

RESEARCH

Open Access



# Nonlinear expression and visualization of nonmetric relationships in genetic diseases and microbiome data

Xianchao Zhu<sup>1</sup>, Xianjun Shen<sup>1\*</sup>, Xingpeng Jiang<sup>1</sup>, Kaiping Wei<sup>1</sup>, Tingting He<sup>1</sup>, Yuanyuan Ma<sup>1</sup>, Jiaqi Liu<sup>1</sup> and Xiaohua Hu<sup>1,2</sup>

From IEEE International Conference on Bioinformatics and Biomedicine 2017  
Kansas City, MO, USA. 13-16 November 2017

## Abstract

**Background:** The traditional methods of visualizing high-dimensional data objects in low-dimensional metric spaces are subject to the basic limitations of metric space. These limitations result in multidimensional scaling that fails to faithfully represent non-metric similarity data.

**Results:** Multiple maps t-SNE (mm-tSNE) has drawn much attention due to the construction of multiple mappings in low-dimensional space to visualize the non-metric pairwise similarity to eliminate the limitations of a single metric map. mm-tSNE regularization combines the intrinsic geometry between data points in a high-dimensional space. The weight of data points on each map is used as the regularization parameter of the manifold, so the weights of similar data points on the same map are also as close as possible. However, these methods use standard momentum methods to calculate parameters of gradient at each iteration, which may lead to erroneous gradient search directions so that the target loss function fails to achieve a better local minimum. In this article, we use a Nesterov momentum method to learn the target loss function and correct each gradient update by looking back at the previous gradient in the candidate search direction. By using indirect second-order information, the algorithm obtains faster convergence than the original algorithm. To further evaluate our approach from a comparative perspective, we conducted experiments on several datasets including social network data, phenotype similarity data, and microbiomic data.

**Conclusions:** The experimental results show that the proposed method achieves better results than several versions of mm-tSNE based on three evaluation indicators including the neighborhood preservation ratio (NPR), error rate and time complexity.

**Keywords:** Multiple maps t-SNE, Data visualization, Non-metric similarities, Nesterov momentum

## Background

A large number of studies have shown that genetic diseases with overlapping phenotypes are closely related to function-related gene mutations [1, 2]. From another perspective, there are similar pathophysiological mechanisms between different clinical features and genetic diseases [3, 4]. In addition, classical methods of

dimensionality reduction and visualization of data have been applied to the analysis of microbial data [5]. However, generally speaking, the integration and analysis of microbiome big data are still in its preliminary stage. There are currently no effective integration techniques and visualization methods to exploit microbiome big data. Some studies have focused on established mathematical models that exploit the complicated correlations between phenotypes and genotypes in isomeric genomic datasets such as genetic expression data, gene ontology annotations [6], and protein-protein interaction

\* Correspondence: [xjshen@mail.ccnu.edu.cn](mailto:xjshen@mail.ccnu.edu.cn)

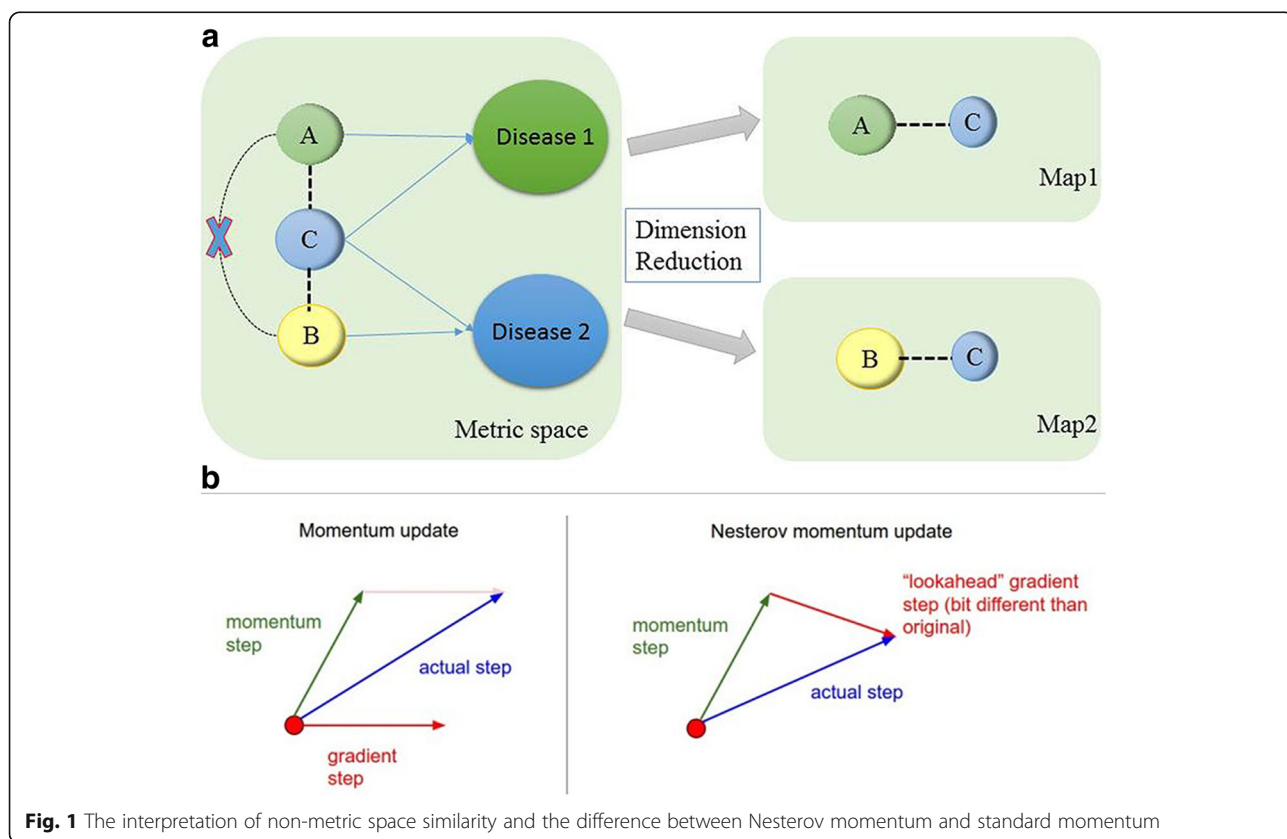
<sup>1</sup>School of Computer, Central China Normal University, Wuhan, China  
Full list of author information is available at the end of the article



networks [7, 8]. In addition, some studies prove that non-metric attributes are important features of microbial data [9]. Researching the associations between diseases not only helps us to discover their mutual hereditary basis [10], but also provides us new insights into the molecular circadian mechanisms [11] and prospective drug target studies [12]. Each person's gut microbiota has a dominant flora in the intestine and can be divided into three different "intestinal types" based on the characteristics of the human intestine. This finding can help us discover the relationship between drugs, diet, microbes and the body in different states of health and disease [13]. These microbes distributed in different parts of the body play a vital role in our health. Lowering the dimensions of data and extracting useful information from data in the analysis of microbiome big data, with the help of statistics and pattern recognition, the structure and characteristics of the microbial community could be analyzed; new biological hypothesis could be proposed and examined.

Before performing computational tasks on a large amount of data, to conduct preliminary visualization and exploration at first will help us understand this task intuitively. By visualizing the relationships between disease phenotypes, we may gain new insights into the relationships between genes and disease. The conventional

method of dimensionality reduction visualizes high-dimensional space objects into two-dimensional or three-dimensional metric space by constructing a single map in low-dimensional space [14]. However, this visualization method suffers from the basic limitations of the metric space. The main limitation of metric space comes from the triangular inequality criterion. For example, from a biological point of view, if phenotype A is associated with phenotype B in the metric space and phenotype B is associated with phenotype C, logically, phenotype A should be associated with phenotype C. As a matter of fact, this restriction is most likely to be ruined by the implicit structure of similarity data. Because these diseases may be interrelated in different categories, they may have overlapping phenotypes in which a cluster of phenotypes may belong to disparate illness categories. The mm-tSNE [15] can properly model non-transitive similarities by assign a significance weight to each point in disparate maps. For example, we imbed three instance phenotypes A, B, and C into two maps in low dimensional space (see Fig.1 (a)), mm-tSNE assigns a significance weight of 1 to the phenotype A on the first map, assign an importance weight 1 for the phenotype B in the second map and assign to the phenotype C a significance weight in both maps is 0.5. As a result, the pairs of similarities between phenotype A and B is 0.



**Fig. 1** The interpretation of non-metric space similarity and the difference between Nesterov momentum and standard momentum

The mm-tSNE approach breaks down the nature of metric-space transitivity similarities by visualizing data points into multiple maps [15]. Nevertheless, mm-tSNE may have some drawbacks, that is, the data points with high significance weights in the uniform map do not accord with the uniform cluster structure. That adds to the difficulty of explaining the implication of each and every map. The mm-tSNE regularization [16] improves the mm-tSNE by introducing the Laplacian penalty term in the target loss function. The Laplacian penalty term has been widely applied to many machine learning models [17, 18]. Compared with mm-tSNE, a preponderance of mm-tSNE regularization is that it adopts clustering structure of variate and offers more sparsity for parameter estimation. These methods use standard momentum updates [19] to evaluate point of the gradient at each iteration. But sometimes the gradient of the previous update is wrong, it would make the current update jump high, which leads to excessive oscillation. This article is an extended version of the mm-tSNE regularization based on NAG from an earlier conference publication [20]. In contrast to these previous papers, this article: (1) contains more detailed technical and experimental descriptions; and (2) includes additional experimental results on some microbial datasets. In this article, we use a Nesterov momentum method [21, 22] to learn the target loss function and correct each gradient update by looking back at the previous gradient in the candidate search direction. The key difference between standard momentum and Nesterov momentum is that standard momentum calculates the gradient before the velocity is applied, while Nesterov momentum calculates the gradient after doing so. Therefore, the calibration gradient can be corrected faster and more accurately. This benign-looking difference seems to allow Nesterov momentum to change velocity in a quicker and more responsive way, letting it behave more stable than momentum in many situations, especially for higher values of momentum coefficient. By indirectly using the information of the second order, the Nesterov momentum method achieves a better convergence rate than the momentum method and further reduces the error rate of the loss function. The results of the present study indicate that the proposed method can obtain comparable performance compared with the original methods and provide a better data visualization framework.

**Methods**

**T-distributed stochastic neighborhood embedding (t-SNE)**

t-Distributed Stochastic Neighborhood Embedding (t-SNE) is a classical multi-dimensional scaling technique [23] It is a non-linear mapping method based on the early work of Stochastic Neighbor Embedding [24].

As data points are mapped from high-dimensional space to low-dimensional space, the distances between data points are maintained and local information and global information are preserved. This method has been applied to the visualization of data in many fields such as literature [25], linguistic data [26], and breast cancer CADx imaging data [27]. In t-SNE, the similarities amongst data points are modeled by probability metrics different from the Euclidean distance decision. The paired distances between data points in a high-dimensional space are transformed by Gaussian distribution into probability distances  $p_{ij}$  to represent the similarities between data points:

$$p_{ij} = \frac{\exp(-\|x_i - x_j\|/2\sigma^2)}{\sum_k \sum_{l \neq k} \exp(-\|x_l - x_k\|/2\sigma^2)}, \text{ for } \forall i \forall j : i \neq j. \tag{1}$$

The aim of t-SNE is to calculate and retain the probabilistic of distances between all object points in low-dimensional space. In t-SNE, the two or three-dimensional “metric space” is defined as a long-tailed distribution  $Q_{ij}$  that centers at each and every point, for purposing of avoiding the “crowding problem [23]”. The paired distances between data points in a low dimensional space is transformed into a probability distance  $q_{ij}$  by t-distribution to represent the similarities between data points:

$$q_{ij} = \frac{\left(1 + \|y_i - y_j\|^2\right)^{-1}}{\sum_k \sum_{l \neq k} \left(1 + \|y_l - y_k\|^2\right)^{-1}}, \text{ for } \forall i \forall j : i \neq j. \tag{2}$$

The difference between the similarity  $q_{ij}$  in the low-dimensional space and the similarity  $p_{ij}$  in the high-dimensional space is measured by calculating the KL divergence between the joint distributions  $P$  and  $Q$ :

$$C = KL(P||Q) = \sum_i \sum_{j \neq i} p_{ij} \log \frac{p_{ij}}{q_{ij}}. \tag{3}$$

**Multiple maps t-SNE**

mm-tSNE is a variant of the t-SNE method that breaks down the traditional limitations of a single metric map by constructing multiple mappings  $M$  in a low-dimensional space to visualize pairwise similarities in non-metric spaces.

Multiple maps t-SNE constructs  $M$  maps in low dimensional space, where each map contains  $N$  data points. In the map with index  $m$ , the data point with index  $i$  has an importance weight  $\pi_i^{(m)}$ , which represents the importance of data point  $i$  in map  $M$ , and the sum of the weights of data point  $i$  in all maps is equal to 1.

Therefore, the pairwise similarity  $q_{ij}$  between data points in a low-dimensional space is measured by a weighted sum of pairwise similarities between data points  $i$  and  $j$  in all the maps. Its mathematical definition is as follows:

$$q_{ij} = \frac{\sum_m \pi_i^m \pi_j^m \left(1 + \|y_i^{(m)} - y_j^{(m)}\|^2\right)^{-1}}{\sum_{m'} \sum_{k \neq i} \pi_k^{(m')} \pi_i^{(m')} \left(1 + \|y_k^{(m')} - y_i^{(m')}\|^2\right)^{-1}} \text{ for } \forall i \forall j : i \neq j, \tag{4}$$

where  $y_i^{(m)}$  indicates that the data point  $i$  in the high-dimensional space is mapped to the  $m$  map in the low-dimensional space. Since it is more difficult to directly calculate the parameter  $\pi_i^{(m)}$ . In order to simplify the calculation, the weight of importance  $\pi_i^{(m)}$  is obtained by

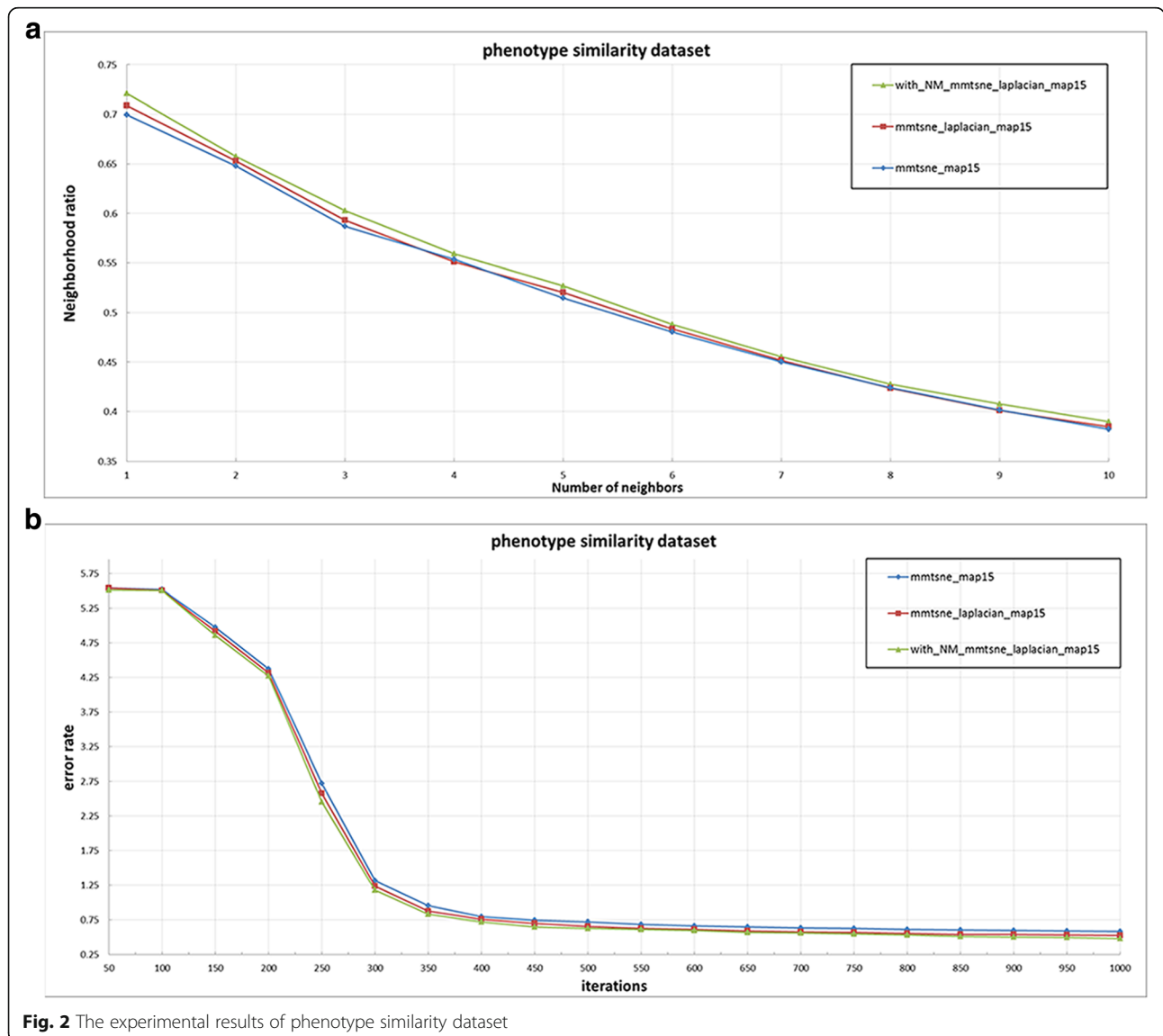
calculating the unconstrained  $\omega_i^{(m)}$ :

$$\pi_i^{(m)} = \frac{e^{-\omega_i^{(m)}}}{\sum_{m'} e^{-\omega_i^{(m')}}}. \tag{5}$$

The objective loss function has the uniform form as Eq. 3, but the cost function minimum is calculated by the location of the point  $y_i^{(m)}$  in all relevant metric maps and the associated unrestrained weight  $\omega_i^{(m)}$ .

**Multiple maps t-SNE with Laplacian regularization**

Multiple maps t-SNE with Laplacian regularization (mm-tSNE regularization) alleviates the problem that the higher-weighted data points in the uniform map do not accord with the uniform clustering structure by



**Fig. 2** The experimental results of phenotype similarity dataset

adding Laplacian penalties to the original mm-tSNE cost function  $C(Y)$ .

$$C(Y) = KL(P||Q) = (1-\lambda) \sum_i \sum_{j \neq i} p_{ij} \log \frac{p_{ij}}{q_{ij}} + \lambda \pi^T L \pi, \tag{6}$$

where  $L = (\text{diag}(\sum_j p_{ij}) - P)_{ij}$ .

The gradient about the mapping point  $y_i^{(m)}$  in the low-dimensional space is calculated by the following equation:

$$\frac{\partial C(Y)}{\partial y_i^{(m)}} = 4(1-\lambda) \sum_j \frac{\partial C(Y)}{\partial d_{ij}^{(m)}} (y_i^{(m)} - y_j^{(m)}), \tag{7}$$

where  $d_{ij}^{(m)} = \|y_i^{(m)} - y_j^{(m)}\|^2$ .

The gradient about the weights  $\omega_i^{(m)}$  in the low-dimensional space is calculated by the following equation:

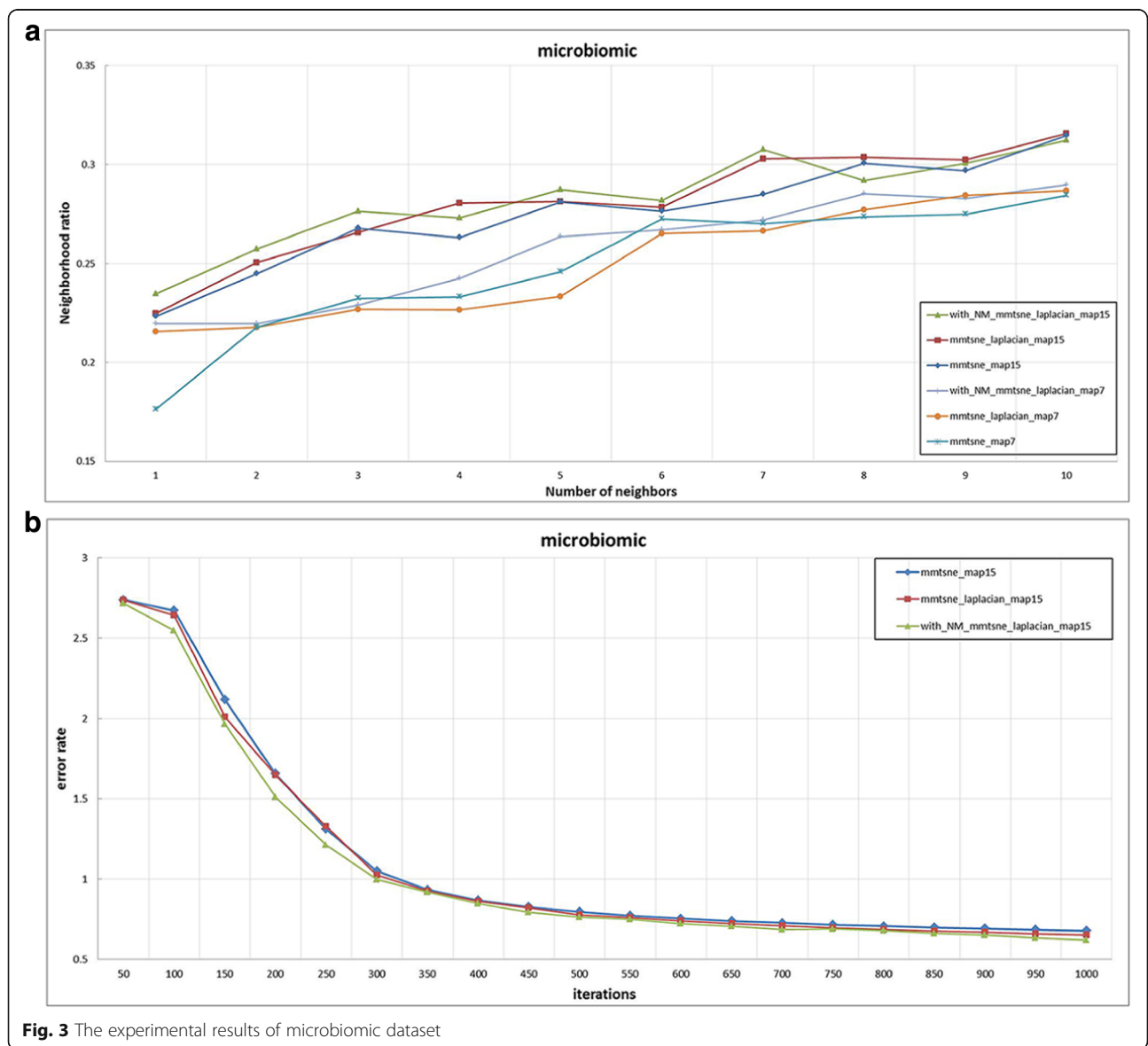
$$\frac{\partial C(Y)}{\partial \pi_i^{(m)}} = \sum_j \left( \frac{2}{q_{ij} Z} (p_{ij} - q_{ij}) \right) \pi_j^{(m)} (1 + d_{ij}^{(m)})^{-1} + \lambda L \pi, \tag{8}$$

where  $Z = \sum_k \sum_{l \neq k} \sum_{m'} \pi_i^{m'} \pi_k^{m'} (1 + d_{kl}^{m'})$ .

Mathematically, the gradient update of the momentum item is given by the following equation:

$$v^{(t)} = \gamma v^{(t-1)} - \eta \frac{\partial C(Y)}{\partial Y}, \tag{9}$$

$$Y = Y + v^{(t)}, \tag{10}$$



**Fig. 3** The experimental results of microbiomic dataset

where  $Y$  are the model parameters, the velocity is  $v^{(t)}$ , the momentum coefficient is  $\gamma \in [0, 1]$  and  $\eta$  is the learning rate at iteration  $t$ ,  $\frac{\partial C(Y)}{\partial Y}$  is the gradient.

**Simplified Nesterov momentum**

Nesterov momentum [21, 22] is a first-order optimization method to improve stability and convergence of regular gradient descent. The algorithm update rules are as follows [28, 29]:

$$v^{(t)} = \mu^{(t-1)}v^{(t-1)} - \varepsilon^{(t-1)} \nabla f\left(\theta^{(t-1)} + \mu^{(t-1)}v^{(t-1)}\right), \tag{11}$$

$$\theta^{(t)} = \theta^{(t-1)} + v^{(t)}, \tag{12}$$

where  $\theta_t$  are the model parameters, the velocity is  $v^{(t)}$ ,  $\mu^{(t)} \in [0, 1]$  is the momentum coefficient and  $\varepsilon^{(t)} > 0$  is the learning rate at iteration  $t$ ,  $f(\theta)$  is the objective function and  $\nabla f(\theta)$  is a shorthand notation for the gradient  $\frac{\partial f(\theta)}{\partial \theta} | \theta = \theta'$ .

The equivalent form is as follows:

$$v^{(\hat{t})} = \mu^{(t-1)}v^{(t-1)} - \varepsilon^{(t-1)} \nabla f\left(\theta^{(\hat{t}-1)} - \varepsilon^{(t-1)}\mu^{(t-1)}[\nabla f\left(\theta^{(t-1)}\right) - \nabla f\left(\theta^{(\hat{t}-2)}\right)]\right). \tag{13}$$

$$\theta^{(\hat{t})} = \theta^{(\hat{t}-1)} + v^{(\hat{t})}. \tag{14}$$

Different from the momentum term, Nesterov momentum renews the parameter vector at some position  $\theta^{(t)}$ , which depends on  $\mu^{(t-1)}v^{(t-1)}$  as well as in the last momentum update of the current parameter position. The gradient correction to the velocity  $v_t$ , with the Nesterov momentum, is calculated at point  $\theta^{(t)} + \mu^{(t-1)}v^{(t-1)}$ , and if  $\mu^{(t-1)}v^{(t-1)}$  is an even worse update,  $\nabla f(\theta^{(t-1)} + \mu^{(t-1)}v^{(t-1)})$  will point reversely  $\theta^{(t)}$  more

forcefully than the gradient computed at  $\theta^{(t)}$ , hence providing a larger and more timely correction to  $v^{(t)}$ . Fig. 1 (b) illustrates the geometric significance of this phenomenon. With the equivalent form of Nesterov momentum, we can observe the difference between Nesterov momentum and standard momentum. The direction of this update has increased by an amount of  $\mu^{(t-1)}[\nabla f(\theta^{(t-1)}) - \nabla f(\theta^{(\hat{t}-2)})]$ , the change is essentially an approximation of the second order of the objective function. Since Nesterov momentum uses the second-order information of the objective function, the Nesterov momentum is more efficient than the standard momentum term in modifying the large and undue velocity in each iteration, which makes it run faster than the momentum method, and can further reduce the error rate of the loss function.

**Multiple maps t-SNE regularization based on Nesterov momentum**

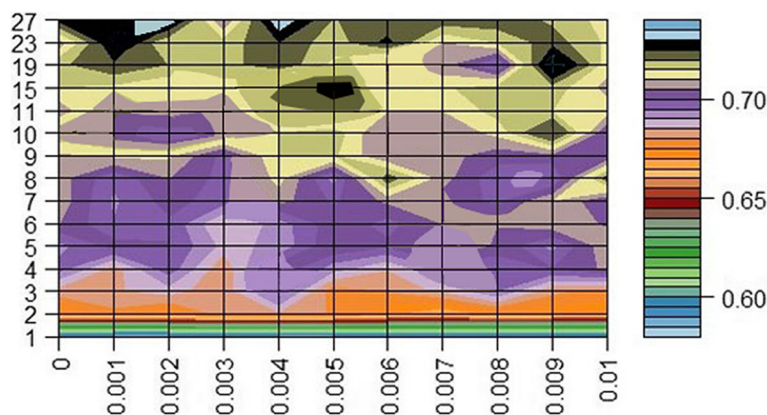
In this article, unlike the original several versions of mm-tSNE, we use the Nesterov momentum method to optimize the target loss function, which lets the loss function reach the optimal value better and faster and obtain a higher neighborhood preservation ratio.

The learning algorithm is as follows:

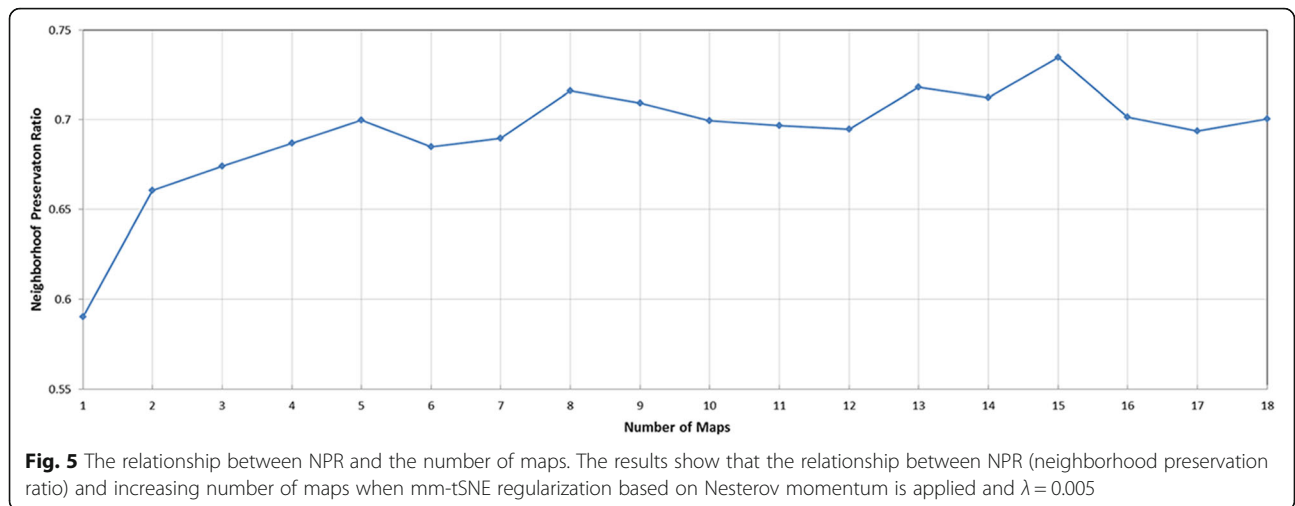
$$v^{(t)} = \gamma v^{(t-1)} - \eta \frac{\partial C(Y)}{\partial} (Y + \gamma v^{(t-1)}). \tag{15}$$

$$Y = Y + v^{(t)}, \tag{16}$$

where  $Y$  represents the model parameter to be optimized,  $v^{(t)}$  represents the velocity of the  $i$  iteration,  $\gamma \in [0, 1]$  represents the momentum coefficient,  $\eta$  represents the learning rate for the  $i$  iteration, and  $\frac{\partial C(Y)}{\partial Y}$  represents the gradient.



**Fig. 4** Heatmap of neighbourhood preservation ratio for mm-tSNE regularization based on Nesterov momentum



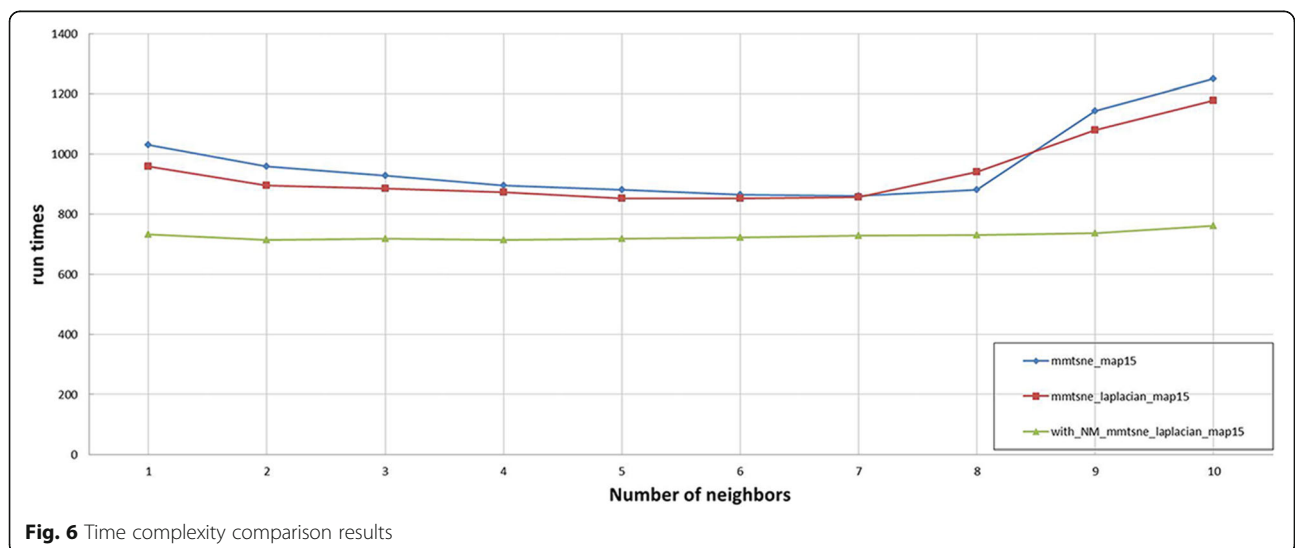
**Datasets**

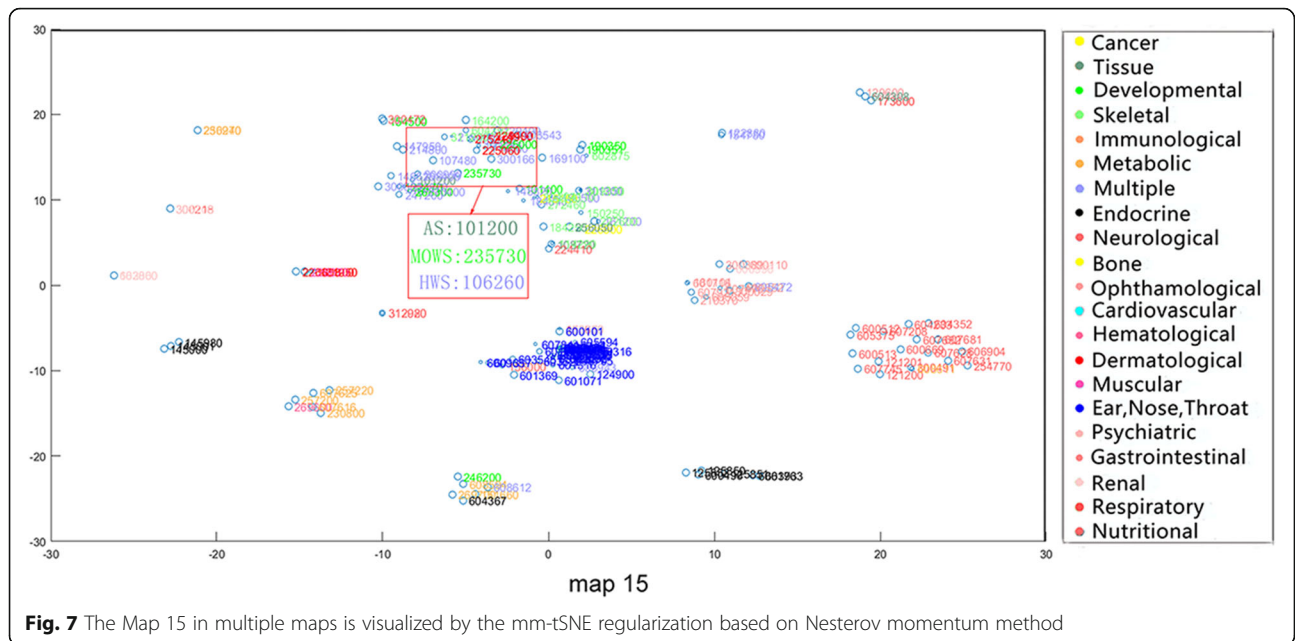
To assess the performance of our approach, we apply our method to several datasets, including phenotypic similarity dataset and microbial dataset. The microbial dataset consisted of 6313 orthologous proteins which are from 345 individual intestinal microorganisms [30]. After data preprocessing, a similarity matrix of 1299 KOs is finally obtained. The phenotypic similarities come from the Online Mendelian Inheritance in Man (OMIM) database [31, 32], which contains 1025 phenotypes related to 21 diseases, respectively, according to the disease classification information from the Human Disease Network [8]. At them in the middle, the value of similarity less than 0.5 is filtered out.

**Evaluation indicators**

**Neighborhood preservation ratio** The ideal state for dimensionality reduction visualization is that the

neighboring point of the sample point  $x_i$  in the high-dimensional space is exactly the same as its neighboring point in the low-dimensional space  $y_i$ . That is, it is assumed that the neighboring points around the sample point  $x_i$  pass through the high-dimensional space. After the dimensional method is projected into a two-dimensional space, the neighboring points around  $y_i$  coincide with the high-dimensional space. The neighborhood preservation ratio is a measure proposed by Laurens van der Maaten [15], which measures similarities in the high-dimensional space are preserved in the low-dimensional space by the mm-tSNE method. For each data point  $i$ , we choose its  $k$  highest  $p_{ij}$ -values in the high-dimensional space as its  $k$  nearest neighbors ( $N^{i1}$  for short), and select the  $k$  highest  $q_{ij}$ -values in the low-dimensional space as its  $k$  nearest neighbors ( $N^{i2}$  for short). By calculating the intersection of  $N^{i1}$  and  $N^{i2}$ , it can be determined whether the reduced-dimensional





**Fig. 7** The Map 15 in multiple maps is visualized by the mm-tSNE regularization based on Nesterov momentum method

visualization method used can maintain the distribution of neighboring points of data in high-dimensional space. Therefore, NPR indicates the average ratio of the number of neighbors to be saved.

$$NPR = \frac{1}{n} \sum_{i=1}^n \frac{|N^{i1} \cap N^{i2}|}{k}, \tag{17}$$

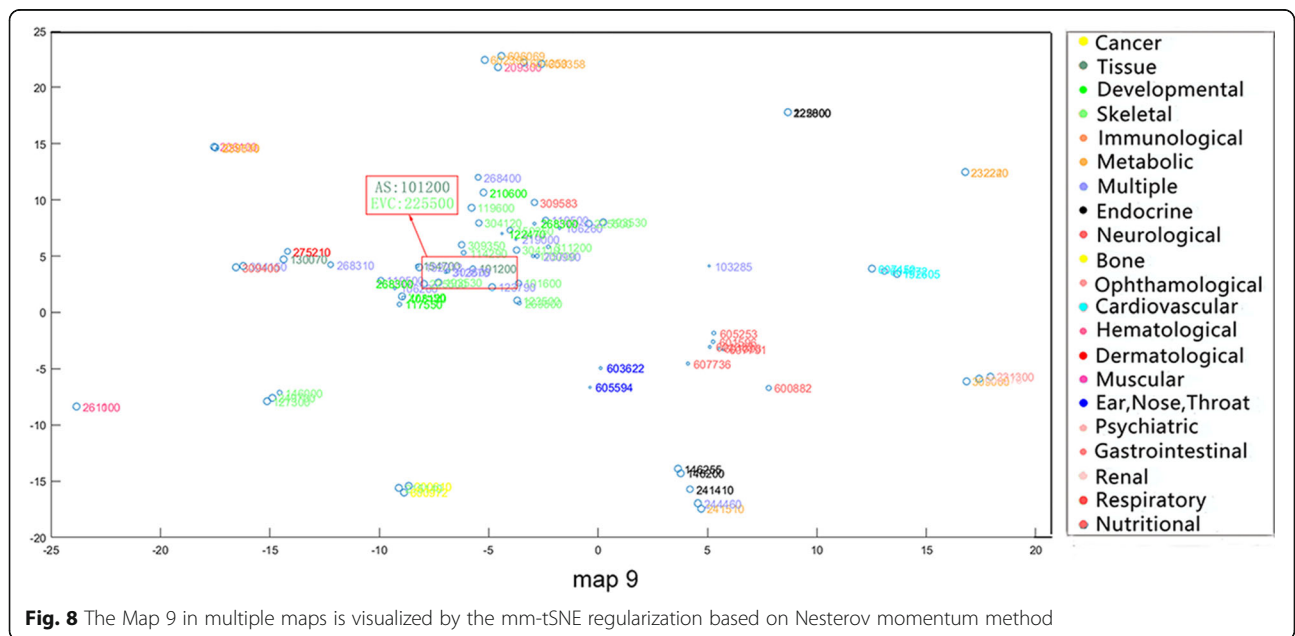
where  $|N^{i1} \cap N^{i2}|$  is the number of points that common points in high-dimensional space and low-dimensional space and  $n$  represent the total number of visualized target data points.

**Error rate** The error rate represents the cost of using the  $KL$  divergence method to model the difference between the  $Q$  distribution and the  $P$  distribution.

**Time complexity** The time complexity of the algorithm is measured by the number of times the basic operations are repeated.

**Results**

We compare the mm-tSNE regularization based on Nesterov momentum method with the original several mm-tSNE methods in the phenotype (Fig. 2) and



**Fig. 8** The Map 9 in multiple maps is visualized by the mm-tSNE regularization based on Nesterov momentum method



**Table 1** Extracted similarities from original matrix

Phenotype With OMIMID	AS (OMIM:101,200)	MOWS (OMIM:235,730)	HWS (OMIM:106,260)	EVAS (OMIM:225,500)
AS (OMIM:101,200)	1	0.5957	0	0.5148
MOWS (OMIM:235,730)	0.5957	1	0.5298	0
HWS (OMIM:106,260)	0	0.5298	1	0.5392
EVAS (OMIM:225,500)	0.5148	0	0.5392	1

microbiome (Fig. 3) dataset respectively using the neighborhood preservation ratio, the error rate and the time complexity as the evaluation indicators.

We then apply the mm-tSNE regularization based on Nesterov momentum to explore the nonmetric relationships on phenotype similarity dataset and microbiomic dataset. The number of model parameters  $m$ —the number of maps and  $\lambda$ —the penalty term are selected according to the neighborhood preservation ratio (NPR) (See methods). Fig. 2 and Fig. 3 show the experimental results on phenotype similarity dataset and microbial dataset, respectively. The mm-tSNE regularization based on Nesterov momentum has performance comparable with mm-tSNE and mm-tSNE regularization. The green line in Fig. 2 and Fig. 3 shows that our proposed models are at an advantage over original mm-tSNE methods of several versions. Fig. 4 is the heat map of NPR in the parametric space of  $m$  and  $\lambda$  when apply mm-tSNE regularization based on Nesterov momentum algorithm. The x-axis represents the value of  $\lambda$  in the experiment, and the y-axis represents the number of maps. The color change in the legend represents a gradual decrease in the preservation ratio of the neighborhood from high to low. When  $\lambda = 0.002$  and the number of maps is 27, the neighbor's preservation ratio is maximized. Nevertheless, according to the experimental results, we choose the number of maps as 15, and set the  $\lambda$  as 15 as our model parameters, because it is sufficient to model the non-metric structure of phenotype similarities and KOs similarities. When the mm-tSNE regularization based on Nesterov momentum is applied, the relationship between the NPR and the number of maps is shown in Fig. 5. When  $\lambda = 0.005$  and  $m = 15$ , we obtain the highest neighborhood preservation ratio. Overall, the mm-tSNE regularization based on Nesterov momentum obtains better performance compared to other methods and reduces the time complexity of algorithm from  $O(1/k)$  (after  $k$  steps) to  $O(1/k^2)$  [21] (See Fig. 6). Since the processed data of the proposed algorithm is a matrix with  $N \times N$  size, the spatial complexity of proposed algorithm does not improve relative to the original algorithms. The space complexity of the proposed algorithm is  $O(N^2)$ .

## Discussion

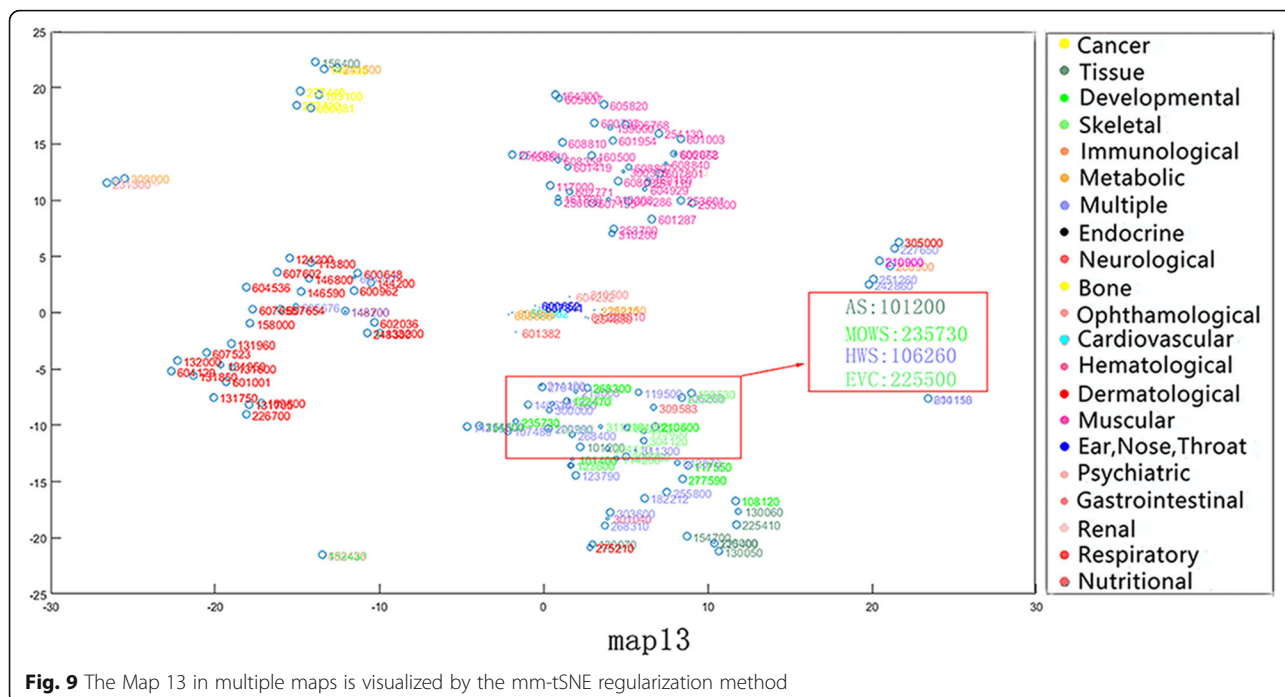
From the phenotypic point of view, similar phenotypes tend to converge into the same class. Nevertheless, some of the phenotypes in the same disease category may exist

in other disease categories as well. In addition, we discover that our method compared to mm-tSNE and mm-tSNE regularization can better appropriately model non-transitive similarities between phenotypes. For example, Apert syndrome (AS, OMIM ID: 101200) has importance weights of 0.5967 and 0.3896 at two maps (Maps 9 and 15, See Fig. 7 and Fig. 8). Removing the phenotype of each map with an importance weight less than 0.1 prevents visualization from being too clutter. In Map 9, Ellis-van Creveld syndrome (EVC, OMIM ID: 225500) is one of the neighbors of the AS, with similarity of 0.5148 (See Table 1) and they have an importance weights of 0.5967 and 0.9474 in the metric space Map 9 severally (See Table 2). In Map 15, AS has a near neighbor Mowat-Wilson syndrome (MOWS, OMIM ID: 235730) with similarity 0.5957. From Table 2, it can be found that MOWS is not displayed on Map 9 and EVAS is not displayed on Map 15, the fact that they are both neighbors in single maps. In other words, the neighbor of AS in Map 9 is not essentially the neighbor of it in Map 15. In fact, the similarity between EVC and MOWS is 0 (See Table 1). Although the initial aim of mm-tSNE regularization and mm-tSNE is to find intransitivity similarity. We find that the mm-tSNE and mm-tSNE regularization combine the four phenotypes in Table 1 into one map (See Fig. 9 and Fig. 10). This result indicates that the mm-tSNE regularization based on Nesterov momentum excavates non-transitive similarity of the original several methods without discovering.

Except MOWS, at Map 15 (see Fig. 7), AS has another near neighbor—Hay-Wells syndrome (HWS, OMIM: 106260) with a similarity 0.5957. AS, MOWS and HWS are all neighbors in Map 15. Nevertheless, astonishing truth is that the similarity between AS and HWS is 0 (See Table 1). Then we have a deep analysis of these three phenotypes. Apert syndrome is a congenital disease; the main symptoms include craniosynostosis, middle facial hypoplasia, hands and feet, with the tendency

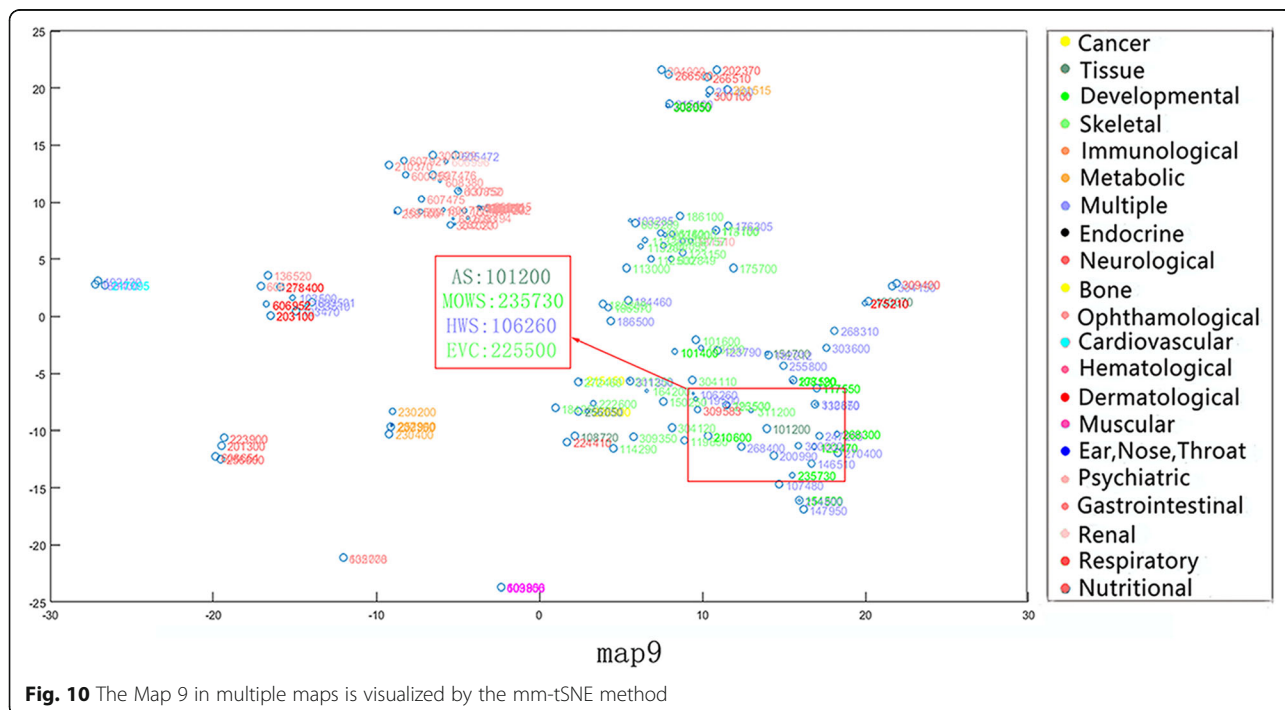
**Table 2** Importance weights for extracted phenotypes

	Map9	Map15
AS (OMIM:101200)	0.5967	0.3896
MOWS (OMIM:235730)	9.0475e-04	0.9920
HWS (OMIM:106260)	0.1436	0.8348
EVC: (OMIM:225500)	0.9474	0.002



of bone structure fusion [33–35]. Mowat-Wilson’s syndrome is an autosomal dominant complex dysplasia, characterized by a variety of clinical symptoms such as mental retardation, motor retardation, epilepsy, vaso-vagal disease and neuropathy, caused by mutations in individual functions [36–38]. HWS is a rare, complex disease characterized by congenital ectodermal dysplasia

with a variety of symptoms including thinning hair, mild hypohidrosis, scalp infection, dental hypoplasia, and maxillary dysplasia [39–41]. Although these three diseases belong to different types of diseases (tissue, developmental and multiple respectively), they have the same symptoms, such as nail and tooth dysplasia and skeletal deformities. The experimental result shows that



**Table 3** The weights for KOs similarity. Large values are shown by bold

	Map1	Map2	Map3	Map5	Map7	Map10	Map12	Map13
<b>K00691</b>	0.006	0.0041	0.0073	0.0082	<b>0.2809</b>	0.0046	<b>0.2498</b>	<b>0.4028</b>
<b>K05340</b>	0.0064	0.0035	0.0056	0.007	<b>0.3366</b>	0.0049	<b>0.5822</b>	0.0061
<b>K06204</b>	0.0030	0.0029	0.0021	0.0934	<b>0.5417</b>	0.0031	0.0034	<b>0.3292</b>

although the text mining method [42] measures the direct similarity between AS and HWS as 0, our method does deduce their true relationship from data. This is different from non-transitive similarity modeling, because they are in the uniform metric space Map 15.

The experimental results demonstrate that our proposed method reveals the non-transitive similarity not found in the original several mm-tSNE methods in microbiomic dataset (See Table 3). K00691 is a maltose phosphorylase involved in glucose metabolism and transcription [43]. Table 3 shows three KOs, of which at least three maps have an importance weight of not less than 0.2, which are respectively close to K00691. K05340 is a transporter involved in signal transduction and glucose uptake of cellular activity. K06204 is a Dnak inhibitor that is involved in the biofilm formation and prokaryotic cell activities of *Escherichia coli* and rRNA transcription [44]. From Table 3 we can see that although these three KOs are similar in Map 7, they are not similar to each other in other maps. For example, K05340 in Map 12 is not similar to K06204. Likewise, K06204 is not similar to K05340 in Map 13. These non-transitive similarities can not be expressed by traditional data visualization methods.

## Conclusions

We propose a new method to optimize the mm-tSNE regularization cost function. Experimental result shows that this method outperforms several versions of mm-tSNE, when measured by neighborhood preservation rate and error rate. In this study, it is shown that non-metric properties are ubiquitous in biological and microbiological data and should be considered in future studies. Traditional visualization techniques are effective when applied to small and medium-scale data, but they still face a huge challenge when applied to large biological and microbiological data. In future research work, we will propose a method to solve the problem of high computational complexity and problems in data visualization caused by the increase of data volume and the high dimensionality.

## Abbreviations

AS: Apert syndrome; EVC: Ellis-van Creveld syndrome; HWS: Hay-Wells syndrome; mm-tSNE regularization: Multiple maps t-SNE with Laplacian regularization; mm-tSNE: Multiple maps t-SNE; MOWS: Mowat-Wilson syndrome; NPR: Neighborhood preservation ratio; OMIM: Online Mendelian Inheritance in Man; t-SNE: t-Distributed Stochastic Neighborhood Embedding

## Acknowledgements

Not applicable.

## Consent to publication

Not applicable.

## Funding

Publication costs are funded by the National Natural Science Foundation of China (61532008) and the National Key Research and Development Program of China (2017YFC0909502).

## Availability of data and materials

The social network dataset used in our experiment can be downloaded in [https://lvdmaaten.github.io/multiplemaps/Multiple\\_maps\\_t-SNE/Multiple\\_maps\\_t-SNE.html](https://lvdmaaten.github.io/multiplemaps/Multiple_maps_t-SNE/Multiple_maps_t-SNE.html). This dataset is available for public and free to use.

The microbial dataset used in our experiment can be downloaded in [ftp://penguin.genomics.cn/pub/10.5524/100001\\_101000/100036/Intermediate\\_results/](ftp://penguin.genomics.cn/pub/10.5524/100001_101000/100036/Intermediate_results/). This dataset is available for public and free to use.

The phenotypic similarity dataset used in our experiment can be downloaded in <http://www.cmbi.ru.nl/MimMiner/cgi-bin/main.pl>. This dataset is available for public and free to use.

## About this supplement

This article has been published as part of *BMC Bioinformatics Volume 19 Supplement 20, 2018: Selected articles from the IEEE BIBM International Conference on Bioinformatics & Biomedicine (BIBM) 2017: bioinformatics*. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-19-supplement-20>.

## Authors' contributions

XS and XJ designed the algorithm based on mm-tSNE regularization. XZ implemented the mm-tSNE regularization based on Nesterov momentum algorithm and run the experiments. KW and YM helped plan the experimental analysis. JL contributed to writing the manuscript. TH and XH supervised and helped conceive the study. All authors read and approved the final manuscript.

## Ethics approval and consent to participate

Not applicable.

## Competing interests

The authors declare that they have no competing interests.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

## Author details

<sup>1</sup>School of Computer, Central China Normal University, Wuhan, China.

<sup>2</sup>College of Computing and Informatics, Drexel University, Philadelphia, PA 19104, USA.

Published: 21 December 2018

## References

1. Brunner HG, Van Driel MA. From syndrome families to functional genomics. *Nat Rev Genet.* 2004;5:545–51.
2. Lim J, et al. A protein–protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell.* 2006;125(4):801–14.

3. Limviphuvadh V, et al. The commonality of protein interaction networks determined in neurodegenerative disorders (NDDs). *Bioinformatics*. 2007; 23(16):2129–38.
4. Oti M, Huynen MA, Brunner HG. Phenome connections. *Trends Genet*. 2008; 24(3):103–6.
5. Wooley JC, Godzik A, Friedberg I. A primer on metagenomics. *PLoS Comput Biol*. 2010;6(2):e1000667.
6. Freudenberg J, Propping P. A similarity-based method for genome-wide prediction of disease relevant human genes. *Bioinformatics*. 2002;18(suppl2): S110–5.
7. Lage K, et al. A human phenome-interactome network of protein complexes implicated in genetic disorders. *Nat Biotechnol*. 2007;25(3): 309–16.
8. Oti M, et al. Predicting disease genes using protein–protein interactions. *J Med Genet*. 2006;43(8):691–8.
9. Xu, W., Jiang, X., Li, G. (2013) Nonmetric property of diabetes-related genes in human gut microbiome, IEEE International Conference on Bioinformatics and Biomedicine.
10. Loscalzo J, Kohane I, Barabasi AL. Human disease classification in the postgenomic era: a complex systems approach to human pathobiology. *Mol Syst Biol*. 2007;3:124.
11. Wang Q, Jia P, Cuenco KT, Feingold E, Marazita ML, Wang L, et al. Multi-dimensional prioritization of dental caries candidate genes and its enriched dense network modules. *PLoS One*. 8:e76666. <https://doi.org/10.1371/journal.pone.0076666>.
12. P. Csermely, T. Korcsmáros, H J M Kiss, G London, R Nussinov, Structure and dynamics of molecular networks: a novel paradigm of drug discovery: a comprehensive review, *Pharmacol Ther* 138 (3) (2013) 333–408.
13. Arumugam M, et al. Enterotypes of the human gut microbiome. *Nature* 2011; 473:174–180. [PubMed: 21508958].
14. Legendre, P., L. Legendre, *Numerical Ecology* Vol. 20. 2012: Elsevier.
15. Van der Maaten L, Hinton G. Visualizing non-metric similarities in multiple maps. *Mach Learn*. 2012;87(1):33–55.
16. Xu W, Jiang X, Hu X, Li G. Visualization of genetic disease-phenotype similarities by multiple maps t-SNE with Laplacian regularization. *BMC Med Genet*. 2014;7(2):1–9.
17. Li C, Li H. Network-constrained regularization and variable selection for analysis of genomic data. *Bioinformatics*. 2008;24(9):1175–82.
18. He X, et al. Laplacian regularized Gaussian mixture model for data clustering. *Knowledge and data engineering. IEEE Transactions on*. 2011; 23(9):1406–18.
19. Qian N. On the momentum term in gradient descent learning algorithms. *Neural networks*. 1999;12(1):145–51.
20. Shen, X., Zhu, X., Jiang, X., Hu, X. (2017) Visualization of disease relationships by multiple maps t-SNE regularization based on Nesterov accelerated gradient, IEEE International Conference on Bioinformatics and Biomedicine.
21. Nesterov Y. A method for unconstrained convex minimization problem with the rate of convergence  $O(1/k^2)$ . *Doklady ANSSSR* (translated as *SovietMathDocl*). 269:543–7.
22. Nesterov Y. *Introductory lectures on convex optimization: a basic course*. Applied optimization. Kluwer academic Publ. London: Boston, Dordrecht; 2004.
23. Van der Maaten L, Hinton G. Visualizing Data using t-SNE. *J Mach Learn Res*. 2008;9(11).
24. Hinton GE, Roweis S. Stochastic neighbor embedding. In *NIPS'2002; 2003*.
25. Lacoste-Julien S, Sha F, Jordan MI. DiscLDA: discriminative learning for dimensionality reduction and classification. In *NIPS*, volume. 2008;22.
26. Mao Y, Balasubramanian K, Lebanon G. Dimensionality reduction for text using domain knowledge. In: *Proceedings of the 23rd international conference on computational linguistics: posters, COLING '10*, Association for Computational Linguistics, Stroudsburg, PA, USA; 2010. p. 801–9.
27. Jamieson AR, et al. Exploring nonlinear feature space dimension reduction and data representation in breast CADx with Laplacian eigenmaps and t-SNE. *Med Phys*. 2010;37:339.
28. Sutskever I. *Training recurrent neural networks*, Ph.D. thesis. Toronto: CS Dept., U; 2012.
29. Bengio Y, Boulanger Lewandowski N, Pascanu R. Advances in optimizing recurrent networks. In *Proceedings of the 38th International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2013)*, May; 2013.
30. Qin J, et al. A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature*. 2012;490(7418):55–60.
31. Hamosh A, et al. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic acids research*. 2005;33(suppl 1):D514–7.
32. Jiang X, et al. Modularity in the genetic disease phenotype network. *FEBS Lett*. 2008;582(17):2549–54.
33. Mantilla-Capacho JM, Arnaud L, Diaz-Rodriguez M, Barros-Nunez PA. Syndrome with preaxial polydactyly showing the typical mutation Ser252Trp in the FGFR2 gene. *Genet Counsel*. 2005;16:403–6.
34. Moloney DM, Slaney SF, Oldridge M, Wall SA, Sahlin P, Stenman G, Wilkie AOM. Exclusive paternal origin of new mutations in Apert syndrome. *Nature Genet*. 1996;13:48–53.
35. Lajeunie E, De Parseval N, Gonzales M, Delezoide AL, Journeau P, Munnich A, Le Merrer M, Renier D. Clinical variability of Apert syndrome. *J Neurosurg*. 2000;90:443.
36. Mowat DR, Wilson MJ, Goossens M. Mowat-Wilson syndrome. *J Med Genet*. 2003;40:305–10.
37. Streng S, Heinritz W, Zweier C, Rauch A, Rolle U, Merckenschlager A, Froster UG. Pulmonary artery sling and congenital tracheal stenosis in another patient with Mowat-Wilson syndrome. (letter). *Am J Med Genet*. 2007;143A: 1528–30.
38. Horn D, Weschke B, Zweier C, Rauch A. Facial phenotype allows diagnosis of Mowat-Wilson syndrome in the absence of Hirschsprung disease. *Am J Med Genet A*. 2004;124A:102–4.
39. Hay RJ, Wells RS. The syndrome of ankyloblepharon, ectodermal defects and cleft lip and palate: an autosomal dominant condition. *Brit J Derm*. 1976;94:287–9.
40. McGrath, J. A., Duijf, P. H. G., Doetsch, V., Irvine, A. D., de Waal, R., Vanmolokot, K. R. J., Wessagowitz, V., Kelly, A., Atherton, D. J., Griffiths, W. A. D., Orlov, S. J., Ausems, M. G. E. M., Yang, A., McKeon, F., Bamshad, M. A., Brunner, H. G., Hamel, B. C. J., van Bokhoven, H. Hay-Wells syndrome is caused by heterozygous missense mutations in the SAM domain of p63. *Hum Mol Genet* 10: 221–229, 2001.
41. Bertola DR, Kim CA, Sugayama SMM, Albano LMJ, Utogawa CY, Gonzalez CH. AEC syndrome and CHAND syndrome: further evidence of clinical overlapping in the ectodermal dysplasias. *Pediatr Derm*. 2000;17:218–21.
42. van Driel MA, et al. A text-mining analysis of the human phenome. *European journal of human genetics*. 2006;14(5):535–42.
43. Zhou J, Ashouian N, Delepine M, Mastsuda F, Chevillard C, Rivlet R, Schildkraut CL, Birshtein BK. The origin of a developmentally regulated Igh replicon is located near the border of regulatory domains for Igh replication and expression. *PNAS*. 2002;99(21):13693–8.
44. Adachi Y, Asakura Y, Sato Y, Tajima T, Nakajima T, Yamamoto T, Fujieda K. Novel SLC12A1 (NKCC2) mutations in two families with Bartter syndrome type 1. *Endocr J*. 12 Nov 2007;54(6):1003–7.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

