IDEE TRANSACTIONS ON VISUALIZATION AND COMPUTER SPATHICS, VOL. 24, NO. 1, JANUARY 2018

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL.24, NO.1, JANUARY 2018

Visual Exploration of Semantic Relationships in Neural Word Embeddings

Shusen Liu, Peer-Timo Bremer, Jayaraman J. Thiagarajan, Vivek Srikumar,

Bei Wang, Yarden Livnat and Valerio Pascucci

Abstract—Constructing distributed representations for words through neural language models and using the resulting vector spaces for analysis has become a crucial component of natural tanguage processing (NLP). However, despite their widespread application, tittle is known about the structure and properties of these spaces. To gain insights into the reliationship where words, the NLP community has begun to adopt high-dimensional visualization techniques. In particular, researchers commonly use t-distributed stochastic neighbor

embeddings () SNE) and principal component analysis (PCA) to create two-dimensional embeddings for assessing the overall structure and exploring linear relationships (e.g., word analogies), respectively. Unfortunately, these techniques often produce medicore or

even misleading results and cannot address domain appeal, we consider the contraction of the contraction of

language processing (NLP) is one of the key components in relationships most interesting to researchers. Consequently, to preserve

I world responsible for everything from sets search to doc-ionion and from machine translation to speech recognition, aktirough that led to the recent surge of AI research in carefully chosen subsets of words, i.e., countries and capitals, nown

neept of neural word embeddings, such as word2vec [27] and their planals, etc. Unfortunately, both the linear (PCA) and nonlin

the encept of secured word embodings, such in word/weig process. We constrained word embodings are the [33]. These systems utilize a large corpus of training article constrained word in the co-occurrence statistics between pairs of words is given context, and employ a neural network to infer a vector. He emboding words. Interestingly, the position and difference in the matter of the emboding space without any consideration for the inharent disturbines appear to encode semantic relationships (see Other most striking examples is analogy pairs such as smooth and content of the projection itself. Given the complex nature of the high-dimensional space, any two-dimensional emboding will never a second and content of the projection itself.

IEEE TRANSACTIONS ON VISUALIZATION AND COMPLITER GRAPHICS, VOL. 24, NO. 1, JANUARY 2016

ACTIVIS: Visual Exploration of Industry-Scale

Deep Neural Network Models

Minsuk Kahng, Pierre Y. Andrews, Aditya Kalro, and Duen Horng (Polo) Chau

and subset-level. II. Our user Susan starts exploring the model archite date node (in yellow) displays its neuron activations (at II). 2. The ne and instance subsets: the projected view displays the 2-O projection of

ctivation patterns across instances, subsets, and classes, revealing

Index Terms-Vausi analytics, deep learning, machine learning, int

# LSTMVis: A Tool for Visual Analysis of Hidden State Dynamics in Recurrent Neural Networks

Hendrik Strobelt, Sebastian Gehrmann, Hanspeter Pfister, and Alexander M. Rush

tations learned by the model. The ma

LSTMVis, a visual analysis tool for recu

OMs require visual space sions, and as such do not

sting this subset to the user. ach is, thus, directly linked to

igness that a scatterplot ntify the above-mentioned

s in a way that is generic

r in scatterplots. I measure. We have our work

te than roughly ten dimensions. attack this problem by exum-atterplots from the d<sup>2</sup> existing

nd potentially interesting plots, thereby making SPLOMs mo

Skeleton-based Scagnostics

José Matute, Alexandru C. Telea, and Lars Linsen

EEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS. VOL. 94. NO. 1, JANUARY 3018

EEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS. VOL. IV. NO. 1. JANUARY 2018

for Designing Deep Neural Networks Nicola Pezzotti, Thomas Höllt, Jan van Gemert, Boudewijn P.F. Lelieveldt, Elmar Eisemann, Anna Vilanova

DeepEyes: Progressive Visual Analytics

are been used for a long-time in dyar, 2D shapes [48]. State-ofa Progressive Visual Analytics system for the analysis of deep neural networks during training. The overview is given by the commonly used lose and accuracy-curves (a) and the Perplexity Histograms (b) a novel visualization detection of stable layers. A detailed analysis per layer is performed in three bighty linked visualizations. Degenerated billed in the Activation Histogram (b), and the activations are visualized on the Import May DE, Finally, in the Elline May.

neural networks are now risating human accuracy in several pattern recognition problems. Compi-relatures are handorated, neural networks learn increasingly complex features directly from the satures, it is now the retwork architecture that is manually engineered. The network architecture pa-ers or the number of litters per layer and their intercentectors are essential for good performance. It exists, designing a resural network is an iterative trial-and-error process that takes days or even-tatisetic used for training. In this paper, we present DeepEyes, a Progressive Visual Analytics sys-ral networks during training. We present novel visualizations, supporting the identification of lay-res and, therefore, are of interver for a detailed analysis. The system isolitates the identification of or itsyers, and information that is not being captured by the network. We demonstrate the effectiven-use cases, showing how a trained network can be compressed, reshaped and adapted to different

s problems, like image and speech recog-

While the results that DNNs can achieve entially remain a black box. An increasin n making the visualization and the analy

While both, the machine learning and the v behaves [27, 37,50], e.g., by showing the pa

example, in networks trained to recognize of first layer generally contains filters that are to

satterns, e.g., grids or stripes. By using hunds n each layer, DNNs allow for more comple

Only recently the training of large DNNs w development of fast parallel hardware, i.e., G

# VIGOR: Interactive Visual Exploration of Graph Query

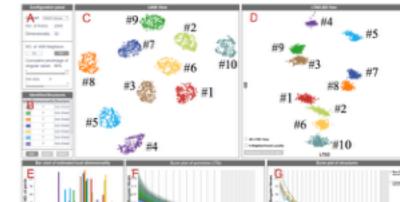
Robert Pienta, Fred Hohman, Alex Endert, Acar Tamersoy fevin Roundy, Chris Gates, Shamkant Navathe, Duen Horng Chau



IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 24, NO. 1, JANUARY 2018 LDSScanner: Exploratory Analysis of Low-Dimensional Structures

Jiazhi Xia, Fenjin Ye, Wei Chen\*, Yusi Wang, Weifeng Chen, Yuxin Ma, and Anthony K.H. Tung

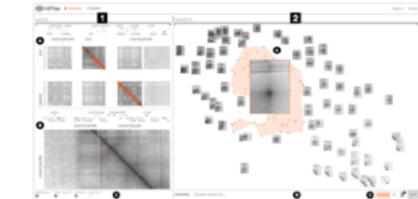
in High-Dimensional Datasets



# HiPiler: Visual Exploration of Large Genome Interaction Matrices with Interactive Small Multiples

EEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. SK, NO. 1, JANUARY 2018

Fritz Lekschas, Benjamin Bach, Peter Kerpedjiev, Nils Gehlenborg, and Hanspeter Pfister



er's interface: the matrix riew (1) with an overview (1A) and detail (1B) matrix. The snippet view (2) presents regions is as interactive small multiples. In this example, snippets are arranged with 1 SNE (2C) and a pile of snippets with a told average pattern is highlighted (2A). View menus for operation are located at the bottom (riC and 2B).

his paper presents an interactive visualization interface—HEPise—for the exploration and visualization of regions of interest one interaction matrices. Genome interaction matrices approximate the physical distance of pairs of regions on the such other and can contain up to 3 million rows and columns with many sparse regions. Regions of interest (RDIss) e.g., by sets of adjacent rows and columns, or by specific visual patterns in the matrix. However, traditional matrix or pan and zoom interfaces tall in supporting search, inspection, and compensor of ROIs in such large matrices. In trix through brushing and linking. The design of HPRer is based on a series of semi-structured interviews with 10 domain ed in the analysis and interpretation of genome interaction matrices. We describe six exploration tasks that are crucial for ors with domain-experts to assess the usability of HiPrier as well as to demonstrate respective findings in the date Interactive Small Multiples, Matrix Comparison, Biomedical Vausization, Genomics

e is about 2 meters long and tightly folded into each

is, bback, ploter) it was harvard of

3D structure is an important factor for regulation of gene expression, replication, DNA repair, and other biological functions. Biologists are interested in succeering in machanisms that drive global and local folding to better understand the vast and complex gene regulation network. This aids comprehension of the functional diversity of cells and how changes in the spatial conformation of the genome can cause diseases [24,32,40]. The probability of two sequences being in close proximity to each

Exploring an entire genome interaction matrix of this size to find and

The probability of two sequences being in close proximity to each other, i.e. intersecting, can be inferred using modern genome sequencing techniques, which yield for every genome a luge symmetric genome interaction matrix with up to 3 million rows and 3 million columns. Both of the 9 trillion matrix cells represents the proximity of two genomic regions. Repetition and hierarchically nested visual patterns can be identified across the matrix, which represent so called regions of interval (BiOls). Those patterns agrees at different scales and range from hundreds of millions down to a few thousand base pairs in size. from hundreds of millions down to a few thousand base pairs in size

received, very calculating various accounts represent associated secretic memors is computationally experienced in this paper, in this paper, we present a machine learning approach to large graph issualization based on computing the topological seminater their consepording seathers. For a given graph, our approach can alrea what the graph would look like indifferent layouts and estimate their consepording seathers. For a given graph, our approach can alrea what the development of a new framework to design graph kinetic. Our experimental study shows that our estimation calculation is considerably leater than computing the actual layouts and he seathers metrics. Also, our graph kernels outperform the state of the art ones in both time and accuracy. In addition, we conducted a user study to demonstrate hat the topological similarity computed with our graph kannel matches perceptual similarity assessed by human users. Index Terms--Orach visualization, graph layout, sesthetics, machine learning, graph kernel, graphlet NTRODUCTION

Fig. 1. A projection of topological similarities between 8,263 graphs measured by our FIW-LOG-LAPLACIAN kernel. Based on the topological similarity, our approach shows what a graph would look like in different layouts and estimates their corresponding aesthetic. topological similarity, our apprisach shows what a graph erous now me. The purpose of the user study. The two graphs in each parameters. The clustered the graphs based on their topological similar tour is comprise, to each other. The projection is computed with 19NE [83], and the highlighted are the most topologically similar, but not isomorphic, to each other. The projection is computed with 19NE [83], and the highlighted

Abstract - Using different methods for laying out a graph can lead to very different visual appearances, with which the viewer perceives different information. Selecting a "good" layout method is thus important for visualizing a graph. The selection can be highly subjective and dependent on the given task. A common approach to selecting a good layout is to use assistants criteria and visual impaction. However, tuly scalulating various layouts and their associated sestions metrics is computationally expensive. In this paper, we present

What Would a Graph Look Like in This Layout? A Machine Learning Approach to Large Graph Visualization Oh-Hyun Kwon, Student Member, IEEE, Tarik Crnovrsanin, and Kwan-Liu Ma, Fellow, IEEE

he are popularly used to represent complex systems, such as social oped to lay out a node-link diagram. A graph's layout results can be voks, power grids, and biological networks. Visualizing a graph thity to bester understand the structure of the data. Many graph adjustion methods have been introduced [16, 43, 86], with the most of the graph [16, 46, 86, 37], it is important to find a "good" layout of and intuitive method being the node-link diagram. r and intuitive method being the node-link diagram.

In the last five decades, a multitude of methods have been devellating the decades, a multitude of methods have been devellating the decades, a multitude of methods have been devellating the decades, a multitude of methods have been devellating the decades, a multitude of methods have been devellating the decades, a multitude of methods have been devellating the decades, a multitude of the graph. Defining a good layout is to use both the esthetic criteria, such as reducing right crossings, and the user's visual

authors are with the University of California, Davis.



Fig. 1. A screenshot of VIGORI showing an analyst exploring a DBLP co-authorship network, looking for nese co-authored papers at the VAST and KDD conferences. (A) The Examplar View visualizes the query, and (B): shows the induced graph formed by joining all query matches. Picking constant node values (e.g., Shixis in the Exit Fusion Graph. (C) Hovering over a node shows its details. (D) The Subgraph Embedding embeds each mallower-dimensional space and clusters them to allow analysis to see patterns and outliers. (E) The Peature Explorer cluster's feature distributions.

SkyLens: Visual Analysis of Skyline on Multi-dimensional Da

View examining the differences between skyline glayers from the attribute and domination perspectives; (d) a Control Panel styline queries; (e) a pop-up window showing a detailed comparison between LeBron James and Chris Paul.

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 24, NO. 1, JANUARY 2018

Visual Diagnosis of Tree Boosting Methods

Shixia Liu, Jiannan Xiao, Junlin Liu, Xiting Wang, Jing Wu, Jun Zhu

6

es and displays the selected one; (d) the feature view displays the feature distributions on the se

Index Terms—tree boosting, model analysis, temporal confusion matrix, tree visualization

ed effective in many applications, such as classification and rank. Ing less comp

60 60 6

coco coco

opob #SSiob

Abstract - Skyline queries have wide-ranging applications in fields that involve multi-criteria decision making, including to: industry, and human resources. By automatically removing incompetent candidates, skyline queries allow users to focus or industry, and human resources. By automatically removing incompetent candidates, styline quiries allow users to focus on a subject of superior data items (i.e., the skylinis), thus reducing the decision-making oserhead. However, users are still required to interp and compare these superior items manually better making a successful chance. This task is challenging because of two issue first, people usually have fuzzy, unstable, and inconsistent preferences when presented with multiple candidates. Second, skyling-points on a multi-dimensional space. To address these issues we propose BayLens, a visual analytic system among at revealing the superiority of styline points from different perspectives and at different scales to all users in their decision making. The socranics demonstrate the usefulness of SkyLens on two datasets with a dozen of attributes. A qualitative study is also conducted to show that users can efficiently accomplish skyline understanding and immediate with Skyl, error.

Index Terms—Skyline query, skyline visualization, multi-dimensional data, visual analytics, multi-criteria decision making

Given a multi-dimensional dataset, skyline queries automatically rute the dataset to a safust of superior points that are not dominated do others; this safust is referred to as skyling [9]. Skyling equents are operated to variesse fields that involve metallicitativity decisions making operated to variesse fields that involve metallicitativity decisions making.

. X. Zhao, Y. Wu, X. Du, Y. Chen, Y. Wang, D. Lee, and H. Ox are with the ong nong Chiteralty of Science and Becknology. T. We is the corresponding author. E-mail: { science; youth, salane, yelvered, vounget, dire, human } (trust let.)

vintelliese org. and reference the Digital Object Mentifier below, what Object Mentifier no. 30.1109/TVCG 2017:2746738



IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICE, VOL.24, NO. 1, JANUARY 2018

## ConceptVector: Text Visual Analytics via Interactive Lexicon Building using Word Embedding

Deokgun Park, Seungyeon Kim, Jurim Lee, Jaegul Choo, Nicholas Diskopoulos, and Niklas Elmqvist, Senior Member, IEEE



Fig. 1. Concept/Vector supports interactive construction of lexicon-based concepts. Here the user creates a new unipolar concept (1) by adding initial keywords related to "tidal flooding" (2). The system recommends related words along with their semantic groupings (5), also shown in a scatterplate (4), rewaiting word- and cluster-level relationships. Invalvant words can be specified to improve recommendation quality (5). Concepts (8) can then be used to rank document curpors (10). Document scores can be visualized in a scatterplat/based on concepts such as 'tidal flooding' and 'money' (7). Users can further refine concepts based on results (8).

Abstract—Central to many test analysis methods is the notion of a soncept' a set of semantically related keywords characterizing a specific object, phenomenon, or theme. Advances in word embedding allow building a concept from a small set of seed terms. However, neive application of such techniques may nesult in false positive errors because of the polysemy of natural language. To emitting the proteins, we present a visual analytics system called ConseptNetor that guides a use intuiting such consists and then using them to analyse documents. Document-analysis case studies with real-world datasets demonstrate the fine-grained analysis provided by ConceptNetor. To support the abbrate modeling of concepts, we introduce a bipolar concept model and support for specifying interference words. The validate the interactive tenion building interface by a user study and expert reviews. Quantitative evaluation shows that the bipolar texicon generated with our methods is comparable to human-generated ones.

Index Terms—Text analytics, visual analytics, word embedding, text summarization, text classification, concepts

E. Ein is with Google Inc. in Mountain West, CR, USA, E-mail: seargered Rycogle.com.

E-mail: nicholas.diakopoulos@pnail.com.

Manuscripe received 31 Max 2017; accepted I Aug. 2017. Date of publication 28 Aug. 2017; date of current sension 1 Oct. 2017.

Generally, building a lexicon for a particular concept requires signi-

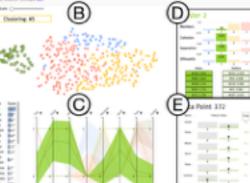
on textual concepts, defined as a set of semantically related keywor

describing a particular object, phenomenon, or theme. For examp

icant human-effort, and thus only a limited number of human-generated concepts have been available, usually with a small number of keywords

IEEE TRANSACTIONS ON VISUALISATION AND COMPUTER GRAPHICS, VOL. 24, NO. 1, JANUARY 2016

Burn Chul Kwon, Ben Eyserbach, Janu Verma, Christopher deFilippi, Walter F. Stewart, and Adam Perer



learning that can be useful for summarizing and aggregating complex multi-dimensional data. However, data can be clustered in many ways, and there exists large body of algorithms designed to neveral different patterns. While having access to a evide variety of algorithms to have the harbor, in granting and difficult for data scientists to stocke and parameterizes algorithms to get the clustering results relevant for their dataset and analytical tasks. To affecte this problem, we built clusterwision, a visual analytics tool that helps ensure

Visual Supervision of Unsupervised Clustering

Distaining #5

# Comparing Visual-Interactive An Experin

Jürgen Bernard, Marco Hutter, Matthias Zej

Fig. 1: Evaluation of four visualization techniques (a)-(d) that support the visual-interactive labeling process. Our study reveals at Class Coloring (b) and Convex Hull (c) are the most useful techniques. Both capture character

lassification model in an intuitive way. Our study shows that they can compete with and even outperform active learning strategic Abstract - Labeling data instances is an important task in machine learning and visual analytics. Both fields provide a broad set

Index Terms—Labeling, Visual-Interactive Labeling, Information Visualization, Visual Analytics, Active Learning, Machine Learning, Classification, Experiment, Dimensionality Reduction

Fig. 1. The exploratory interface of LDSScanner, after the analyst has identified structures. (a)The configuration panel. (b) The identified-structures view. (c) The r-SNE view. (d) The LTSO-GD view. (r) Bar chart of estimated local dimensionality. (f) Sorse plot of

space, manifold, visual exploration

structures, such as clusters in linear subspaces or non-linear manifolds. A large number of automatic approaches have been proposed for

the intrinsic structs propose LDSScans

of labeling strategies, whereby machine learning [and in particular active learning (billows a rather model-centered approach and visual analytics employs rather see-centered approaches [visual-interactive labeling). Both approaches have individual steingates and weaknesses. In this work, we conduct an experiment with three parts to assess and compare the performance of these different three parts to assess the property of the performance of these different and weaknesses. In this work, we conduct an experiment with three parts to assess and compare the performance of these different parts to assess the property of the performance of the performa and westnesses, in this work, we conduct an experiment with three parts is assess and compare the performance of these different labeling strategies, in our study, we (1) identify different visual labeling strategies for user centered labeling, (2) investigate strengths and westnesses of labeling strategies for different labeling tasks and task complexities, and (2) shed light on the effect of using different visual encodings to guide the visual-invessors between the further compare tabeling of single versus multiple instances at a time, and quantify the impact on efficiency. We systematically compare the performance of visual interactive labeling with that of active laterning. Our main findings are that visual-interactive labeling can outperform active isteming, plain the candition that differently reduction separates well the class distributions. Mirrorever, using dimension reduction in combination with additional visual encodings that expose the internal state of the learning model turns out to improve the performance of visual-interactive labeling.

beling follows the principle of attaching information to some object. neural networks require large amounts of such labeled data to learn

then skyline queries will remove A from the candidate list becau-

or A is preferred. B is always a better choice under any circ

ances. Thus, slyline queries may significantly reduce the number of andidates for the tourist without affecting bis/her final choice.

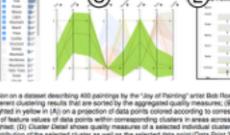
However, skyline queries only solve half of the problem, because users still have to select the most ideal item manually based on their personal preference. In the aforementioned example, travel agents generally cannot decide which city is the best for the marist. Bostead.

the agents can only present all superior cities with their pros and con

to the tourist to decide. To make a successful decision, users need to

ious skyline points, which is rather difficult, especially when the data





Clustering Results shows: 15 different clustering results that are sorting by the aggregated quadrant measures; (8) Projection shows a selected discharing result triplings at selected discharing result triplings at selected discharing result triplings at a selected discharing result triplings at a selected discharing result triplings at the selection of the sel Point shows the feature value distribution of the selected cluster as well as the selected data point (Data Point 372 within Cluster 2)

We live in a world that mutinely produces more tennal data on a daily busis than can be comfortably viewed—let alone analyzed—by medicine, politics, and business, to name just a few. As a result, automatic text analysis an embdos, such as sentiment analysis [34], document summarization [4], and probabilistic topic modeling [3] are becoming increasingly important. Central in most of these methods is the focus

MD USA E-mail: {intainer, elm}@umd.edu.

J. Lee and J. Choo, the corresponding author, are with Konu University in Send, Republic of Korea. E-mail: {jurint)101. jchoo}@konu.ac.ke

enserroung a particular object, phenomenon, or theme. For example, sentiment analysis can be viewed as analyzing documents according to two concepts: positive and negative sentiment. Similarly, the trajec de-rived in topic modeling can be thought of as document-driven concepts. The benefit of this sunfied view is that concepts, once created, can ten be shared and reason among times, similarly to widely applicable lesicon sets such as Linguistic Inquiry and Word Count (LEWC) [37] or General Impairer (GD) (29). or information on obtaining reprints of this article, please send e-mail to reprinted lines.org, and reference the Digital Object Identifier below. Digital Object Identifier no. 10.1109/TVCG.2017.2144478

chading cost, climate, quality of sorvice, and safety. If city A is less desirable in every attribute than city B (i.e., A is dominated by B).

data scientists find the right clustering among the large amount of techniques and parameters available. Our system clusters data using a variety of distatering techniques and garameters and then ranks distancing results utilizing the equity-metrics. In addition, users an guide the system to produce more relevant results by providing task-relevant constraints or mind data. Our visual user interface flows users to find high quality clustering results, explore the clusters using several coordinated visualization techniques, and select te cluster result that best suits their task. We demonstrate this novel approach using a case study with a team of researchers in the

Visual metaphors for exploring high-dimensional datasets come in a variety of forms, each with their own strengths and weaknesses in both visualization and interaction [37,69]. In particular, datasets with high Indications from previous studies [8,33] have shown that analyses

Towards a Systematic Combination of Dimension Reduction Abstract— Dimension reduction algorithms and clustering algorithms are both frequently used techniques in visual analytics. Both families of algorithms assist analysis in performing related tasks regarding the similarity of observations and finding groups in datasets. Though initially used independently, recent works have incorporated algorithms from each family into the same visualization systems. However, these algorithms combinations are often and hoo or inconnensed, working in independently and in parallel nather than integrating some degree of intendependence. A number of design decisions must be addressed when employing dimension neduction and clustering algorithms concurrently in a visualization system, including the selection of each algorithm, and order in which they are processed, and how to present and internal with the resulting projections. This speed contributes an overness of combining dimension reduction and clustering into a visualization system, discussing the challenges inherent in developing a visualization system that makes use of both families of assorbitms.

een the basis for our visual design. In the second phase, we verified our propo

Index Terms.... Hierarchical Date, HSNE, Single-Cell Analysis, Visual Quitlance

limensionality present tractability challenges for computation, design, use a complex combination of both developing clusters and organizing and interaction [29]. One frequently used method of visual abstraction observations in space in the sensemaking process [76] as they explore is to reduce a high-dimensional dataset into a low-dimensional space while preserving properties of the high-dimensional structure (e.g., retain or respect pairwise relationships from the higher dimensions in the display, as well as clusters that naturally develop due to expressive the lower dimensional projection). Such dimension reduction algointeractions updating the underlying layout (these interaction types are

interaction types are

TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. DL. NO. 1, JANUARY 2018

CyteGuide: Visual Guidance for Hierarchical Single-Cell Analysis

IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, VOL. 24, NO. 1, JANUARY 2016

Index Terms-Dimension reduction, clustering, algorithms, visual analytics

and Clustering in Visual Analytics John Wenskovitch, Student Member, IEEE, Ian Crandell, Naren Ramakrishnan, Member, IEEE,

Leanna House, Scotland Leman, Chris North



# \* Dimensionality Reduction





