Radial Clustergrams: Visualizing the Aggregate Properties of Hierarchical Clusters

Dimitris K. Agrafiotis,* Deepak Bandyopadhyay, and Michael Farnum

Johnson & Johnson Pharmaceutical Research & Development, L.L.C., 665 Stockton Drive, Exton, Pennsylvania 19341

Received October 3, 2006

A new radial space-filling method for visualizing cluster hierarchies is presented. The method, referred to as a radial clustergram, arranges the clusters into a series of layers, each representing a different level of the tree. It uses adjacency of nodes instead of links to represent parent−child relationships and allocates sufficient screen real estate to each node to allow effective visualization of cluster properties through color-coding. Radial clustergrams combine the most appealing features of other cluster visualization techniques but avoid their pitfalls. Compared to classical dendrograms and hyperbolic trees, they make much more efficient use of space; compared to treemaps, they are more effective in conveying hierarchical structure and displaying properties of nodes higher in the tree. A fisheye lens is used to focus on areas of interest, without losing sight of the global context. The utility of the method is demonstrated using examples from the fields of molecular diversity and conformational analysis.

INTRODUCTION

Clustering is a common technique used to partition a set of data points into groups (clusters), so that the points in each group share some common characteristics—typically proximity according to some distance or similarity measure. The technique has found numerous applications in chemoinformatics, as it allows the analyst to reduce the complexity of large chemical data sets to a more manageable size.[1−4] Clustering algorithms fall into two broad categories: hierarchical, which partition the data by successively applying the same process to clusters discovered in previous iterations, and partitional, which determine the clusters in a single step.[5] Hierarchical methods can be agglomerative (bottom-up) or divisive (top-down). The former begin with each point in a separate cluster and merge them into successively larger clusters, whereas the latter start with the whole set and successively split it into smaller and smaller groups. Hierarchical algorithms appear to be more widespread, perhaps because they assume very little a priori knowledge on the part of the analyst. The output of a hierarchical algorithm is a tree structure, representing the hierarchical classification of the data.

A good visualization of a hierarchical clustering tree should engage the pattern recognition capabilities of the human visual system and enable viewers to see at a glance the size of a particular cluster and its proximity to other clusters and to distinguish related and unrelated data items. Many efforts to visualize the results of hierarchical clustering have drawn from the extensive literature on graph drawing and tree layout algorithms. Lacking the space to properly survey these methods here, we refer the reader to two relatively recent reviews. Herman[6] surveys algorithms for graph and tree visualization and navigation, including a section on hierarchical clusters; Strehl and Ghosh[7] include a complementary overview of algorithms for visualizing nonhierarchical clusters, such as self-organizing maps,[8] and the authors' own CLUSION permuted similarity matrix plot.

Dendrograms have classically been the most popular cluster visualization method. This layout visually emphasizes both the neighbor relationship between data items in clusters (horizontal) and the number of levels in the cluster hierarchy (vertical). Several software packages have been released that use dendrograms for cluster visualization, including Spotfire,[9] Treeview,[10] GeneMaths,[11] and Expression Profiling/EP-CLUST.[12] As in other tree layout methods, the difficulty in displaying a dendrogram increases with the number of nodes; for instance, the statistical package Stata will draw a dendrogram only if it has 100 or fewer nodes. The Hierarchical Clustering Explorer program[13,14] implements a compressed dynamic dendrogram view where users can restrict the number of clusters they view at a time and show those clusters more clearly.

Treemaps[15] are a technique for visualizing large trees (originally directory trees) in a space-filling manner. They recursively slice and dice the screen space available using vertical and horizontal rectangles at alternating levels of the hierarchy, each with a thickness proportional to the size of the node that it represents. Treemaps have been used to visualize hierarchical clusters of gene expression data[16] and gene ontologies.[17] A variant of treemaps called quantum treemaps[18] underlies a rectangle-packing hierarchy visualization called HeiankyoView[19] that has found use in chemoinformatics.[20,21]

The treemap concept was applied to radial space-filling visualization of hierarchical clusters in the sunburst,[22] and independently in the aggregate treemap[23] and the information slice.[24] These methods draw a concentric pie chart where each successive pie is split into sectors that correspond to the clusters at that level of the hierarchy. In the application to directory tree visualization described by Stasko and Zhang,[25] sector sizes were proportional to relative file or

* Corresponding author tel.: (610) 458-6045; fax: (610) 458-8249; e-mail: dagrafio@prdus.jnj.com.

10.1021/ci600427x CCC: $37.00
Published on Web 12/24/2006

directory sizes, radii were constant, and colors indicated file type or age. The authors found that users were able to complete simple tasks faster and more accurately using the sunburst than when using treemaps. A related technique called the pietree[26,27] separates sectors by thick white borders, varies sector radii depending on hierarchy depth, and maps file size/count directly into sector area, as treemaps do for rectangles.

One limitation of radial space-filling and linear tree visualizations is the decreasing size and resolution of clusters with many nodes or deep down in the hierarchy. This limitation has been addressed by a separate detailed display of the cluster in focus,[24,28] by using hyperbolic geometry to enlarge areas of interest,[29] by showing the focus area angularly expanded or placed concentrically inside or outside the entire visualization within the same space,[25] by allowing user-controlled radial and angular distortion of the focus area,[30] and by expansion/contraction of the hierarchy levels on demand.[30,27]

Our current method, the radial clustergram, was developed independently from these older radial space-filling systems, though it has certain similarities in look and feel. Our sector angles and radii mirror those in Schonlau's linear clustergram,[31] a specialized structure for intuitive visualization of hierarchical clusters. A distinguishing feature of our implementation is the use of a magnification technique known as fisheye distortion to enlarge an area of interest without losing the broader context around it. Our color scheme, data display within sectors, and navigation user interface are tuned to the needs of chemical informatics and computational chemistry; to the best of our knowledge, this is the first application of interactive radial space-filling visualization of cluster hierarchies in this field. A radial plot illustrating a classification of natural products is found in the supporting information of reference 52. This plot was constructed manually in Excel, uses a different color-coding strategy, and is non-interactive. Last, we describe the integration of radial clustergrams in a larger data mining and visualization application known as Third Dimension Explorer (3DX).[32]

IMPLEMENTATION

Layout. The radial clustergram is a circular space-filling layout divided into a series of layers of width *dr*, each representing a different level (depth) of the tree. The plot is constructed using the following rules: (1) the center represents the root of the hierarchy; (2) deeper nodes in the hierarchy are drawn further from the center; (3) child nodes are drawn within the arc subtended by their parents; (4) the sweep angle of a node is proportional to its size (i.e., to the number of points that fall under that node), and (5) the sweep angle of a nonleaf node is the sum of all its children. Consider, for simplicity, a simple binary classification tree. The center of the plot represents the root of the tree, and the innermost layer (0, *dr*) represents its two children. The area is divided into two segments, one for each child node. The relative size of the two segments is proportional to the number of points in the corresponding nodes. Each of the segments that does not correspond to a terminal node is projected out to the next layer (*dr*, *dr* + *dr*) and is further subdivided into two narrower segments, on the basis of the relative sizes of their respective children. This process is repeated recursively until all the terminal nodes in the tree have been processed. Thus, the maximum radius of the clustergram is equal to $(d - 1)dr$, where d is the depth of the tree. Although the preceding description pertains to binary trees, extension to *n*-ary trees is straightforward.

Color Coding. Each segment can be color-coded by a user-defined property of the elements in that node, typically using an aggregation function like the average, minimum, maximum, and so forth. The advantage of this approach is that the property used for color-coding need not be related to the method used to construct the cluster hierarchy. For example, one can cluster a set of compounds on the basis of their molecular similarity and color-code the resulting clusters on the basis of the average activity of the compounds in a particular biological assay system. Nodes with null properties are typically colored in gray (e.g., clusters which consist exclusively of compounds whose biological activity has not been measured).

A radial clustergram can display both unpruned and pruned trees. An unpruned tree is a tree whose terminal nodes are all singletons. Since the purpose of clustering is to divide the data points into a set of relatively homogeneous groups, several pruning algorithms have been devised to obtain the optimal set of clusters from the complete hierarchy. A popular algorithm in chemoinformatics is Kelley's cluster level selection method, which calculates, for all levels of the hierarchy, a penalty value that is a tradeoff between intracluster and intercluster variance. The level of hierarchy that has the lowest penalty value is selected and used to partition the data set into the corresponding number of clusters. To allow the interactive exploration of pruned hierarchies, our radial clustergrams provide three viewing modes: (1) display and color-code by property every node in the full, unpruned tree, (2) display and color-code by property only the nodes in the pruned tree, and (3) display all the nodes in the tree, but color-code by property only the nodes in the pruned tree and gray out the ones that were pruned. These modes are illustrated in Figure 1a−c, respectively.

Navigation. As the number of nodes in the hierarchy increases, any visualization of the clustering is challenged to display all the information effectively. In the radial clustergram, the effect is that terminal slices near the periphery of the visualization can become too narrow to distinguish. We address this problem using a technique known as fisheye distortion. The idea, which was first introduced by Furnas,[33] is to use a zoom lens to simultaneously provide both a global and a detailed view of a structure. In many information visualization systems, contextual information is often lost when a user zooms on a point of interest. In most cases, this loss of information also causes loss of orientation. The fisheye belongs to a family of methods known as *focus-plus-context*, which allow the user to focus on a special item while capturing the broader context around it. An application of this technique in visualizing conventional tree layouts has been reported.[34]

The fisheye effect is illustrated in Figure 2. A lens is used to enlarge the area of interest, while showing other portions of the image in progressively less detail. Vertices that fall inside the lens are magnified, while the rest are extended. This is accomplished using a nonlinear distortion function, which transforms the distance of a vertex from the focal point

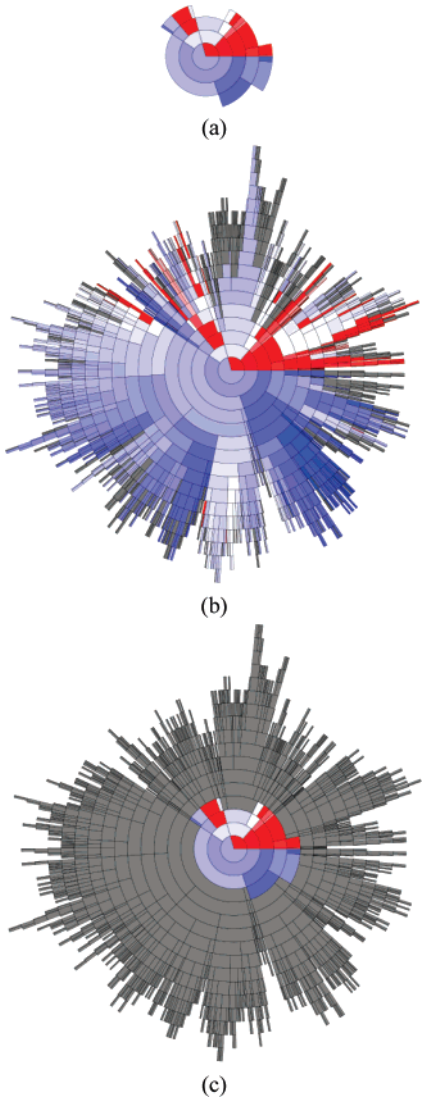

Figure 1. Three viewing modes of a radial clustergram. (a) Only the nodes in the pruned tree are displayed and color-coded by property; pruned nodes are not shown. (b) All nodes in the tree, both pruned and unpruned, are displayed and color-coded by property. (c) Pruned nodes are displayed in gray; unpruned nodes are color-coded by property.

of the lens. The function should be concave, mapping the distances monotonically in such a way that points located near the focal point of the lens are mapped further away from it. Here, we use a simple distortion function that provides a smooth magnification effect and preserves fairly well the edges of the context. If r, m, and $\mathbf{x_c}$ denote the radius, magnification, and focal point of the lens, $\mathbf{x}$ denotes the coordinates of the vertex to be transformed, and d denotes the Euclidean distance between $\mathbf{x}$ and $\mathbf{x_c}$, the transformed coordinates $\mathbf{x_t}$ of the vertex $\mathbf{x}$ are computed by

$$\mathbf{x_t} = \mathbf{x_c} + m(\mathbf{x} - \mathbf{x_c}) \qquad \text{if } d \le r$$

and

$$\mathbf{x_t} = \mathbf{x_c} + [1 + r(m - 1)/d](\mathbf{x} - \mathbf{x_c}) \qquad \text{if } d > r$$

(Interestingly, if the magnification factor is less than 1, the lens behaves like a "black hole" consuming the space around it; although this is a useless effect from a practical point of view, it is quite entertaining to watch.) Other types of lenses are described in the work of Carpendale and Montagnese.[35]

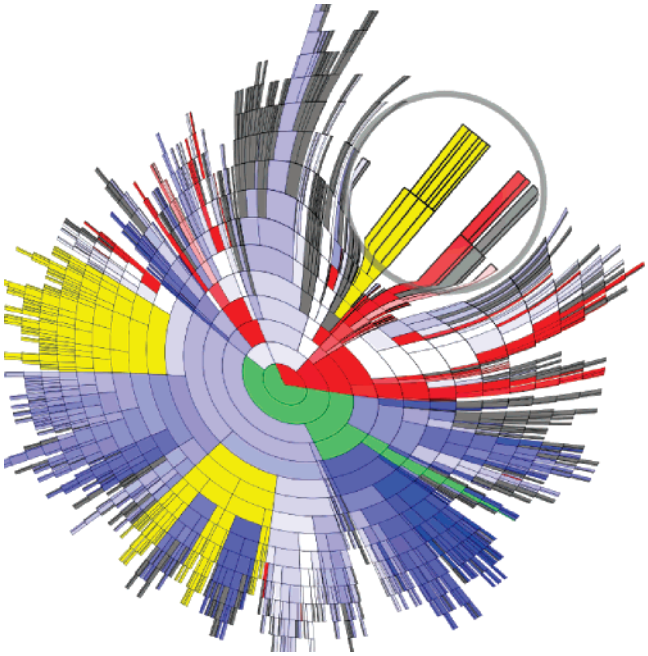

Figure 2. Magnification of an area of interest using the fisheye lens.

While fisheye is an extremely powerful technique, proper implementation requires attention to detail. The fisheye distorts the position of each vertex. If the distortion is to be applied faithfully, the edges connecting the vertices must also be distorted. Mathematically, the result of that distortion is a general curve. To achieve the correct result, long line segments must be divided into short ones, a process known as convex segmentation. These transformations are implemented extremely efficiently in the underlying graphics library (vide infra), resulting in very fast redraw rates and a smooth transition as the user interactively explores different areas of the clustergram.

User Interaction. Our radial clustergram offers a full set of navigation and selection options through event bindings (see Table 1). Core to the radial clustergram, and indeed any data mining and visualization technique, is the ability to browse through the records and select a subset of them for further analysis and/or processing. The user interface provides a mechanism to select multiple nodes at any level of the hierarchy. A selection event (selection/deselection) is simultaneously applied to a node and all of its children. Selecting a parent node and then deselecting some of its children provides a simple and fast mechanism to select any subset of clusters in the hierarchy. Another useful concept for data browsing is that of the current record. The current record is the data object that has the focus and is highlighted with a specific color in all (linked) visualizations of the same data set. In radial clustergrams, only singleton nodes can be set as current. The current node is highlighted in green (along with its entire path to the root); selected nodes are highlighted in yellow, and if the current node is also selected, it is highlighted with a color between yellow and green (Figure 3).

Implementation. Our implementation is based on Shemanarev's Anti-Grain Geometry (AGG) library.[36] AGG is a

Table 1. Event Bindings

event	action
MOUSE_DOWN	clear previous selection and select the node under the mouse and all of its offspring
CTRL_MOUSE_DOWN	select/unselect the node under the mouse and all of its offspring without changing the pre-existing selection
"S"	select the terminal node containing the centroid of the cluster under the mouse
"C"	set as current the terminal node containing the centroid of the cluster under the mouse, and highlight the entire path leading to it from the root in green
CTRL_M	loop through the three viewing modes (show pruned tree only, show entire tree with pruned nodes in gray, show entire tree with all nodes color-coded by property)
CTRL_Z	toggle fisheye lens on/off
SHIFT_MOUSE_DOWN[a]	move lens to new mouse position
SHIFT_MOUSE_MOVE[a]	drag lens to new mouse position
UP_ARROW[a]	increase the radius of the lens by 10% (20% if CTRL is pressed)
DOWN_ARROW[a]	decrease the radius of the lens by 10% (20% if CTRL is pressed)
LEFT_ARROW[a]	increase the magnification of the lens by 10% (20% if CTRL is pressed)
RIGHT_ARROW[a]	decrease the magnification of the lens by 10% (20% if CTRL is pressed)
DOWN[a]	reset lens to default size and magnification

[a] These bindings are only available when the fisheye lens is on.

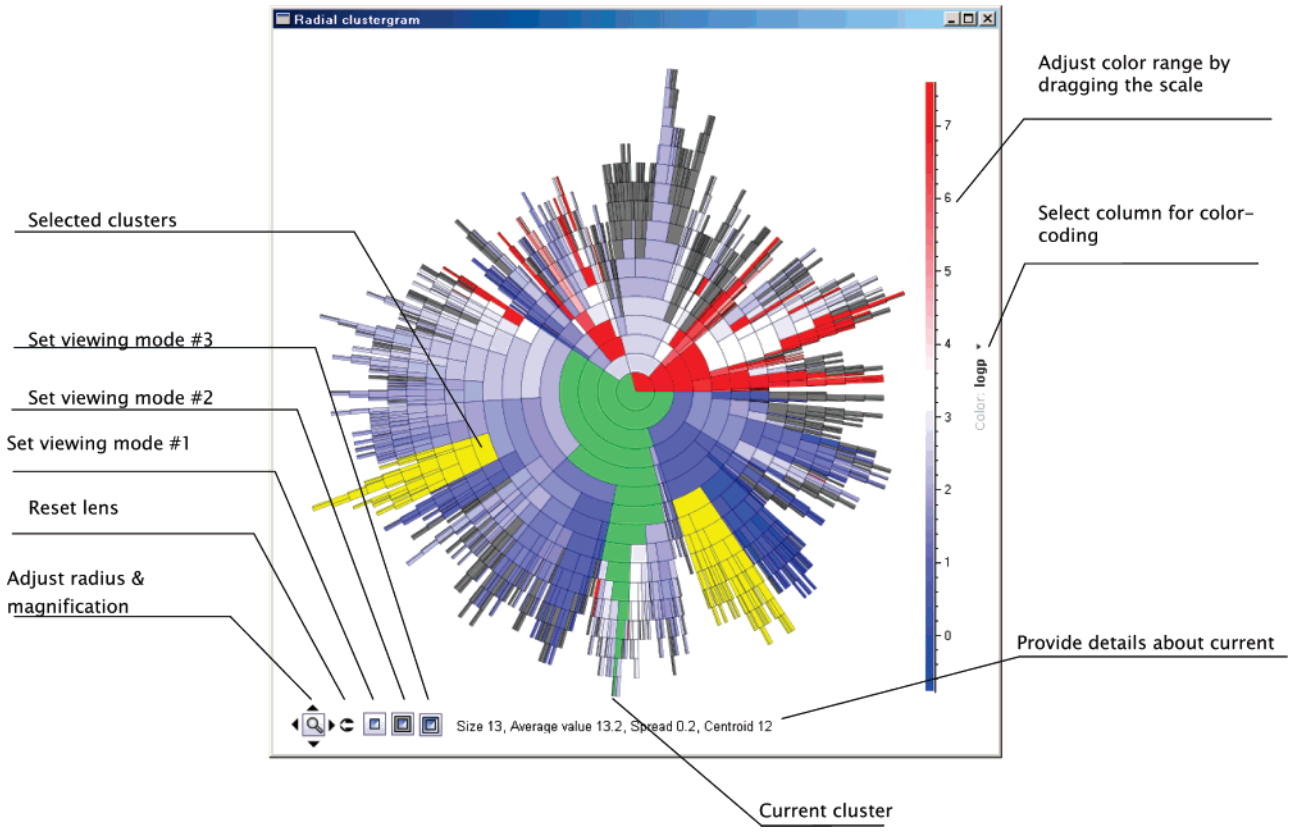

Figure 3. .NET control, including UI components for adjusting the color scale, setting lens properties, and switching between viewing modes.

high-quality, lightweight, extensible, and platform-independent rendering engine written in standard ANSI C++, which provides very fast antialiased graphics with subpixel accuracy. The radial clustergram is implemented as a .NET control (Figure 3) and is built upon a .NET version of the AGG library written in C++/CLI (available with .NET 2.0). The control is fully integrated with 3DX,[32] a state-of-the-art application that facilitates the reduction, upload, retrieval, mining, visualization, and reporting of drug discovery data, and is an integral part of J&JPRD's new pharmaceutical research informatics platform known as ABCD.[37]

DISCUSSION

We apply radial clustergrams to visualize two different types of chemical data: the diversity of chemical libraries[38,39] and the conformational space of small organic molecules.[40] Visualization is an important component of diversity analysis. The most common techniques are simple histograms and multivariate scatter plots, Kohonen networks,[41,42] and nonlinear maps.[43−45] Kohonen networks are perhaps the most similar to the current method, as they are space-filling plots that convey cluster proximity. Unlike radial clustergrams or treemaps, however, they do not convey hierarchical relationships and cannot visualize more than one property at a time because the cells are constrained on a fixed lattice. A radial clustergram uses sweep angle and color to represent two properties (one of them typically being, but not restricted to, the size of the cluster), angular proximity to represent cluster similarity, and radial extension to represent hierarchy.

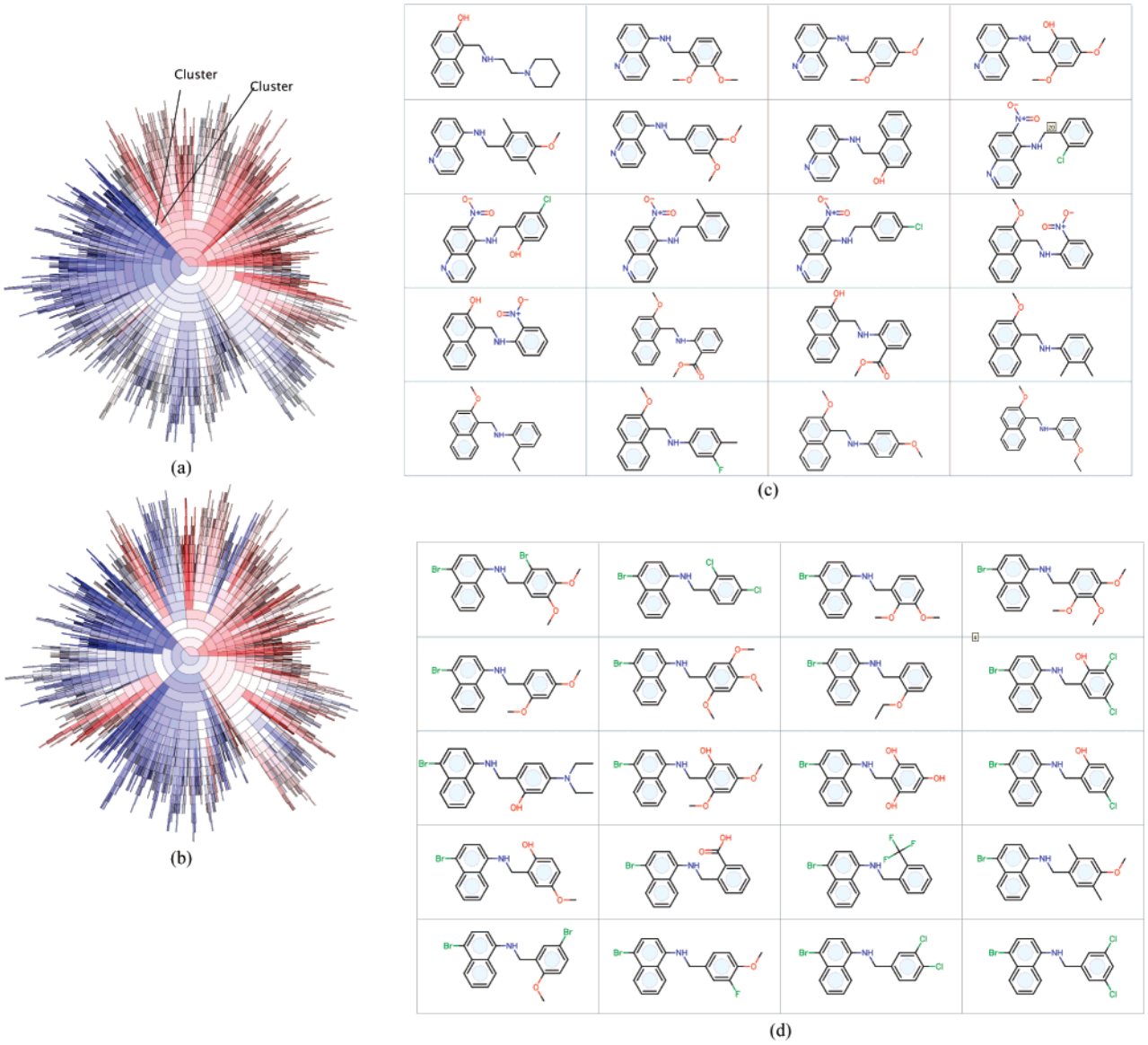

Figure 4. Unpruned radial clustergrams of the amination library, using the mean for aggregation. (a) Color-coding by molecular weight; (b) color-coding by logP; (c) 20 representative structures in cluster "A"; (d) 20 representative structures in cluster "B".

Despite arguments by Grady et al. that similar pie-chart views are inefficient,[46] our radial clustergram greatly improves the information density in the available space compared to other viewers. The interface provides the flexibility to display as many levels as needed, even if the underlying data has deeper structure. Additionally, the fisheye lens allows the viewing of details even when the number of nodes is quite large.

Our first example in Figure 4 shows the radial clustergram of a combinatorial library containing 2500 compounds assembled by combining 50 amines and 50 aldehydes by reductive amination. Each of the products was described by 117 topological descriptors including molecular connectivity indices, κ shape indices, subgraph counts, information-theoretic indices, Bonchev−Trinajstić indices, and topological state indices.[47] These descriptors were normalized and decorrelated using principal component analysis, resulting in an orthogonal set of 10 latent variables, which accounted for 95% of the total variance in the data. The data was clustered using a very fast implementation of the bisecting *k*-means clustering algorithm,[48,49] using the Euclidean distance in the 10-dimensional principal component space as a measure of similarity.

The unpruned radial clustergram of the resulting cluster hierarchy is shown in Figure 4, color-coded by the average molecular weight (a) and logP (b). Color-coding ranges from red (high values) to blue (low values) through white (average values). The two plots have a very similar appearance, which reflects a significant correlation between molecular weight and logP ($R^2 = 0.43$). One can easily distinguish major families and subfamilies of molecules with related structures and properties at all levels of the hierarchy. Sharp color changes across cluster boundaries reveal structurally related chemical families with distinctly different properties. An example is the two clusters designated as "A" and "B" in Figure 4. Twenty-five representative structures from each of these clusters are shown in Figure 4c and d, respectively. While all these molecules share a common topology, which explains their proximity in diversity space, the ones in the second cluster contain several halogens and at least one bromine atom, which increases both their molecular weight and logP. None of the molecules in the first cluster contain

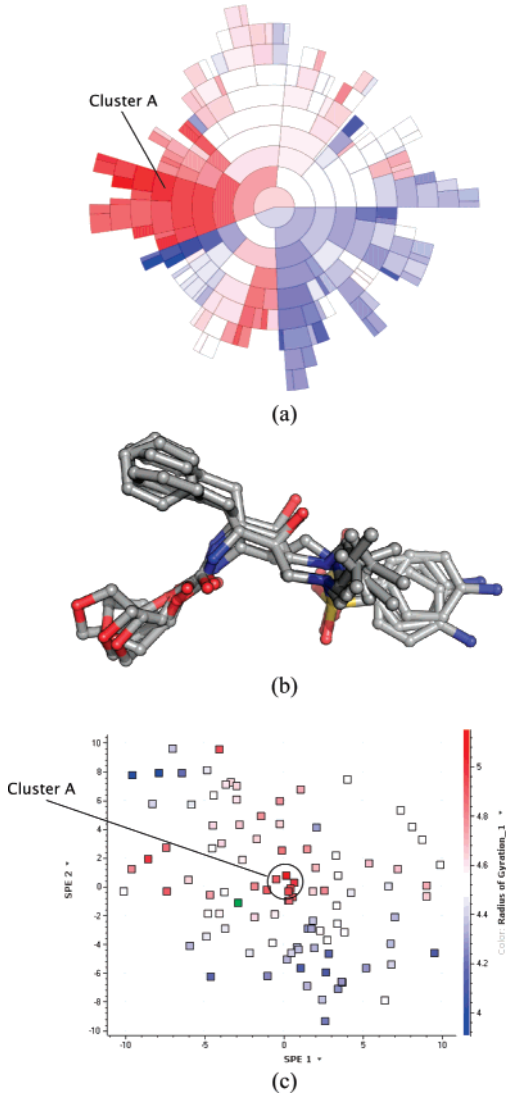

Figure 5. Unpruned radial clustergrams of the conformational space of amprenavir, using the mean for aggregation and color-coded by the radius of gyration. (a) Radial clustergram; (b) aligned conformations that belong to the highlighted cluster "A"; (c) SPE map obtained with a neighborhood radius of 1.7 RMSD; the compounds in the circle belong to cluster "A".

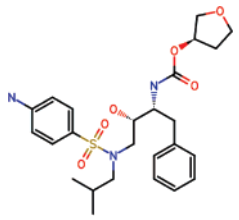

Figure 6. Structure of amprenavir.

bromine, and none of them carries more than one light halogen (F or Cl).

The second example in Figure 5 illustrates the use of radial clustergrams for visualizing conformational space. The data set consists of 100 random conformations of the HIV protease inhibitor amprenavir (Figure 6), generated by a distance geometry algorithm known as stochastic proximity embedding (SPE)[50,51] and minimized using the MMFF94s force field. Each pair of conformers was superimposed using a least-squares fitting procedure, and the resulting root-mean-square deviation (RMSD) was used as a measure of the similarity between the two conformations. The radial clustergram color-coded by the radius of gyration (a measure of the extendedness or compactness of the conformation) is shown in Figure 5a. Figure 5c shows a SPE (nonlinear) map of the resulting conformations, derived by embedding the conformations in a two-dimensional space in such a way that the distances of the points on the map are proportional to the RMSD distances of the respective conformations. Only local proximities (RMSD < 1.7) are preserved, in order to better capture the nonlinear geometry of the data.[43] A representative cluster of four extended conformations with a large radius of gyration (bright red) is highlighted in both visualizations, and the corresponding aligned structures are shown in Figure 5b. Again, the major conformational families are clearly delineated, with color-coding providing valuable additional context.

CONCLUSIONS

Radial clustergrams are an intuitive method for visualizing cluster hierarchies that make very efficient use of real estate. Unlike conventional dendrograms where the terminal nodes are arranged in a linear fashion, radial clustergrams arrange the nodes circularly and thus require significantly less footprints. Parent−child relationships are clearly discerned, as is the relative size of the clusters at all levels of the hierarchy. Color-coding allows the user to understand how aggregate properties are propagated up the hierarchical structure, and a fisheye lens is used to magnify areas of interest without losing sight of the greater context. We believe that this component, when properly implemented and integrated into an interactive data analytic application with dynamically linked displays, can be very effective in visualizing large cluster hierarchies.

ACKNOWLEDGMENT

We thank Dr. Fangqiang Zhu for many useful discussions and for critically reviewing this manuscript.

REFERENCES AND NOTES

(1) Willett, P. *Similarity and Clustering in Chemical Information Systems*; Research Studies Press: Letchworth, England, 1987.
(2) Downs, G. M.; Willett, P. Similarity Searching and Clustering of Chemical-Structure Databases Using Molecular Property Data. *J. Chem. Inf. Comput. Sci.* **1994**, *34*, 1094−1102.
(3) Brown, R. D.; Martin, Y. C. Use of Structure−Activity Data To Compare Structure-Based Clustering Methods and Descriptors for Use in Compound Selection. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 572−584.
(4) Bocker, A.; Derksen, S.; Schmidt, E.; Teckentrup, A.; Schneider, G. A Hierarchical Clustering Approach for Large Compound Libraries. *J. Chem. Inf. Model.* **2005**, *45*, 807−815.
(5) Jain, A. K.; Murty M. N.; Flynn P. J. Data Clustering: A Review. *ACM Computing Surveys*, **1999**, *31* (3), 264−323. Available at http://citeseer.ist.psu.edu/jain99data.html (accessed Sept 10, 2006).
(6) Herman, I. Graph Visualization and Navigation in Information Visualization: A Survey. *IEEE Trans. Vis. Comp. Graph.* **2000**, *6* (1), 24−43.
(7) Strehl, A.; Ghosh, J. Relationship-Based Clustering and Visualization for High-Dimensional Data Mining. *INFORMS J. Comput.* **2003**, *15* (2), 208−230. Available at http://citeseer.ist.psu.edu/strehl02relationshipbased.html (accessed Sept 20, 2006).
(8) Kohonen, T. *Self-Organizing Maps*; Springer: Berlin, Germany, 1995.
(9) Spotfire. *www.spotfire.com* (accessed Sept, 2006).
(10) Eisen M. B.; Spellman, P. T.; Brown, P. O.; Botstein, D. Cluster Analysis and Display of Genome-Wide Expression Patterns. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 14863−14868.

(11) GeneMaths, Applied Maths. http://www.applied-maths.com/ge/ge.htm (accessed Sept 20, 2006).

(12) Expression Profiler for Clustering, Analysis and Visualization of Gene Expression and Other Genomic Data. http://ep.ebi.ac.uk (accessed Sept 22, 2006).

(13) Seo, J.; Shneiderman, B. *Understanding Hierarchical Clustering Results by Interactive Exploration of Dendrograms: A Case Study with Genomic Microarray Data*; University of Maryland Technical Report HCIL-2002-10, CS-TR-4370, UMIACS-TR-2002-50, ISR-TR-2005-29; University of Maryland: College Park, MD, 2001.

(14) Seo, J.; Shneiderman, B. Interactively Exploring Hierarchical Clustering Results. *IEEE Comput.* **2002**, *35* (7), 80−86.

(15) Shneiderman, B. Tree Visualization with Tree-Maps: 2-d Space-Filling Approach. *ACM Trans. Graphics* **1992**, *11* (1), 92−99.

(16) McConnell, P.; Johnson, K.; Lin, S. Applications of Treemaps to Hierarchical Biological Data. *Bioinformatics*, **2002**, *18* (9), 1278−1279.

(17) Babaria, K. Using Treemaps to Visualize Gene Ontologies. Preliminary report, 2001. Available at http://www.cs.umd.edu/hcil/treemap/GeneOntologyTreemap.pdf (accessed Sept 18, 2006).

(18) Bederson, B. B.; Shneiderman, B.; Wattenberg, M. Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies. *ACM Trans. Graphics* **2002**, *21* (4), 833−854.

(19) Itoh, T.; Yamaguchi, Y.; Ikehata, Y.; Kajinaga, Y. Hierarchical Data Visualization Using a Fast Rectangle-Packing Algorithm. *IEEE Trans. Vis. Comput. Graphics* **2004**, *10* (3), 302−313.

(20) Itoh, T.; Yamashita, F. Visualization of Multi-Dimensional Data of Bioactive Chemicals Using a Hierarchical Data Visualization Technique "Heiankyo View". In *Proceedings of the 60th Conference on Information Visualization*, Tokyo, Japan, 2006; pp 23−29.

(21) Yamashita, F.; Itoh, T.; Hara, H.; Hashida, M. Visualization of Large-Scale Aqueous Solubility Data Using a Novel Hierarchical Data Visualization Technique. *J. Chem. Inf. Model.* **2006**, *46* (3), 1054−1059.

(22) Stasko, J.; Catrambone, R.; Guzdial, M.; McDonald, K. An Evaluation of Space-Filling Information Visualizations for Depicting Hierarchical Structures. *Int. J. Human-Comput. Stud.* **2000**, *53*, 663−694.

(23) Chuah, M. C. Dynamic Aggregation with Circular Visual Designs. In *Proceedings of the 1998 IEEE Information Visualization Symposium*, Raleigh Durham, NC, 1998; pp 35−43. Available at http://citeseer.ist.psu.edu/chuah98dynamic.html (accessed Sept 14, 2006).

(24) Andrews, K.; Heidegger, H. Information Slices: Visualising and Exploring Large Hierarchies Using Cascading, Semicircular Discs. *Proceedings of the 1998 IEEE Information Visualization Symposium*, Raleigh Durham, NC, 1998; pp 9−12.

(25) Stasko, J.; Zhang, E. Focus+Context Display and Navigation Techniques for Enhancing Radial, Space-Filling Hierarchy Visualizations. *Proceedings of the 2000 IEEE Information Visualization Conference*, Salt Lake City, UT, 2000.

(26) Dix, A.; Beale, R.; Wood, A. Architectures to Make Simple Visualisations Using Simple Systems. In *Proceedings of the Working Conference on Advanced Visual Interfaces*, Palermo, Italy, 2000; ACM Press: New York, 2000; pp 51−60.

(27) O'Donnell, R.; Dix, A.; Ball, L. J. Exploring the Pietree for Representing Numerical Hierarchical Data. In *Proceedings of the ACM Conference on Human-Computer Interaction (HCI 2006*), London, 2006; in press.

(28) Cava, R. A.; Luzzardi, P. R. G.; Freitas, C. M. D. S. The Bifocal Tree: A Technique for the Visualization of Hierarchical Information Structures. In *Workshop on Human Factors in Computer Systems (IHC2002*), Fortaleza, Brazil, 2002.

(29) Lamping, J.; Rao, R.; Pirolli, P. A Focus+Context Technique Based on Hyperbolic Geometry for Visualizing Large Hierarchies. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Denver, Colorado, May 7−11, 1995; ACM Press/Addison-Wesley: New York, 1995; pp 401−408.

(30) Yang, J.; Ward, M. O.; Rundensteiner, E. A. InterRing: An Interactive Tool for Visually Navigating and Manipulating Hierarchical Structures. In *Proceedings of the IEEE Symposium on Information Visualization (INFOVIS 2002*); IEEE Computer Society: Washington, DC, 2002; p 77.

(31) Schonlau. M. The Clustergram: A Graph for Visualizing Hierarchical and Nonhierarchical Cluster Analyses. *Stata J.* **2002**, *2* (4), 391−402.

(32) Manuscript in preparation.

(33) Furnas, G. W. Generalized Fisheye Views. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '86*); ACM Press: Boston, Massachusetts, 1986; pp 16−23.

(34) Tominski, C.; Abello, J.; van Ham, F.; Schumann, H. Fisheye Tree Views and Lenses for Graph Visualization. In *Proceedings of the Conference on Information Visualization (July 05 − 07, 2006). IV*; IEEE Computer Society: Washington, DC, 2006; pp 17−24.

(35) Carpendale, S.; Montagnese, C. A Framework for Unifying Presentation Space. In *Proceedings of the 14th Annual ACM Symposium on User Interface, Software and Technology (UIST 2001*); ACM Press: Orlando, Florida, 2001; pp 61−70.

(36) Shemanarev, M. The Anti-Grain Geometry Project. http://www.antigrain.com (accessed June 1, 2006).

(37) Uehling, M. D. ABCD: How to Spell Discovery. BioIT World Magazine **2004**, *6* (June). Available online at http://www.bio-itworld.com/archive/061704/discovery.html (accessed June 20, 2006).

(38) Agrafiotis, D. K. Diversity of Chemical Libraries. In *The Encyclopedia of Computational Chemistry*; Schleyer, P. v. R., Allinger, N. L., Clark, T., Gasteiger, J., Kollman, P. A., Schaefer, H. F., III; Schreiner, P. R., Eds.; John Wiley and Sons: Chichester, U. K., 1998; Vol. 1, pp 742−761.

(39) Agrafiotis, D. K.; Lobanov, V. S.; Salemme, F. R. Combinatorial Informatics in the Post-Genomics Era. *Nat. Rev. Drug Discovery* **2002**, *1*, 337−346.

(40) Agrafiotis, D. K.; Rassokhin, D. N.; Lobanov, V. S. Multidimensional Scaling and Visualization of Large Molecular Similarity Tables. *J. Comput. Chem.* **2001**, *22* (5), 488−500.

(41) Bauknecht, H.; Zell, A.; Bayer, H.; Levi, P.; Wagener, M.; Sadowski, J.; Gasteiger, J. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 1205−1213.

(42) Wagener, M.; Sadowski, J.; Gasteiger, J. *J. Am. Chem. Soc.* **1995**, *117*, 7769−7775.

(43) Agrafiotis, D. K.; Xu, H. A Self-Organizing Principle for Learning Non-Linear Manifolds. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15869.

(44) Agrafiotis, D. K. Stochastic Proximity Embedding. *J. Comput. Chem.* **2003**, *24*, 1215−1221.

(45) Agrafiotis, D. K.; Xu, H. A Geodesic Framework for Analyzing Molecular Similarities. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 475−484.

(46) Grady, N.; Flanery, R.; Donato, J.; Schryver, J. Issues in Time-Series and Categorical Data Exploration. In *Information Visualization in Data Mining and Knowledge Discovery*; Fayyad, U. M., Grinstein, G. G., Wierse, A., Eds.; Morgan Kaufmann: San Francisco, CA, 2001; pp 229−235.

(47) Hall, L. H.; Kier, L. B. The Molecular Connectivity Chi Indexes and Kappa Shape Indexes in Structure-Property Modeling. In *Reviews in Computational Chemistry*; Boyd, D. B., Lipkowitz, K. B., Eds.; VCH Publishers: New York, 1991; pp 367−422.

(48) Steinbach, M.; Karypis, G.; Kumar, V. *A Comparison of Document Clustering Techniques*; Technical Report 00-034; Department Computer Science & Engineering: University of Minnesota, Minneapolis, MN, 2000.

(49) Downs, G. M.; Barnard, J. M. Clustering Methods and Their Uses in Computational Chemistry. In *Reviews in Computational Chemistry*; Lipkowitz, K. B., Boyd, D. B., Eds.; Wiley-VCH: Weinheim, Germany, 2002; Vol. 18, pp 1−40.

(50) Xu, H.; Izrailev, S.; Agrafiotis, D. K. Conformational Sampling by Self-Organization. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 1186−1191.

(51) Izrailev, S.; Zhu, F.; Agrafiotis, D. K. A Distance Geometry Heuristics for Biasing Conformational Sampling towards Extended or Compact Conformations. *J. Comput. Chem.* **2006**, *27* (16), 1962−1969.

(52) Koch, M. A., et al. Charting biologically relevant chemical space: a structural classification of natural products (CONP). *Proc. Natl. Acad. Sci.* **2006**, *103* (28), 10606−10611.

CI600427X