



# PageRank Network Analysis of Website Usability Stage Two Presentation

Group 2: Andrew Braunagel & Dominic Mathiang





# Our Mission:

Analyze and Create Visualizations of the page structure and page ranking for university websites in order to assist web developers and administrators with making improvements to the usability and likeability of their websites.



# Look back on Stage One...

# Stage One - phases

- Phase 1 & 2: Project Scope and Research
- Phase 3: Web Crawler and Graph Creation
- Phase 4: Visualization and Statistical Analysis
- Phase 5: Random Walk
- Phase 6: Website Creation and Data Displays

# Phase 1 & 2: Project Scope and Research

#### Resources:

- Python 3.10.0 | Documentation
- Urllib Package | <u>Documentation</u>
- BeautifulSoup Package | <u>Documentation</u>
- ➤ Gephi | <u>Homepage</u>
- HTML/CSS for Website
- Discord for Communication

#### Project Scope:

- Crawl provided URL and harvest the pages and links
- > Provide clear visualizations of the website's page structure
- > Offer quantitative information based on statistical analysis and random walk results
- > Offer qualitative analysis based of objective rating scales
- > Provide suggestions for web developers on how they can improve their website

# Phase 3: Web Crawler and Graph Creation

#### Web Crawler:

- Utilize urllib python package to send HTTP GET requests to web server
- > Parse the **HTML page contents** for all link tags (<a></a>) using BeautifulSoup package
- Analyze the href text for each link to remove external links and format internal links to work with crawler
- > Append all unique links to the crawl list, to ensure crawling of entire website
- > Output: .txt file containing list of all links in **Source Page, Destination Page** format

#### Graph Creation:

- > Using the networkx package, the program generates a **directed graph** with pages as **nodes** and links as **edges**
- Output: .gexf file to be used with Gephi

https://www.oswego.edu, https://www.oswego.edu https://www.oswego.edu, https://www.oswego.edu/admissions/apply https://www.oswego.edu,https://www.oswego.edu/admissions/visit https://www.oswego.edu, https://www.oswego.edu/news https://www.oswego.edu,https://www.oswego.edu/mvoswego https://www.oswego.edu, https://www.oswego.edu/library https://www.oswego.edu,https://www.oswego.edu/registrar/college-calendar https://www.oswego.edu, https://www.oswego.edu/dining-menu https://www.oswego.edu.https://www.oswego.edu/about 10 https://www.oswego.edu,https://www.oswego.edu/about/fast-facts https://www.oswego.edu, https://www.oswego.edu/about/history-and-traditions https://www.oswego.edu, https://www.oswego.edu/about/our-region https://www.oswego.edu,https://www.oswego.edu/about/maps-parking-directions 14 https://www.oswego.edu,https://www.oswego.edu/about/diversity-and-inclusion https://www.oswego.edu,https://www.oswego.edu/about/environment-and-sustainability https://www.oswego.edu, https://www.oswego.edu/about/leadership 16 17 https://www.oswego.edu,https://www.oswego.edu/about/contact-us https://www.oswego.edu, https://www.oswego.edu/tuition-and-aid 19 https://www.oswego.edu,https://www.oswego.edu/academics https://www.oswego.edu, https://www.oswego.edu/academics/graduate-programs https://www.oswego.edu,https://www.oswego.edu/online https://www.oswego.edu, https://www.oswego.edu/extended-learning https://www.oswego.edu,https://www.oswego.edu/academics/colleges-and-schools 24 https://www.oswego.edu.https://www.oswego.edu/admissions https://www.oswego.edu, https://www.oswego.edu/admissions/undergraduate-admissions https://www.oswego.edu, https://www.oswego.edu/admissions/graduate-admissions https://www.oswego.edu,https://www.oswego.edu/admissions/request-information https://www.oswego.edu,https://www.oswego.edu/student-life 29 https://www.oswego.edu, https://www.oswego.edu/student-life/living-oswego https://www.oswego.edu, https://www.oswego.edu/student-life/dining-experience https://www.oswego.edu,https://www.oswego.edu/student-life/student-activities 32 https://www.oswego.edu,https://www.oswego.edu/student-life/health-and-wellness https://www.oswego.edu,https://www.oswego.edu/student-life/athletics-and-recreation 34 https://www.oswego.edu,https://www.oswego.edu/financial-aid/content/office-financial-aid 35 https://www.oswego.edu,https://www.oswego.edu/career-services/career-services-0 36 https://www.oswego.edu,https://www.oswego.edu/center-for-experiential-learning/excel-experiential-courses-and-engaged-learning https://www.oswego.edu,https://www.oswego.edu/fitness-centers/health-well-being-resources-students-faculty-and-staff https://www.oswego.edu,https://www.oswego.edu/artswego/artswego-0 https://www.oswego.edu,https://www.oswego.edu/center-for-experiential-learning/mentor-scholar 39 40 https://www.oswego.edu.https://www.oswego.edu/financial-aid/scholarships 41 https://www.oswego.edu,https://www.oswego.edu/research 42 https://www.oswego.edu, https://www.oswego.edu/admissions/plan-your-visit 43 https://www.oswego.edu,https://www.oswego.edu/safety-on-campus 44 https://www.oswego.edu, https://www.oswego.edu/about/index

ormal text file

length: 68,801,454 lines: 562,323 Ln: 62 Co

Example .txt Output from Web Crawler

lraft" xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xsi:schemaLocation="http://www.gexf.net/1.. '>

node="static" name="">

lu" label="https://www.oswego.edu" /> iu/admissions/apply" label="https://www.osweqo.edu/admissions/apply" /> lu/admissions/visit" label="https://www.oswego.edu/admissions/visit" /> lu/news" label="https://www.oswego.edu/news" /> u/myoswego" label="https://www.oswego.edu/myoswego" /> lu/library" label="https://www.oswego.edu/library" /> u/registrar/college-calendar" label="https://www.oswego.edu/registrar/college-calendar" /> lu/dining-menu" label="https://www.oswego.edu/dining-menu" /> iu/about" label="https://www.oswego.edu/about" /> lu/about/fast-facts" label="https://www.oswego.edu/about/fast-facts" /> lu/about/history-and-traditions" label="https://www.oswego.edu/about/history-and-traditions" /> lu/about/our-region" label="https://www.oswego.edu/about/our-region" /> lu/about/maps-parking-directions" label="https://www.oswego.edu/about/maps-parking-directions" /> lu/about/diversity-and-inclusion" label="https://www.oswego.edu/about/diversity-and-inclusion" /> lu/about/environment-and-sustainability" label="https://www.oswego.edu/about/environment-and-sustainabi. lu/about/leadership" label="https://www.oswego.edu/about/leadership" /> iu/about/contact-us" label="https://www.oswego.edu/about/contact-us" /> iu/tuition-and-aid" label="https://www.oswego.edu/tuition-and-aid" /> lu/academics" label="https://www.oswego.edu/academics" /> lu/academics/graduate-programs" label="https://www.oswego.edu/academics/graduate-programs" /> lu/online" label="https://www.oswego.edu/online" /> iu/extended-learning" label="https://www.oswego.edu/extended-learning" /> iu/academics/colleges-and-schools" label="https://www.oswego.edu/academics/colleges-and-schools" /> lu/admissions" label="https://www.oswego.edu/admissions" /> u/admissions/undergraduate-admissions" label="https://www.oswego.edu/admissions/undergraduate-admission lu/admissions/graduate-admissions" label="https://www.oswego.edu/admissions/graduate-admissions" /> lu/admissions/request-information" label="https://www.oswego.edu/admissions/request-information" /> lu/student-life" label="https://www.oswego.edu/student-life" /> lu/student-life/living-oswego" label="https://www.oswego.edu/student-life/living-oswego" /> iu/student-life/dining-experience" label="https://www.oswego.edu/student-life/dining-experience" /> lu/student-life/student-activities | label="https://www.oswego.edu/student-life/student-activities | /> lu/student-life/health-and-wellness" label="https://www.oswego.edu/student-life/health-and-wellness" /> lu/student-life/athletics-and-recreation" label="https://www.oswego.edu/student-life/athletics-and-recre u/financial-aid/content/office-financial-aid" label="https://www.oswego.edu/financial-aid/content/office-financial-aid/content/office-financial-aid/content/office-financial-aid/content/office-financial-aid/content/content/office-financial-aid/content lu/career-services/career-services-0" label="https://www.oswego.edu/career-services/career-services-0" u/center-for-experiential-learning/excel-experiential-courses-and-engaged-learning" label="https://www

Example .gexf Output from Web Crawler

### Phase 4: Visualization and Statistical Analysis



				_
Context	×			-
Nodes: Edges: Directed	7282 554406 d Graph			
Filters	Statistics ×			_
Settings				
🗉 Netw	ork Overview			
Average	werage Degree 32.77			
Avg. Weighted Degree			Run	0
Network Diameter		12	Run	0
Graph Density			Run	0
HITS			Run	0
Modularity		0.207	Run	0
PageRank			Run	0
Connected Components			Run	0
Node	Overview			
Avg. Clustering Coefficient			Run	0
Eigenveo	ctor Centrality		Run	۲
🗷 Edge	Overview			
Avg. Pat	h Length	3.554	Run	0
🖻 Dyna	mic			
# Nodes	5		Run	0
# Edges			Run	0
Degree			Run	0
Clusterin	ng Coefficient		Run	0



# Phase 5: Random Walk

- Organize lines into dictionary with Source Page as key, and Destination Pages as list of values
- Starting from website homepage, select random destination
- Split URL to pull hub name and increment hub count by 1
- Repeat previous two steps as many times as necessary
- Output: HTML table with hub name, step count, and percentage of times hit

HUB NAME	TIMES HIT	PERCENTAGE
Admissions	1466559	14.666
About	1431040	14.31
Student-Life	707141	7.071
News	575908	5.759
Academics	442046	4.42
Extended-Learning	265336	2.653
Communications-And-Marketing	235401	2.354
Programs	218651	2.187
Registrar	210238	2.102
Safety-On-Campus	198298	1.983
Library	198033	1.98
Institutional-Research	182384	1.824



### Phase 6:

# Website Creation and Data Displays

Link To Project Website

# Stage One Difficulties

- Was initially using MatPlotLib for visualization of graph; however, it was slow and not very user-friendly
- Time consuming to run programs via Wifi connection
- Determining how to objectively rate websites based on data and statistical findings

# Stage Two Plans

- Develop a consultation company to work with university web developers on website improvement projects.
  - > Development team to improve software and implement additional features
  - Marketing and Sales team for customer acquisitions
  - Research and Development team to run mass student surveys and experiments to develop precise objective rating scales.



# **Questions?**