted graph cuts without eigenvectors: a multilevel approach," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 11, pp. 1944 –1957, Nov. 2007.

[18] K. Deng, P. Mehta, and S. Meyn, "Optimal Kullback-Leibler aggregation via spectral theory of Markov chains," *IEEE Trans. on Automatic Control*, vol. 56, no. 12, pp. 2793–2808, Dec. 2011.