

|                                  |
|----------------------------------|
| Network Theory and Dynami        |
| Course Team                      |
| BigBlueButton: NTDS              |
| Module Overview                  |
| Books and Other Resources        |
| Lecture Material                 |
| Assignments                      |
| Assignment 01: Introducti        |
| Assignment 02: Network El        |
| Assignment 03: Small Worl        |
| Assingment 04: Hubs              |
| Assignment 05: Strong and        |
| <b>Assignment 06: Graph-base</b> |
| Assignment 07-08: Network        |
| Assignment 09: Dynamics I        |
| Assignment 10: Dynamics I        |
| Exam Eligibility Assignme        |
| Demo Exam Paper                  |
| Forum                            |

## Assignment 06: Graph-based View of the Web



### Performance summary

✓ Assessed

#### Success status



Undefined

#### Score

18.5 of 100 points

#### Attempts

1

#### Results

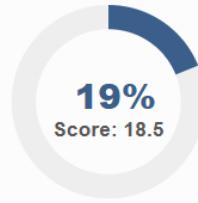
|        |  |
|--------|--|
| Course | Network Theory and Dynamic Systems (SS 2025)<br>ID: 4669112833 / 109697249570546 |
| Test   | Assignment 05: Directions and Weights<br>ID: 4562682569                          |

### This are your test results

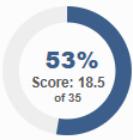
Duration 1h 5m 54s 5/22/2025, 11:02 AM - 5/22/2025, 12:07 PM

Answered 9 of 9 questions (100%)

Your score 18.5 of 100 points (19%)



#### 01. Knowledge Tasks (35 points) 4

[go to section >](#)

#### 02. Practical Tasks (30 points) 3

[go to section >](#)

#### 03. Programming Tasks (35 points) 2

[go to section >](#)

### 01. Knowledge Tasks (35 points) 18.5 of 35 points (53%)

#### True/false (20 points)

Status Answered

Your score 18.5 / 20  93%

#### Response

Each right answer takes 1 point.

Wrong answers take -0.5 points.

| Unanswered               | Right                               | Wrong                               |
|--------------------------|-------------------------------------|-------------------------------------|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input type="checkbox"/>            |

A web crawler can be used to index the content of websites for a search engine.

|                          |                                     |                                     |
|--------------------------|-------------------------------------|-------------------------------------|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> |
|--------------------------|-------------------------------------|-------------------------------------|

All web crawlers respect the rules set in the robots.txt file of a website.

|                          |                          |                                     |
|--------------------------|--------------------------|-------------------------------------|
| <input type="checkbox"/> | <input type="checkbox"/> | <input checked="" type="checkbox"/> |
|--------------------------|--------------------------|-------------------------------------|

Web crawlers only download the homepages of websites.

|                          |                                     |                          |
|--------------------------|-------------------------------------|--------------------------|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
|--------------------------|-------------------------------------|--------------------------|

PageRank is completely independent of the web page content.

|                          |                                     |                          |
|--------------------------|-------------------------------------|--------------------------|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/> |
|--------------------------|-------------------------------------|--------------------------|

The higher the teleportation parameter in PageRank, the more random the browsing model becomes.

|                          |                                     |                                     |   |
|--------------------------|-------------------------------------|-------------------------------------|---|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input checked="" type="checkbox"/> | PageRank values are determined only by the number of direct links to a node.                  |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | The teleportation factor in PageRank prevents the algorithm from getting stuck in loops.      |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | PageRank is only applicable to web pages.   |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | All links in a retweet network indicate positive endorsements of the content.                 |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Diffusion networks are static and do not change over time.                                    |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Every node in a retweet network represents an individual human user.                          |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | Networks with segregated clusters do not facilitate diverse information exposure.             |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | All nodes with high out-strength are genuine and not bots.                                    |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The direction of links in diffusion networks is irrelevant to understanding information flow. |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | The structure of a diffusion network can reveal the spread pattern of specific memes.         |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | All misinformation is political in nature.  |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The presence of echo chambers can decrease the spread of misinformation.                      |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | All nodes in a diffusion network contribute equally to the spread of information.             |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Larger nodes in a retweet network always represent more credible sources.                     |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Nodes with zero out-strength are not influential in diffusion networks.                       |

▼ Solution

Each right answer takes 1 point.

Wrong answers take -0.5 points.

| Unanswered               | Right                               | Wrong                               |   |
|--------------------------|-------------------------------------|-------------------------------------|---|
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | A web crawler can be used to index the content of websites for a search engine.                 |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | All web crawlers respect the rules set in the robots.txt file of a website.                     |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Web crawlers only download the homepages of websites.   |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | PageRank is completely independent of the web page content.                                     |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | The higher the teleportation parameter in PageRank, the more random the browsing model becomes. |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | PageRank values are determined only by the number of direct links to a node.                    |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | The teleportation factor in PageRank prevents the algorithm from getting stuck in loops.        |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | PageRank is only applicable to web pages.   |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | All links in a retweet network indicate positive endorsements of the content.                   |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Diffusion networks are static and do not change over time.                                      |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Every node in a retweet network represents an individual human user.                            |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | Networks with segregated clusters do not facilitate diverse information exposure.               |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | All nodes with high out-strength are genuine and not bots.                                      |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The direction of links in diffusion networks is irrelevant to understanding information flow.   |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | The structure of a diffusion network can reveal the spread pattern of specific memes.           |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | All misinformation is political in nature.  |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | The presence of echo chambers can decrease the spread of misinformation.                        |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | All nodes in a diffusion network contribute equally to the spread of information.               |
| <input type="checkbox"/> | <input type="checkbox"/>            | <input checked="" type="checkbox"/> | Larger nodes in a retweet network always represent more credible sources.                       |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> | <input type="checkbox"/>            | Nodes with zero out-strength are not influential in diffusion networks.                         |

### ☰ Misinformation on Social Media Platforms (5 points)

|            |          |
|------------|----------|
| Status     | Answered |
| Your score | 0 / 5    |

#### Response

Examine the impact of misinformation on social media platforms and discuss how network analysis techniques, such as the study of diffusion networks, can be used to identify and combat the spread of false information. Consider the role of bots and echo chambers in the propagation of misinformation.

Misinformation in social media can cause public confusion, political skepticism, and may even damage public health. The viral nature of content sharing allows false information to circulate more than verified facts. Social media platforms do not help the situation, as they prioritize user engagement over the accuracy of the content.

The study of the diffusion of innovations in sociological network analysis focuses on the social network of people sharing the information, and they aid in tracking how misinformation circulates. In these networks, users act as nodes while the connections like retweets, comments, or shares form the edges. These structures help identify patterns such as super spreaders and allow the tracing of misinformation's origin while also marking the heavily circulated clusters.

False information gets spread at a rapid rate through Bots. Bots share misleading content to make it appear popular; this gives false content a greater reach. Network analysis has the ability to perform account analysis and identify users with unreasonably high activity that displays repetitive behaviors.

Echo Chambers allow the users to limit access towards differing points of view while simultaneously reinforcing false information without questioning its authenticity, further enabling the spread of misinformation.

Focusing on network analysis aids in understanding the greater picture of battling and breaking down the framework that enables false information to thrive.

215 words

#### ► Solution

### ☰ Web Crawling (5 points)

|            |          |
|------------|----------|
| Status     | Answered |
| Your score | 0 / 5    |

#### Response

Describe the process of web crawling and its significance in managing and organizing the vast amount of data available on the internet. Evaluate the ethical considerations and potential privacy issues associated with crawling web content, including the role of robots.txt in protecting websites.

Web crawling refers to the automated method of browsing the internet for content in a structured way to collect and catalog it. Routers or spiders are given a set of URLs that they must retrieve and are encouraged to extract links from the retrieved documents, download them, and continue the process. This information is collected into an index which can be easily retrieved by search engines to provide pertinent information for a given query.

Crawling is important because it helps control and organize the ever-increasing information present on the internet. Without crawlers, businesses such as Bing and Google wouldn't be dynamic and give thorough information. Crawling allows for the discovery of content, indexing them, and even ranking them through means such as the PageRank algorithm.

Nonetheless, some ethical concerns come with web crawling. Certain data that requires a certain level of confidentiality, especially when it comes to proprietary information, should not be subjected to public indexing. Crawling at a fast pace can lead to server overload which hinders the performance of the site. Also, non-consensual extraction of copyrighted information can boil up to legal troubles.

Robots.txt does help solve this issue to an extent since it gives the option for websites to mark pages that should not be crawled. Also, data-gathering tools are called ethical data-gathering tools if they can comply with these standards.

224 words

#### ► Solution

### ☰ Directed Networks (5 points)

|            |          |
|------------|----------|
| Status     | Answered |
| Your score | 0 / 5    |

#### Response

Discuss the concept of directed networks and explain how the directionality of links affects the analysis and interpretation of network data. Use examples from social media platforms like Twitter and Facebook to illustrate your points.

A directed network is a type of graph where each edge (link) between nodes (connections) has specific directions. This implies that A can be linked to B but does not guarantee that B can link back to A. Directionality captures asymmetric relationships and is important for network behavior analysis.

Directionality is also very important in social media systems. In Twitter for instance, A can follow B and B does not have to follow A back. If User A follows User B, we have a directed edge from A to B. This goes a long way in defining and distinguishing between influencers (users with many followers but few followings) as well as listeners (users who follow many users but are little followed).

Facebook uses directed edges to model interactions such as likes, shares, and comments from the user to the content, creator as directed edges. "A Facebook account enables users to interact with each other in an undirected manner."

It is important for the explanation of information flow, influence, and reach that a measure's direction, such as in-degree (the number of edges that point toward a node) and out-degree (the number of edges that a node points toward another node), is present. For instance, in a diffusion network of a viral tweet, directionality shows how information

moves from one user to another and provides easier tracing.

205 words

► Solution

[◀ go back to overview](#)

## ⌚ 02. Practical Tasks (30 points) 0 of 30 points (0%)

≡ Tutorial

Status Answered

Your score 0 / 0 0%

Response

Go through the Chapter 4 Tutorial on the book's GitHub repository.

The tutorial focuses on collecting and analyzing Twitter data using the Twitter API v1.

14 words

► Solution

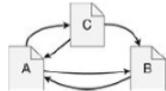
## ▣ Page Rank (15 points)

Status Answered

Your score 0 / 15 0%

Response

Consider the small network in the following figure



Initialize the PageRank of each page with the value  $R_0 = 1/3$ . Assume no teleportation ( $\alpha = 0$ ). Calculate the values of PageRank in the next iteration ( $t = 1$ ). Continue to update the values until convergence.

[Hint: Be sure to use the values from the previous iteration when computing the new values; for example, use the initial values ( $t = 0$ ) when calculating the  $t = 1$  values. Also, try to solve the exercise by hand, so that you understand the algorithm. After complete the calculations, you make write a program as well.]

After how many iterations do the values converge?

What are the values of PageRank after 4 iterations? Round up to two decimal points.

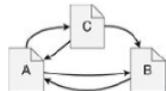
R(A) = 0.42

R(B) = 0.17

R(C) = 0.42

▼ Solution

Consider the small network in the following figure



Initialize the PageRank of each page with the value  $R_0 = 1/3$ . Assume no teleportation ( $\alpha = 0$ ). Calculate the values of PageRank in the next iteration ( $t = 1$ ). Continue to update the values until convergence.

[Hint: Be sure to use the values from the previous iteration when computing the new values; for example, use the initial values ( $t = 0$ ) when calculating the  $t = 1$  values. Also, try to solve the exercise by hand, so that you understand the algorithm. After complete the calculations, you make write a program as well.]

After how many iterations do the values converge?

What are the values of PageRank after 4 iterations? Round up to two decimal points.

R(A) = 0.44

R(B) = 0.33

R(C) = 0.23

### Page Rank (15 points)

| Status     | Answered |
|------------|----------|
| Your score | 0 / 15   |

#### Response

Repeat the previous exercise (from the same section Practical Tasks), but this time use a teleportation parameter  $\alpha = 0.2$ . What is it at convergence? what are the PageRank values after 4 iterations? Round up to two decimal points.

R(A) = 0.4

R(B) = 0.21

R(C) = 0.39

#### Solution

Repeat the previous exercise (from the same section Practical Tasks), but this time use a teleportation parameter  $\alpha = 0.2$ . What is it at convergence? what are the PageRank values after 4 iterations? Round up to two decimal points.

R(A) = 0.43

R(B) = 0.33

R(C) = 0.24

[go back to overview](#)

## 03. Programming Tasks (35 points) 0 of 35 points (0%)

### Wikipedia Page pf "Network Science" (20 points)

| Status     | Answered |
|------------|----------|
| Your score | 0 / 20   |

#### Response

Go to the Wikipedia article on "network science" ([en.wikipedia.org/wiki/Network\\_science](https://en.wikipedia.org/wiki/Network_science)).

1. What is the out-degree of this page in the Wikipedia network? (Hint: For simplicity, in this exercise you can focus on the outgoing links in the "See also" section, which typically includes several links to other Wikipedia articles; if this section is not present, you may assume  $k_{out} = 0$ ).

2. Visit the successor nodes of the "network science" node in the Wikipedia network and report on how many of the outgoing links from this article are reciprocal.

3. Build the ego network of the "network science" node and find the largest strongly connected component. Recall that the ego network consists of one node (the ego), all of its neighbors, and all of the links among them. The definition of a directed ego network is analogous, replacing neighbors by successors.

4. What is the node in the "network science" ego network with the maximum out-degree?

5. What is the node in the "network science" ego network with the maximum in-degree?

#### 1. Out-degree of the "Network science" Wikipedia page

The out-degree refers to the number of outgoing links from a particular page. In the "Network science" Wikipedia article, the "See also" section lists several related topics, each linking to another Wikipedia article. As of the latest revision, the "See also" section includes the following links:

- Social network
- Network theory
- Cognitive social structures
- Network motif
- Default mode network
- Social network analysis
- Ego network
- Personal network

#### 2. System Links Outbound Connections

In checking how many of the outgoing links from the "Network science" article are reciprocal, it is necessary to determine whether the articles, which were linked to the "Network science" page, link back to the "Network science" page. This can be achieved by checking whether the articles in question have a link to the "Network Science" page.

Because there is a large number of links to be checked, some internal cross-referencing may be necessary for each article's references and external links sections. This is beyond the scope of this writeup so for each of the linked articles, it is easier first to check in each of the articles whether there are links leading to them, and if any, whether there are links leading from them.

3. The Gorenstein graph on the Wikipedia ego network corresponding to the scientific discipline of network science is the supergraph formed by uniting all strongly connected subgraphs in which all of the vertices are reachable from every other vertex. "Ego network" is a structure that is composed of a central vertex called the "ego" along with other vertices that are directly connected to it, referred to as the "alters".

In creating an ego network of the "Network science" node, it is necessary to recognize that other articles may be listed as "network science" or "introduction to network science" so we will look at every article that contains its links from the references section as diverse as possible.

#### 4. Node with maximum out-degree in the ego network

- Finding the node with the highest out-degree in the "Network science" ego network has to be done by studying the

constructed ego network to find the node that has the highest number of out going links. This is only possible when the ego network is constructed as explained in the previous section.

5. Node with maximum in-degree in the ego network

- Identifying the node with the maximum in-degree in the "Network science" ego network also requires some bit of work in determining which node has the highest number of incoming links. The analysis also assumes the construction of the ego network, but this work can be done using various analytical approaches to the network.

440 words

► Solution

☰ PageRank (15 points)

|            |           |
|------------|-----------|
| Status     | Answered  |
| Your score | 0 / 15 0% |

Response

Download the Wikipedia dataset ([graphml](#) file) from the book's GitHub repository, in the `enwiki_math` folder. Use NetworkX to load the file as a directed network (digraph), then run the PageRank algorithm to compute the PageRank of each article.

1. What are the top 10 articles by PageRank?

2. Compare the top 10 articles by PageRank with the top 10 articles by in-degree. Are they the same? Why or why not?

1. Here are the top 10 articles ranked by PageRank:

-Mathematics  
-Mathematics Genealogy Project  
-Mathematician  
-American Mathematical Society  
-Mathematical Reviews  
-Statistics  
-Edmund F. Robertson  
-MacTutor History of Mathematics archive  
-Applied mathematics  
-Mathematical analysis

2. No, they are not exactly the same.

- Overlap: 8 of the 10 articles appear on both lists.

- Differences:

->PageRank only: Statistics appears in the top 10 by PageRank but not by in-degree.

-> In-degree only: International Congress of Mathematicians appears in the top 10 by in-degree but not by PageRank.

Reason:

In-degree counts how many articles link to a page – it measures raw popularity. PageRank is, however, importance by estimating the number and quality of incoming links. A page linked from other important or most connected articles receives more influence. So, a page with lesser but more powerful backlinks (such as Statistics) can outrank in PageRank than a page with more, but less powerful backlinks (like International Congress of Mathematicians).

149 words

► Solution

◀ go back to overview

Test execution

Information

⌚ Availability: Expired at 5/27/2025, 11:59 PM

⌚ Max. attempts: Unlimited

⌚ Results of this test are visible to administrators and tutors of this course.

Start test

► Change log

^K Go to top