Social Network Analysis

Giri Iyengar

Cornell University

gi43@cornell.edu

March 14, 2018

Overview

1. Social Networks

2. HITS

3. Page Rank

Overview

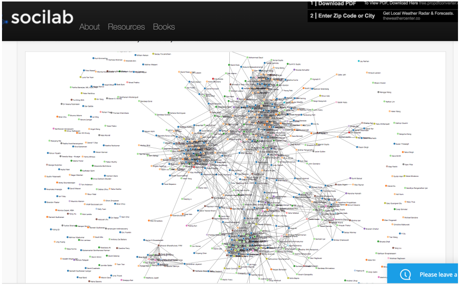
1. Social Networks

2. HITS

3. Page Rank

Social Networks

- Facebook
- LinkedIn
- Twitter
- Instagram
- Snapchat
- WhatsApp


socilab
About
Resources
Books
1 | Download PDF
To View PDF, Download Here free.propdfconverter.
2 | Enter Zip Code or City
Get Local Weather Radar & Forecasts.
theweathercenter.co
Please leave a

Social Networks: Some common questions asked

- Bridge nodes
- Node Centrality
- Between-ness
- Closeness
- Hierarchy
- Density
- Network open-ness

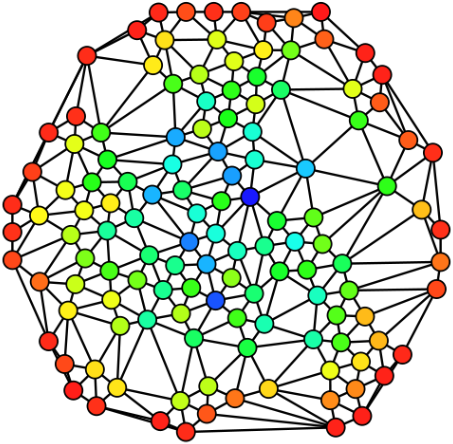


Figure: By Claudio Rocchini

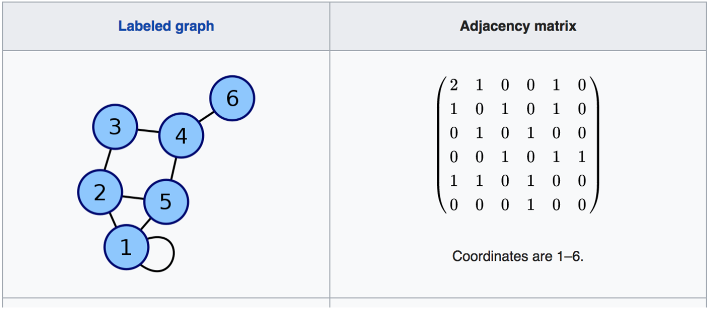
Other types of networks

- Co-Authorship (undirected)
- Citation (directed)
- Disorder - Gene associations (directed)
- Protein - Protein interaction networks (undirected)
- Disease networks (directed)

Representing Networks by Matrices

Adjacency Matrix

In graph theory and computer science, an adjacency matrix is a square matrix used to represent a finite graph. The elements of the matrix indicate whether pairs of vertices are adjacent or not in the graph. If the graph is undirected, the adjacency matrix is symmetric



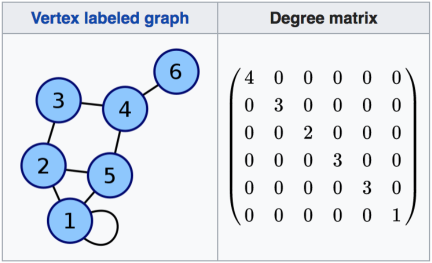
Labeled graph	Adjacency matrix
	$\begin{pmatrix} 2 & 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$ Coordinates are 1–6.

Figure: from Wikipedia

Representing Networks by Matrices

Degree Matrix

In the mathematical field of graph theory the degree matrix is a diagonal matrix which contains information about the degree of each vertex. That is, the number of edges attached to each vertex. It is used together with the adjacency matrix to construct the Laplacian matrix of a graph.



Vertex labeled graph	Degree matrix
	$\begin{pmatrix} 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}$

Figure: from Wikipedia

Closeness Centrality

Closeness of a Vertex

Closeness of a vertex is defined as the inverse of the distance between that vertex and all other vertices in a Graph: $C(x) = \frac{1}{\sum_y d(x,y)}$. In directed graphs, a more central node has more links pointing to it.

Extension to weakly connected graphs

$C(x) = \sum_{y \neq x} \frac{1}{d(x,y)}$

We can also replace the **Arithmetic mean** by **Harmonic mean** above to get **Harmonic Centrality**

Degree Centrality

Degree Centrality of a Vertex

$C(x) = deg(x)$ where $deg(x)$ is either incoming (popularity) or outgoing (gregariousness) degree

Degree Centrality of a whole graph

- Find the most central node, (v^*)
- For any graph of same number of vertices, what is the maximum possible $H = \max \sum_y C(y^*) - C(y)$
- This happens when you have a star-connected graph. One central node and all other nodes connected to it
- Compute $C(G) = \frac{\sum_v C(v^*) - C(v)}{H}$

Overview

1 Social Networks

2 HITS

3 Page Rank

HITS: A Measure of node centrality

- For each node, give it two scores
- Hub score
- Authority score
- Invented by Jon Kleinberg, Cornell University

HITS

Hyperlink-Induced Topic Search (HITS; also known as hubs and authorities) is a link analysis algorithm that rates Web pages, developed by Jon Kleinberg. The idea behind Hubs and Authorities stemmed from a particular insight into the creation of web pages when the Internet was originally forming; that is, certain web pages, known as hubs, served as large directories that were not actually authoritative in the information that they held, but were used as compilations of a broad catalog of information that led users direct to other authoritative pages. In other words, a good hub represented a page that pointed to many other pages, and a good authority represented a page that was linked by many different hubs. The scheme therefore assigns two scores for each page: its authority, which estimates the value of the content of the page, and its hub value, which estimates the value of its links to other pages.

Source: Wikipedia

HITS Intuition

Journals such as *Science* and *Nature* receive a lot of citations and are therefore considered *high impact*. Consider two relatively obscure journals that receive roughly the same number of citations. The one that has *more* citations from Science or Nature should be more authoritative.

HITS Intuition

In the Internet, the number of incoming links represents roughly how *important* a site is. However, even if it has only a few links coming in, if they are coming from *Google*, then it probably is more important.

HITS: Algorithm Sketch

- Perform a query and get a **root set**
- Expand the root set into a **base set** by collecting all pages pointed by this set and some of the links that point to this set

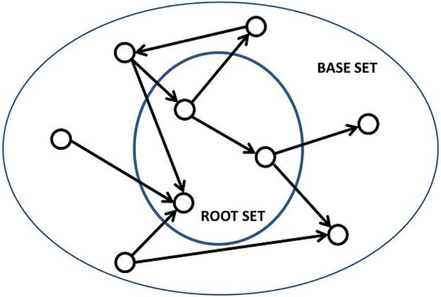


Figure: https://commons.wikimedia.org/wiki/File:SetsEN.jpg

HITS: Algorithm Sketch

- On this focused sub-graph, the HITS computation is performed. Initialize all scores to 1.
- **Authority Update**: Update each node's Authority score to be equal to the sum of the Hub Scores of each node that points to it. That is, a node is given a high authority score by being linked from pages that are recognized as Hubs for information.
- **Hub Update**: Update each node's Hub Score to be equal to the sum of the Authority Scores of each node that it points to. That is, a node is given a high hub score by linking to nodes that are considered to be authorities on the subject.
- Normalize the values and repeat as necessary

Overview

1. Social Networks

2. HITS

3. Page Rank

Page Rank

Page Rank

PageRank is an algorithm used by Google Search to rank websites in their search engine results. PageRank was named after Larry Page, one of the founders of Google. It is a way of measuring the importance of website pages. PageRank works by counting the number and quality of links to a webpage to determine a rough estimate of how important the website is. The underlying assumption is that more important websites are likely to receive more links from other websites [1].

Rank of a URL

$$r(P_i) = \sum_{P_j \in B_{P_i}} \frac{r(P_j)}{|P_j|} \tag{1}$$

Where B_{P_i} is the set of pages pointing into P_i and $|P_i|$ is the number of outlinks from P_i.

Page Rank

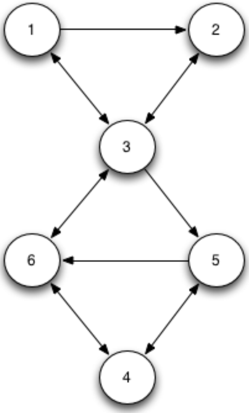


Figure: Toy Example with 6 URLs.

Page Rank for Toy Universe

Table 1 shows the PageRank results after first two iterations of the calculation.

Table: First few iterations of PageRank calculations on a tiny web with 6 Urls

Iteration 0	Iteration 1	Iteration 2	Rank at Iteration 2
$r_0(P_1) = \frac{1}{6}$	$r_1(P_1) = \frac{1}{24}$	$r_2(P_1) = \frac{1}{12}$	6
$r_0(P_2) = \frac{1}{6}$	$r_1(P_2) = \frac{1}{8}$	$r_2(P_2) = \frac{5}{48}$	5
$r_0(P_3) = \frac{1}{6}$	$r_1(P_3) = \frac{1}{3}$	$r_2(P_3) = \frac{1}{4}$	1
$r_0(P_4) = \frac{1}{6}$	$r_1(P_4) = \frac{1}{6}$	$r_2(P_4) = \frac{1}{6}$	3
$r_0(P_5) = \frac{1}{6}$	$r_1(P_5) = \frac{1}{8}$	$r_2(P_5) = \frac{1}{6}$	4
$r_0(P_6) = \frac{1}{6}$	$r_1(P_6) = \frac{5}{24}$	$r_2(P_6) = \frac{11}{48}$	2

Page Rank

Iterative calculation of Page Rank

If we define a matrix A such that each element $A_{ji} = \frac{1}{|P_i|}$ if there is a link from page i to page j and 0 otherwise. Then, we get $r_{i+1} = A \times r_i$. Let r be the final probability vector. We want r such that $r = A \times r$

Random browser extension - to make things work

What happens if the graph is disconnected? Or there are some one-way paths that prevent you from coming back? Let's assume that most of the time the user will follow one of the links in the page but sometimes, they get bored and start at a new URL somewhere. This makes the matrix A well-behaved. Removes all the zeros. Updates the matrix to

$$\mathbf{B} = 0.85 \times \mathbf{A} + \frac{0.15}{n}$$

Page Rank: Perron Frobenius Theorem

Perron Frobenius Theorem states

Any square matrix A with positive entries has a unique eigenvector with positive entries (up to a multiplication by a positive scalar), and the corresponding eigenvalue has multiplicity one and is strictly greater than the absolute value of any other eigenvalue.

Markov matrices

If all entries of a Markov matrix A are positive then A has a unique equilibrium: there is only one eigenvalue 1. All other eigenvalues are smaller than 1.

Page Rank: The eigenvector corresponding to the eigenvalue of 1.

рахмат
danke
謝謝
ngiyabonga
Баярлалаа
spas
teşekkür ederim
спасибо
mersi
kia ora
barka
welalin
tack
mahalo
tapadh leat
faafetai lava
vinaka
спасибі
blagodaram
dank je
misaotra
matondo
paldies
grazzi
kiitos
dankie
thank you
gracias
хвала
nanni
nandi
dhanyavad
asante
manana
bedankt
bayarlalaa
gracie
hvala
maururu
köszönöm
dankon
aciü
akun
obrigada
murakoze
tenki
enkosi
djiere dieuf
tau
mochchakkeram
дякую
mamnun
chnorakaloutioun
gratias ago
gràcies
sulpáy
dziękuję
sobodi
dĕkuji
taiku
go raibh maith agat
chokrane
obrigado
mèsi
sagolun
sukriya
kop khun krap
arigatō
takk
dakujem
trugarez
ありがとう
didi madloba
najis tuke
tanemirt
grazie
shukriya
merce
мерси
kam sah hamnida
terima kasih
rahmat
rahmet
ευχαριστώ
diolch
dhanyavadagalu
তোমাকে ধন্যবাদ
감사합니다
xiexie
merci