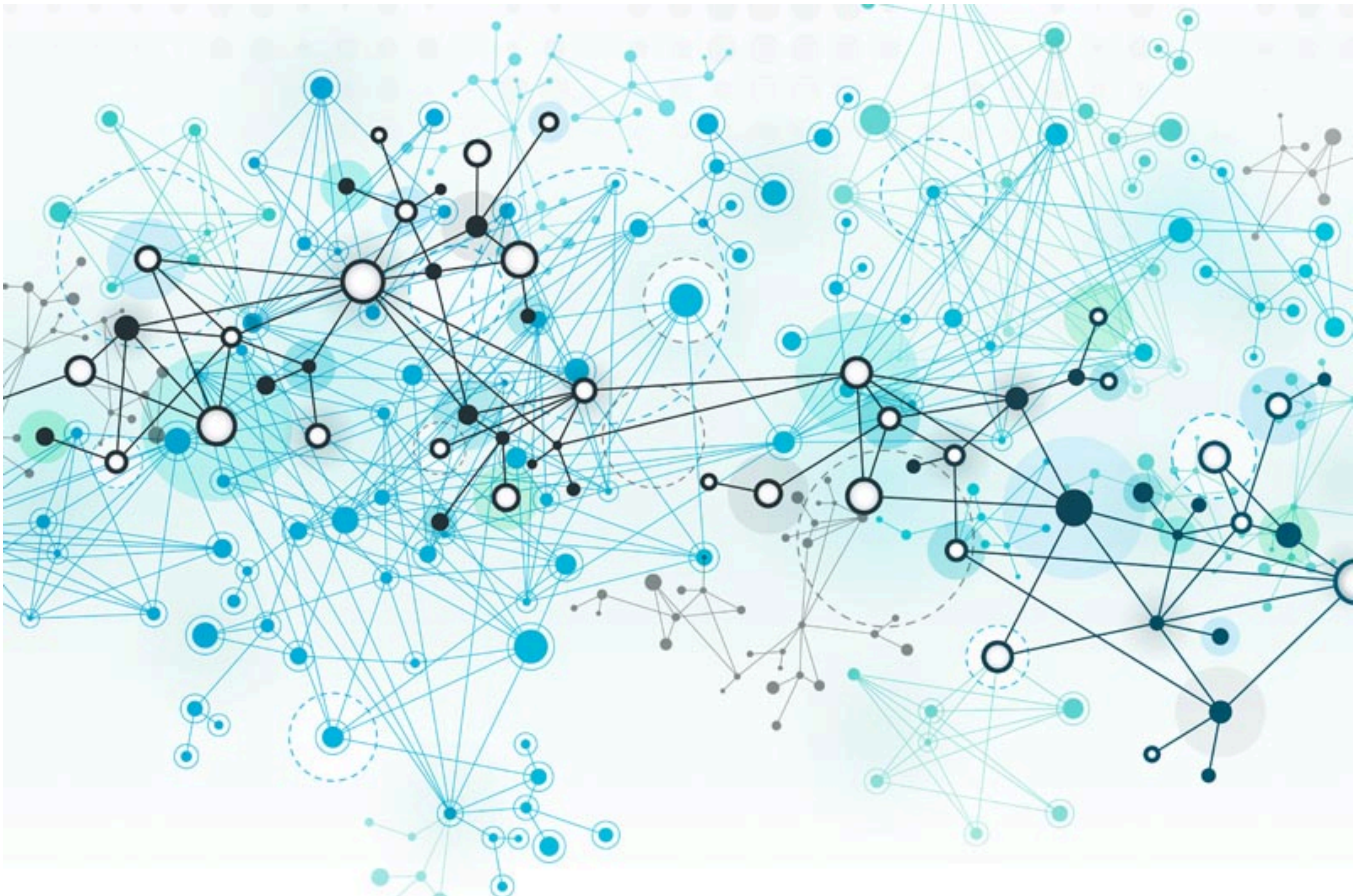
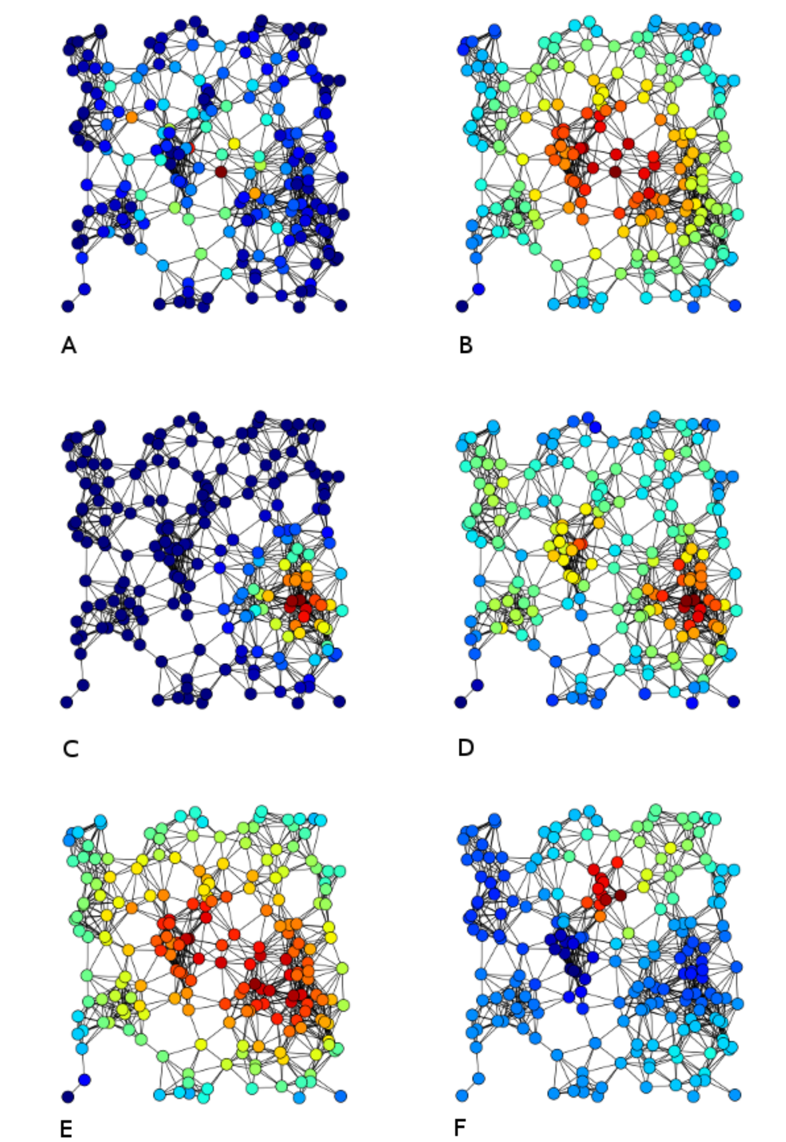
BIG NETWORK DATA
ECE 227 UC San Diego



Node centrality



What are the "most important" nodes in a network?"

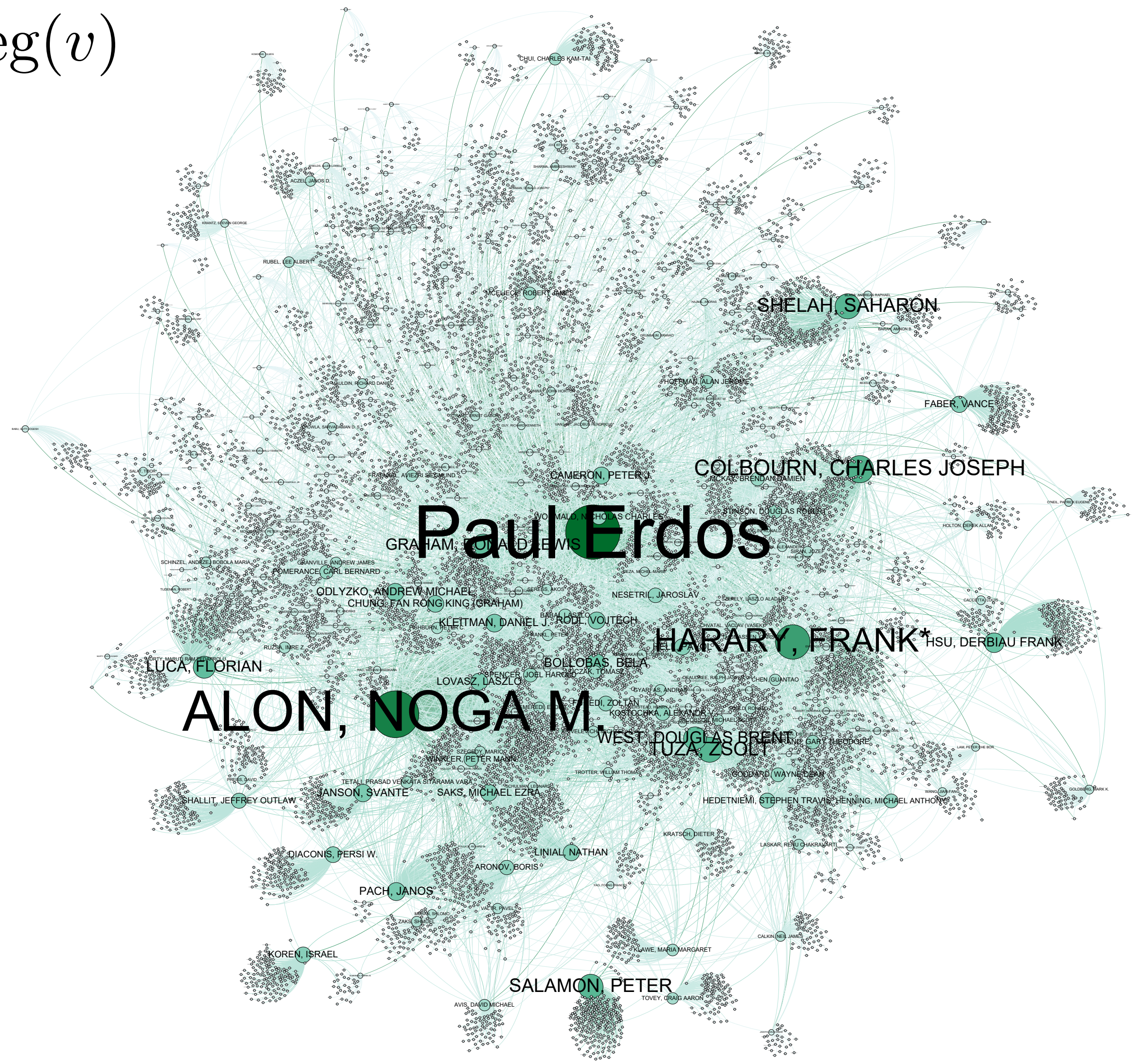
Define a real-valued function indicating the "importance" according to some metric

Subjective measure that may differ from application to application

Degree centrality

A node is more important if it is more connected

$$C(v) = \deg(v)$$



Closeness centrality

A node is more important if it closer (on average) to the other nodes.

Closeness centrality is the (inverse) average length of the shortest path between the node and all other nodes in the network.

$$C(v_i) = \frac{N-1}{\sum_{j=1}^{N} d_{i,j}} \approx \frac{N}{\sum_{j=1}^{N} d_{i,j}}$$

$d_{i,j}$ indicates the length of the shortest path between v_i and v_j

Closeness centrality requires computation of the shortest paths between all pairs of vertices

$O(N^3), O(N^2 \log N), O(NE)$

Betweeness centrality

A node is more important if it must be used more often in communication among all pairs of nodes

The number of times a node acts as a bridge along the shortest path between two other nodes.

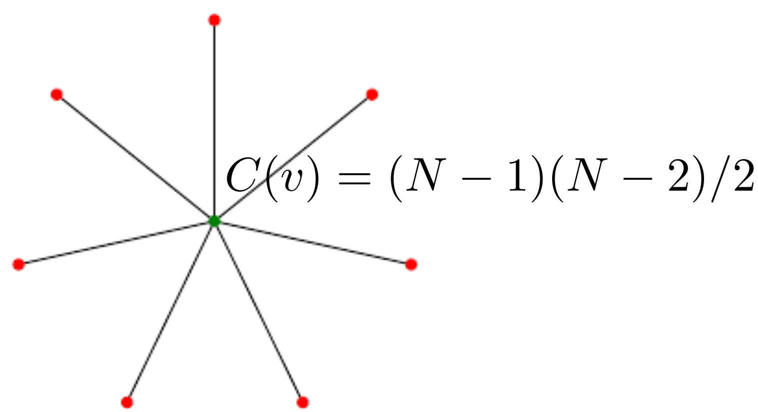
$$C(v) = \sum_{s,t} \mathbb{1}[v \in \text{Shortest Path}(s,t)]$$

It also requires computation of the shortest paths

Betweeness centrality

It may be normalized by dividing by the number of pairs of vertices not including v, that is $(N-1)(N-2)/2$

This corresponds to normalizing by the largest value of betweeness centrality for a network of N nodes, that is achieved by a star network



Eigenvector centrality

A node is more important if has more links to other important nodes

It extends the notion of degree centrality, which gives to all neighbors equal weights, to weighted neighbors according to their degrees

Note: a node with large degree does not necessarily have a high eigenvector centrality (it might be that all neighbors have low eigenvector centrality). A node with high eigenvector centrality does not necessarily have high degree (the node might have few but important neighbors).

Eigenvector centrality

A node is more important if has more links to other important nodes

$$C(v_i) = \frac{1}{\lambda} \sum_{v_j \in \mathrm{Ne}(v_i)} C(v_j)$$

λ is some (non-unique) proportionality constant

Eigenvector centrality

$$C(v_i) = \frac{1}{\lambda} \sum_{v_j \in \mathrm{Ne}(v_i)} C(v_j) = \frac{1}{\lambda} \sum_{j=1}^{N} a_{i,j} C(v_j)$$

$\mathbf{X} \equiv \{C(v_j)\}_{j=1 \ldots N}$ node centralities column vector

$\mathbf{A} \equiv \{a_{i,j}\}_{i,j=1 \ldots N}$ graph adjacency matrix

$$\mathbf{A}\mathbf{X} = \lambda \mathbf{X}$$

Node centralities form an eigenvector of $\mathbf{A}$

We choose the eigenvector corresponding to the largest eigenvalue

Eigenvector centrality

Assume if $\mathbf{A}$ is irreducible, (i.e. the graph is connected)

$$\mathbf{A}\mathbf{X} = \lambda\mathbf{X}$$

In general there will be many different eigenvalues for which a non-zero eigenvector solution exists

But since the entries of $\mathbf{A}$ are non-negative, and $\mathbf{A}$ is irreducible, by the Perron-Frobenius theorem, there is a unique, real, largest eigenvalue

The components of the corresponding eigenvector are taken as the centrality values of the nodes in the network

This approach also generalizes to a weighted adjacency matrix

Recursive interpretation

Degree centrality considers only connectivity at distance one

Eigenvalue centrality considers connectivity at infinite distances

$$\mathbf{X}_0 = [1, 1, \ldots, 1]^T$$

$$\mathbf{X}_1 = \frac{\mathbf{A}\mathbf{X}_0}{\|\mathbf{A}\mathbf{X}_0\|}$$ Start with normalized node degree vector

Refine accounting for node degrees of neighbors

$$\mathbf{X}_2 = \frac{\mathbf{A}\mathbf{X}_1}{\|\mathbf{A}\mathbf{X}_1\|} = \mathbf{A} \times \mathbf{A} \times \mathbf{X}_0 \frac{1}{\|\mathbf{A}\mathbf{X}_1\|\|\mathbf{A}\mathbf{X}_0\|}$$

Refine accounting for node degrees of neighbors of neighbors…

$$\mathbf{X}_n = \frac{\mathbf{A}\mathbf{X}_{n-1}}{\|\mathbf{A}\mathbf{X}_{n-1}\|} = \overbrace{\mathbf{A} \times \mathbf{A} \times \cdots \times \mathbf{A}}^{n} \times \mathbf{X}_0 \frac{1}{\|\mathbf{A}\mathbf{X}_{n-1}\| \cdots \|\mathbf{A}\mathbf{X}_0\|}$$

Recursive interpretation

If $\mathbf{A}$ has a unique largest eigenvalue and $\mathbf{X}_0$ has a non-zero component in the direction of the eigenvector $\mathbf{X}^*$ associated to this dominant eigenvalue, then we have

$$\mathbf{X}_n \to \mathbf{X}^*$$

Principal component interpretation

Assume A admits a spectral decomposition

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^{-1}$$

Consider any vector in the range-space of A

$$\mathbf{Y} = \mathbf{A}\mathbf{X}$$

In the new coordinate system defined by the eigenvectors, we have

$$\mathbf{Y}' = \mathbf{\Lambda}\mathbf{X}'$$

$$\mathbf{Y}' = \mathbf{Q}^{-1}\mathbf{Y}$$

$$\mathbf{X}' = \mathbf{Q}^{-1}\mathbf{X}$$

Principal component interpretation

$$\mathbf{Y}' = \mathbf{\Lambda X}'$$

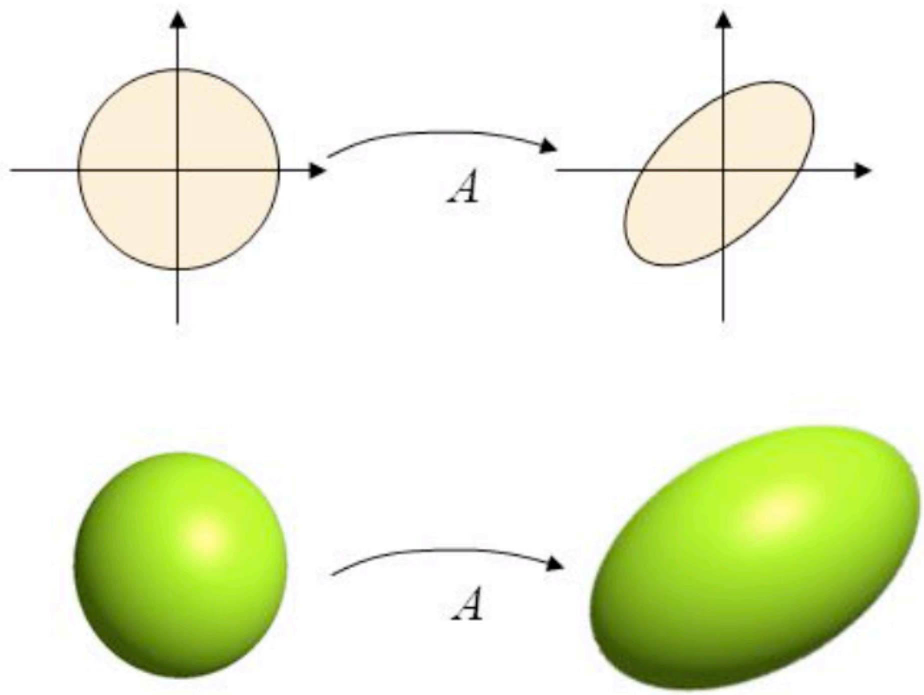
In the new coordinate systems Y has a higher magnitude (i.e. it is stretched) along the direction of the principal eigenvector

By applying A repeatedly eventually Y lies along the direction of the principal eigenvector, as the other components become negligible

Geometric interpretation

Any linear (non-singular) transformation A maps hyperspheres into hyperellipses

This corresponds to squeezing and stretching the hypersphere along some directions

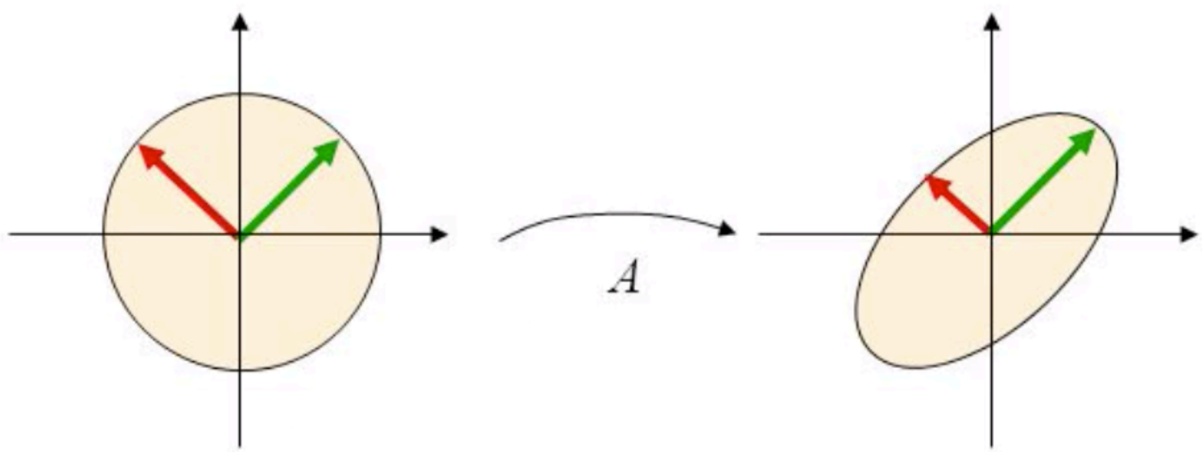


Geometric interpretation

If A is symmetric, the main axes of the ellipse, corresponding to the axes along which the sphere is scaled, are the eigenvectors

The eigendecomposition tells us which axes it scales and by how much

$$\mathbf{A}\mathbf{Q}_i = \lambda_i \mathbf{Q}_i$$



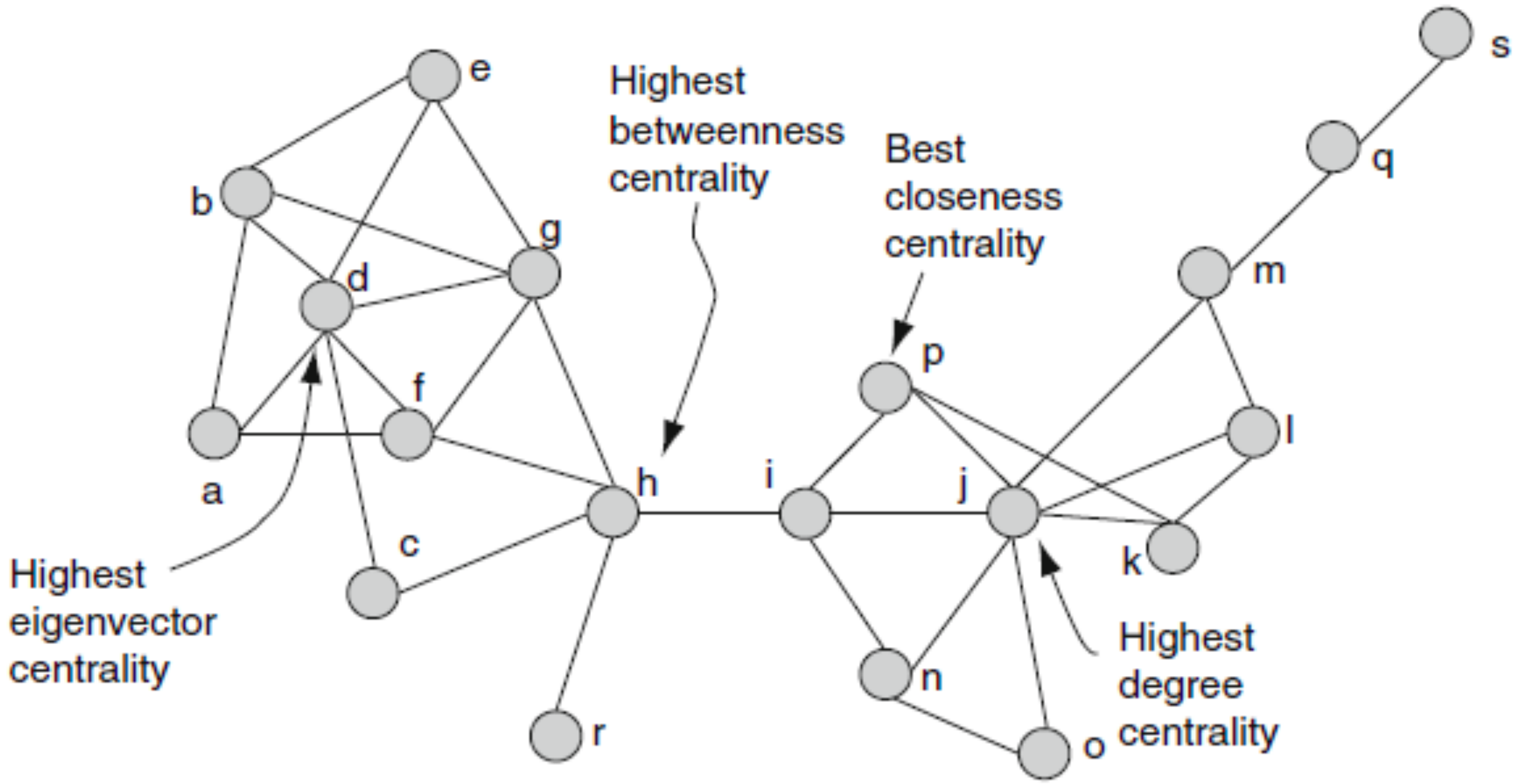
The axis along which the largest scaling occurs, is the principal dimension, identified by the eigenvector of degree centrality

Some Caveats

Centrality indicates node important according to some rationale

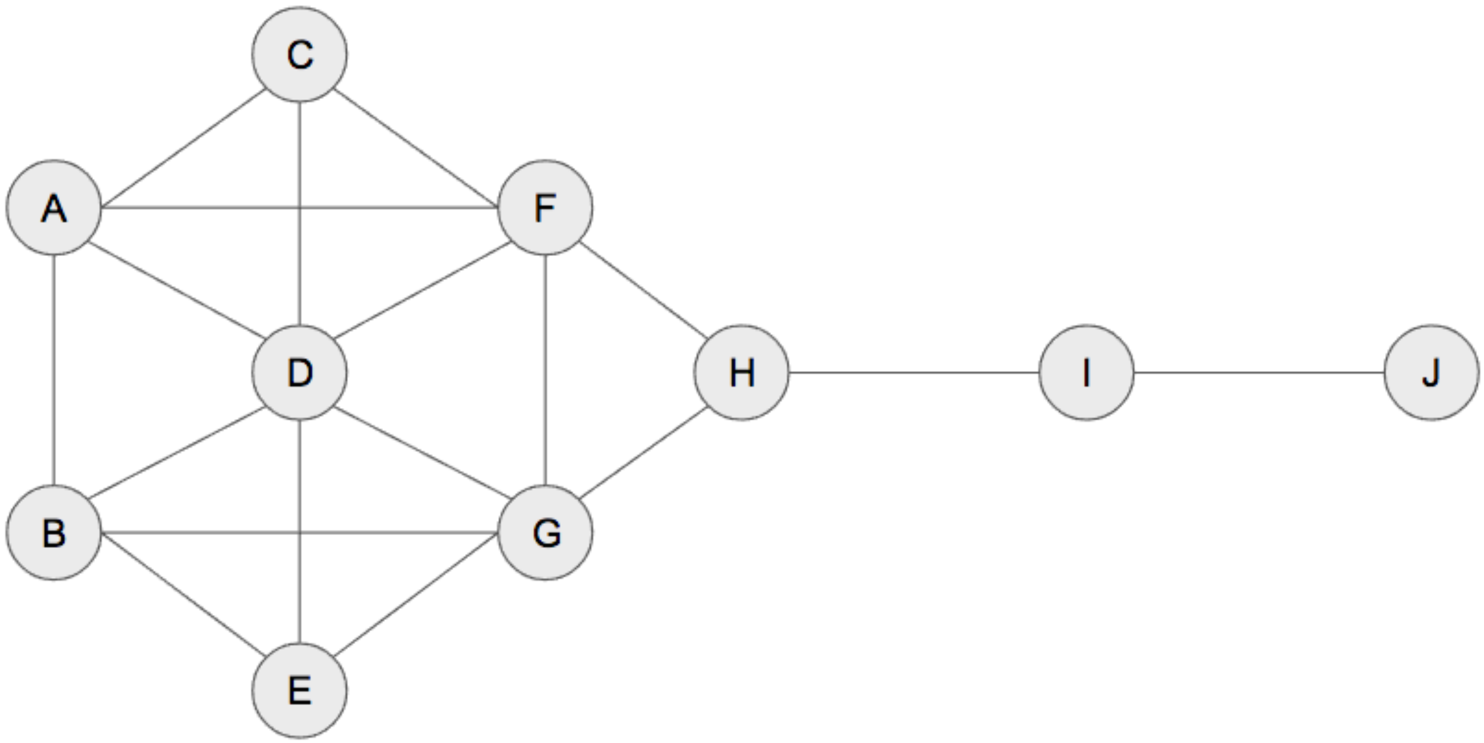
Different notions can or cannot be the same

Centrality considers only network structure: determining the most influential nodes in a social network, for example, depends not only on the structure but also on the influence model



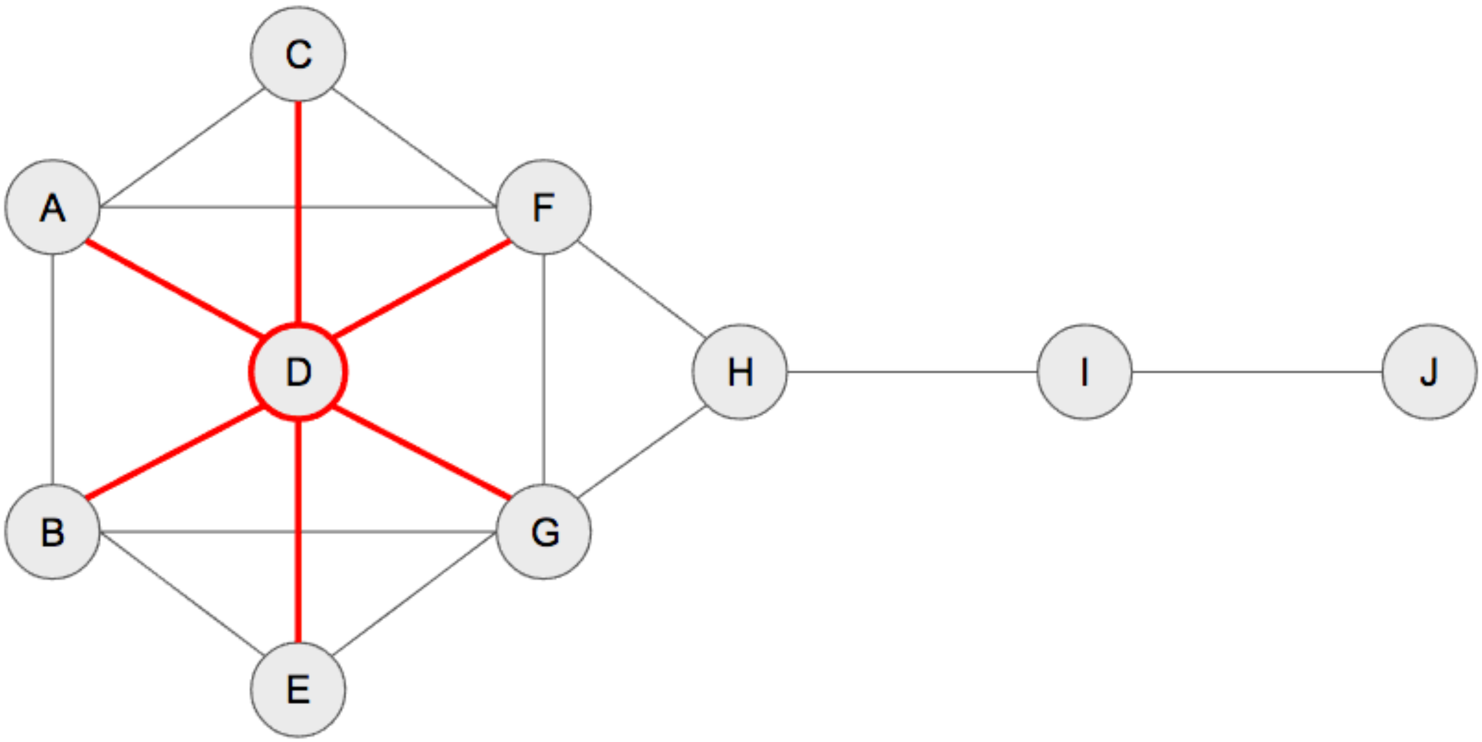
The Krackhardt Kite

Which is the most central node?



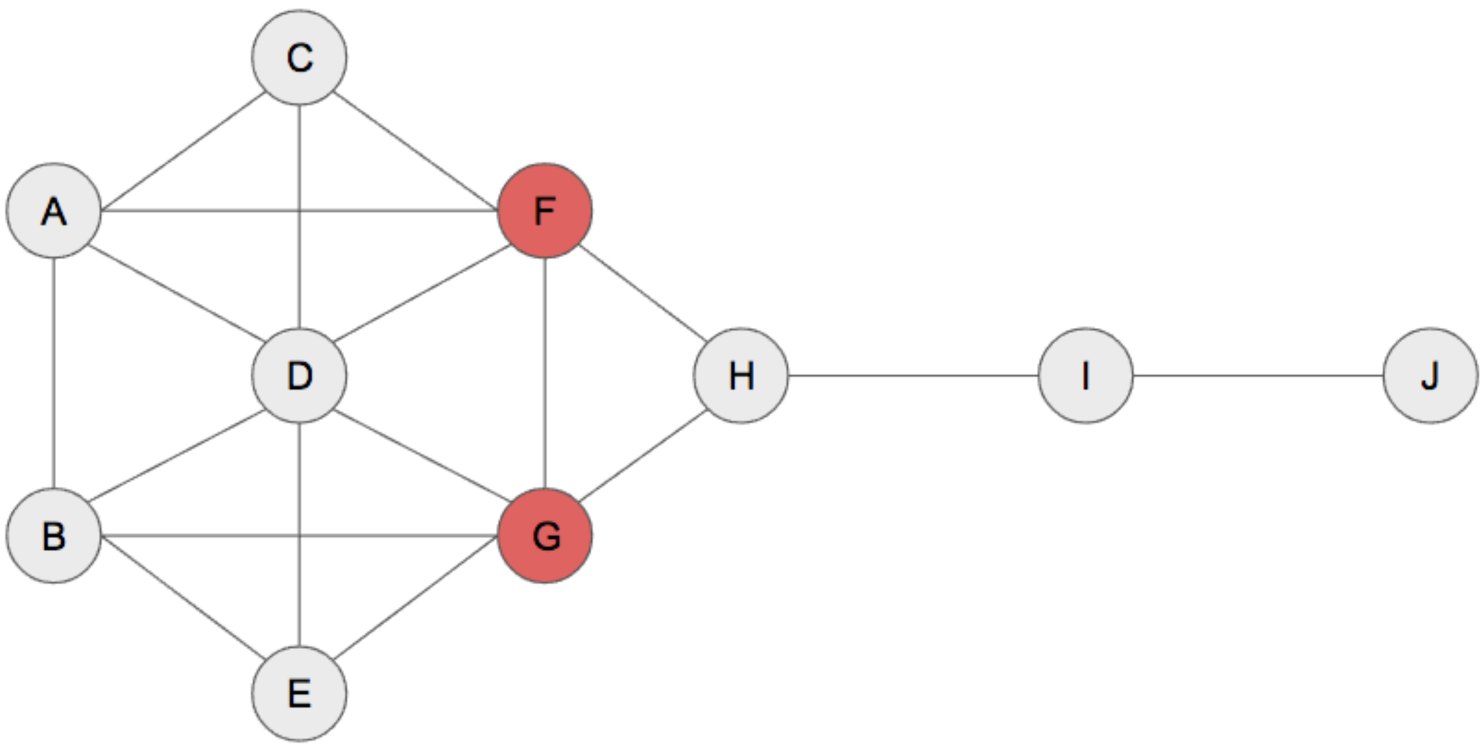
The Krackhardt Kite

Which is the most central node?



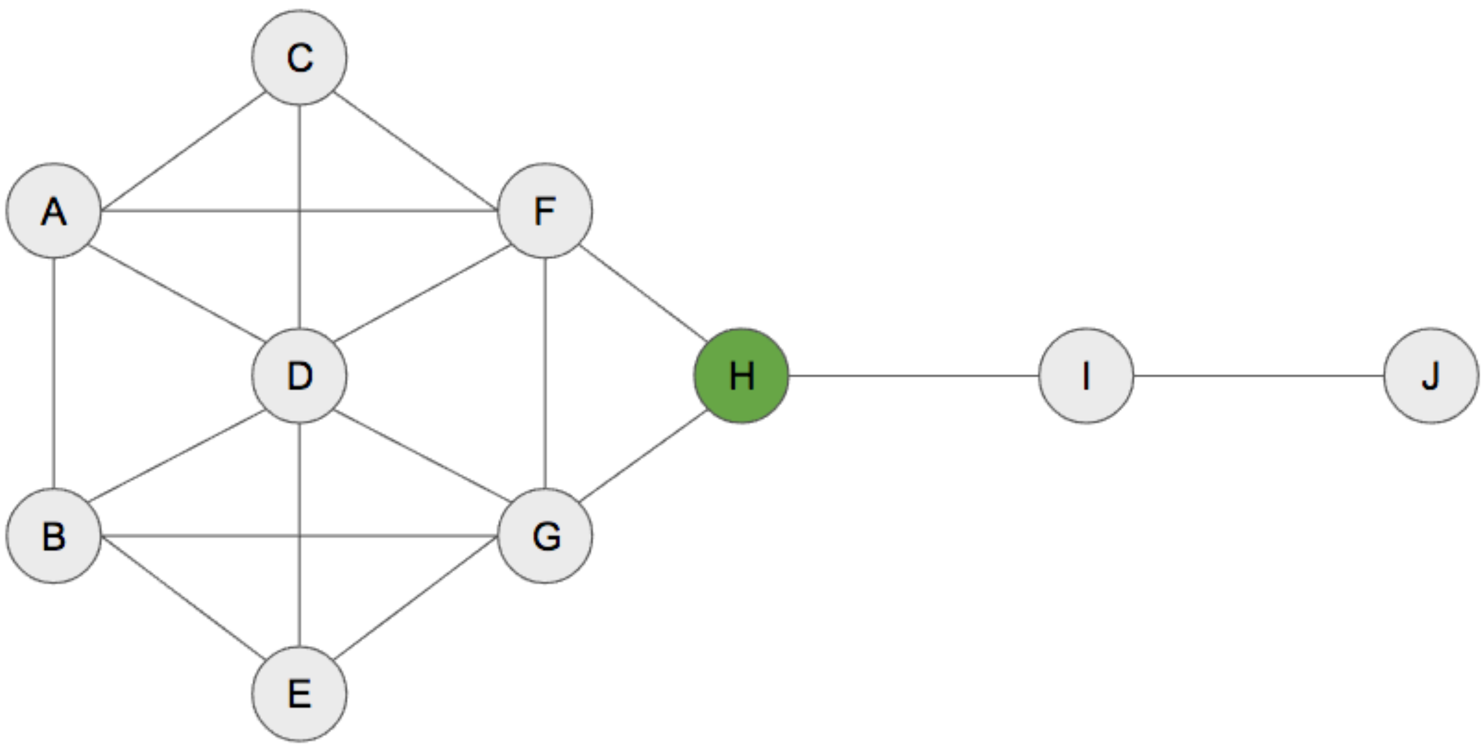
The Krackhardt Kite

Which is the most central node?



The Krackhardt Kite

Which is the most central node?



Network centrality

How central is the most central node?

Compute the sum-differences in centrality between the most central node and all other nodes

Divide by the largest sum-difference of any network of the same size

Network centrality

The notion of centrality can be extended to the amount of centralization of the entire network

$$C(G) = \sum_{i=1}^{N} C(v^*) - C(v_i)$$

Normalizing by the largest centralization for a graph of N nodes (achieved by a star graph), we have

$$C(G) = \frac{\sum_{i=1}^{N} C(v^*) - C(v_i)}{(N-1)[(N-1)-1]} = \frac{\sum_{i=1}^{N} C(v^*) - C(v_i)}{N^2 - 3N + 2}$$