

# Vicinity Exploration: Enabling User-Driven Visual Search of Multiple Machine Learning Models for Precision Medicine

Suresh K. Bhavnani PhD<sup>1</sup>, Archana Ayyaswamy MS<sup>1</sup>, Tianlong Chen BS<sup>2</sup>,  
Shyam Visweswaran MD PhD<sup>3</sup>, Gowtham Bellala PhD<sup>4</sup>, Kevin E. Bassler PhD<sup>2</sup>

<sup>1</sup>Institute for Translational Sciences, University of Texas Medical Branch, Galveston, TX;

<sup>2</sup>Department of Physics, University of Houston, Houston, TX, <sup>3</sup>Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, PA, <sup>4</sup>C3 IoT, Redwood City, CA

## Abstract

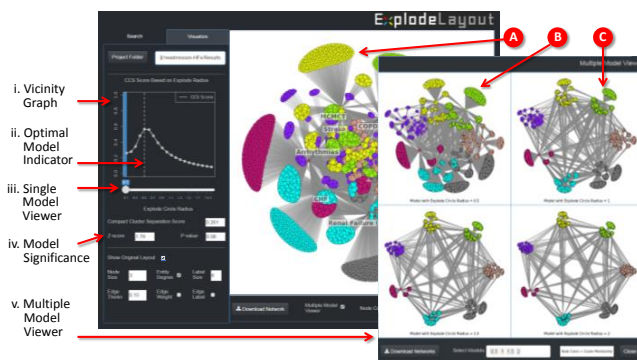
Although machine learning methods (e.g., cluster analysis) are increasingly being integrated into visual analytical applications for identifying complex patterns (e.g., patient subgroups) in large data sets, such approaches typically generate a single best model based on optimizing an objective function. However, comparison of models in the vicinity of the optimal model can enable analysts to explore tradeoffs among key model parameters, important when exploring large datasets. Here we describe the user interface features critical for *vicinity exploration*, and demonstrate its efficacy in the use of a network layout algorithm to explore a large dataset related to precision medicine.

## Introduction

Machine learning (ML) methods such as cluster analysis and feature selection are increasingly being integrated into visual analytical applications to enable the rapid identification and comprehension of patterns in large datasets. However, such methods typically output a single best model determined by optimizing an objective function. For example, the *ExplodeLayout*<sup>1</sup> algorithm attempts to enhance the comprehension of networks that have significant but highly overlapped clusters (Fig. 1A). The algorithm takes as input a network layout (e.g., generated from a layout algorithm such as Fruchterman Reingold<sup>2</sup>), and node clusters (e.g., generated by a modularity algorithm<sup>2</sup>), and maps each cluster's centroid onto equidistant points on an imaginary circle. The algorithm searches for a circle radius such that the ratio of the total non-overlapped area among the clusters, and the total space required for the entire network layout is maximized, resulting in an optimal cluster separation within a compact area (Fig. 1B). Here we describe the user interface (UI) design features and utility of *vicinity exploration*, an interaction approach that enables users to explore multiple ML models in the vicinity of the optimal model such as that generated by *ExplodeLayout*, to better comprehend complex patterns in large datasets.

## Method and Results

As shown in Fig. 1, we designed five UI features for enabling vicinity exploration of an optimal model generated by ML algorithms such as *ExplodeLayout*: (i) **vicinity graph** implemented as a line graph of how the objective function changes with respect to the circle radius, (ii) **optimal model indicator** implemented as a dotted line on the graph to help recall the circle radius of the optimal model, (iii) **single model viewer** implemented as a scroll bar to select a circle radius and display a high resolution image of the corresponding exploded network, (iv) **model significance** implemented as progressive analysis and display of significance of vicinity models to address the high computation time in large datasets, and (v) **multiple model viewer** implemented as side-by-side views of user-selected models in the vicinity of the optimal model to reduce working memory loads during model comparisons. The above features were used by a geriatrician to comprehend patient subgroups within all 30-day readmitted hip fracture patients (n=6150), extracted from the 2010 Medicare database that had at least one of the 8 significant comorbidities shown. The optimal layout (Fig. 1B, radius=0.5) revealed 2 levels of heterogeneity in each of 7 patient subgroups: patients with one comorbidity on the outer side, and patients with more than one comorbidity on the inner side of each cluster. A higher separation of the clusters (Fig. 1C, radius=1), while not optimal based on the objective function and with lower resolution of the clusters, enabled the



**Figure 1.** Five UI features of *vicinity exploration* implemented for the *ExplodeLayout* algorithm to help comprehend patient subgroups.

comprehension of inter-cluster edges showing how patients shared comorbidities across clusters. This led to hypotheses of how renal disease and diabetes interacted with other comorbidities precipitating hospital readmission in each patient subgroup.<sup>1</sup> Future research will test the generality and usability of the above features in other ML-integrated visual analytical applications.

## References

1. Bhavnani SK, et al. Enabling Comprehension of Patient Subgroups and Characteristics in Large Bipartite Networks: Implications for Precision Medicine. Proc. of AMIA Summit on TBI (2017).
2. Newman M. Networks: An Introduction. Oxford University Press; 2010.