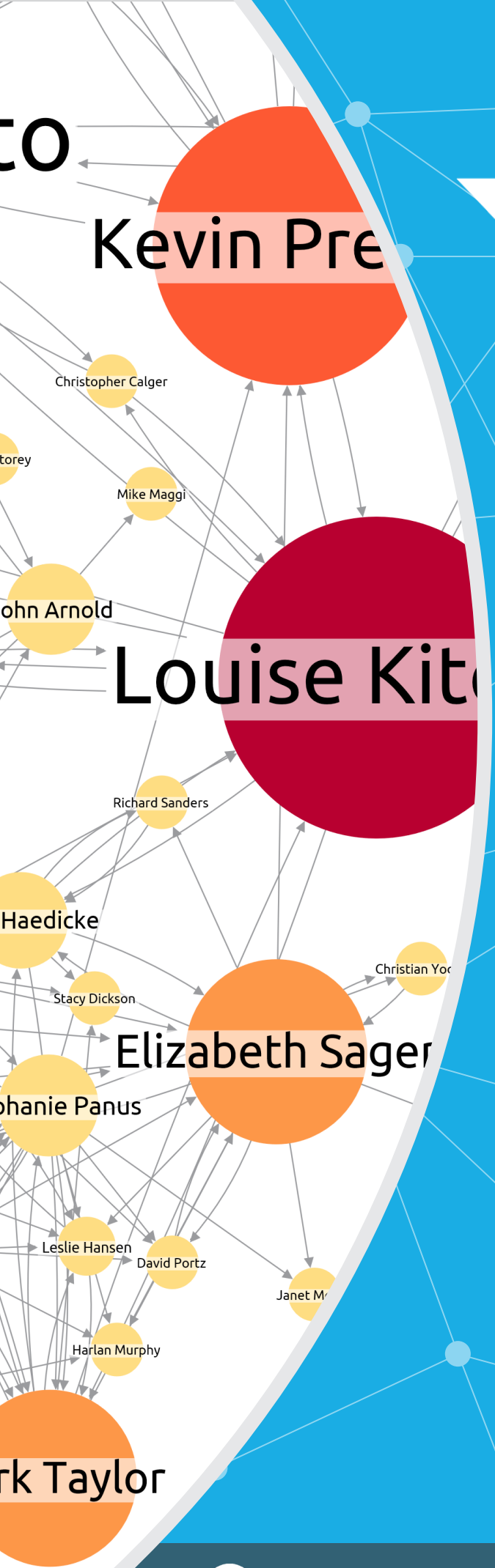


KeyLines

White Paper:

Using KeyLines to Visualize Social Networks

Contents

www.cambridge-intelligence.com **USA:** +1 (775) 842-6665 **UK:** +44 (0)1223 362 000
Cambridge Intelligence Ltd, 59 – 61 Regent Street, Cambridge, CB2 1AB, UK.

Using KeyLines to visualize Social Networks

1 What is a social network?

A social network is most easily understood as a structure of social actors joined together by connections. By visualizing and analyzing these social networks, we can observe behaviors, understand network flow, identify influence, and make predictions about how individuals or groups may behave.

These actors do not necessarily need to be people. The techniques we describe in this document can be easily applied to any kind of connected system.

We have listed some interesting use cases for visualizing social networks on our website: http://keylines.com/social-networks.

2 Breaking down a social network

Broadly speaking, social graphs can be analyzed at three different levels. Each of these will offer different kinds of insight:

- **Network** – viewing the entire network as a whole. This is useful for seeing the overall patterns in a network and to understand the network's connectivity and balance.
- **Group** – this is when a graph is viewed at a sub-network looking at one segment of the graph at a time. From this perspective, clusters of connections are visible and dynamics between groups start to become apparent. It is also at this level that we can best identify strategically important and well-connected nodes.
- **Entity** – when the social graph is viewed at a micro level – node by node. At this level, we can identify small clusters of nodes, discover bottlenecks and understand individual entities within their connected environment.

To best exploit these three levels of insight, a social network visualization tool should allow users to zoom and interact with the graph, accessing the data at the scale required.

3 Social network analysis (SNA)

Social network analysis (SNA) is the process of understanding social networks through the use of network (or 'graph') theories. Put simply, SNA seeks to explain social behavior by analyzing the structure of these networks, rather than just looking at individuals in isolation.

The practice has its roots in academic social science, but has many real-world applications. Including:

- Finding important and influential actors in a large data set
- Understanding and explaining network dynamics, at an individual and cluster level
- Implying and qualifying connections that may not be explicit in the data
- Improving the effectiveness of network flow
- Beginning to predict future behaviors
- Predicting consequences of particular scenarios

3.1 Centrality Measures

A node's centrality is a measure of its prominence, or structural importance, within a network in terms of power, communication, influence, control or status.

The KeyLines Network Visualization Toolkit includes a range of centrality measures that can be integrated into your data visualization applications.

3.1.1 Degree Centrality

Degree centrality is the simplest measure of a node's connectivity within a network. This measure can be useful in recognizing important nodes, as it quickly highlights the players that transmit a large amount of information.

KeyLines calculates a degree score based on the number of direct links a node has to other nodes – incoming, outgoing or both.

Questions that a degree measure can help to answer include:

- Who is the most / least popular person in this network?
- Who can call upon the most resource in this network?

4.1.2 Betweenness

Betweenness centrality is a way of understanding how important a node is in connecting different parts of the network.

KeyLines calculates the betweenness centrality score by identifying all of the shortest paths within a network, and then counting how many times a node falls on one. Identifying these 'bridges' between clusters allows us to disrupt or improve information flow through a network.

Betweenness centrality should be used with care: a high score could indicate an important node connecting disparate clusters, or just that a node is on the periphery of several clusters.

The measure is useful for answering questions such as:

- Who or what can most strongly control information flow around the network?
- Who or what would cause the most disruption to flow if they were removed?

4.1.3 Closeness

This measure is similar to betweenness, but instead of calculating the number of paths through each node, it calculates a node's proximity to other nodes. It does this by calculating all of the shortest paths in a network, and then assigning each node a score based on the sum of its shortest paths.

Closeness is most insightful when a network is sparsely connected. In a highly connected network you will often find all nodes have a similar score.

Closeness centrality helps you answer questions including:

- Who can most efficiently obtain information on other nodes in the network?
- Who could most quickly spread information in a network?

4.1.4 EigenCentrality

EigenCentrality is a measure of influence in a network. High EigenCentrality score indicates a strong influence over other nodes in the network. It is useful because it indicates not just direct influence,

but also implies influence over nodes more than one 'hop' away.

A node may have a high degree score (i.e. many connections) but a relatively low EigenCentrality score if many of those connections are with similarly low-scored nodes. Also, a node may have a high betweenness score (indicating it connects disparate parts of a network) but a low EigenCentrality score because it is still some distance from the centers of power in the network.

EigenCentrality can help you answer the question:

- Who or what holds wide-reaching influence in my network?
- Who or what is important in my network on a macro scale?

4.1.5 PageRank

Like EigenCentrality, PageRank can help uncover influential or important nodes whose reach extends beyond just their direct connections.

PageRank identifies the significance of nodes by assigning a score based upon the number of in-coming links (its 'indegree'). The mechanism is named after Larry Page of Google fame and was the original algorithm behind the Google search engine.

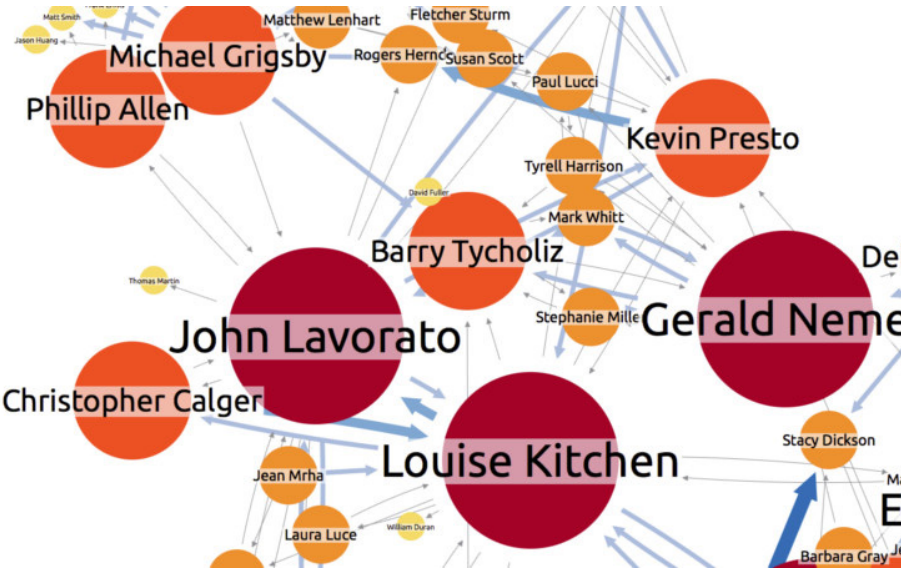
PageRank is another way of answering the same questions as EigenCentrality, i.e.:

- Who or what holds wide-reaching influence in my network?
- Who or what is important in my network on a macro scale?

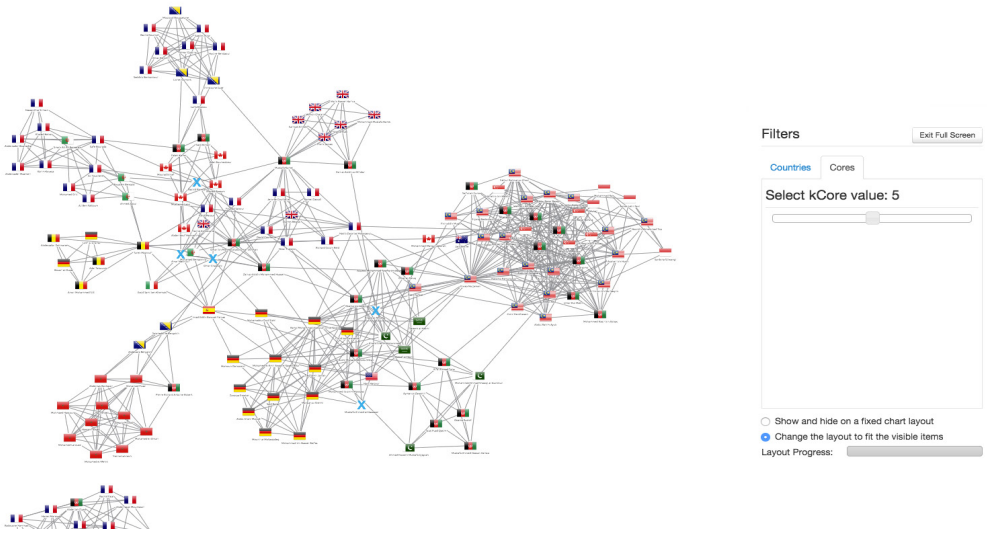
4.2 Communicating and enhancing SNA measures

Centrality information needs to be succinctly communicated to users. KeyLines offers various methods for doing this, including:

- **Formatting** – key players can be highlighted in a network by coloring or re-sizing nodes.



- **Filters** – filtering helps eliminate noise and allows focus on the connections of interest. Or by filtering by degree it's possible to reveal clusters of highly inter-connected nodes (known as kCores) – a fast and easy way of discovering sub-networks in a busy graph.



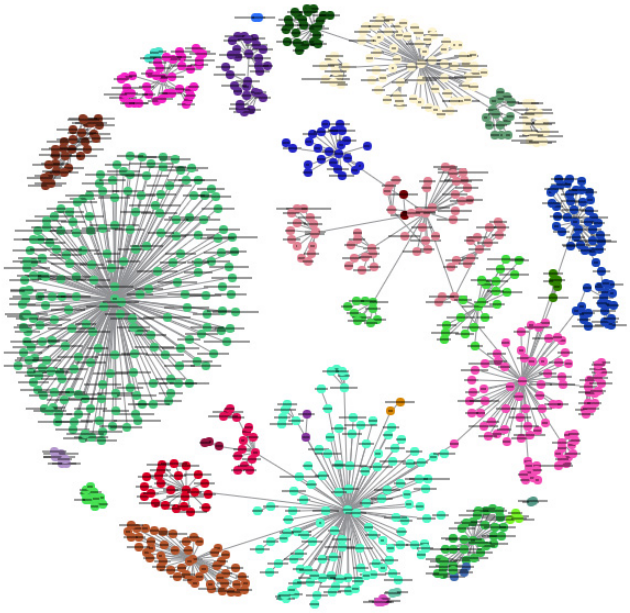
- **Animation** – animation can be applied to highlight nodes of importance, or show changes of centrality.
- **Glyphs, labels and annotations** – additional information can be added in the format of glyphs, labels or annotations.

5 Community finding / clustering

The clustering function allows you to uncover community structures within large datasets. It is based on a concept known as modularity – a way to measure how readily a network can be divided into sub-networks, or modules.

There are many use cases in which understanding network clusters is beneficial, including:

Cyber security – studying clusters can help you to model how something can move through a network. For example, malicious software will propagate more quickly through a dense community, compared to a sparse one.



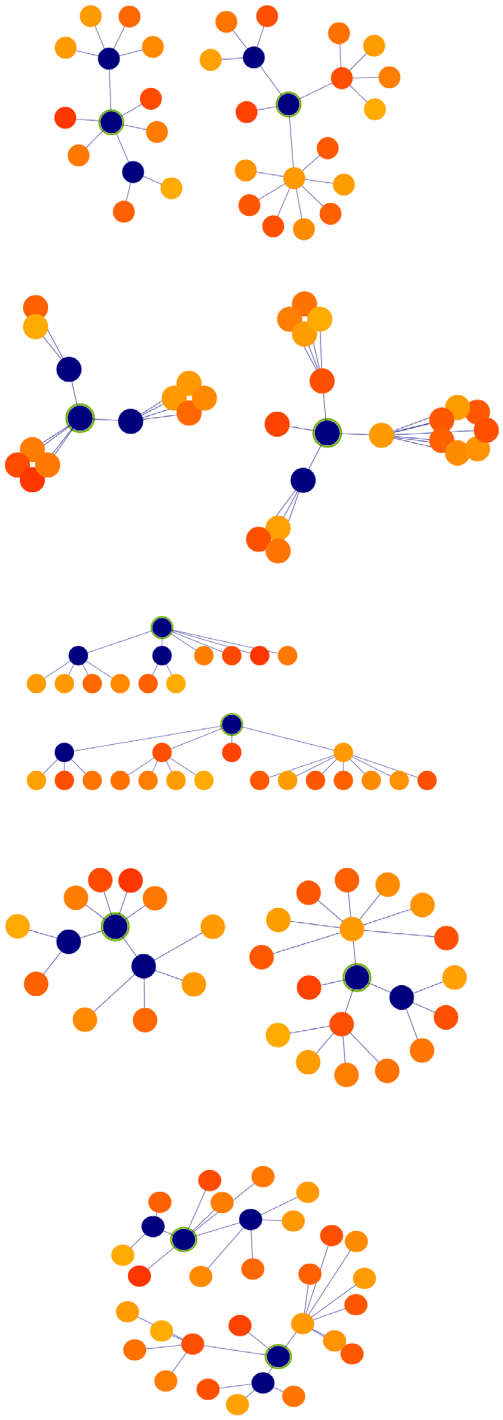
Security – law enforcement and security agencies often need to extrapolate insight about organizational structures from complex communications meta-data.

Clustering algorithms make it much easier to find communities and start learning more about the organizations being investigated.

Anti-fraud – inevitably, organized gangs perform the most damaging and costly fraud. By looking at clusters of fraudulent activity, investigators can find and shut down fraud rings.

6 Understanding layouts

KeyLines offers several different automatic layouts to use when visualizing social networks. Each of these provides a precise and consistent ways to untangle nodes and links and present them in a meaningful way.



6.1 Standard

Our Standard Layout rearranges elements to reduce overlap and distribute elements evenly around the chart. This is a good way to 'overview' a network before focusing in on areas of interest.

6.2 Structural

This layout finds and groups nodes that share similar properties. These similar nodes are displayed in fans around a central node or cluster. This is useful for understanding the overall structure of a network, highlighting distinct clusters of nodes and discovering similarities.

6.3 Hierarchical

Using the hierarchical layout, child nodes are shown in horizontal layers below their parents, in the format of a family tree. This is useful for showing dependencies and hierarchies in a social network.

6.4 Radial

This layout arranges child nodes in concentric circles around the parent, with generations grouped in new orbits. This layout helps when the social network contains too many child-parent relationships to be suitable for a hierarchical layout, or in situations where users need to see the number of hops from one node to another.

6.5 Lens

The lens layout arranges nodes in a circular shape, pushing densely-connected nodes to the center and less densely-connected nodes to the edge. It is a particularly useful layout when dealing with large social datasets where the users need to quickly find well-connected nodes.

6.6 Tweak

The Tweak layout is an adaptation of the standard layout, designed to help you visualize dynamic or evolving networks. The layout adapts itself as links are created and destroyed, allowing the viewer to see clearly where and how the changes happen.

What is KeyLines?

KeyLines is a toolkit for software developers to quickly and easily build powerful network visualization applications.

The visualization applications built with KeyLines utilize the HTML5 canvas element, meaning they run completely in a web browser and can be easily integrated into existing systems and dashboards, or as standalone applications. Its toolkit approach means every deployment of KeyLines is 100% customized to the users, their data and the questions they need to answer.

Want to learn more?

We have extra resources and information available to download from our website. www.keylines.com.

If you have any questions about pricing, or would like a free trial, just get in touch. We would be delighted to help!

www.keylines.com/contact.

www.cambridge-intelligence.com **USA:** +1 (775) 842-6665 **UK:** +44 (0)1223 362 000
Cambridge Intelligence Ltd, 59 – 61 Regent Street, Cambridge, CB2 1AB, UK.