LaNet-vi in a Nutshell

Ignacio Alvarez-Hamelin [*†], Luca Dall'Asta [*], Alain Barrat [*], and Alessandro Vespignani [*‡]

January 12, 2006

Abstract

This paper contains a short introduction to the use of the LaNet-vi visualization tool for the analysis of large sparse networks. LaNet-vi is based on the k-core decomposition, whose characteristics are summarized in this paper along with the main features of the developed algorithm and layout. We elucidate the relation among the layout's features and the graph topological properties. Finally we show how LaNet-vi's visualizations can be used to exploit the network organizational features: hierarchical relationships, shells interconnectivity or nodes clustering provide global structural information at a glance, yielding a rapid method for graphs fingerprinting.

1 Why k-core decomposition?

LaNet-vi is a visualization tool for large scale networks based on the k-core decomposition [1].

Formally, the *k-core* of a graph $\mathcal{G}$ is the connected maximal induced subgraph which has minimum degree greater than or equal to k [2]. Roughly speaking, it is the maximal subgraph $\mathcal{H}$ of $\mathcal{G}$ with the property that the minimum number of edges from any vertex in $\mathcal{H}$ towards other vertices of $\mathcal{H}$ is at least k.

Starting from $k = 1$ (for graphs without isolated vertices), a simple recursive algorithm allows to obtain all k-cores of a graph.

- Every vertex of a connected graph belongs to the 1-core. In Fig.1, we have highlighted the different core using closed lines of different types. A dashed line encloses all the vertices in the 1-core (the entire graph).
- Then, all vertices of degree $d < 2$ are recursively cut out. In Fig.1 all these vertices are colored in blue. The other vertices maintain a degree $d \geq 2$ also after the pruning of the blue ones, therefore they are not eliminated. The remaining vertices form the 2-core, enclosed by a dotted line.
- Further pruning allows to identify the innermost set of vertices, the 3-core. One can check that all red vertices in Fig.1 have internal degree (i.e. between red vertices) at least 3. This core is highlighted by a dash-dotted line.

[*]Laboratoire de Physique Théorique (UMR du CNRS 8627), Bâtiment 210, Université de Paris-Sud, 91405 ORSAY Cedex France.

[†]Facultad de Ingeniería, Universidad de Buenos Aires, Paseo Colón 850, C 1063 ACV Buenos Aires, Argentina.

[‡]School of Informatics and Department of Physics, Indiana University, Bloomington, IN 47408, USA.

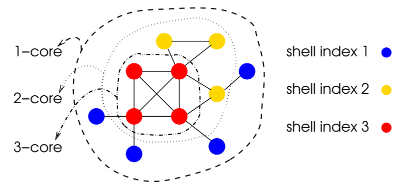

Fig.1. Sketch of the k-core decomposition for a small graph. Each closed line contains the set of vertices belonging to a given k-core, while colors on the vertices distinguish different k-shells.

The use of different colors for the vertices is useful to stress another important property: the *shell index* of a vertex. A vertex has shell index k if it belongs to the k-core but not to the $k+1$-core. A k-shell collects all vertices with the same shell index, i.e. those vertices that are pruned at the same stage of the procedure. Blue vertices in Fig.1 belong to the 1-shell, green ones to the 2-shell and the red vertices compose the 3-shell that, being the highest one, coincides with the 3-core.

In Graph Theory there are many other definitions that are usefully exploited in the analysis of social networks [3] as cliques, n-cliques, n-clans, n-clubs, k-plexes, ls-sets, etc. Many of these notions can be, in principle, used to draw a reduced representation of a graph by means of pruning or renormalization algorithms. The k-core analysis is particularly indicated for two main situations:

- the analysis of large sparse networks, since the visualization algorithm runs in a time $O(n+e)$ (where n and e are the total number of vertices and edges respectively);
- the analysis of networks with a non trivial hierarchical structure, since the vertices are displayed in different $2D$ annular shells in relation with their actual shell index, that provides a centrality measure and a manner to gain information on the hierarchical order in the network.

2 Basic features of the visualization's layout

The following list is a set of simple rules for a complete understanding of the visualization layout. They should provide the reader with a method to extract information on the original graph starting from the visual observation of the representation's properties and a way to compare different graphs. All points are illustrated with figures.

The main features of the layout's structure listed below are visible in Fig.2 where, for the sake of simplicity, we don't show any edge. The leftmost panel displays the case in which all k-cores have a single component, while in the rightmost one an example of k-core fragmentation is reported. Indeed, it is possible that, during the pruning procedure, the remaining nodes forming a k-core do not belong to the same connected component. When such a fragmentation occurs, LaNet-vi computes the separed components of the core and displays all of them in a coherent way (see below).

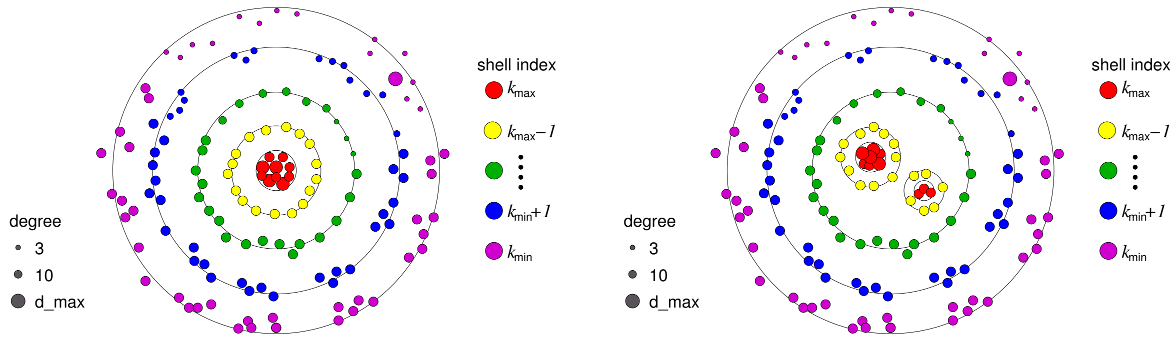

Fig.2. The two drawings show the structure of a typical LaNet-vi's layout in two important cases: on the left, all k-cores are connected; on the right, some k-cores are composed by more than one connected component. The vertices are arranged in a series of concentric shells, each one corresponding to a particular shell index. The diameter of the k-shell depends on both the index value and, in case of multiple components (right) also on the relative fraction of vertices belonging to the different components. The color of the vertices corresponds to their shell index, while their size is logarithmically proportional to their original degree.

- The **visualization's layout** is two-dimensional, composed of a series of concentric circular shells (see the five different shells in Fig.2).
- Each **shell** corresponds to a single shell index and all vertices in it are therefore drawn with the same color.
- A **color scale** allows to distinguish different values of shell index: in LaNet-vi's images, as in Fig.2, the violet is used for the minimum value of shell index k_{min}, then nuances of blue, green and yellow compose a graduated scale for higher and higher values of shell index up to the maximum value k_{max} that is colored in red.
- The **diameter** of each k-shell depends on the shell index k, and is proportionnal to $k_{max} - k$, (In Fig.2, the position of each shell is schematized by a circle having the corresponding diameter). The presence of a trivial order relation in the values of shell index ensures that all shells are placed in a concentric arrangement. On the other hand, when a k-core is fragmented in two or more components, the diameter of the different components depends also on the relative number of vertices belonging to each of them, i.e. the fraction between the number of vertices belonging to that component and the total number of vertices in that k-shell. This is a very important information, providing a way to distinguish between multiple components at a given coreness value. Looking at the two central components for high coreness values in Fig.2 (right), we immediately realize that the bigger one contains a larger fraction of vertices.
- Finally, the **size** of each node is proportional to the original degree of that vertex; we use a logarithmic scale for the size of the drawn bullets.

3 Network fingerprinting with LaNet-vi

The k-core decomposition peels the network layer by layer, revealing the structure of the different shells from the outmost one to the more internal ones. LaNet-vi provides a direct way to distinguish their different hierarchies and structural organization by means of some simple quantities: the radial width of the shells, the presence and size of clusters of vertices in the shells, the correlations between degree and shell index, the distribution of the edges interconnecting vertices of different shells, etc. The following features are useful to extract this structural information out of the laNet-vi visualization.

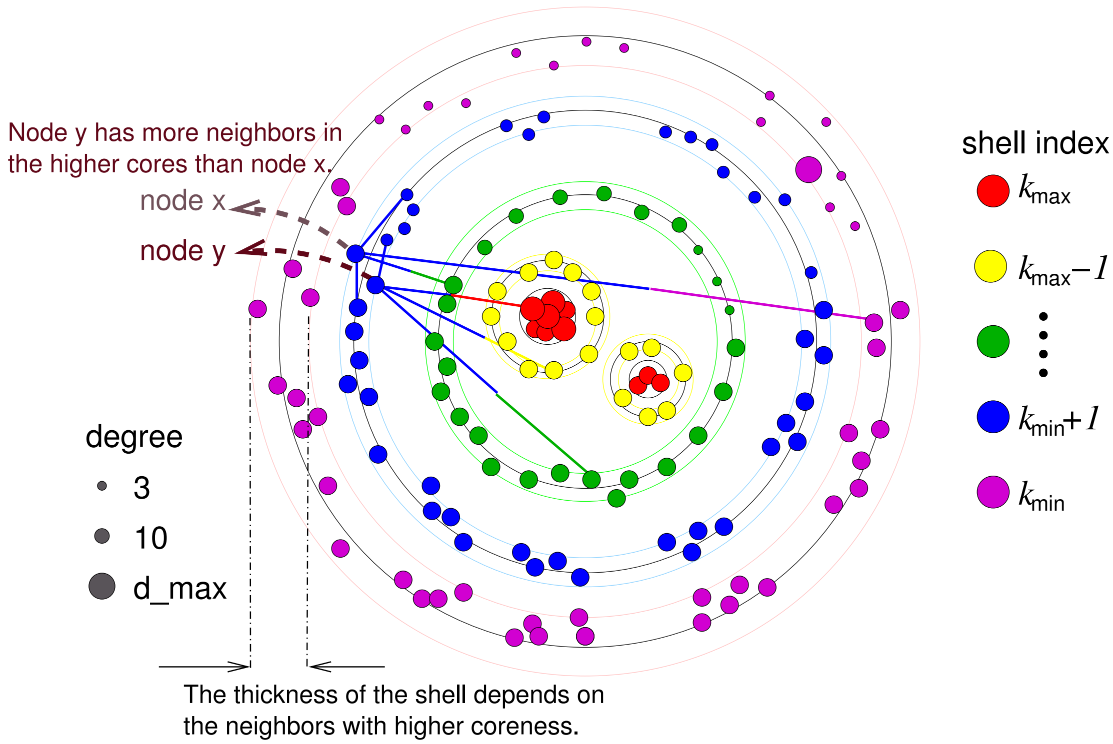

Fig.3. Each shell has a certain radial width around its diameter's values. This width depends on the correlation's properties of the vertices in the k-shell. The dashed lines in the figure point out the width of the outmost shell, that corresponds to the lowest k-shell. In the second shell, we have pinpointed two nodes x and y. The node y is more internal than x because a larger part of its neighbors belongs to higher k-shells compared to x's neighbors. Indeed, y has three links to nodes of higher shell index, while x has only one.

1) ***Shells Width***
In the LaNet-vi's representations the width can change considerably from shell to shell. The thickness of a shell depends on the shell indices of the neighbors of the vertices in the corresponding k-shell. For a given shell-diameter (corresponding to the black circle in the median position of shells in Fig.3), each vertex can be placed more internal or more external with

respect to this reference line. Nodes with more neighbors in higher k-shells are closer to the center and viceversa, as shown in Fig.3. Node y is more internal than node x because it has three edges towards nodes if higher shell index compared to the single edge emerging from x towards inner shells. The maximum thickness of the shells is controlled by the ϵ parameter (i.e., see Eq.1 of [1]).

2) ***Shell Clusters***

The angular distribution of vertices in the shells is not completely homogeneous. Fig.3 shows that clusters of vertices can be observed. The idea is that of grouping together all nodes of the same k-shell that are directly linked in the original graph and of representing them close one to another in the shell. Thus, a shell is divided in many angular sectors, each one containing a cluster of vertices. This feature allows to figure out at a glance if the k-shells are composed of a single large connected component rather than divided into many small clusters, or even if there are isolated vertices (i.e. disconnected from all other nodes in the shell, not from the rest of the k-core!).

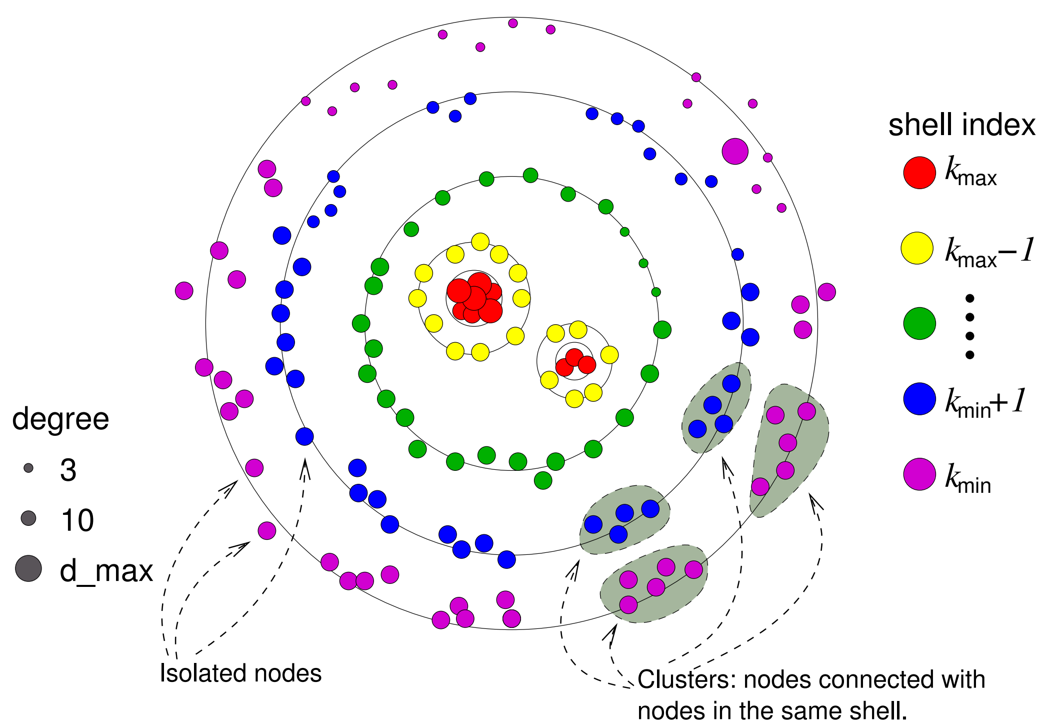

Fig.3. The figure shows the clustering properties of nodes in the same k-shell. In each k-shell, nodes that are directly connected between them (in the original graph) are drawn close one to the other, as in a cluster. Some of these sets of nodes are circled and highlighted in gray. Three examples of isolated nodes are also indicated; these nodes have no connections with the others of the same k-shell.

3) *Degree-Coreness Correlation*

Another property that can be studied from LaNet-vi's images is the correlation between the degree of the nodes and the shell index. In fact, both quantities are centrality measures and the presence or the absence of correlations between them is a very important feature characterizing a network's topology. The nodes displayed in the most internal shells are those forming the central core of the network; the presence of correlations between degree and shell index then corresponds to the fact that the central nodes are most likely high-degree hubs of the network. This effect is indeed observed in many real communication networks with a clear hierarchical structure, as the Internet at the Autonomous System level or the World Wide Air-transportation network. On the contrary, the presence of hubs in external shells is typical of networks without a clear global hierarchical structure as the World-Wide Web or the Internet Router Level. In this case, emerging star-like configurations appear with high degree vertices connected only to very low degree vertices. These vertices are rapidly pruned out in the k-core decomposition even if they have a very high degree, leading to the presence of local hub in the external k-shells. These examples are shown in the Image Gallery (http://xavier.informatics.indiana.edu/lanet-vi/#examples).

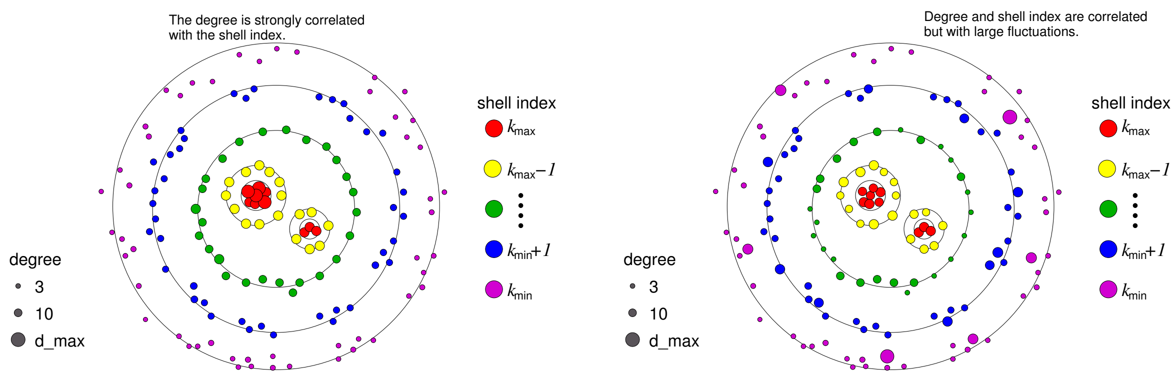

Fig.3. The two figures show different correlation properties between the shell index and the degree of the nodes. On the left, we report a graph with strong degree-shell index correlation: the size of the nodes grows going from the periphery to the center, in strong correlation with the shell index. The right-hand drawing shows a graph in which there is the degree-shell index correlations are blurred by large fluctuations, as stressed by the presence of some hubs in the external shells.

4) ***Edges***

Finally, only a homogeneously randomly sampled fraction of the edges is shown. We can tune the percentage of drawn edges in order to get the better trade-off between the clarity of visualization and the necessity of giving information on the way the nodes are mainly connected. Edge-reduction techniques can be implemented to improve the algorithm's capacity in representing edges; however, a homogeneous sampling does not alter the extraction of topological information, ensuring a low computational cost. Finally, the two halves of each edge are colored with the color of the corresponding extremities to make more evident the connection among vertices in different shells.

5) ***Disconnected components:*** The fragmentation of any given k-shell in two or more disconnected components is represented by the presence of a corresponding number of circular shells with different centers (Fig. 2). The diameter of these circles is related with the number of nodes of each component and modulated by the γ parameter (see Eqs. 9 and 10 of [1]). The distance between components is controlled by the δ parameter (Eqs. 4 and 5 of [1]).

In summary, LaNet-vi makes possible a direct, visual investigation of a series of properties:

- hierarchical structures of networks;
- connectivity and clustering properties inside a given k-shell;
- relations and interconnectivity between different levels of the hierarchy;
- correlations between degree and shell index, i.e. between different measures of centrality.

Remarks and suggestions are welcome.

References

[1] J.-I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, 'k-core decomposition: a tool for the visualization of large scale networks,' *arXiv.org:cs/0504107*.

[2] S. Seidman, 'Internal cohesion of LS sets in graphs,' *Social Networks* **5**, 97-107 (1983).

[3] S. Wasserman and K. Faust, *Social Network Analysis: Methods and Applications*, Cambridge University Press, New York and Cambridge (1994).