

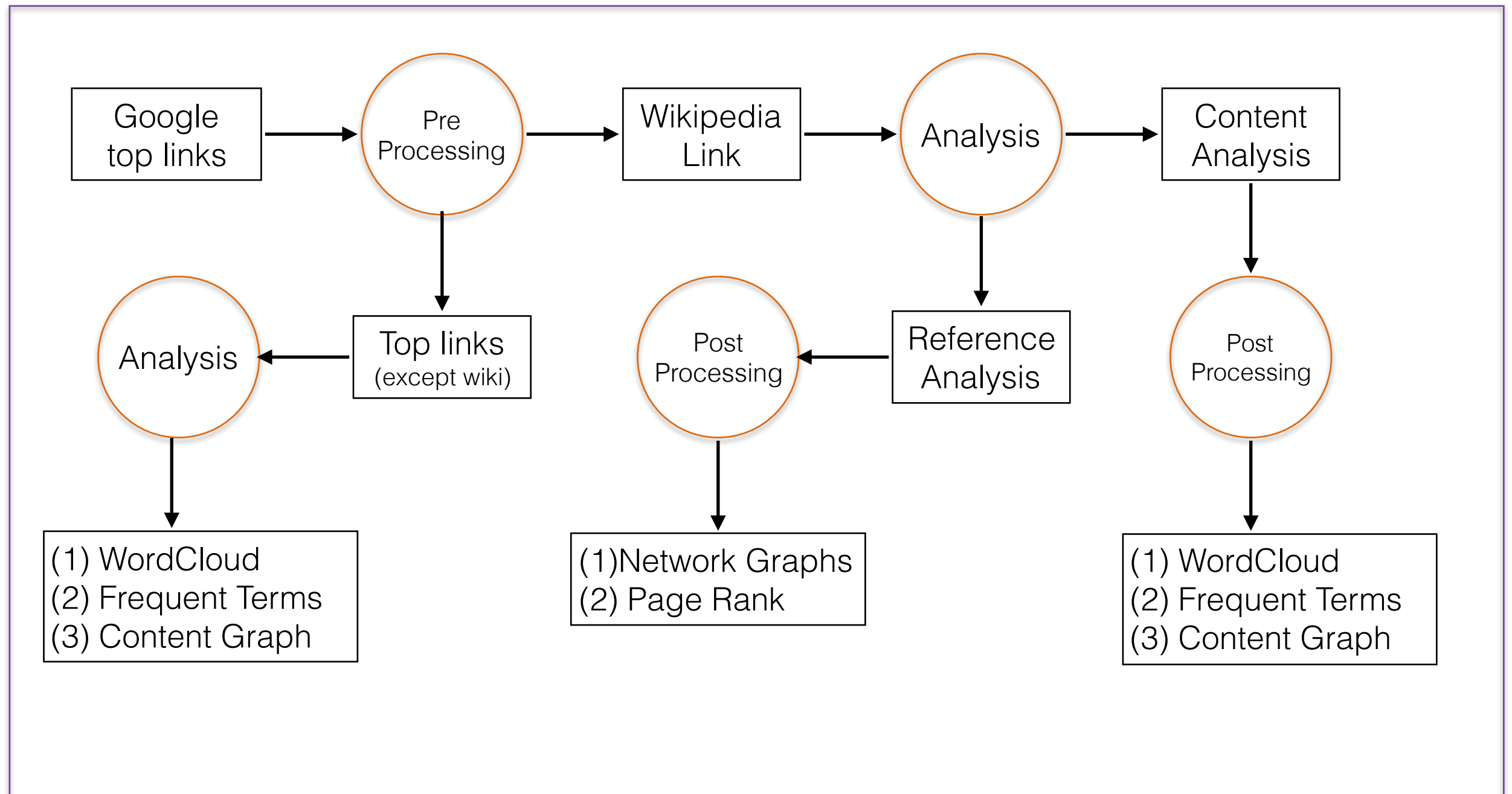
Page Rank and Content Analysis

Ravi K Rajendran
Grad Student, Computer Science
CS688 Web Analytics and Text Mining

Content Analysis

Flow of the project

Shiny



Content Analysis

Google Search results

Query-1: Microsoft

data

/url?q=https://www.microsoft.com/&sa=U&ved=0ahUKEwivvPTunYbRAhUI4oMKHWdDBlwQFggjMAA&usg=AFQjCNHRxMvFJHpKyV9IY0hlnEiQ72QKFg

/url?q=https://www.microsoft.com/en-us/download&sa=U&ved=0ahUKEwivvPTunYbRAhUI4oMKHWdDBlwQjBAIKjAB&usg=AFQjCNGTrUUL9vd0rPrGQ1v8wO1SfR4Ong

/url?q=https://support.microsoft.com/en-us/help/17777/downloads-for-windows&sa=U&ved=0ahUKEwivvPTunYbRAhUI4oMKHWdDBlwQjBAILDAE&usg=AFQjCNErVk1cEviX-PDR0c3Fi0Od9UGewA

/url?q=https://support.microsoft.com/en-us&sa=U&ved=0ahUKEwivvPTunYbRAhUI4oMKHWdDBlwQjBAILjAC&usg=AFQjCNF6aRph0bOTt3cg0Y2KV32v-lrpIA

/url?q=https://www.microsoft.com/en-us/surface&sa=U&ved=0ahUKEwivvPTunYbRAhUI4oMKHWdDBlwQjBAIMDAF&usg=AFQjCNEdQO5O6sGieU9993YxyAe3RkAwmg

Query-2: Apple

data

/url?q=http://www.apple.com/&sa=U&ved=0ahUKEwj97nvnYbRAhXoz4MKHehFDMsQFgghMAA&usg=AFQjCNFf-ryUHIE9AY4RteHfcyP7KaHQYg

/url?q=https://support.apple.com/&sa=U&ved=0ahUKEwj97nvnYbRAhXoz4MKHehFDMsQjBAIKDAB&usg=AFQjCNFJd2wCgEOSMQKwLVAR9RV2OKBG5A

/url?q=http://www.apple.com/iphone/&sa=U&ved=0ahUKEwj97nvnYbRAhXoz4MKHehFDMsQjBAIKjAD&usg=AFQjCNGQawLTu4JOsD5slu1yGAB222XaLg

/url?q=http://www.apple.com/mac/&sa=U&ved=0ahUKEwj97nvnYbRAhXoz4MKHehFDMsQjBAILDAC&usg=AFQjCNH_J82DTQR9KpsErLtZbSQ8wToTRQ

/url?q=http://www.apple.com/retail/&sa=U&ved=0ahUKEwj97nvnYbRAhXoz4MKHehFDMsQjBAILjAE&usg=AFQjCNEwWGxkc6yS-yrWvzEoqyrvP1jxmg

/search?q=apple&ie=UTF-8&prmd=ivnsz&source=univ&tbn=nws&tbo=u&sa=X&ved=0ahUKEwj97nvnYbRAhXoz4MKHehFDMsQqAlIMA

Content Analysis

Google Search results

Query-1: Microsoft

data

<https://www.microsoft.com/>

<https://www.microsoft.com/en-us/download>

<https://support.microsoft.com/en-us/help/17777/downloads-for-windows>

<https://support.microsoft.com/en-us>

<https://www.microsoft.com/en-us/surface>

Query-2: Apple

data

<http://www.apple.com/>

<https://support.apple.com/>

<http://www.apple.com/iphone/>

<http://www.apple.com/mac/>

<http://www.apple.com/retail/>

Content Analysis

Wiki - References:

Query1

/wiki/Wikipedia:Protection_policy#semi

/wiki/File:Microsoft_logo_(2012).svg

/wiki/File:Microsoft_building_17_front_door.jpg

/wiki/Microsoft_Redmond_campus

/wiki/Redmond,_Washington

/wiki/Types_of_business_entity

/wiki/Public_company

/wiki/Ticker_symbol

/wiki/NASDAQ

/wiki/NASDAQ-100

/wiki/Dow_Jones_Industrial_Average

/wiki/S%26P_100

/wiki/S%26P_500

/wiki/Software

/wiki/Computer_hardware



Query1

Wikipedia Protection policy semi

File Microsoft logo 2012 svg

File Microsoft building 17 front door jpg

Microsoft Redmond campus

Redmond, Washington

Types of business entity

Public company

Ticker symbol

NASDAQ

NASDAQ 100

Dow Jones Industrial Average

S 26P 100

S 26P 500

Software

Computer hardware

Content Analysis

Wiki - References:

Query2
/wiki/Wikipedia:Protection_policy#semi
/wiki/Apple
/wiki/Apple_(disambiguation)
/wiki/Geographic_coordinate_system
/wiki/File:Apple_logo_black.svg
/wiki/File:Apple_Headquarters_in_Cupertino.jpg
/wiki/Apple_Campus
/wiki/Infinite_Loop_(street)
/wiki/Cupertino,_California
/wiki/Types_of_business_entity
/wiki/Public_company
/wiki/Ticker_symbol
/wiki/NASDAQ
/wiki/NASDAQ-100
/wiki/Dow_Jones_Industrial_Average



Query2
Wikipedia Protection policy semi
Apple
Apple disambiguation
Geographic coordinate system
File Apple logo black svg
File Apple Headquarters in Cupertino jpg
Apple Campus
Infinite Loop street
Cupertino, California
Types of business entity
Public company
Ticker symbol
NASDAQ
NASDAQ 100
Dow Jones Industrial Average

Content Analysis

Common Wikipedia Link References:

ZDNet

International Standard Serial Number

Joshua Topolsky

Engadget

AOL

Wikipedia Wikimedia sister projects

OpenCorporates

Apple Computer, Inc v Microsoft Corp

Template Major information technology companies

Template talk Major information technology companies

Information technology

List of the largest information technology companies

List of largest Internet companies

Semiconductor sales leaders by year

Atos

Booz Allen Hamilton

BMC Software

BT Global Services

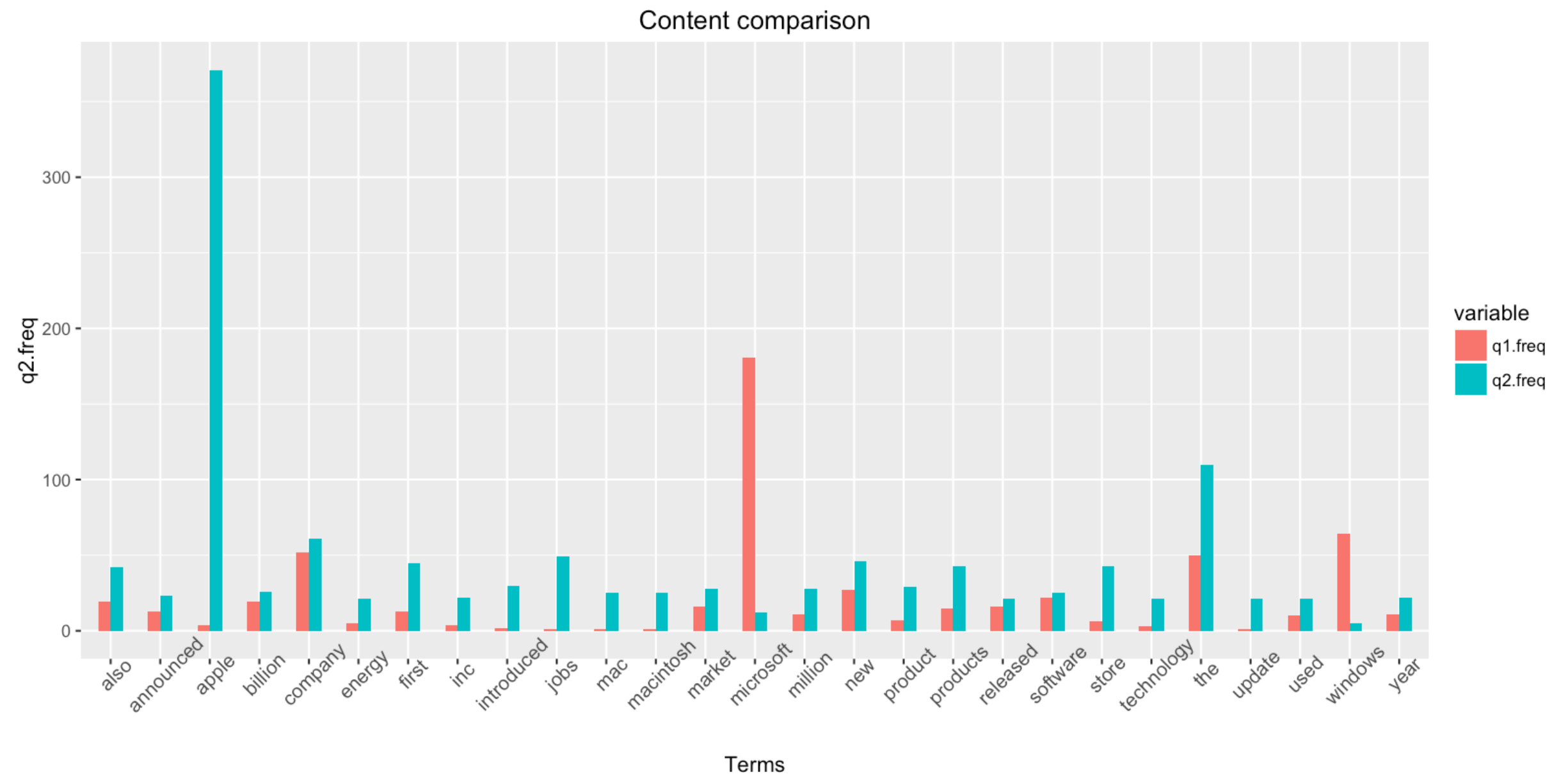
CACI

Capgemini

Content Analysis

Results:

Common contents

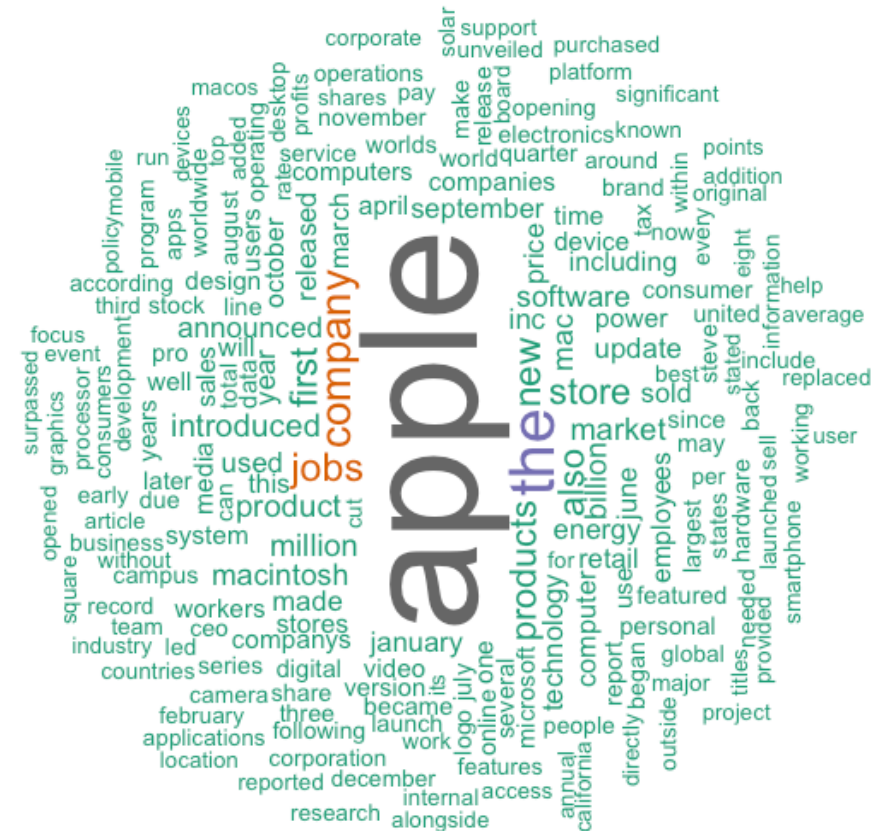
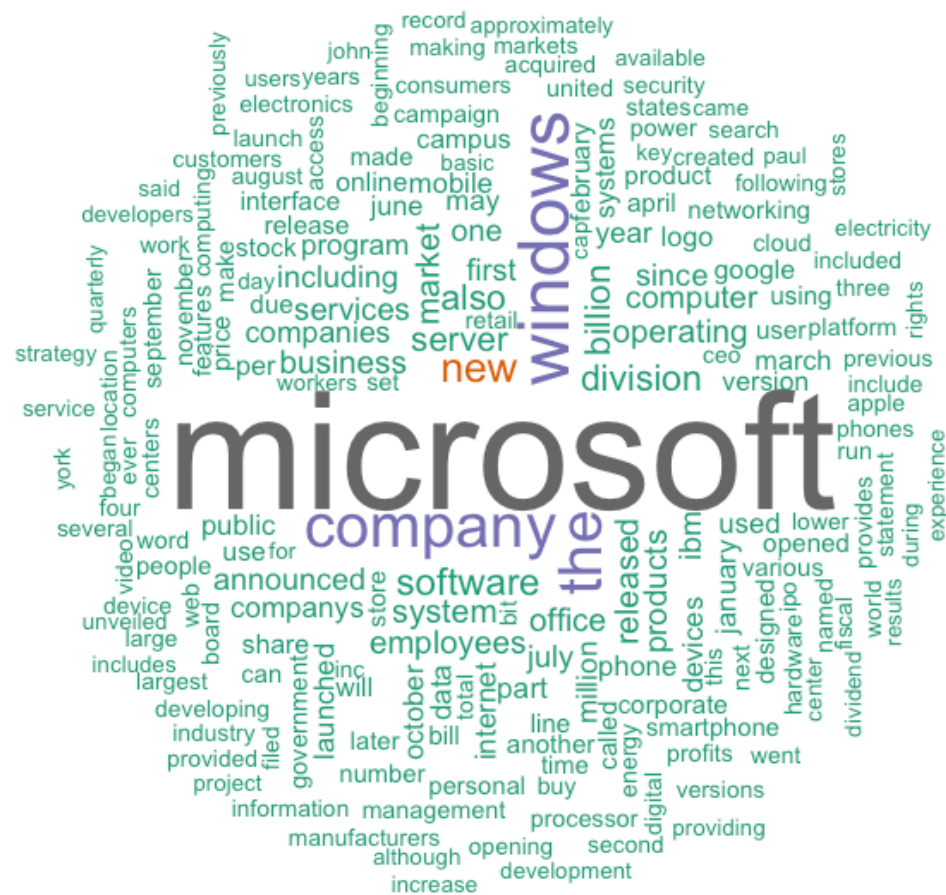


Content Analysis

Results:

Query-1: Microsoft

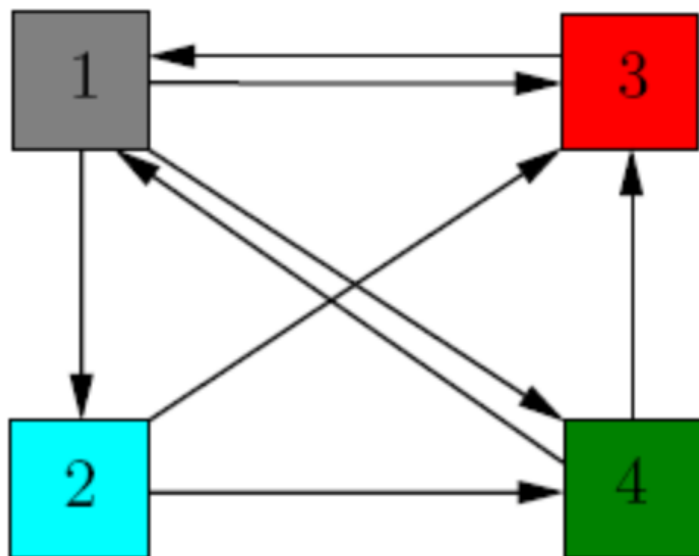
Query-2: Apple



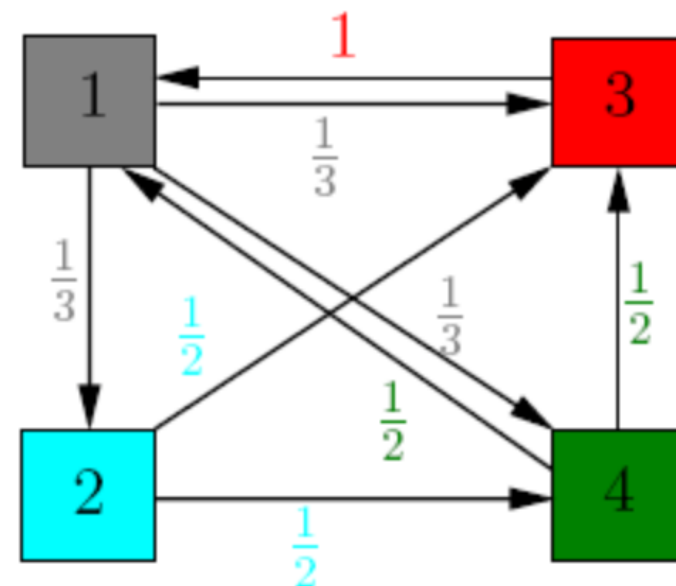
Page Ranking

Pages : 1 - 4

Page Links



Weighted

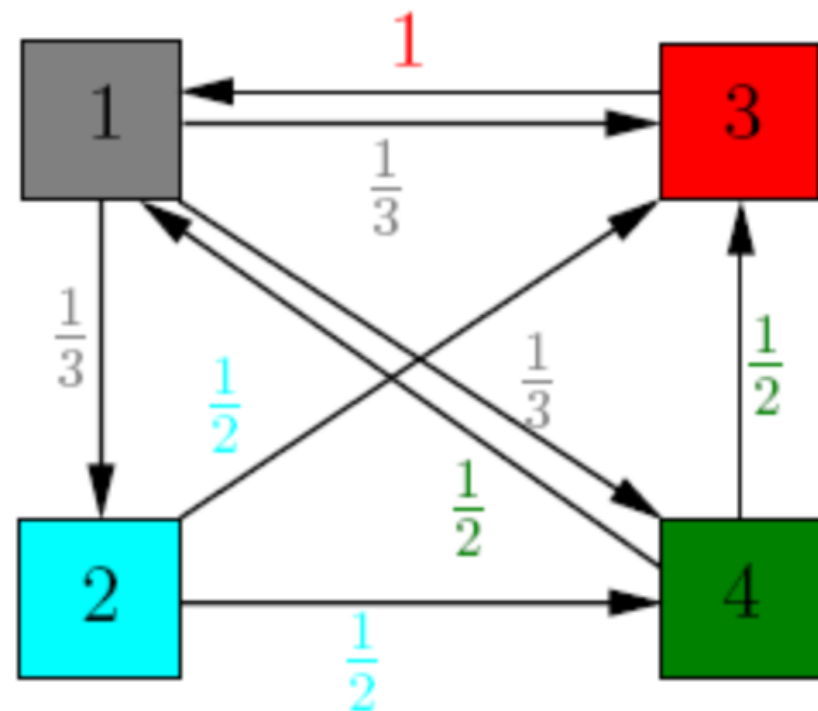


Reference: <http://www.math.cornell.edu/~mec/Winter2009/RalucaRemus/Lecture3/lecture3.html>

Page Ranking

Pages : 1 - 4

Weighted



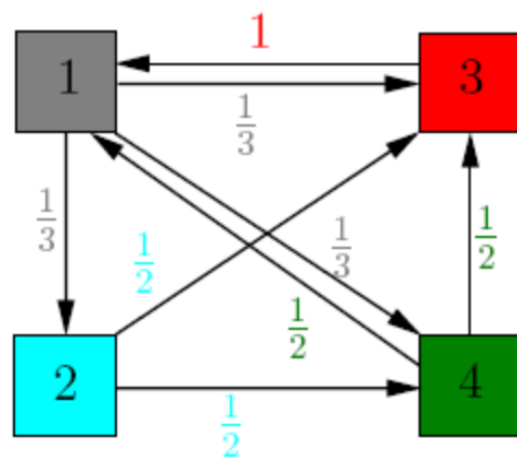
Transition Matrix

$$A = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

Page Ranking

Page Vector

Weighted



Transition Matrix

$$A = \begin{bmatrix} 0 & 0 & 1 & \frac{1}{2} \\ \frac{1}{3} & 0 & 0 & 0 \\ \frac{1}{3} & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{2} & 0 & 0 \end{bmatrix}.$$

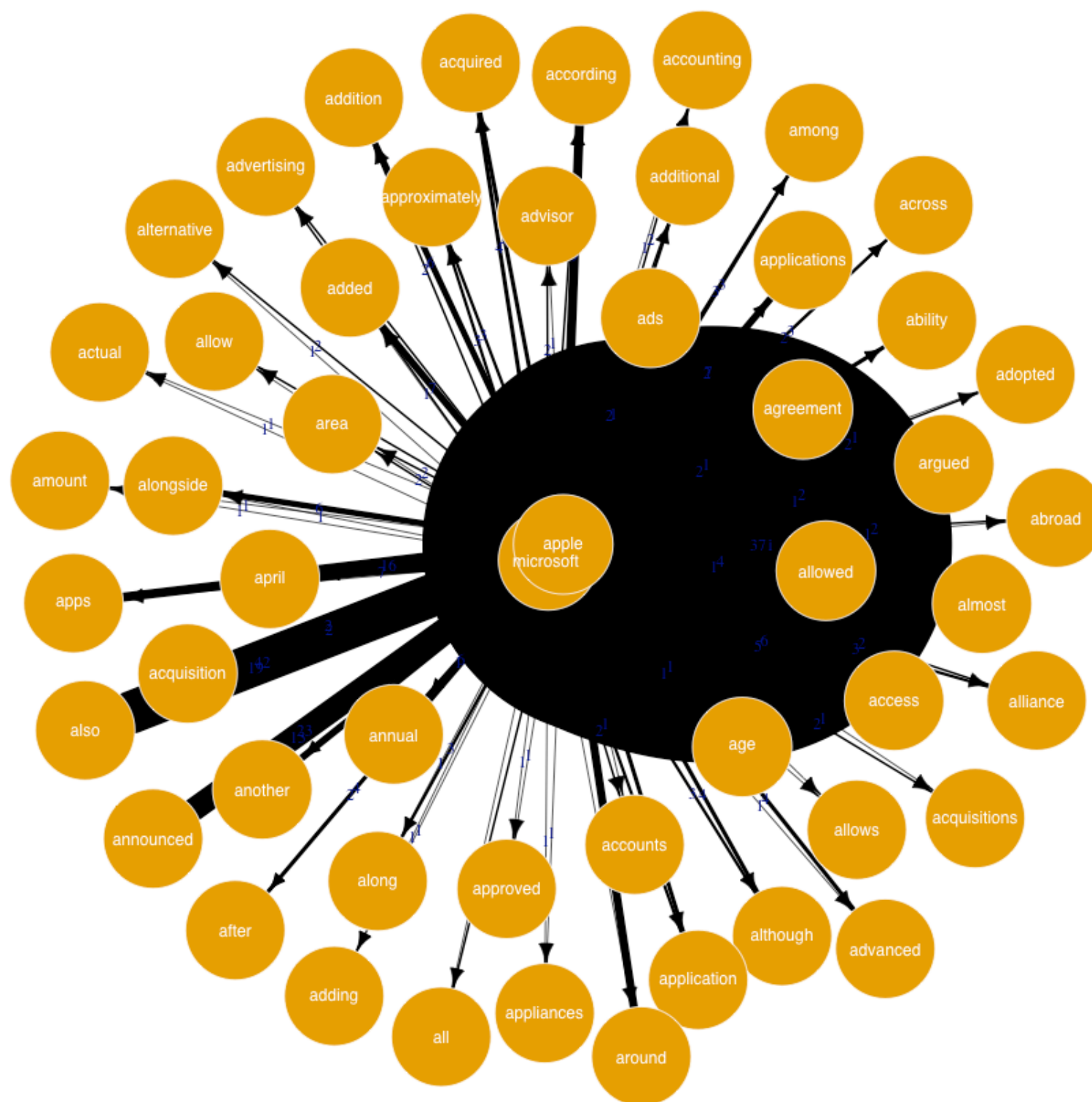
Page rank Calculation

$$\mathbf{v} = \begin{pmatrix} 0.25 \\ 0.25 \\ 0.25 \\ 0.25 \end{pmatrix}, \quad A\mathbf{v} = \begin{pmatrix} 0.37 \\ 0.08 \\ 0.33 \\ 0.20 \end{pmatrix}, \quad A^2\mathbf{v} = A(A\mathbf{v}) = A \begin{pmatrix} 0.37 \\ 0.08 \\ 0.33 \\ 0.20 \end{pmatrix} = \begin{pmatrix} 0.43 \\ 0.12 \\ 0.27 \\ 0.16 \end{pmatrix}$$

$$A^3\mathbf{v} = \begin{pmatrix} 0.35 \\ 0.14 \\ 0.29 \\ 0.20 \end{pmatrix}, \quad A^4\mathbf{v} = \begin{pmatrix} 0.39 \\ 0.11 \\ 0.29 \\ 0.19 \end{pmatrix}, \quad A^5\mathbf{v} = \begin{pmatrix} 0.39 \\ 0.13 \\ 0.28 \\ 0.19 \end{pmatrix}$$

$$A^6\mathbf{v} = \begin{pmatrix} 0.38 \\ 0.13 \\ 0.29 \\ 0.19 \end{pmatrix}, \quad A^7\mathbf{v} = \begin{pmatrix} 0.38 \\ 0.12 \\ 0.29 \\ 0.19 \end{pmatrix}, \quad A^8\mathbf{v} = \begin{pmatrix} 0.38 \\ 0.12 \\ 0.29 \\ 0.19 \end{pmatrix}$$

Network Graph



- Thank you