This session will be taught in Hybrid format

When joining on Zoom

- When your microphone or camera are switched on you will be able to be heard in the teaching room and may be seen on the screen in the room.
- If the screen is recorded, you may be captured on the recording if you choose to contribute and have your camera switched on.
- Chat messages to everyone and privately to the host may be visible on screen in the room.

When joining in the Teaching room

- Academic teaching staff should let you know when Zoom is connected and you may see the Zoom meeting on the screen.
- Once connected to Zoom, microphones may capture sound from the whole room. Your voice may be shared to Zoom when speaking at a normal talking volume.
- Once connected to Zoom, cameras will capture video images from the room. These may be shared to Zoom. The camera capture will be a wide shot of the whole room or block of seating. It will not capture close-up images of individual students.
- If the session is recorded, video from the room will be captured on the recording. Your voice will be captured on the recording if you choose to contribute.
- If you connect to the Zoom meeting from within the room (for example to use the chat), please keep your audio on mute or use headphones to avoid feedback issues with the room audio.

The recordings may be stored in the cloud ad any personal information within the recording will be processed in accordance to the General Data Protection Regulations 2018.
A substantial contribution is considered to be anything more than merely answering questions or participating in a group discussion. Where you make a substantial contribution to the delivery of the recorded events, a signed consent form will be obtained prior to the recording being made available for viewing. The Consent Form will address your personal informationand any copyright or other intellectual property in the recording.

(Complex) Networks Analysis

Centrality Analysis (Part I): Local Measures

Felipe Orihuela-Espina

November 8, 2024

Table of Contents

Section 1

Centrality analysis (Part I): Local Measures

Centrality analysis

Centrality analysis aims at identifying the most prominent nodes or edges according to some measure of importance.

Some applications include (but are not limited to):

- Understand influence and power.
- Find critical points e.g. for attack or resilience.
- Establish popularity
- Rank web pages

Remark

Centrality is **not** about the location of the node. Remember, in a graph the location of the node is irrelevant.

Remark

The focus here is on node centrality. However, there is equally edge centrality. Sometimes, the same scores can be used interchangably or analogous definition exist for edges [3]. Occasionally, some centrality score is specific to either network constituent, whether vertices or edges.

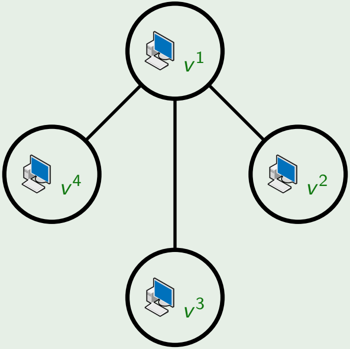
Subsection 1

Degree centrality

Centrality analysis

Example

Assuming all computers and network connections to be equal; Which computer will you attack in order to cause more disruption to the network? and more importantly, why?



Degree centrality: Node Degree

Perhaps, the simplest figure to quantify the relative importance of a node is to count the number of edges it connects to.

Definition

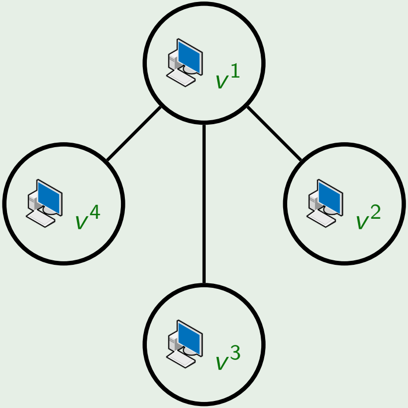
The number of edges incident of on a vertex $v \in V$ is called the **degree** of the node.

- There's no analogous concept for edges as edges always connect nodes, but that does not mean we cannot quantify the importance of edges e.g. we shall learn later of the concept of **bridges**.

Going back to our scenario...

Example

Assuming all computers and network connections to be equal; Which computer will you attack in order to cause more disruption to the network? and more importantly, why?



$$
\begin{aligned}
degree(v^1) &= 3 \\
degree(v^2) &= 1 \\
degree(v^3) &= 1 \\
degree(v^4) &= 1
\end{aligned}
$$

Degree centrality: Node Degree

Exercise

Calculate the degree of each node in the following graph:

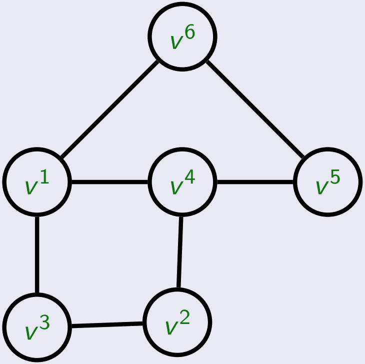


Figure: A simple graph labelled.

Degree centrality: Node Degree

Exercise

Calculate the degree of each node in the following graph:

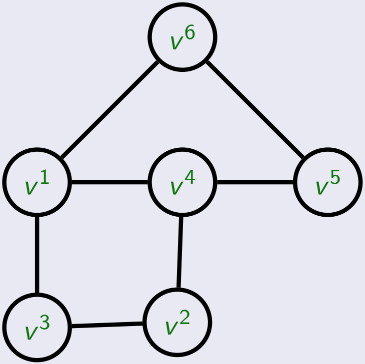


Figure: A simple graph labelled.

Degrees:

- $degree(v^1) = 3$
- $degree(v^2) = 2$
- $degree(v^3) = 2$
- $degree(v^4) = 3$
- $degree(v^5) = 2$
- $degree(v^6) = 2$

Degree centrality: Node Degree

The measure of centrality yielded by the node degree is referred to as **degree centrality**,

- The rationale behind degree centrality is that the more connected a node is, the more central it is considered.
- But watch out! The concept is not as simple as it looks. Degree centrality can be affected by
  - the network density,
  - whether the network is directed or not,
  - whether the network is weighted or not,
  - whether the network is a single graph or a multi-graph,
  - etc

Degree centrality: Node Degree

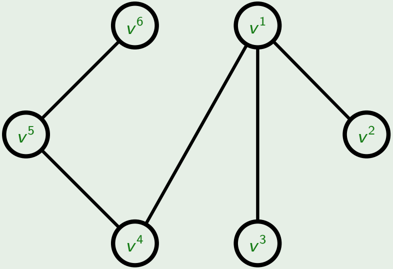
Affectance of degree centrality by density[1]

The value of the degree on its own is not as critical as it is relative to the degrees of other nodes.

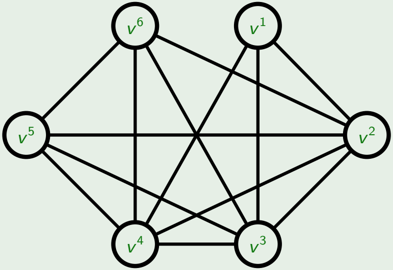
Example

Let be a node v^1 in a certain network with say degree 3

If most other nodes have lower degrees, then this is a highly central node..



If most other nodes have bigger degrees, then this node has low centrality.



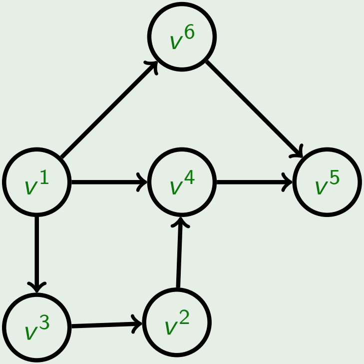
[1] We shall learn the formal concept of network density later, but by now it suffices to say that density is related to the number of edges; given a fix number of nodes, the more edges, the denser the graph, and consequetly, the higher the average degrees.

Degree centrality: Node Degree

Affectance of degree centrality by the directionality of the edges

For directed graphs, the degree of a node can be further split into the **in-degree** and the **out-degree**.

Example



Node	In-degree	Out-degree	Degree
v1	0	3	3
v2	1	1	2
v3	1	1	2
v4	2	1	3
v5	2	0	2
v6	1	1	2

Weighted Graphs

Affectance of degree centrality by the weighting of edges

Definition

Let $G = (V, E)$ be a graph and W a relation between edges and a number set $\mathbb{X}$[a]. $G_W = (V, E, W)$ is a **weighted graph** if every $e \in E$ has an associated numerical value $w(e)$, $\forall e \in E : w : e \mapsto x$ with $w(e) \in W$ and $x \in \mathbb{X}$.

[a] A **number set** is a set for which all of its elements are numbers. Common number sets are $\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, etc

- $w(e)$ is called the **weight** or **length** of e[2].
- If $w(e) < 0$, then e is said to be a **negative edge**.

...It goes without saying graphs can be both directed and weighted at the same time.

[2] Do not confuse the **length** of and edge, with the **length** of a chain!

Affectance of degree centrality by the weighting of edges

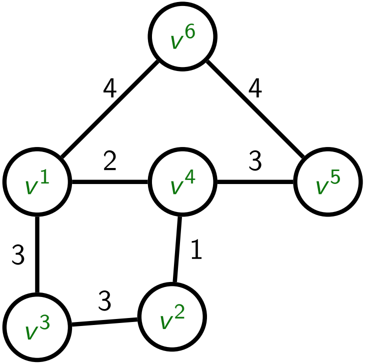


Figure: Weighted graphs.

Degree centrality

Affectance of degree centrality by the weighting of edges

- The sum of weights of all edges connected to that node is called the **strength** of the node.

$$strength(v^i) = \sum_j w_{ij}$$

where $w_{ij} = weight(e(v^i, v^j))$

Degree centrality

Affectance of degree centrality by the weighting of edges

How would you establish the node's degree centralilty in weighted graphs?

Affectance of degree centrality by the weighting of edges

How would you establish the node's degree centralilty in weighted graphs?

Some options already in the literature include (but are not limited to):

- (naive) Ignore weights and count each edge equally regardless of the weight.
- Use the node *strength* as a proxy of its degree.

...and of course there are variants for when your graph is directed.

Affectance of degree centrality by the multiplicity of edges

- A collection of *repeated* edges in E is called a **multi-side**, a **multiple edge** or **parallel edges**.

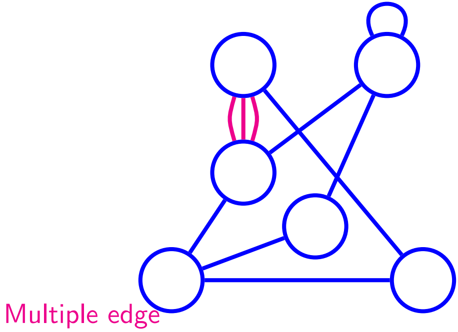


Figure: Multiple edges.

Affectance of degree centrality by the multiplicity of edges

Definition

A **multigraph** G is an ordered pair $G = (V, E)$ where V is a set of **nodes** or **vertices**, and $E \subseteq V \times V$ is a multiset[a] of (not ordered) pairs of vertices referred to as **edges**.

[a] In a set, the multiplicity of the elements is irrelevant; all elements may be considered different for all purposes. A **multiset** is a set that can contain many copies of the same element. In a multiset, multiplicity of objects is explicit and relevant.

- A graph without multiple edges (or loops) is referred to as **simple graph**
- ...and yes, there are weighted and directed variants.

Affectance of degree centrality by the multiplicity of edges

How would you establish the node's degree centralilty in multigraphs?

Degree centrality

Affectance of degree centrality by the multiplicity of edges

How would you establish the node's degree centralilty in multigraphs?

Some options already in the literature include (but are not limited to):

- Dichotomize the network to remove the multiple edges (count every parallel edge as just one) and proceed as in a case of a single graph.
- Count every parallel edge independentlly to yield a weight proportional to the number of parallel edges and proceed as in a case of weighted graphs

...and of course there are variants for when your graph is weighted and/or directed.

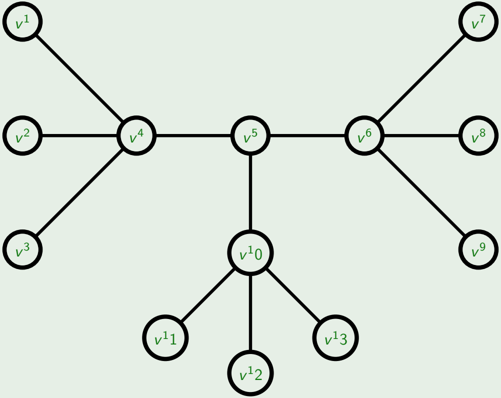
Subsection 2

Betweenness centrality

Centrality analysis

Example

You work for the Royal Mail. Assuming all nodes and edges to be equal; Which distribution center supports more traffic? and more importantly, why?



Centrality analysis

Degree centrality is not the only way to measure centrality.

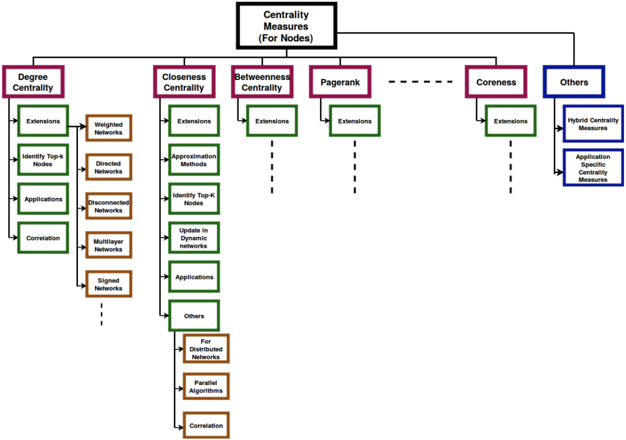


Figure: Measures of centrality. Figure reproduced from [4].

A different notion of importance is related to *network flows*.

Under this notion of importance, a node is more central the more flows (transfers across the network) it sustains. This is referred to as **betweenness centrality**.

But how would you measure flows then?

Assuming no weights, or directions, it is reasonable to assume that flows or transfers occur across the **shortest paths**.

So let's define (shortest) paths...

Graphs: Chains and paths

Definition

Let $G = (V, E)$ be a graph and $S_G(v^i, v^{j+1}) = \{v^i, e^i, v^{i+1}, e^{i+1}, \dots, v^{j+1}\}$ be a sequence[a] of alternating vertices in V and edges, starting and finishing on vertices, and where $e^k = (v^k, v^{k+1}) \in E$. The sucession[b] of S_G, $s_G = \{S_G[n]\}_{n=1}^{|S_G|}$ is known as **chain**.

- v^i is called the **start** vertex.
- v^{j+1} is called the **end** vertex.

[a] A **sequence** is an ordered set.

[b] A **sucession** is the *act* of following of a sequence. For most practical purposes, the terms sucession and sequence are interchangeable.

- If a graph is simple, sometimes you will see the chain described only by the set of vertices in the sequence s_G; i.e. $S_G(v^i, v^{j+1}) = \{v^i, e^i, v^{i+1}, e^{i+1}, \dots, v^{j+1}\} \equiv \{v^i, v^{i+1}, \dots, v^{j+1}\}$

Graphs: Chains and paths

- If the start and end vertex are the same, $v^i = v^{j+1}$, the chain is said to be **closed**. Otherwise, it is said to be **open**.
- Two vertices $v^i, v^{j+1} \in V$ are:
  - **connected** if there is a chain (or path) between them; $\exists s_G(v^i, v^{j+1}) : v^i, v^{j+1}$ are connected
  - **disconnected** if there isn't a chain (or path) between them; $\nexists s_G(v^i, v^{j+1}) : v^i, v^{j+1}$ are disconnected

Graphs: Chains and paths

Let $s_G(v^i, v^j)$ be a chain. Some types of chains are summarised in Table 1.

Type	Vertices can repeat	Edges can repeat	Open / Close[3]	Directed
Walk	✓	✓	Either	Either
Trail	✓	✗	Either	Either
Circuit	✓	✗	Closed	Either
(Simple) Path	✗	✗	Open	✗
Directed Path or Trajectory	✗	✗	Open	✓
Cycle	✗	✗	Closed	✗
Directed Cycle	✗	✗	Closed	✓

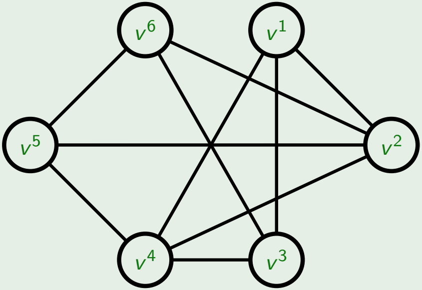
Table: Types of chain in a graph.

[3] For closed sequences, the start and end vertices may repeat.

Betweenness centrality

Example

Consider the following graph:



Here are the paths to go from v^1 to v^6;

$$S_G(v^1, v^6) = \{v^1, v^2, v^6\}$$
$$S_G(v^1, v^6) = \{v^1, v^3, v^6\}$$
$$S_G(v^1, v^6) = \{v^1, v^4, v^5, v^6\}$$
$$S_G(v^1, v^6) = \{v^1, v^4, v^2, v^6\}$$
$$S_G(v^1, v^6) = \{v^1, v^3, v^4, v^5, v^6\}$$
$$S_G(v^1, v^6) = \{v^1, v^2, v^4, v^5, v^6\}$$
$$S_G(v^1, v^6) = \{v^1, v^4, v^5, v^2, v^6\}$$

The following however is not a path (as some nodes are repeated);

$$S_G(v^1, v^6) = \{v^1, v^2, v^4, v^5, v^2, v^6\}$$

Betweenness centrality

Now that we know what a path is, we can define the **shortest path** (a.k.a geodesic) as the path requiring the sequence traversing less number of edges in the chain (**length of the path**).

For weighted graphs:

- The sum of the lengths of the edges in a chain is the **weight** or **length** of the (weighted) chain.
- The shortest path between two connected vertices $v^i, v^j \in V$ in a weighted graph is the path with the least weight among all the paths between those vertices.

Betweenness centrality

Example

In the previous example;

$\{v^1, v^2, v^6\}$ is shorter than $\{v^1, v^3, v^4, v^5, v^6\}$.

Note that there may be more than one path sharing the same length and hence there may be more than one shortest path, e.g. $\{v^1, v^2, v^6\}$ has the same length than $\{v^1, v^3, v^6\}$ and since no other path between v^1 to v^6 is shorter, then both of these paths are the shortest path from v^1 to v^6.

Betweenness centrality

Remember; in betweenness centrality, a node is more central the more flows it sustains.

Under the assumption that flows occur across the shortest path, betweenness centrality can be established by simply counting the number of shortest paths that cross a certain node.

The higher the number of shortest paths that cross a certain node, the more central the node is.

Betweenness centrality

Perhaps, the most common estimator of the betweenness centrality of a node v^k, $g(v^k)$, is Freeman's betweenness centrality, given by Eq (1);

$$g(v^k) = \sum_{\substack{v^i, v^j \in V \\ v^i \neq v^k, v^j \neq v^k}} \frac{\sigma(v^i, v^j | v^k)}{\sigma(v^i, v^j)} \tag{1}$$

where:

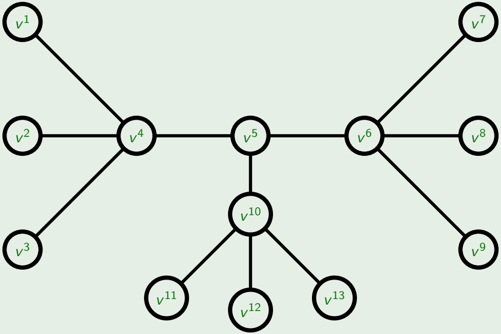
- $\sigma(v^i, v^j)$ is the number of shortest paths between nodes v^i and v^j
- $\sigma(v^i, v^j | v^k)$ is the number of shortest paths between nodes v^i and v^j that pass through v^k.
- Note 1: if $v^i = v^j$ then $\sigma(v^i, v^j) = 1$
- Note 2: if $v^k = v^i$ or $v^k = v^j$ then $\sigma(v^i, v^j | v^k) = 0$
- Note how we do *not* count the cases where the node is the start or the end of the path, i.e. $v^i \neq v^k, v^j \neq v^k$.

Betweenness centrality

We are now ready to go back to our scenario.

Example

You work for the Royal Mail. Assuming all nodes and edges to be equal; Which distribution center supports more traffic? and more importantly, why?



Example

(Cont.) First, let's make a list of all shortest paths; we can do this by creating a large table (larger than what fits in this slide) where the rows represent origins and columns represent ends.

	v^1	v^2	$\dots$	v^n
v^1	NA	$sp_{1,2}$	$\dots$	$sp_{1,n}$
v^2	$sp_{2,1}$	NA	$\dots$	$sp_{2,n}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
v^n	$sp_{n,1}$	$sp_{n,2}$	$\dots$	NA

The content of each cell in this table, is the set of *all* shortest paths between the corresponding nodes $sp_{i,j}$.

In this case, because this is a non-directed graph, the table will be symmetric i.e. the set shortest paths from v^i to v^j is the same as the set shortest paths from v^j to v^i, $sp_{i,j} = sp_{j,i}$.

Example

(Cont.) Let's do a couple of these by hand for the purposes of exemplification:

$$sp_{1,2} = \{\{v^1, v^4, v^2\}\}$$

Note the double curly braces as this is a set of sets; each shortest path is a set and $sp_{i,j}$ is a set of shortest paths.

$$sp_{3,13} = \{\{v^3, v^4, v^5, v^{10}, v^{13}\}\}$$

Can you do the rest on your own?

Example

(Cont.) The next step just requires doing a loop; For each node, simply go over the table and find how many times the node appears in all cells ignoring the row and column of the node itself, i.e. across all possible pairs of nodes but where the node itself is neither the start or the end of the path.

For instance, for node v^4:

$$
\begin{aligned}
v^4 :& sp_{1,2}, sp_{1,3}, sp_{1,5}, sp_{1,6}, sp_{1,7}, sp_{1,8}, sp_{1,9}, sp_{1,10}, sp_{1,11}, sp_{1,12}, sp_{1,13} \\
& sp_{2,3}, sp_{2,5}, sp_{2,6}, sp_{2,7}, sp_{2,8}, sp_{2,9}, sp_{2,10}, sp_{2,11}, sp_{2,12}, sp_{2,13} \\
& sp_{3,5}, sp_{3,6}, sp_{3,7}, sp_{3,8}, sp_{3,9}, sp_{3,10}, sp_{3,11}, sp_{3,12}, sp_{3,13}
\end{aligned}
$$

Note how v^4 does *not* participate in any short path between nodes "below" or on the "right" of v^5.

Betweenness centrality

Example

(Cont.) Hence, the Freeman's betweeness centrality for v^4 is:

$$g(v^4) = \sum_{v^i, v^j \in V} \frac{\sigma(v^i, v^j | v^k)}{\sigma(v^i, v^j)} = 30$$

In this naive example there does not happen to be any $sp_{i,j}$ with more than 1 shortest path i.e. $\sigma(v^i, v^j) = 1$ in all cases, which makes calculations a bit easier. Should there be more than one shortest path in one set, the value of that path ought to be divided by the number of shortest paths in $sp_{i,j}$ as per Freeman's betweeness centrality estimator.

Betweenness centrality

Example

(Cont.) We can proceed equally for each node to find the one with greatest Freeman's betweeness centrality.

You can verify the betweeness centrality for each node:

$$\begin{array}{l|l|l} g(v^1) = 0 & g(v^6) = 30 & g(v^{11}) = 0 \\ g(v^2) = 0 & g(v^7) = 0 & g(v^{12}) = 0 \\ g(v^3) = 0 & g(v^8) = 0 & g(v^{13}) = 0 \\ g(v^4) = 30 & g(v^9) = 0 & \\ g(v^5) = 48 & g(v^{10}) = 30 & \end{array}$$

Therefore, in this example, the node with the highest betweeness centrality is v^5 which does *not* coincide with the node(s) with greater degree centrality $\{v^4, v^6, v^{10}\}$.

Betweenness centrality

Advanced material

The calculation of Freeman's betweenness centrality by brute force is simple but computationally expensive.

Brandes' algorithms [1, 2] are a more efficient option to calculate Freeman's betweenness centrality by they are bit more difficult to understand given that some calculations are reused and hence it is difficult to keep track of what's happening. Different variants exist for different types of graphs but the rationale is the same; aggregate pairwise dependencies and hence eliminating the need to calculate all dependencies explicitly.

To know more: https://www.cl.cam.ac.uk/teaching/1617/MLRD/slides/slides13.pdf

Betweenness centrality

Similar to degree centrality, betweenness centrality also is affected by many factors e.g.

- directionality of edges,
- presence of edge weights,
- etc

Hence a considerable number of estimators exists for the different types of graphs.

Advance material

We won't get into further details here, but if you are interested you can find a nice review in [4].

References I

Ulrik Brandes.
A faster algorithm for betweenness centrality.
Journal of mathematical sociology, 25(2):163–177, 2001.

Ulrik Brandes.
On variants of shortest-path betweenness centrality and their generic computation.
Social networks, 30(2):136–145, 2008.

Timo Bröhl and Klaus Lehnertz.
A straightforward edge centrality concept derived from generalizing degree and strength.
Scientific Reports, 12(1):4407, 2022.

Akrati Saxena and Sudarshan Iyengar.
Centrality measures in complex networks: A survey.
arXiv preprint arXiv:2011.07190, 2020.

Questions

Thank you! Questions?