# COMMON RANDOMIZED SHORTEST PATHS (C-RSP): A SIMPLE YET EFFECTIVE FRAMEWORK FOR MULTI-VIEW GRAPH EMBEDDING

*Anuththari Gamage*[⋆] *Brian Rappaport*[†] *Shuchin Aeron*[‡] *Xiaozhe Hu*[§]

[⋆] Department of ECE, University of Illinois Urbana-Champaign, Urbana, IL
[†] School of ECE, Cornell University, Ithaca, NY
[‡] Department of ECE & [§] Department of Mathematics, Tufts University, Medford, MA

## ABSTRACT

Real-world data sets often provide several types of information about the same set of entities, showing us how they interact from different viewpoints. These data sets are well represented by multi-view graphs, which consist of multiple edge sets across the same set of nodes. Combining multiple views improves the quality of inferences drawn from the underlying data, which has led to increased interest in developing efficient multi-view graph embedding methods. We propose an algorithm, C-RSP, that generates a common (C) embedding of a multi-view graph using Randomized Shortest Paths (RSP). This algorithm generates a dissimilarity measure between nodes by minimizing the expected cost of random walks between any two nodes across all views of the graph, in doing so encoding both the local and global structure of the graph. We test C-RSP on both real and synthetic data and show that it outperforms benchmark algorithms at embedding and clustering tasks.

***Index Terms***— multi-view graphs, graph embedding, graph distances, randomized shortest paths, graph clustering

## 1. INTRODUCTION

To model and understand complex systems, we must consider how different entities within a system relate to one another. Modeling a system as a standard single-view graph is a natural first step, but many relational data sets provide more than just one view of the same underlying set of entities: examples include multiple modes of interaction through social networks [1, 2], multi-omics measurements in single cell RNA sequencing data [3, 4], and 2D projections of a single 3D object captured from multiple vantage points for 3D reconstruction [5, 6]. These systems are more accurately modeled using multi-view graphs, which contain several distinct sets of edges over the same nodes. Graph embeddings map the entities in the data set to vectors in Euclidean space, which can be used for various applications such as data visualization, clustering, and link prediction. These embeddings are normally done on single-view graphs, but this idea can also be extended to multi-view data. Generally, incorporating multiple views of data with complementary information about the nodes improves the accuracy of the embedding, increasing recent interest in exploring multi-view graph embedding algorithms.

In this paper, we propose a generalized distance on multi-view graphs called the **Common Randomized Shortest Path (C-RSP)** dissimilarity, based on the RSP dissimilarity on single-view graphs, and we introduce a novel method of embedding multi-view graphs using this new measure. We show experimentally that C-RSP encodes the overall structure of a multi-view graph more effectively than other benchmark multi-view graph algorithms, as shown through better embedding and clustering performance on both synthetic and real-world data sets.

**Relation to Existing Work**: In a multi-view graph embedding, each node of the graph is assigned a vector that incorporates data from all views of the graph. This type of embedding has been accomplished through algorithms based on techniques such as matrix factorization [7–9], tensor factorization [10, 11], and spectral embedding [12–14]. Most of these algorithms focus on clustering multi-view graphs, a specific application of the embeddings generated. High clustering accuracy is an indicator of a good embedding since relative similarity between nodes should be correctly reflected in the embedding. The similarity between nodes of a graph is usually quantified by a distance measure, such as the shortest path or geodesic distance or the commute time distance [15–17]. The commute time distance in particular is known to encode the clusters of a graph better than the shortest path distance [16, 18]. However, it has been shown that for certain graphs, the commute time distance fails to capture the global structure of the graph accurately [19]. Since these measures are ill-suited for capturing both the local and global structure of a graph simultaneously, there has been increased interest in alternative distance measures that generalize these two [20–22].

The Randomized Shortest Path (RSP) dissimilarity [21] generalizes the two distances by computing an intermediate measure defined by a tunable parameter $\beta$. This measure, which is based on random walks on a graph, is particularly suitable for graph embedding as it preserves in the embedding space both the local and global features of the manifold from which the data set is sampled. In this paper we extend this distance measure for multi-view graphs as a simple but highly effective algorithm for multi-view graph embedding.

## 2. C-RSP DISSIMILARITY

**Mathematical Preliminaries**: Let $G = \{V, E\}$ be a simply connected single-view graph, where $V = \{1, \dots, n\}$ denotes the set of nodes of the graph and $E = \{(i, j) \mid i, j \in V\}$ denotes the set of edges between nodes. This graph can be represented by its affinity matrix $A \in \mathbb{R}^{n \times n}$, where $a_{ij} \neq 0$ for $(i, j) \in E$, $a_{ij} = 0$ if $(i, j) \notin E$. The degree matrix $D \in \mathbb{R}^{n \times n}$ is the diagonal matrix containing the degrees of the nodes on the diagonals.

We can compute the transition probability matrix $P^{ref} = D^{-1}A$ of the graph $G$, which is row-stochastic and defines a probability distribution on the edges of the graph. A random walk on

This work was supported by the US National Science Foundation (NSF) grants CCF-1319653 and CCF-1553075. The work of X. Hu was supported in part by National Science Foundation grant DMS-1812503. E-mails: ⋆gamaral2@illinois.edu, †br434@cornell.edu, ‡shuchin@ece.tufts.edu, §xiaozhe.hu@tufts.edu.

the graph follows a sequence of nodes with the order determined by these transition probabilities. Consider a particular path on this graph starting at a source node $s$ and a destination node $t$, denoted by $p_{s \to t} = \{s, v_1, v_2, \ldots, v_{t-1}t\}$. Then the probability of the path is given by the product $P^{ref}_{s,v_1} P^{ref}_{v_1,v_2} \ldots P^{ref}_{v_m,t}$, denoted $P^{ref}(p_{s \to t})$. We define the cost of each edge by a cost matrix $C$ with elements $c_{ij} = a_{ij}^{-1}$ where $c_{ij} > 0$. The total cost for path $p_{s \to t}$ is given by $C(p_{s \to t}) = c_{s,v_1} + c_{v_1,v_2} + \ldots + c_{v_{t-1},t}$.

An absorbing path is a path where the destination node $t$ has no outgoing edges except to itself ($c_{t,t} = 1, c_{t,k} = \infty$ for $t \neq k \in V$). Let the set of all such absorbing paths be $\mathcal{P}_{s \to t}$. For C-RSP, we consider only absorbing paths from $s$ to $t$, and our path is denoted $p_{s \to t}$, with the probability of the path under a given probability distribution $P$ denoted by $P(p_{s \to t})$ and the cost of traversing the path denoted by $C(p_{s \to t})$. The expected cost of a random walk from source node $s$ to destination node $t$ over distribution $P$ is $\sum\limits_{p \in \mathcal{P}_{s \to t}} P(p)C(p)$.

**Randomized Shortest Paths Dissimilarity**: The Randomized Shortest Path (RSP) is defined to be the path between two nodes with the minimum expected cost over all transition probability matrices [21]. In order to constrain a random walk between two nodes to an RSP, we compute a new probability distribution $P^{RSP}(p)$ which minimizes the expected cost among all possible probability distributions while maintaining fixed relative entropy or Küllback-Leibler divergence with respect to the reference distribution $P^{ref}(p)$:

$$\begin{aligned} P^{RSP} &= \underset{P}{\operatorname{argmin}} \sum_{p \in \mathcal{P}_{s \to t}} P(p)C(p) \\ \text{subject to} & \sum_{p \in \mathcal{P}_{s \to t}} P(p) \ln \frac{P(p)}{P^{ref}(p)} = J_0, \\ & \sum_{p \in \mathcal{P}_{s \to t}} P(p) = 1 \end{aligned} \tag{1}$$

where $J_0 \in \mathbb{R}$ is the relative entropy between the distributions.

The solution to this constrained optimization is given by the following expression for any $p_{s \to t} \in \mathcal{P}_{s \to t}$ [21]:

$$P^{RSP}(p_{s \to t}) = \frac{P^{ref}(p_{s \to t}) e^{-\beta C(p_{s \to t})}}{\sum\limits_{p \in \mathcal{P}_{s \to t}} P^{ref}(p) e^{-\beta C(p)}} \tag{2}$$

where as $\beta \to \infty$, the RSP dissimilarity reduces to the shortest path distance and as $\beta \to 0$, it reduces to the commute time distance.

The RSP dissimilarity between nodes $s$ and $t$, denoted $\Delta^{RSP}_{st}$, is then calculated as follows:

$$\begin{aligned} \overline{C}_{st} &= \sum_{p \in \mathcal{P}_{s \to t}} P^{RSP}(p)C(p) \\ \Delta^{RSP}_{st} &= \frac{\overline{C}_{st} + \overline{C}_{ts}}{2}. \end{aligned} \tag{3}$$

Note that the computed RSP dissimilarity measure is termed a "measure" instead of a "metric" since it does not follow the triangle inequality for certain ranges of $\beta$ used [21, 22]. Following the derivation of the RSP dissimilarity by Yen et al. [21], an efficient closed-form expression for its computation was found by Kivimäki et al. [22]. The result of this algorithm was the symmetric matrix $\Delta^{RSP} \in \mathbb{R}^{n \times n}$, consisting of all pairwise RSP dissimilarities.

**Deriving a Common RSP Probability Distribution**: Here, we extend the core RSP framework to generate a multi-view graph distance measure. If we represent a single-view graph by $G = \{V, E\}$, then a multi-view graph is denoted $\mathcal{G} = \{V, (E_1, \ldots, E_m)\}$ where each view is given by $G_i = \{V, E_i\}$. We represent this graph with an $n \times n \times m$ affinity tensor, where each $n \times n$ slice of the tensor $A_i$ represents the affinity matrix for that edge set. Note that each $G_i$ is assumed to be a simply connected graph.

We first derive a common probability distribution, $P^{CRSP}$, over all views of the graph. This is accomplished by minimizing the expected cost for all possible paths on all views, with the condition that the common distribution $P^{CRSP}$ and the reference probability distribution of each view, $P^{ref}_i$, have the same fixed relative entropy. This constrained optimization is represented as follows, with reference probability distributions $P^{ref}_1, \ldots, P^{ref}_m$ and cost matrices $C_1, \ldots, C_m$:

$$\begin{aligned} P^{CRSP} &= \underset{P}{\operatorname{argmin}} \sum_{i=1}^{m} \sum_{p \in \mathcal{P}_{s \to t}} P(p)C_i(p) \\ \text{subject to} & \sum_{i=1}^{m} \sum_{p \in \mathcal{P}_{s \to t}} P(p) \ln \frac{P(p)}{P^{ref}_i(p)} = J_0, \\ & \sum_{p \in \mathcal{P}_{s \to t}} P(p) = 1 \end{aligned} \tag{4}$$

To derive the common distribution (which we will call $P$ for ease of notation) under the given constraints, we use the following Lagrange function:

$$\begin{aligned} \mathcal{L} = &\sum_{i=1}^{m} \sum_{p \in \mathcal{P}_{s \to t}} P(p)C_i(p) + \mu \left[ \sum_{p \in \mathcal{P}_{s \to t}} P(p) - 1 \right] \\ &+ \lambda \left[ \sum_{p \in \mathcal{P}_{s \to t}} P(p) \ln \frac{P(p)}{P^{ref}_i(p)} - J_0 \right] \end{aligned}$$

Considering only one path $p$, we obtain the following:

$$\frac{\partial \mathcal{L}}{\partial P(p)} = \sum_{i=i}^{m} C_i(p) + \lambda \ln \frac{P^m(p)}{\prod_{i=1}^{m} P^{ref}_i(p)} + \lambda m + \mu = 0$$

or equivalently

$$\ln \frac{P^m(p)}{\prod_{i=1}^{m} P^{ref}_i(p)} = -\frac{1}{\lambda} \sum_{i=1}^{m} C_i(p) - \frac{\mu}{\lambda} - m$$

which gives

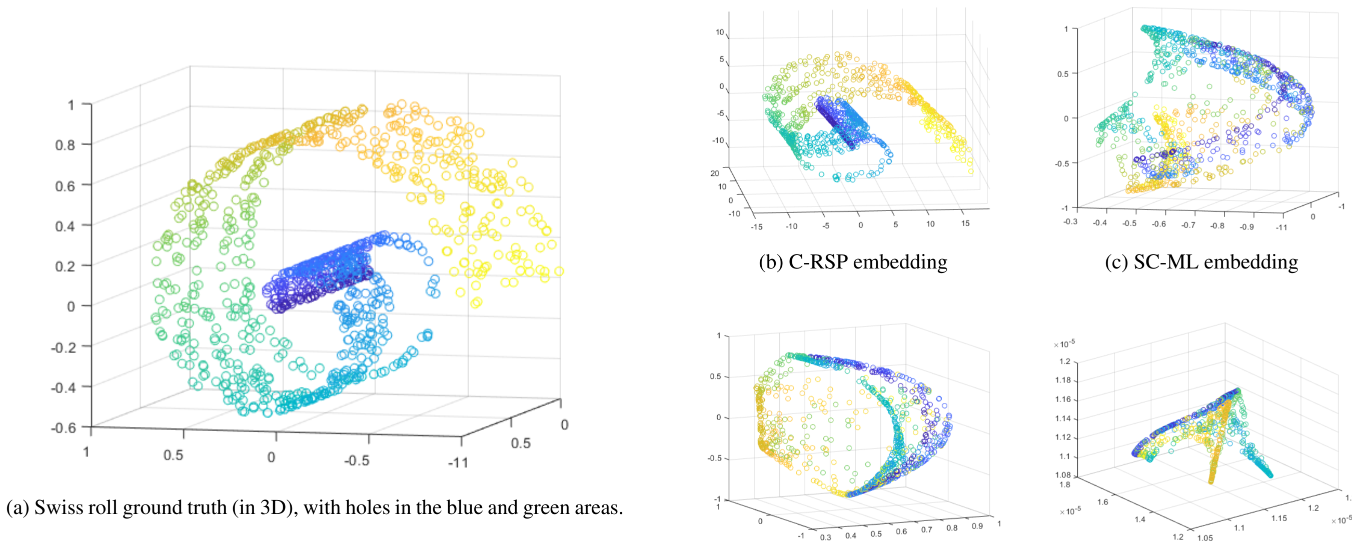
$$\begin{aligned} P(p) &= \sqrt[m]{\prod_{i=1}^{m} P^{ref}_i(p)} \cdot e^{-\frac{1}{m\lambda} \sum\limits_{i=1}^{m} C_i(p) - \frac{\mu}{m\lambda} - 1} \\ &= \overline{\mathbf{P}} \cdot e^{-\beta \overline{\mathbf{C}}} e^{-\frac{\beta\mu}{m} - 1} \end{aligned}$$

where $\overline{\mathbf{P}}(p) = \sqrt[m]{\prod_{i=1}^{m} P^{ref}_i(p)}$ and $\overline{\mathbf{C}}(p) = \frac{1}{m} \sum_{i=1}^{m} C_i(p)$, the geometric and arithmetic means (taken element-wise).

Normalizing this to make it a probability distribution (which is the same as RSP, detailed in [22]), we obtain the following expression for the C-RSP probability distribution for a single path:

$$P^{CRSP}(p_{s \to t}) = \frac{\overline{\mathbf{P}}(p_{s \to t}) \cdot e^{-\beta \overline{\mathbf{C}}(p_{s \to t})}}{\sum\limits_{p \in \mathcal{P}_{s \to t}} \overline{\mathbf{P}}(p) \cdot e^{-\beta \overline{\mathbf{C}}(p)}} \tag{5}$$

With this combined probability distribution, we can use the same algorithm for calculating the C-RSP dissimilarity matrix $\Delta^{CRSP}$ as in the RSP case, which this reduces to in the single layer case. To obtain a multi-view graph embedding from the C-RSP dissimilarity, we use Multi-dimensional Scaling on $\Delta^{CRSP}$. For clustering, we use Spectral Clustering [18] on $(\Delta^{CRSP})^{-1}$.



(a) Swiss roll ground truth (in 3D), with holes in the blue and green areas.

(b) C-RSP embedding

(c) SC-ML embedding

(d) CSC embedding

(e) MultiNMF embedding

**Fig. 1**. Embeddings of the Swiss roll generated by C-RSP and other benchmark algorithms. C-RSP retains the Swiss roll shape as well as the relative distances between nodes. SC-ML and CSC both use spectral embedding methods which preserve the relative distances of the nodes but lose the overall structure, while MultiNMF uses a nonnegative matrix factorization to obtain the embedding vectors.

## 3. EXPERIMENTAL RESULTS

In order to evaluate the quality of the embedding produced by C-RSP, we first tested it on a standard Swiss roll data set, a canonical example used in manifold learning. Lacking a way of quantifying the embedding accuracy, we visually compared it to the embeddings generated by other multi-view algorithms. We also tested the performance of our algorithm quantitatively at clustering tasks against a number of benchmark multi-view graph clustering algorithms: SC-ML [12], CSC [13], and MultiNMF [9]. As metrics, we used the Normalized Mutual Information (NMI) between the labels generated by each algorithm and the ground truth, as well as the correct classification rate (CCR). We used one synthetic data set generated using the Stochastic Block Model (SBM) [23, 24] and three real-world data sets commonly used for multi-view graph clustering tasks: UCI Handwritten Digits[1], 3Sources[2], and Multi-view Twitter data[3]. Our code is available online at https://github.com/Anu-Gamage/C-RSP.

**Results on Synthetic Data**: We first tested the quality of the embeddings generated by C-RSP using the Swiss roll data set in figure 1. To obtain a multi-view graph from this data, the Swiss roll was projected into two dimensions at various angles. An affinity matrix for each view was constructed using the pairwise Euclidean distances between points. In the case of C-RSP, the embeddings were generated from the output C-RSP dissimilarity matrix using Classical Multidimensional Scaling. For the benchmark algorithms, the embedding vectors generated via each algorithm prior to the clustering step were used to visualize the Swiss roll embedding.

As seen in figure 1b, the C-RSP embedding accurately captures the curvature of the Swiss roll and produces a slightly flattened version of the original spiral structure. The embedding also retains the two holes present in the original Swiss roll data. The benchmark algorithms SC-ML, CSC, and MultiNMF all fail to recover the spiral structure and the holes, but retain the relative distances between nodes with some accuracy. This is shown by the grouping of similarly colored nodes in figures 1c-1e.

Next, we tested C-RSP on synthetic multi-view graphs generated using the Stochastic Block Model [23, 24], which is used to simulate graphs with a latent cluster structure. To generate a graph under this model, the nodes are partitioned into $k$ equally sized clusters. Intra-cluster edges are assigned with a probability of $\frac{c}{n}$, to give an average degree of $c$ within the cluster; and inter-cluster edges are assigned with probability $\frac{c(1-\lambda)}{n}$, where $\lambda$ is a parameter used to determine how distinct the clusters should be: $\lambda = 0.9$ results in almost disjoint clusters. To simulate multi-view graphs, we generate $m$ independent SBM graphs with the same set of parameters $n, k, c$, and $\lambda$, and the same partition of nodes for each view. Since this process does not necessarily produce connected graphs, we cull nodes that are not connected in all of the views. In figure 2, we report the variation in NMI as the number of views increases, as well as the variation across multi-view graphs of different sparsities. In each case, C-RSP equals or exceeds the clustering accuracy of the benchmark algorithms. We note that Multi-NMF likely fares the worst on this data set as the graphs are in binary format rather than real-valued data matrices.

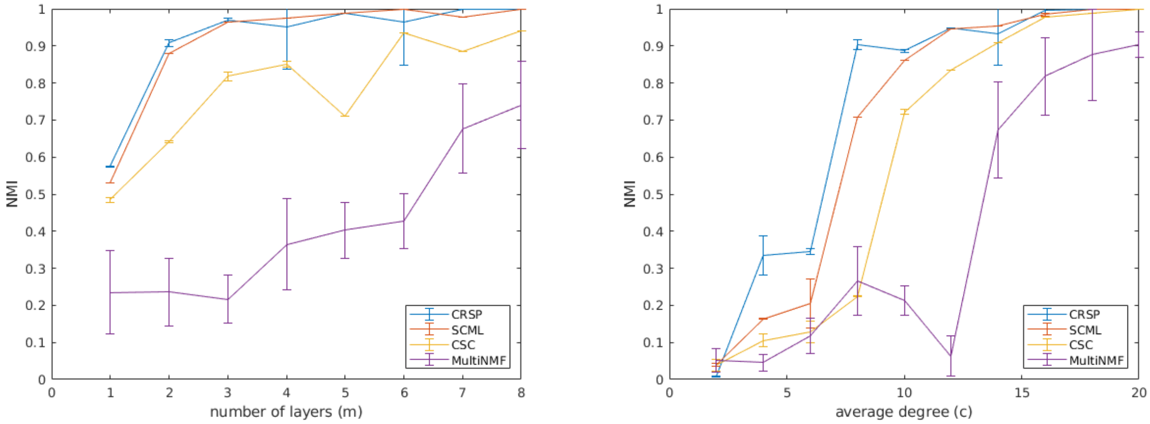
**Results on Real-world Data**: We ran C-RSP and the benchmark algorithms on three widely used real multi-view data sets. In Table 1, we report the NMI and CCR values for each algorithm.

The UCI data set analyzes 2000 images of handwritten digits 0-9, giving feature matrices of Fourier coefficients, pixel averages, and several other aspects of the images. In our experiments, we used 5 of 6 provided feature types as our views and used a Gaussian kernel to construct affinity matrices: the affinity $a_{ij}$ between nodes $i, j$ with feature vectors $x_i, x_j$ is $a_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\tilde{\sigma}^2}\right)$, with $\tilde{\sigma}$ the

[1] https://archive.ics.uci.edu/ml/datasets/Multiple+Features

[2] http://mlg.ucd.ie/datasets/3sources.htm

[3] http://mlg.ucd.ie/aggregation/index.html




(a) NMI, varying number of views, $c = 10$

(b) NMI, varying graph sparsity, $m = 2$

**Fig. 2**. Simulation results on Stochastic Block Model graphs with $n = 500, c = 10, k = 3, m = 2, \lambda = 0.9$ unless otherwise specified, averaged over 10 runs. As the number of views and average degree increases, all algorithms increase in accuracy, with C-RSP outperforming all benchmark algorithms.

| Metric | Data Set | C-RSP | SC-ML | CSC | MultiNMF |
|---|---|---|---|---|---|
| CCR | **UCI** | 86.96% (4.10%) | 80.09 (6.76) | 82.11 (7.99) | 92.33 (0.03) |
| | **3Sources** | 58.22 (4.14) | 51.48 (3.55) | 43.55 (2.59) | 34.50 (0.02) |
| | **MultiviewTwitter** | 82.84 (3.96) | 69.86 (1.50) | 49.54 (3.19) | 56.67 (0.01) |
| NMI | **UCI** | 0.80 (0.01) | 0.76 (0.03) | 0.78 (0.04) | 0.88 (2e-4) |
| | **3Sources** | 0.56 (0.04) | 0.42 (0.02) | 0.31 (0.02) | 0.07 (1e-4) |
| | **MultiviewTwitter** | 0.60 (0.04) | 0.42 (0.01) | 0.28 (0.01) | 0.45 (1e-4) |

**Table 1**. Clustering results on various real-world data sets over 10 runs, reporting the mean and the standard deviation in parentheses. CCR is the correct classification rate and NMI is the normalized mutual information (in the range $[0, 1]$).

median pairwise Euclidean distance between all feature vectors.

The 3Sources data set contains information about a set of news articles reported by three different news sources: the BBC, the Guardian, and Reuters. It covers 416 distinct news stories, of which 169 are reported on by all three agencies. These stories are classified under 6 disjoint clusters: business, entertainment, health, politics, sport, and technology. The three sources provide different views of the same news story, which can be represented as different views of a multi-view graph. We extracted the common stories and constructed an affinity matrix for each source using the same Gaussian kernel on the feature vectors provided.

The Multiview Twitter data set consists of five Twitter user networks and records of their methods of interaction on Twitter. We chose the *politics-uk* data set, 419 user accounts belonging to UK political figures and organizations and 3 views of their interactions: follows, mentions, and retweets. The user accounts form the nodes of the graph, partitioned into 5 disjoint clusters based on political party affiliation: Labour, Conservative, Scottish National Party, Liberal Democrats, or other.

C-RSP significantly outperforms the benchmark algorithms on the both the Multiview Twitter data set and the 3Sources data set with respect to both CCR and NMI. On the UCI Handwritten Digits data set, MultiNMF outperforms C-RSP by a decent margin, but C-RSP still outperforms the other benchmarks in both metrics. The dip in performance of C-RSP is likely due to the choice of the tuning parameter $\beta$: in all experiments we chose $\beta = 0.02$ for C-RSP since this was shown to be the optimal $\beta$ value for RSP [22]. At this tuning, the C-RSP dissimilarities tend towards the commute time distances, which may capture the graph structure less effectively for graphs with more than 1000 nodes [19], such as the UCI data set. Overall, C-RSP provides superior clustering results, confirming that good multi-view graph embeddings improves clustering accuracy. Furthermore, C-RSP is highly robust, with relatively high accuracy on many different types of data.

## 4. CONCLUSION

This paper introduced a novel distance measure for multi-view graphs named C-RSP (Common Randomized Shortest Paths), an extension of the RSP dissimilarity for single-view graphs. The C-RSP measure is a generalization of the commute time distance and the shortest path distance, which allows it to encode both local and global structure of a multi-view graph. This leads to more accurate graph embeddings, which can be used to improve performance in downstream applications such as data visualization, clustering, and link prediction. We tested C-RSP at both embedding and clustering tasks and showed that it produces superior results compared to other benchmark embedding algorithms.

## 5. REFERENCES


[1] Derek Greene and Pádraig Cunningham, "Producing a unified graph representation from multiple social network views," in *Proceedings of the 5th Annual ACM Web Science Conference*, New York, NY, USA, 2013, WebSci '13, pp. 118–121, ACM.

[2] Yu Shi, Myunghwan Kim, Shaunak Chatterjee, Mitul Tiwari, Souvik Ghosh, and Rómer Rosales, "Dynamics of large multi-view social networks: Synergy, cannibalization and cross-view interplay," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2016, KDD 2016, pp. 1855–1864, ACM.

[3] Yifeng Li, Fang-Xiang Wu, and Alioune Ngom, "A review on machine learning principles for multi-view biological data integration," *Briefings in Bioinformatics*, vol. 19, no. 2, pp. 325–340, 2018.

[4] Maria Colomé-Tatché and Fabian J. Theis, "Statistical single cell multi-omics integration," *Current Opinion in Systems Biology*, vol. 7, pp. 54 – 59, 2018.

[5] Hang Su, Subhransu Maji, Evangelos Kalogerakis, and Eric Learned-Miller, "Multi-view convolutional neural networks for 3D shape recognition," in *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015, pp. 945–953.

[6] Kalin Kolev, Maria Klodt, Thomas Brox, and Daniel Cremers, "Continuous global optimization in multiview 3D reconstruction," *International Journal of Computer Vision*, vol. 84, no. 1, pp. 80–96, Aug. 2009.

[7] Wei Tang, Zhengdong Lu, and Inderjit S. Dhillon, "Clustering with multiple graphs," in *Proceedings of the 2009 Ninth IEEE International Conference on Data Mining*, Washington, DC, USA, 2009, ICDM '09, pp. 1016–1021, IEEE Computer Society.

[8] Subhadeep Paul and Yuguo Chen, "Consistent community detection in multi-relational data through restricted multi-layer stochastic blockmodel," *Electronic Journal of Statistics*, vol. 10, no. 2, pp. 3807–3870, 2016.

[9] Jialu Liu, Chi Wang, Jing Gao, and Jiawei Han, "Multi-view clustering via joint nonnegative matrix factorization," in *Proceedings of the 2013 SIAM International Conference on Data Mining*. SIAM, 2013, pp. 252–260.

[10] Xinhai Liu, Wolfgang Glänzel, and Bart De Moor, "Hybrid clustering of multi-view data via Tucker-2 model and its application," *Scientometrics*, vol. 88, no. 3, pp. 819–839, Sept. 2011.

[11] Evangelos E. Papalexakis, Leman Akoglu, and Dino Ience, "Do more views of a graph help? Community detection and clustering in multi-graphs," in *Proceedings of the 16th International Conference on Information Fusion*, July 2013, pp. 899–905.

[12] Xiaowen Dong, Pascal Frossard, Pierre Vandergheynst, and Nikolai Nefedov, "Clustering on multi-layer graphs via subspace analysis on Grassmann manifolds," *IEEE Transactions on Signal Processing*, vol. 62, no. 4, pp. 905–918, Feb 2014.

[13] Abhishek Kumar, Piyush Rai, and Hal Daumé, III, "Co-regularized multi-view spectral clustering," in *Proceedings of the 24th International Conference on Neural Information Processing Systems*, USA, 2011, NIPS'11, pp. 1413–1421, Curran Associates Inc.

[14] Abhishek Kumar and Hal Daumé III, "A co-training approach for multi-view spectral clustering," in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, USA, 2011, ICML'11, pp. 393–400, Omnipress.

[15] Pavel Chebotarev, "The walk distances in graphs," *Discrete Applied Mathematics*, vol. 160, no. 10, pp. 1484 – 1500, 2012.

[16] Ulrike Von Luxburg, Agnes Radl, and Matthias Hein, "Hitting and commute times in large random neighborhood graphs," *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1751–1798, Jan. 2014.

[17] László Lovász, "Random walks on graphs: A survey," *Combinatorics, Paul Erdos is Eighty*, vol. 2, no. 1, pp. 1–46, 1993.

[18] Ulrike Von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, no. 4, pp. 395–416, Dec. 2007.

[19] Ulrike Von Luxburg, Agnes Radl, and Matthias Hein, "Hitting and commute times in large graphs are often misleading," *ArXiv e-prints*, Mar. 2010.

[20] Tatsunori B. Hashimoto, Yi Sun, and Tommi S. Jaakkola, "From random walks to distances on unweighted graphs," in *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, Cambridge, MA, USA, 2015, NIPS'15, pp. 3429–3437, MIT Press.

[21] Luh Yen, Marco Saerens, Amin Mantrach, and Masashi Shimbo, "A family of dissimilarity measures between nodes generalizing both the shortest-path and the commute-time distances," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, New York, NY, USA, 2008, KDD '08, pp. 785–793, ACM.

[22] Ilkka Kivimäki, Masashi Shimbo, and Marco Saerens, "Developments in the theory of randomized shortest paths with a comparison of graph node distances," *Physica A: Statistical Mechanics and its Applications*, vol. 393, pp. 600–616, 2014.

[23] Paul Holland, Kathryn Laskey, and Samuel Leinhardt, "Stochastic blockmodels: First steps," *Social Networks*, vol. 5, pp. 109–137, 06 1983.

[24] Emmanuel Abbe and Colin Sandon, "Community detection in general stochastic block models: Fundamental limits and efficient algorithms for recovery," in *Proceedings of the 2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, Washington, DC, USA, 2015, FOCS '15, pp. 670–688, IEEE Computer Society.