HITS

Description

[From Wikipedia](#)

Applications

The network to analyze can be either directed or undirected. In order to indicate that some edges are more important, an attribute for edge weight can be used. Hence the network can be weighted or un-weighted. It produces two measures for each node in the network, Authority and Hub score. It was originally developed for ranking web pages on the basis of these two measures. But it can be used as effectively on more general problems involving networks of different kinds like citation and other scientometric, bibliometric networks.

Implementation Details

The algorithm requires 3 inputs, the *file* storing the information about the network, *number of iterations* the Authority and Hub values will be calculated (by default it is 20) and the column name that represents the edge weight. The iterations facilitate convergence of the authority and hub scores. But empirically it is found that authority and hub scores tend to converge after 20 iterations, hence the default value. To indicate appropriate relative importance of edges, edge weight attribute can be used. The name of the edge weight column can be selected from the drop down box. The column type should be either *int, real or float*. To indicate un-weighted network the default option of "Unweighted" can be selected.

In the first read-in, the submitted file is validated against the **NWB format**. Once it is validated, the number of nodes *(n)* in the network is passed on. This is used to initialize the matrices that will be used for the HITS computation. An Adjacency matrix *(n x n)*, Authority matrix *(n x 1)*, Hub matrix *(n x 1)* are initialized with default values (0 for adjacency, 1 for authority and hub matrices).

In the second read-in, the adjacency matrix is populated with either the default edge weight (1.0) or the value from the file. When the network is directed, the matrix element corresponding to *[ALGDOC:source node, target node]* is populated. In case of an undirected network, the matrix elements corresponding to *[ALGDOC:node1, node2]* and *[ALGDOC:node2, node1]* are populated with the same edge weight.

Next transpose of the adjacency matrix is created. Now the Authority and Hub matrices are updated in the following manner,

1. Authority Matrix = (Transpose of Adjacency Matrix) x (Hub Matrix)
2. Hub Matrix = (Adjacency Matrix) x (Authority Matrix)
3. Normalize Authority and Hub Matrices by dividing each authority value by the sum of all authority values, and dividing each hub value by the sum of all hub values. This is done to facilitate convergence.

This algorithm is run for *number of iterations* provided by the user.

Then the output file is created in the **NWB** format. In addition to the original file, the node section will have 2 extra attributes, *authority_score and hub_score* of type float. As the name suggests, values from the Authority and Hub matrices corresponding to each node will be inserted.

Links

- Source Code: [Link](#)

Acknowledgments

The original HITS algorithm authored by [Jon Kleinberg](#) was implemented, integrated and documented by Chintan Tank. Also thanks to Micah Linnemeier and Russell Duhon for providing guidance during implementation. For the description I acknowledge Wikipedia.

References

1. Kleinberg, Jon. Authoritative sources in a hyperlinked environment. In Journal of ACM, pages 604-632, September, 1999. [Link](#)
2. Chakrabarti, Soumen., Dom, Byron., Ravi Kumar, S., Raghavan, Prabhakar., Rajagopalan, Sridhar., Tomkins, Andrew., Gibson, David., Kleinberg, Jon. Mining the Web's Link Structure. In Computer, pages 60-67, August, 1999. [Link](#)
3. Bharat, Krishna., Henzinger, Monika R. Improved algorithms for topic distillation in a hyperlinked environment. At Proceedings of the 21st annual international ACM SIGIR conference on Research and development in information retrieval, pages 104-111, 1998. [Link](#)
4. Dean, Jeffrey., Henzinger, Monika. Finding Related Pages in the World Wide Web. At Conference, pages 389-401, 1999. [Link](#)

See Also

ⓘ The license could not be verified: License Certificate has expired! [Generate a Free license now.](#)