Social Network Analysis

Centrality

Dr. Chun-Hsiang Chan

Department of Geography,
National Taiwan Normal University

Photo credit: midjourney

Outline

– Centrality
– Degree Centrality
– Closeness Centrality
– Betweenness Centrality
– Information Centrality
– Eigenvector Centrality
– HITS algorithm
– PageRank
– Paper Reading
– References

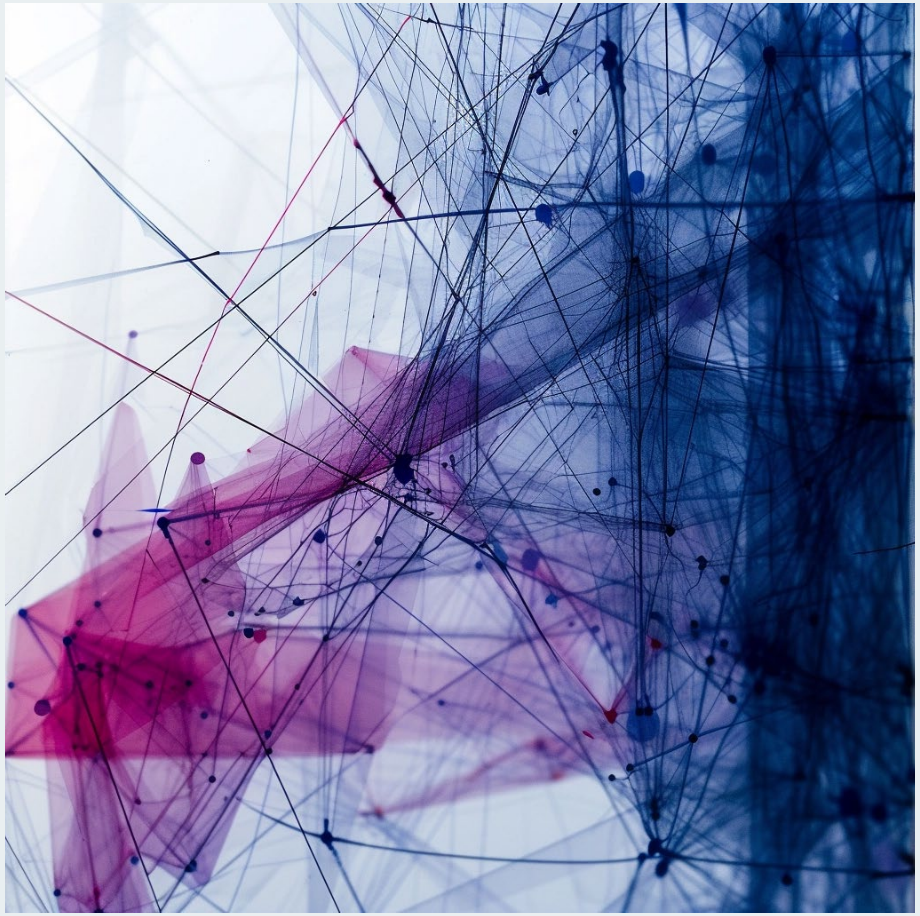


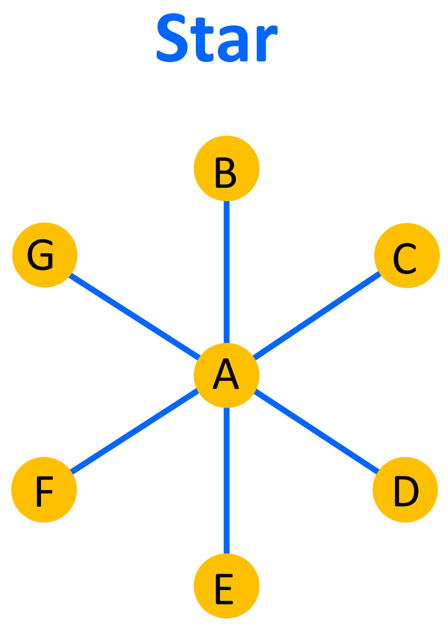
Photo credit: midjourney

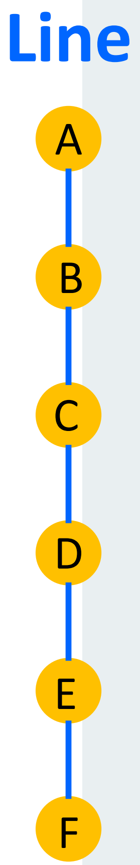
Centrality

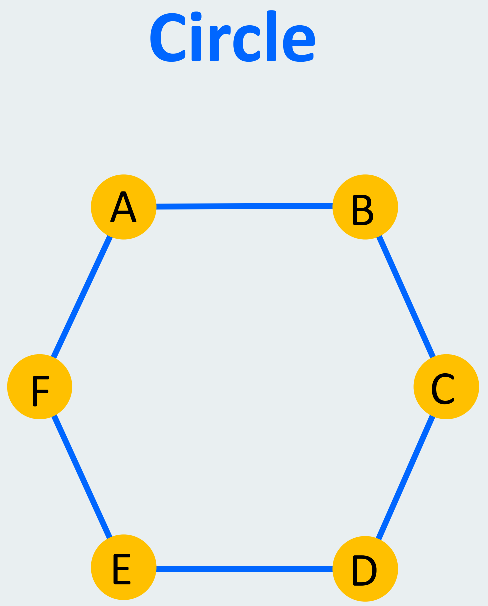
– All sociologists would agree that power is a fundamental property of social structures. There is much less agreement about power and how we can describe and analyze its causes and consequences. We will look at some of the main approaches that social network analysis has developed to study power and the closely related concept of centrality.

– To understand the basic concept of various centrality measurements, we thereby implement three simple networks to demonstrate them: **star**, **line**, and **circle**.

Centrality

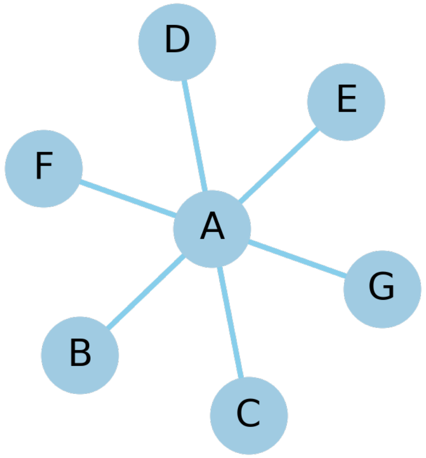


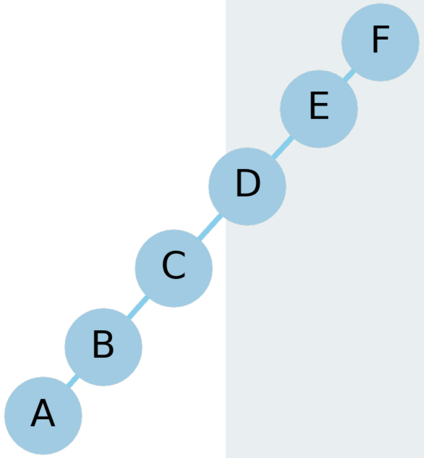


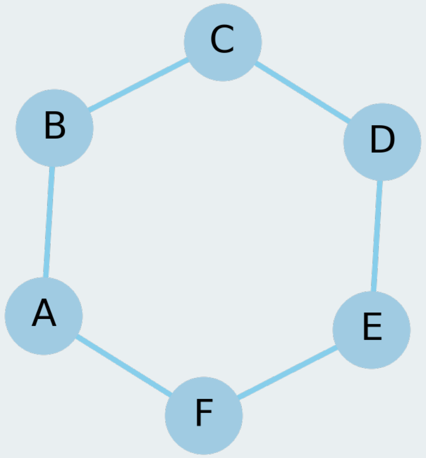


Degree Centrality

– Actors who have more ties to other actors may be in advantaged positions. Because they have many ties, they may have alternative ways to satisfy needs and are less dependent on other individuals. Because they have many ties, they may have access to and can call on more of the network's resources as a whole.







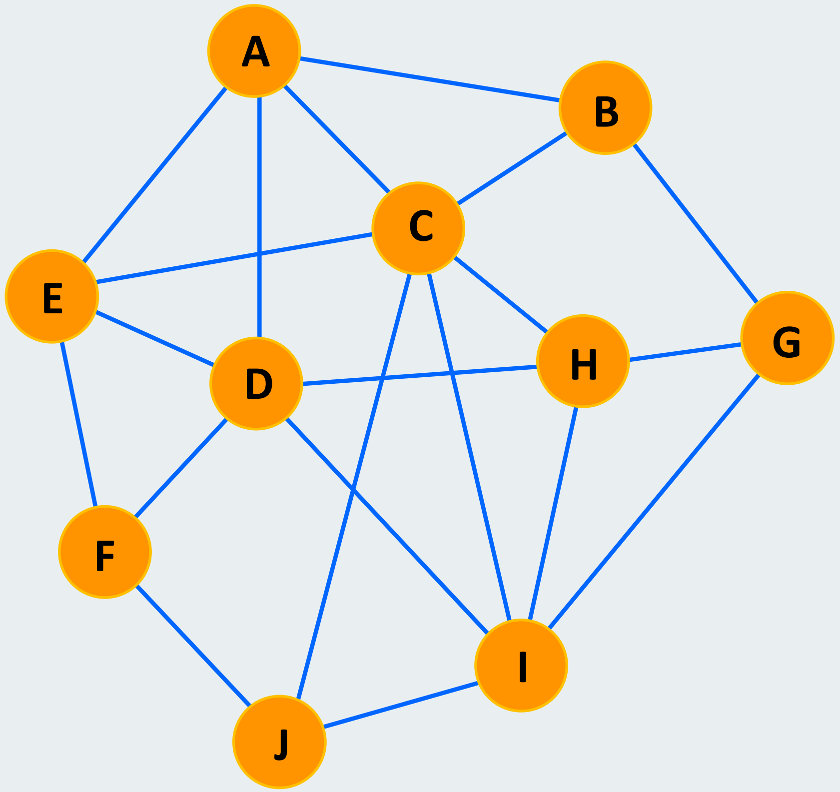
Degree Centrality

– **Degree**

$$C_D = d(n_i) = \sum_j X_{ij}$$

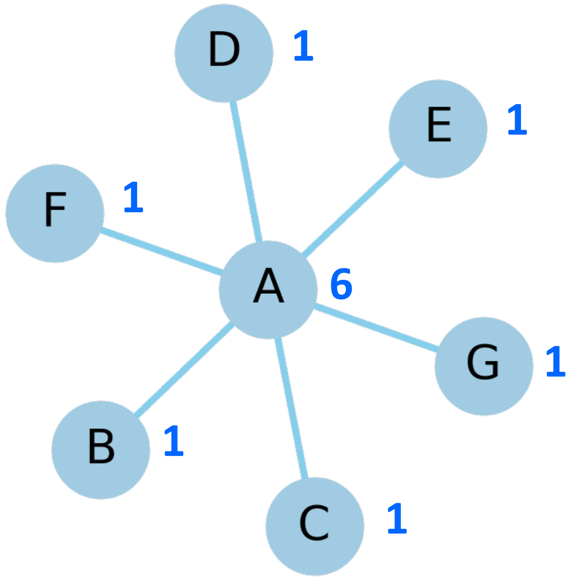
– **Normalized Degree**

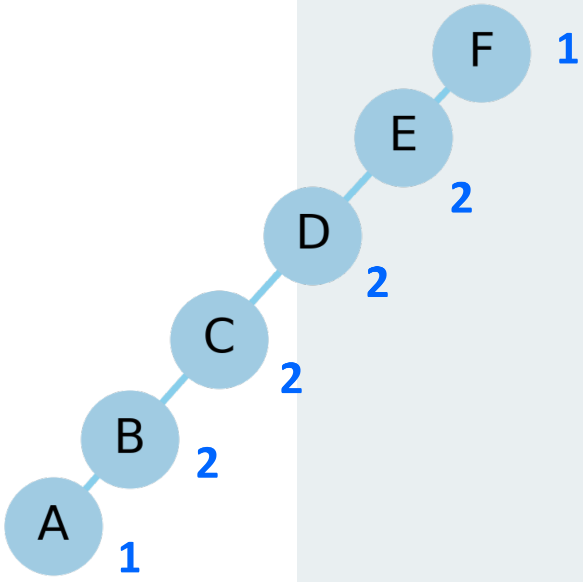
$$C_D = \sum \frac{d_i}{N-1}$$

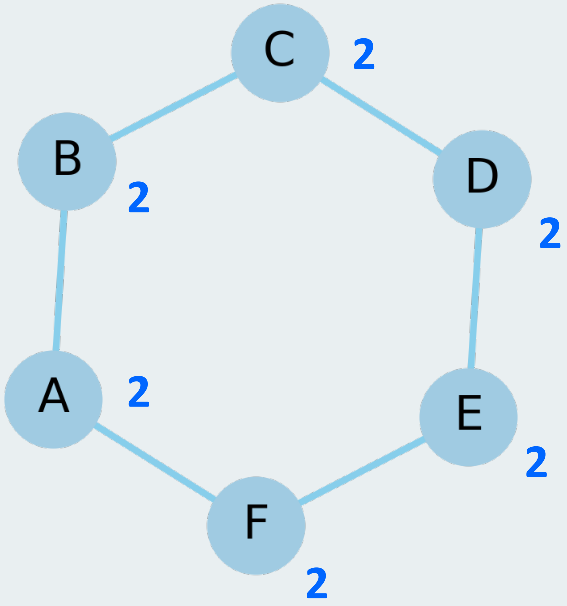


Degree Centrality

– Degree Measurement

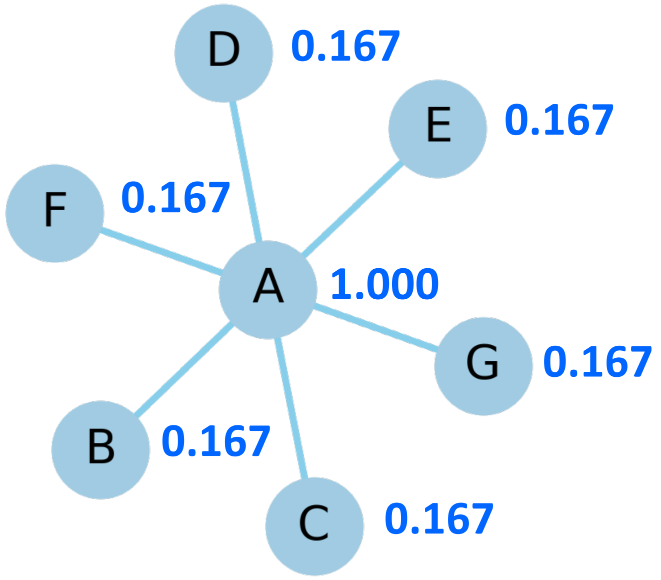


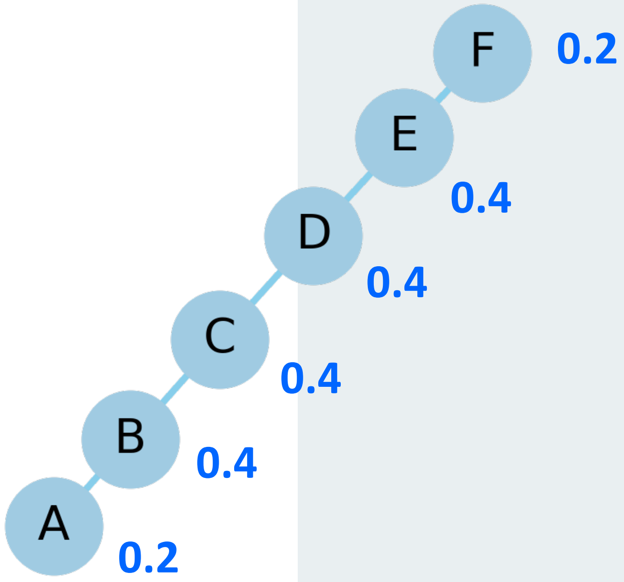


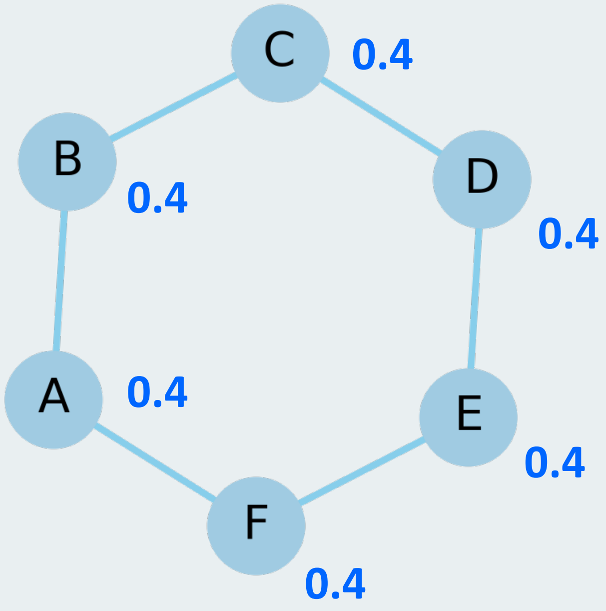


Degree Centrality

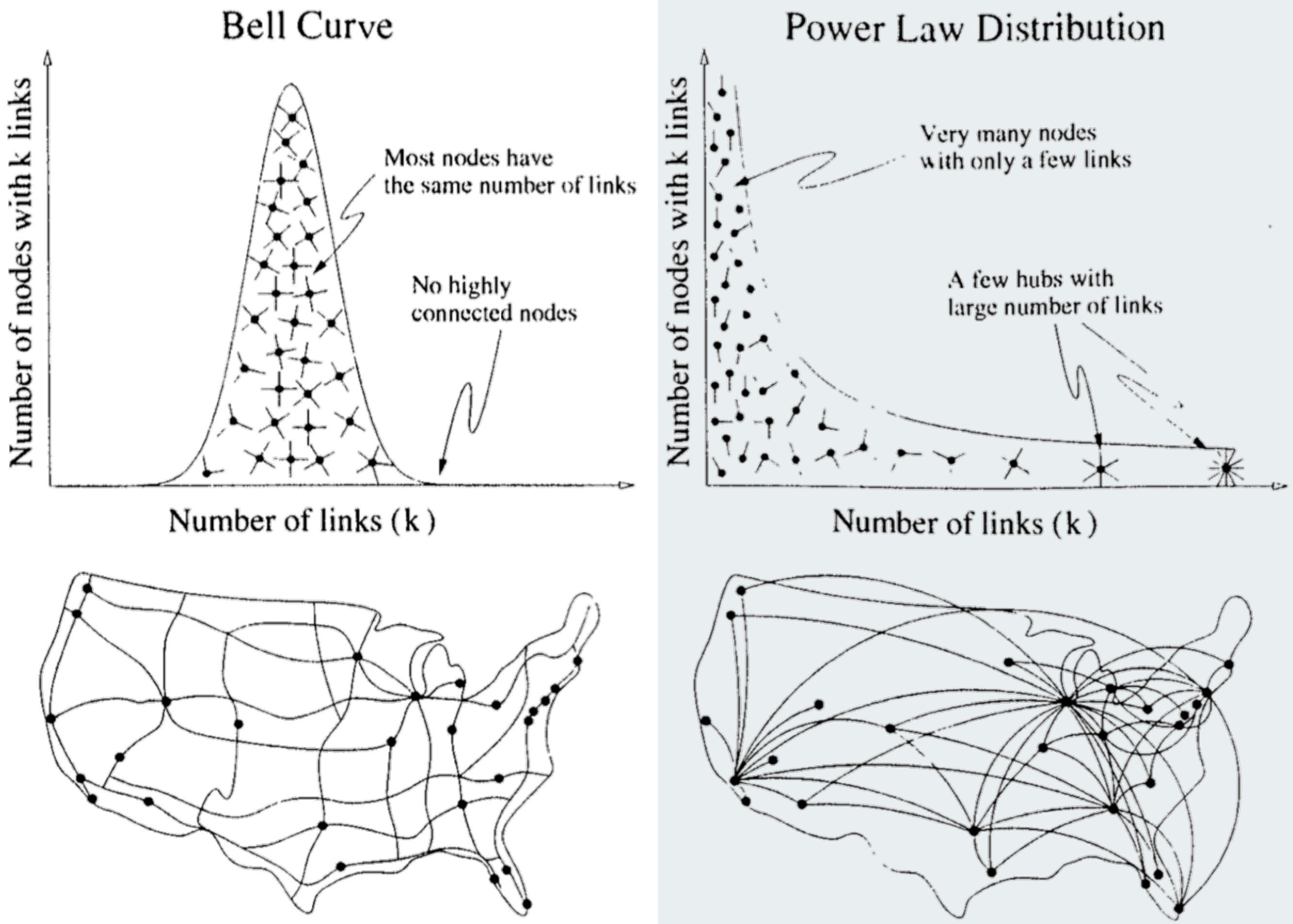
– Degree Centrality Measurement







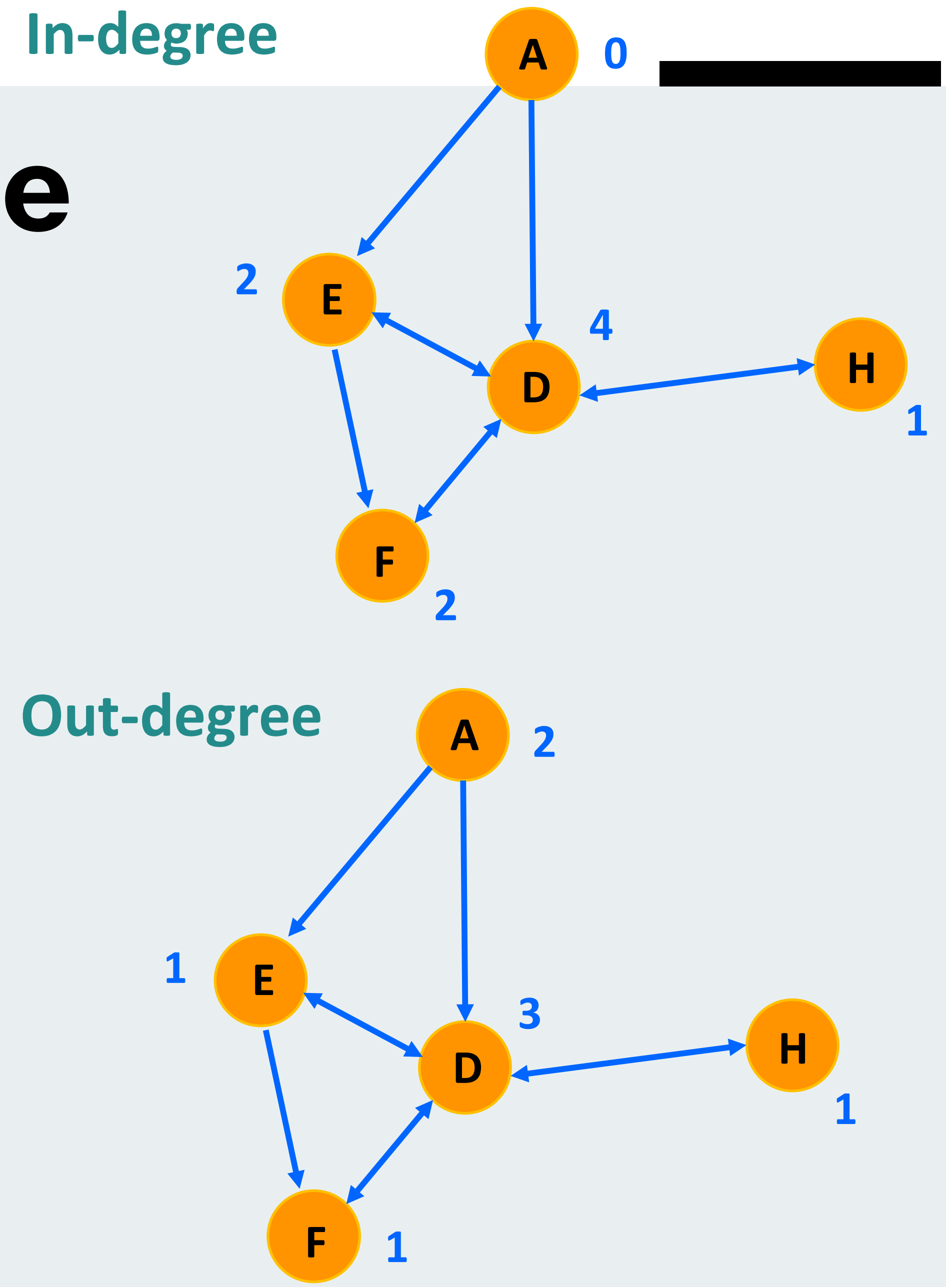
Degree Centrality Application



Bruno Gonçalves (2021) Preferential Attachment and the Barabási-Albert model. https://graphs4sci.substack.com/p/preferential-attachment-and-the-barabasi

In-degree & Out-degree Centrality

– For a directed network, we may connect to others by different linkage directions.
  - **In-degree** is the number of ties received.
  - **Out-degree** is the number of ties sent.



Closeness Centrality

– Degree centrality measures might be criticized because they only take into account the immediate ties that an actor has or the ties of the actor’s neighbors rather than indirect ties to all others.

– One actor might be tied to a large number of others, but those might be rather disconnected from the network as a whole. In a case like this, the actor could be quite central, but only in a local neighborhood.

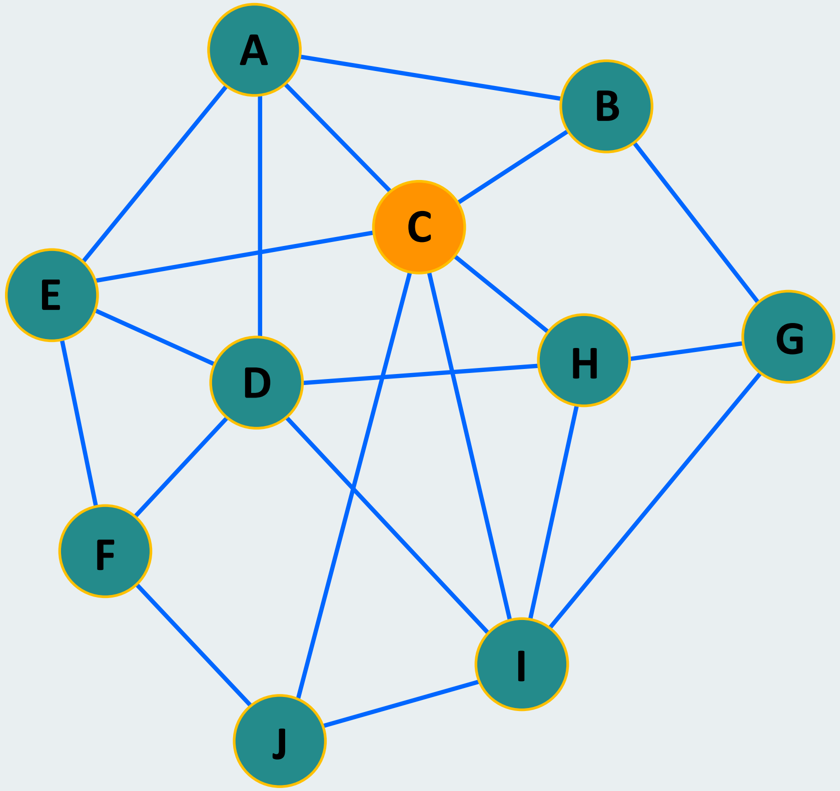
Closeness Centrality

- A further distance indicates a lower impact.
- **Closeness Centrality**

$$C_c(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j)\right]^{-1}$$

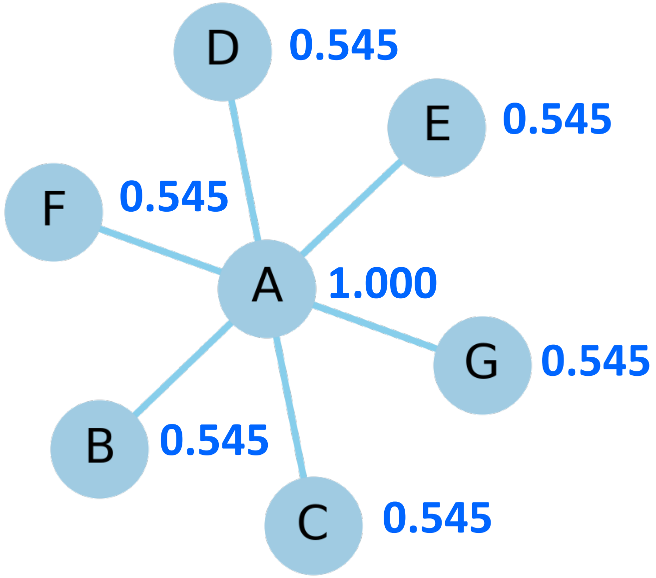
- **Normalized Closeness Centrality**

$$C_c(n_i) = (g-1)\big(C_c(n_i)\big)$$



Closeness Centrality

– Star network

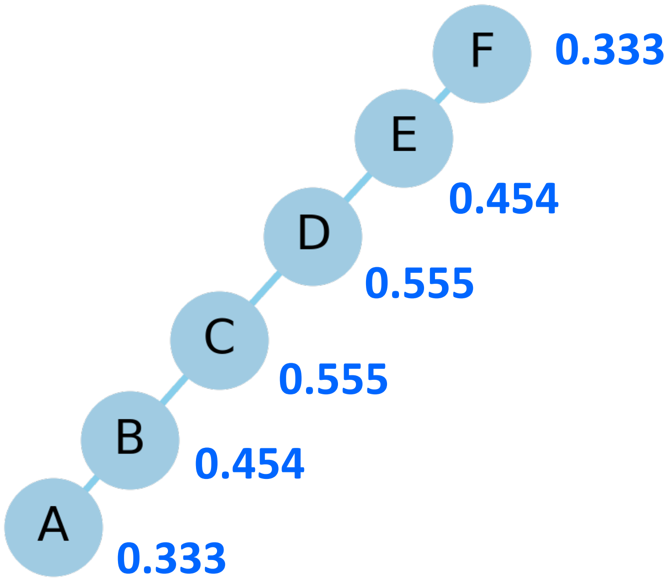


	A	B	C	D	E	F	G
A	0	1	1	1	1	1	1
B	1	0	2	2	2	2	2
C	1	2	0	2	2	2	2
D	1	2	2	0	2	2	2
E	1	2	2	2	0	2	2
F	1	2	2	2	2	0	2
G	1	2	2	2	2	2	0

Closeness	Normalized
0.1666	1.0000
0.0909	0.5454
0.0909	0.5454
0.0909	0.5454
0.0909	0.5454
0.0909	0.5454
0.0909	0.5454

Closeness Centrality

– Line network

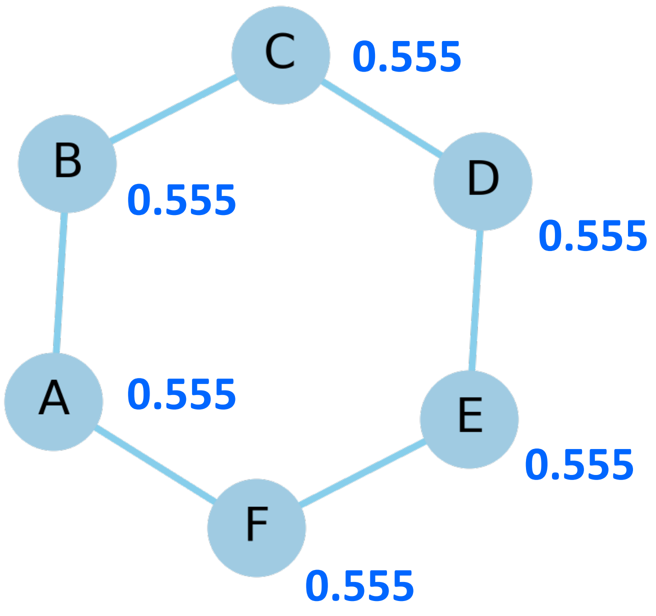


	A	B	C	D	E	F
A	0	1	2	3	4	5
B	1	0	1	2	3	4
C	2	1	0	1	2	3
D	3	2	1	0	1	2
E	4	3	2	1	0	1
F	5	4	3	2	1	0

Closeness	Normalized
0.0666	0.3333
0.0909	0.4545
0.1111	0.5555
0.1111	0.5555
0.0909	0.4545
0.0666	0.3333

Closeness Centrality

– Circle network



	A	B	C	D	E	F
A	0	1	2	3	2	1
B	1	0	1	2	3	2
C	2	1	0	1	2	3
D	3	2	1	0	1	2
E	2	3	2	1	0	1
F	1	2	3	2	1	0

Closeness	Normalized
0.111	0.555
0.111	0.555
0.111	0.555
0.111	0.555
0.111	0.555
0.111	0.555

Closeness Centrality

– How about measuring closeness centrality for a directed network?

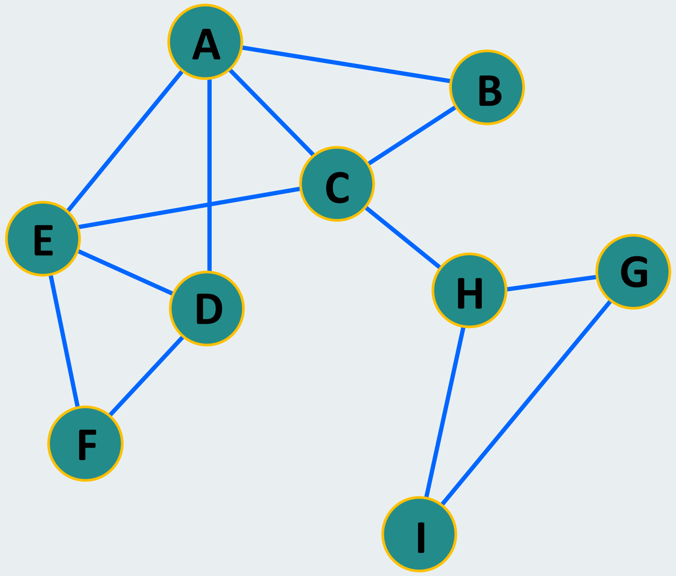
– The closeness centrality measurement in Networkx, a Python package for network analysis, is based on the in-flow distance.

– If you want to measure the out-flow-based closeness centrality, then you may use the function “G.reverse()” instead of.

https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.centrality.closeness_centrality.html

Betweenness Centrality

– Suppose that you are the leader of entertainment in a university and have to hold a campus activity; however, you do not have the contact information for some departments.
– Therefore, you have to contact these departments through some classmates (agents) in order to achieve your goal. These classmates have agent characteristics because they have different friends from various departments.
– To quantify this characteristic, we introduce a new network indicator, "betweenness Centrality."

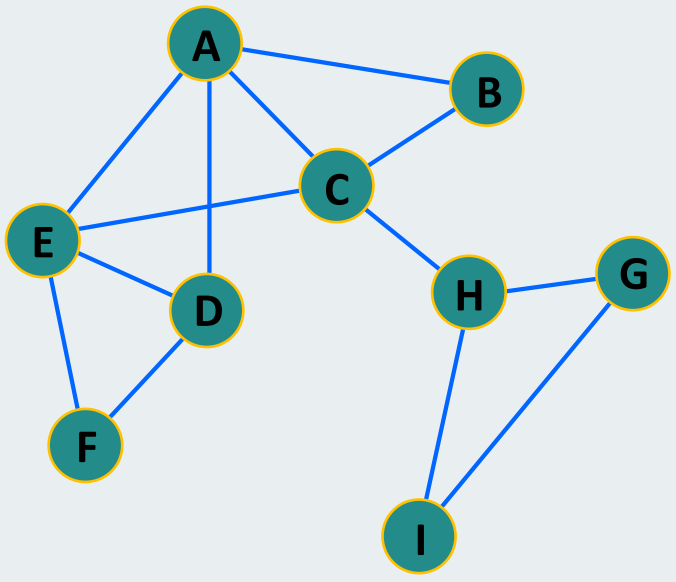


Betweenness Centrality

- The mathematical definition of betweenness centrality as follows,

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

- where σ_{st} is the total number of shortest paths from node s to node t and $\sigma_{st}(v)$ is the number of those paths that pass through v (not where v is an end point).



Betweenness Centrality

– For a normalized betweenness centrality, the betweenness centrality of a node scales with the number of pairs of nodes as suggested by the summation indices. Therefore, the calculation may be rescaled by dividing through by the number of pairs of nodes not including n, so that $g \in [0,1]$. The division is done by $(N-1)(N-2)$ for directed graphs and $(N-1)(N-2)/2$ for undirected graphs, where N is the number of nodes in the giant component.

$$normal\big(g(n)\big) = \frac{g(n) - \min(g)}{\max(g) - \min(g)}$$

Betweenness Centrality

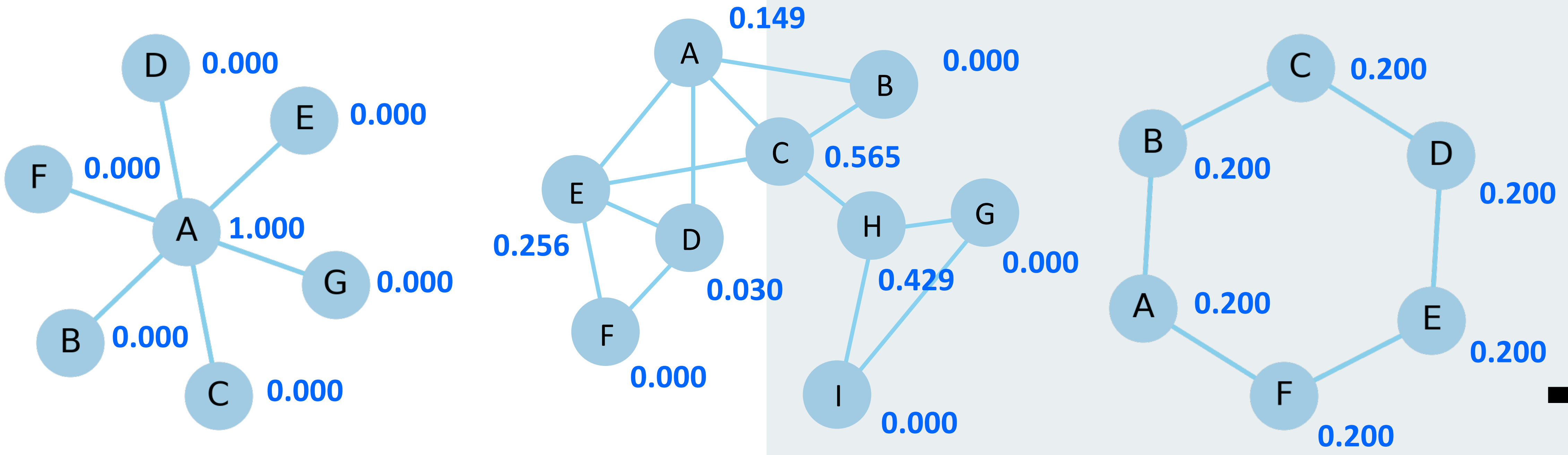
– In a weighted network, the links connecting the nodes are no longer treated as binary interactions but are weighted in proportion to their capacity, influence, frequency, etc., which adds another dimension of heterogeneity within the network beyond the topological effects. A node's strength in a weighted network is given by the sum of the weights of its adjacent edges.

$$s_i = \sum_{j=1}^{N} a_{ij} w_{ij}$$

With a_{ij} and w_{ij} being adjacency and weight matrices between nodes i and j, respectively.

Betweenness Centrality

– Examples



Freeman Discussion

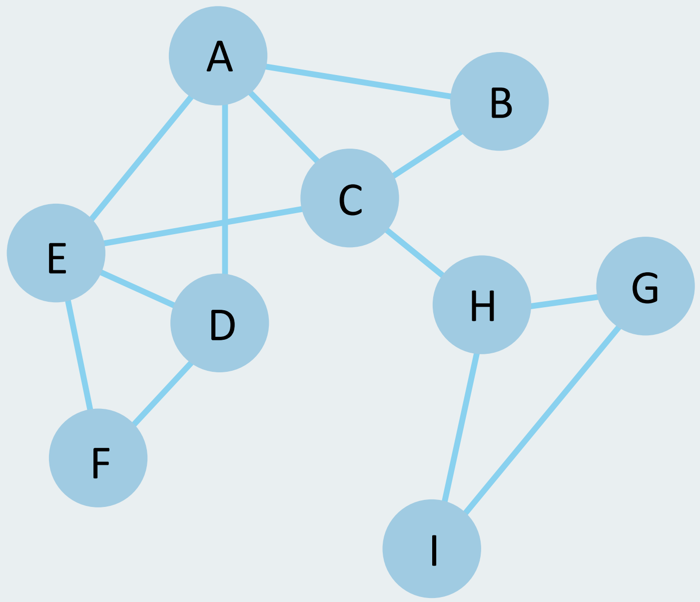
- Degree centrality is to measure the initialization of communication within a network.
- Closeness and betweenness centrality are to measure the impact ability of your neighbors to the others within a network, rather than directly evaluating your connections (i.e., degree centrality).

Network-level Degree Centralization

– **Degree Centralization**

$$C_D = \frac{\sum_{i=1}^{g}[C_D(n^*) - C_D(i)]}{(N-1)(N-2)}$$

$C_D(n^*)$ is the maximum value within the network.



Code Practice:
Please utilize the left-hand-side network example to calculate the degree centralization value of this network.

Network-level Degree Centralization

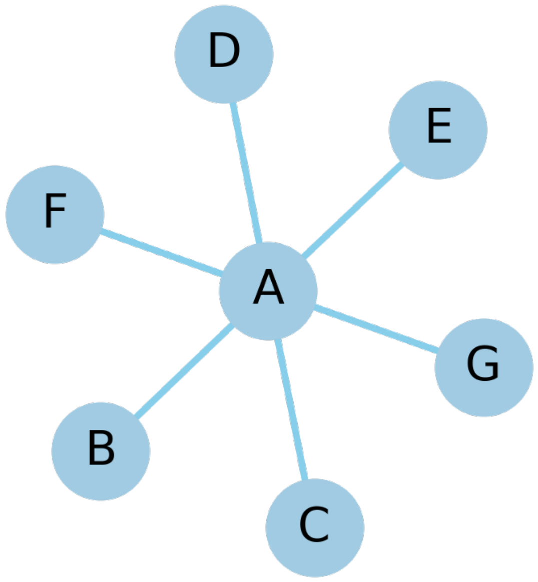
– In Snijders (1981), the scientist adopted the concept of variance for measuring the centralization of a network.

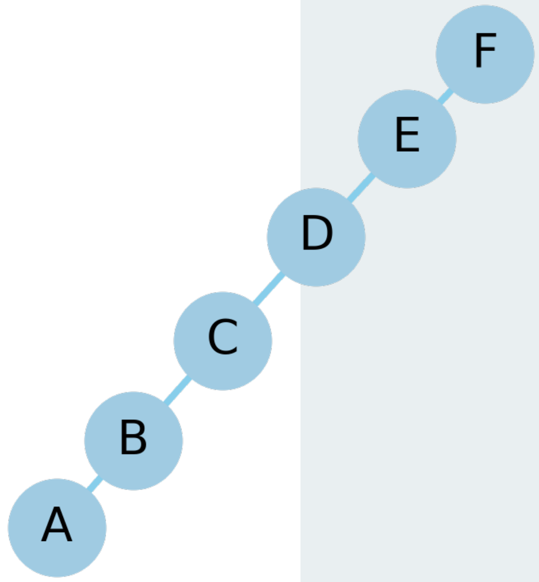
$$S_D^2 = \frac{\sum_{i=1}^{g}(C_D(n_i) - \overline{C_D})^2}{g}$$

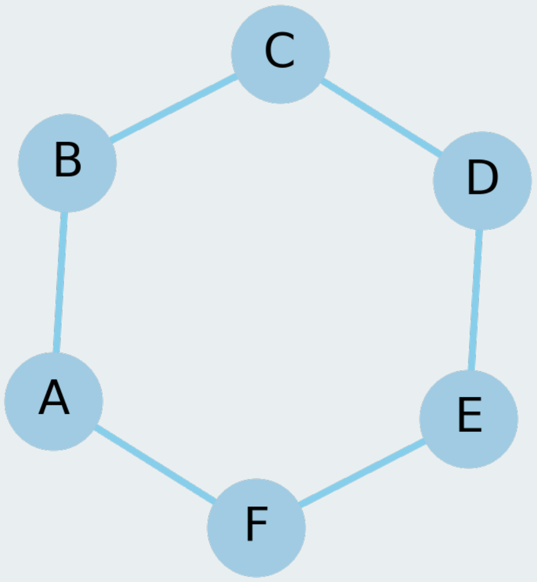
– This indicator equips the concept of variance, which represents the dispersion and heterogeneity between actors (nodes) within the network.

Network-level Degree Centralization

– Please calculate these two network-level degree centralization indicators of the following three networks.







Information Centrality

- Information centrality further considers the information flow between nodes within a network.
- There are two reasons:
  - Assume that the chosen probability of each shortest path is equal
  - Only consider the shortest path and neglect other paths
- Stephenson & Zelen (1989) developed “information centrality,” considering all paths between two nodes with weighting parameters of path distance.
- The further path distance indicates a lower quality of information flow.

Information Centrality

– Current-flow closeness centrality is a variant of closeness centrality based on effective resistance between nodes in a network. This metric is also known as information centrality.

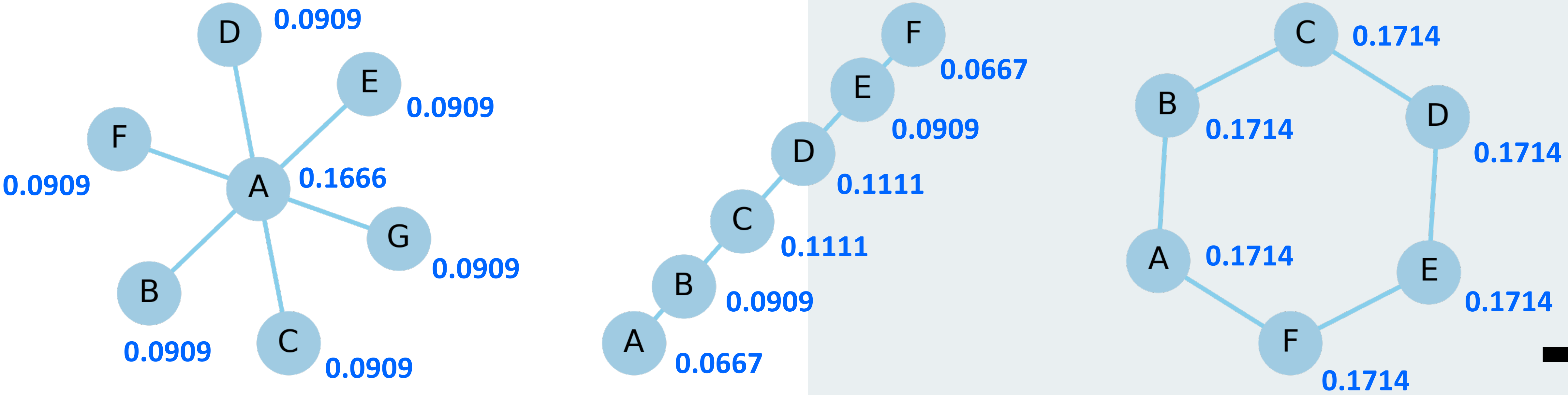
$$C_I = \frac{n}{\sum_{j=1}^{n} \frac{1}{I_{ij}}}$$

– where n is the number of nodes and I_{ij} the Laplacian matrix calculated from the information centrality of a path from node i to j.

– The information centrality of a node is measured by an average of all paths originating from the specific node.

Information Centrality

– Examples



Eigenvector Centrality

– **Eigenvector centrality** (**eigencentrality** or **prestige score**) measures a node's influence in a network. Relative scores are assigned to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes. A high eigenvector score means that a node is connected to many nodes that themselves have high scores.

Eigenvector Centrality

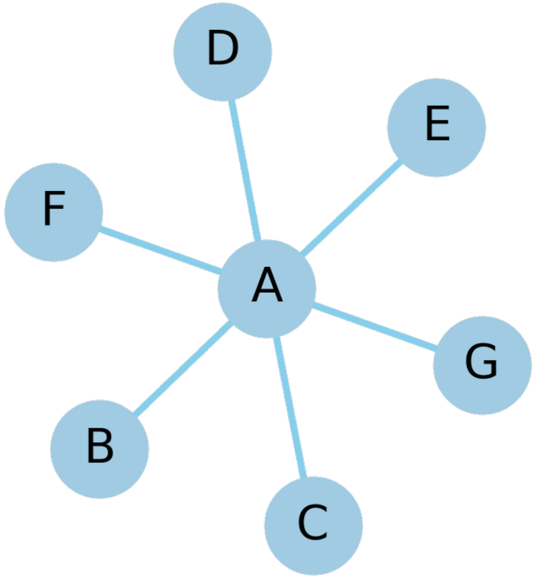
– For a given graph $G \coloneqq (V, E)$ with $|V|$ vertices, let $A = \left(a_{v,t}\right)$ be the adjacency matrix, i.e., $a_{v,t} = 1$ if vertex v is linked to vertex t, and $a_{v,t} = 0$ otherwise. The relative centrality score, x_v of vertex v can be defined as:

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in \mathrm{V}} a_{v,t} x_t$$

– Where $M(v)$ is the set of neighbors of v and λ is a constant. With a small rearrangement this can be rewritten in vector notation as the eigenvector equation.

Eigenvector Centrality

– Matrix Multiplication Meaning



$$\begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 6 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

Eigenvector Centrality

$$Ax = \lambda x$$

– In general, there will be many different eigenvalues λ for which a non-zero eigenvector solution exists. However, the additional requirement that all the entries in the eigenvector be non-negative implies (by the Perron-Frobenius theorem) that only the greatest eigenvalue results in the desired centrality measure.

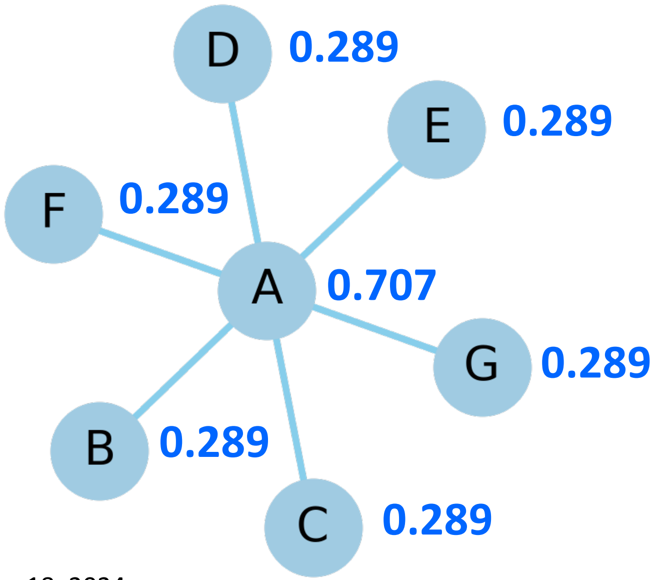
– The v^{th} component of the related eigenvector then gives the relative centrality score of the vertex v in the network. The eigenvector is only defined up to a common factor, so only the ratios of the centralities of the vertices are well defined.

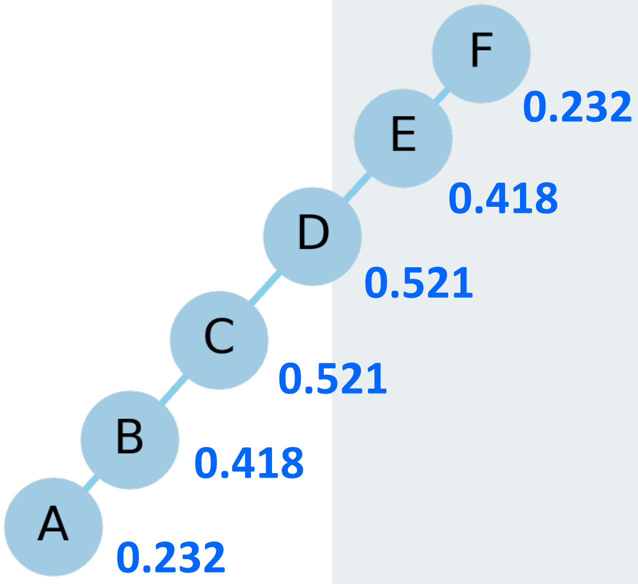
 https://en.wikipedia.org/wiki/Eigenvector_centrality

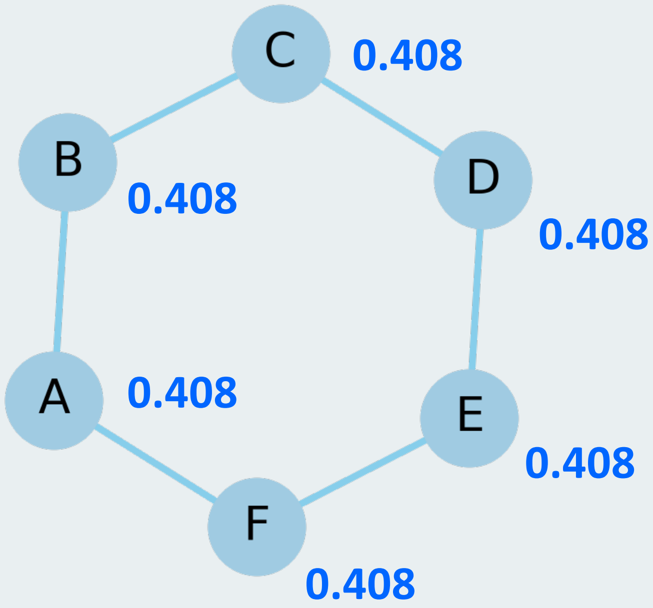
Eigenvector Centrality

– **Examples:**

In the code part, we will tell you how to get the eigenvector centrality from Eigen decomposition.

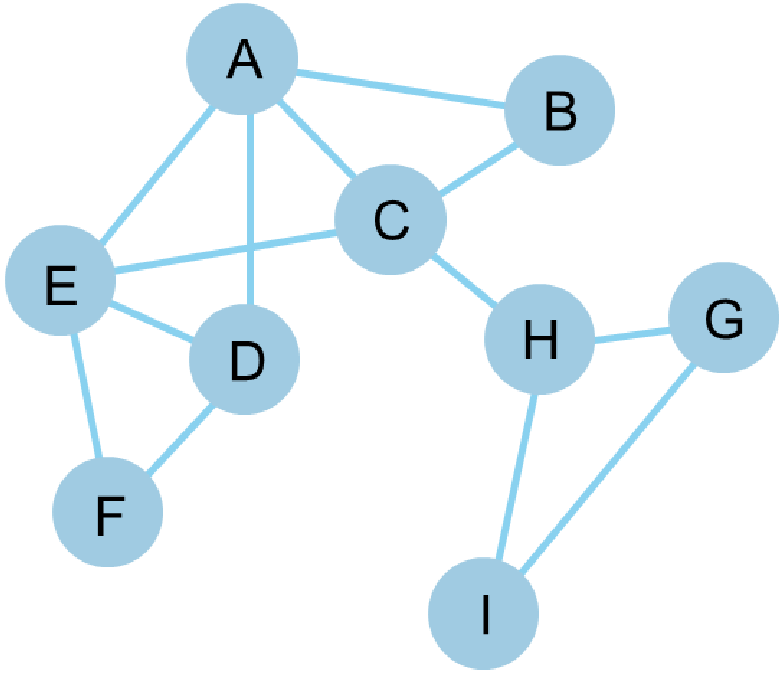






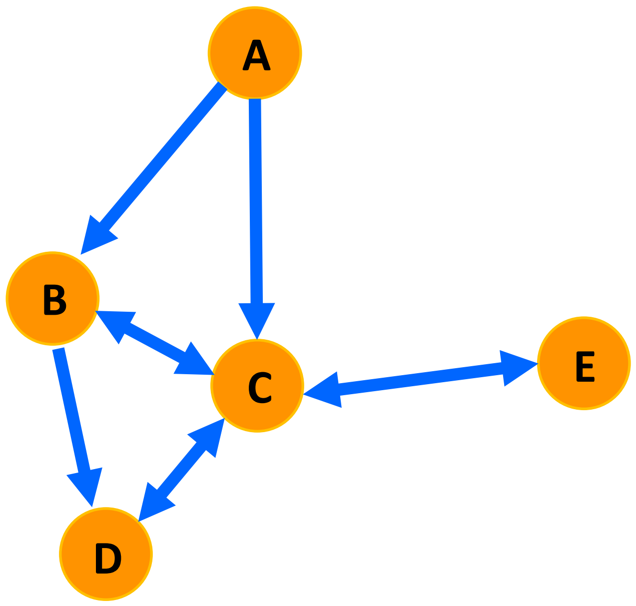
Eigenvector Centrality

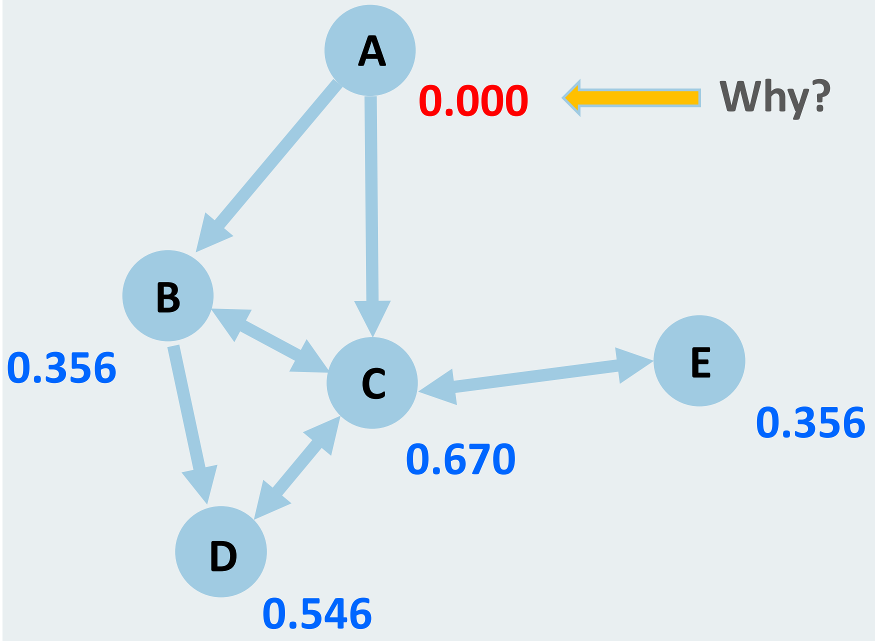
– Please calculate the eigenvector centrality of the following network.



Eigenvector Centrality

– If the given network is **directed**, then what do you expect the eigenvector centrality?





HITS Algorithm

- **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 initially forming; that is, specific web pages, known as hubs, served as large directories that were not 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.

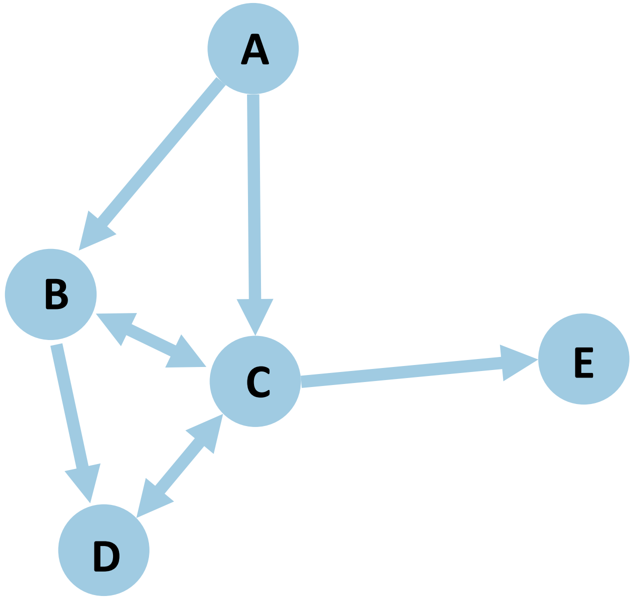
HITS Algorithm

- In other words, a good hub represents a page that points to many other pages, while a good authority represents a page that is linked by many different hubs.
- The algorithm performs a series of iterations with two basic steps:
  - **Authority update**: Update each node's *authority score* to equal the sum of the *hub scores* of each node that points to it. A node is given a high authority score by being linked from pages recognized as Hubs for information.
  - **Hub update**: Update each node's *hub score* to equal the sum of the authority scores of each node it points to. A node is given a high hub score by linking to nodes considered authorities on the subject.

HITS Algorithm

- The Hub score and Authority score for a node is calculated with the following algorithm:
  - Start with each node having a hub score and authority score of 1.
  - Run the authority update rule
  - Run the hub update rule
  - Normalize the values by dividing each Hub score by the square root of all Hub scores and dividing each Authority score by the square root of the sum of the squares of all Authority scores.
  - Repeat the second step as necessary.

HITS Algorithm

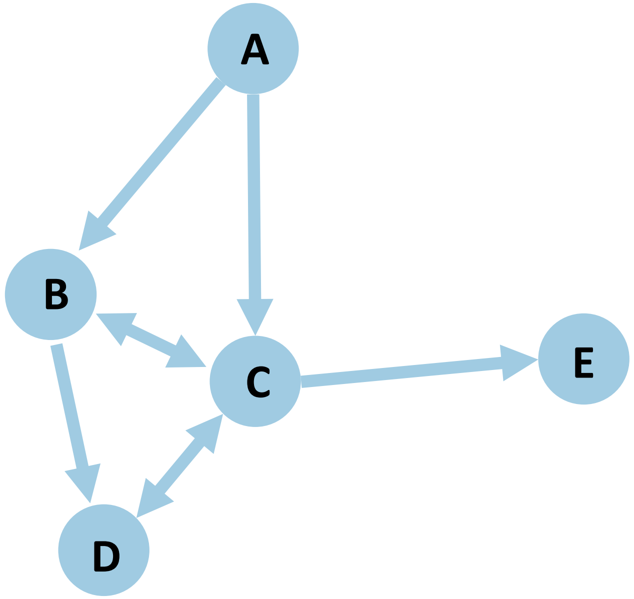


Normalization

	O_Auth	O_Hub	N_Auth	N_Hub	N_Auth	N_Hub
A	1	1	0	2	0	2/8
B	1	1	2	2	2/8	2/8
C	1	1	3	3	3/8	3/8
D	1	1	2	1	2/8	1/8
E	1	1	1	0	1/8	0

Sum = 8 8

HITS Algorithm

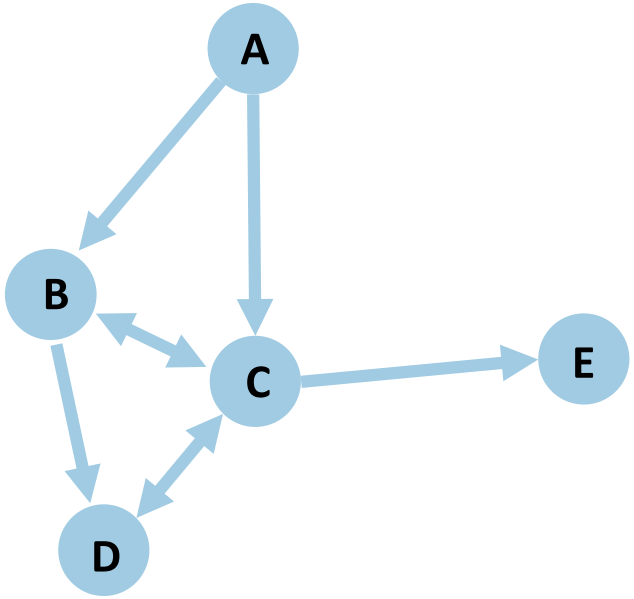


Normalization

	O_Auth	O_Hub	N_Auth	N_Hub	N_Auth	N_Hub
A	0	2/8	0	5/8	0	5/18
B	2/8	2/8	5/8	5/8	5/18	5/18
C	3/8	3/8	5/8	5/8	5/18	5/18
D	2/8	1/8	5/8	3/8	5/18	3/18
E	1/8	0	3/8	0	3/18	0

Sum = 18/8 18/8

HITS Algorithm



Practice …

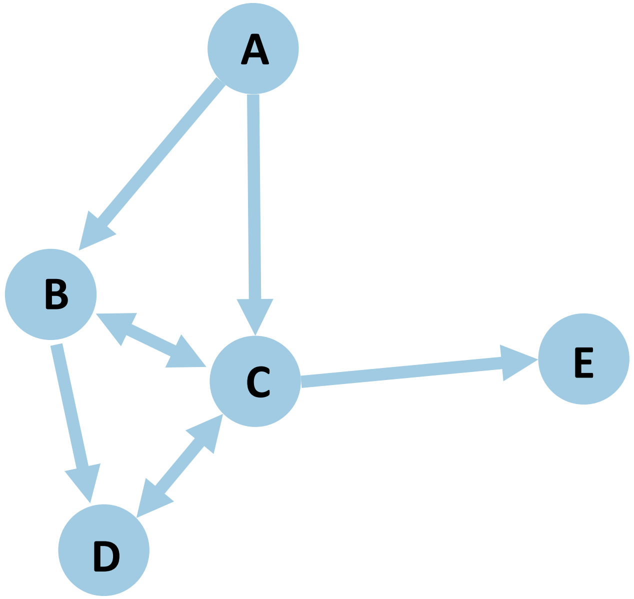
Normalization

	O_Auth	O_Hub	N_Auth	N_Hub
A	0	5/18	0	
B	5/18	5/18		
C	5/18	5/18		
D	5/18	3/18		
E	3/18	0		0

Sum = / /

N_Auth	N_Hub

HITS Algorithm



Practice …

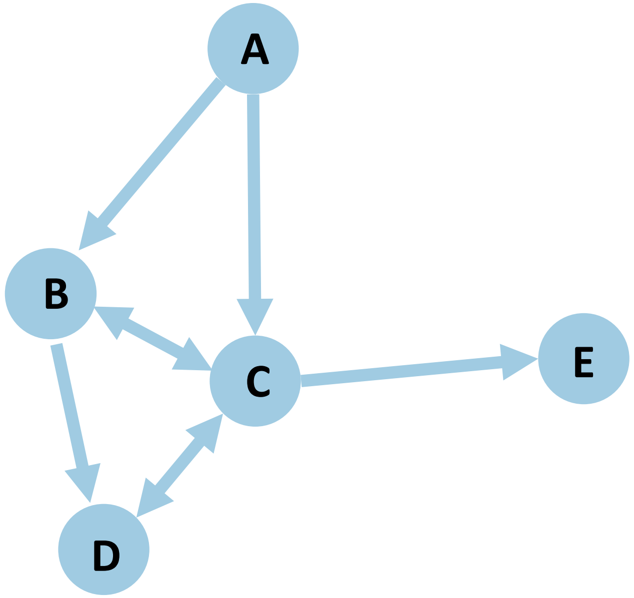
	O_Auth	O_Hub	N_Auth	N_Hub
A	0		0	
B				
C				
D				
E		0		0

Sum = / /

Normalization

N_Auth	N_Hub

HITS Algorithm

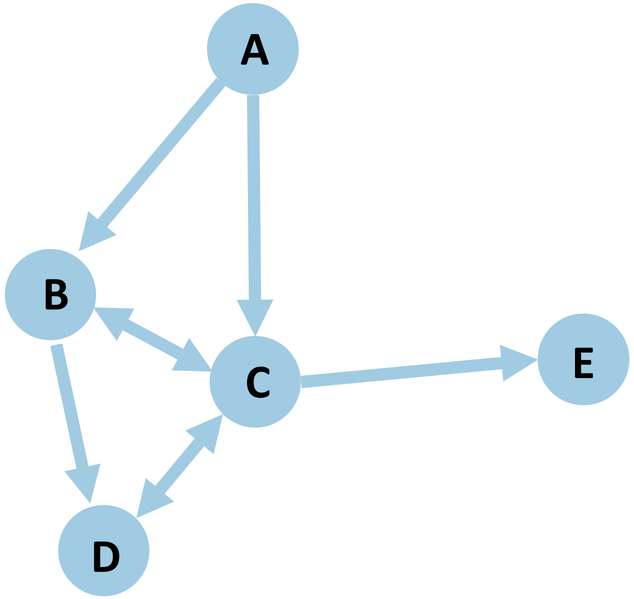


Calculating with NetworkX

	Auth	Hub
A	0.000	0.270
B	0.270	0.270
C	0.315	0.315
D	0.270	0.145
E	0.145	0.000

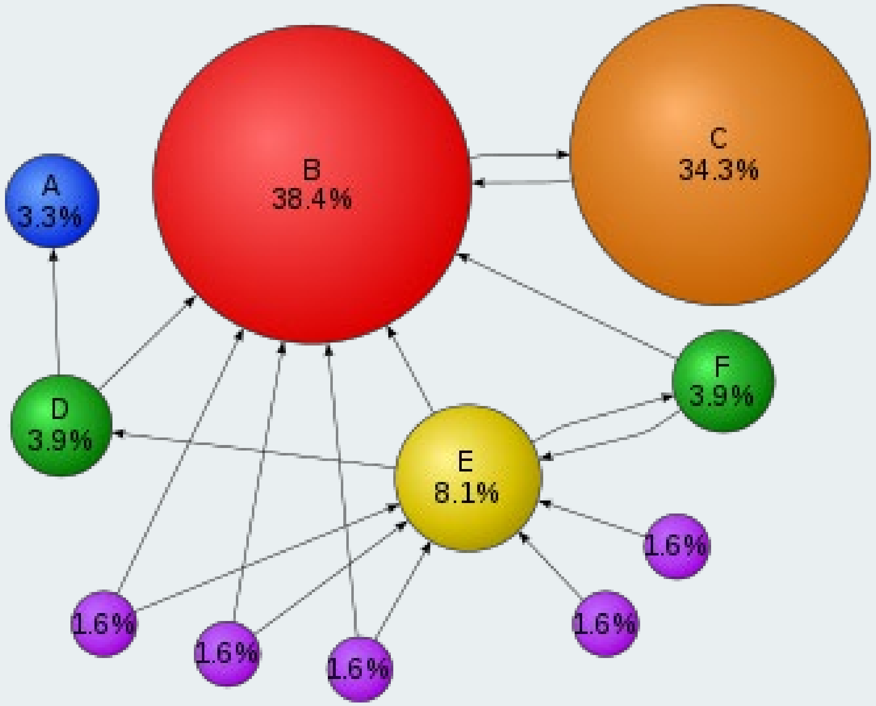
HITS Algorithm

– Design the HITS algorithm code without using networkx and compute the following network.



PageRank Algorithm

- PageRank works by counting the number and quality of links to a page to determine a rough estimate of how important the website is.
- The underlying assumption is that more important websites will likely receive more links from other websites.



https://en.wikipedia.org/wiki/PageRank#:~:text=PageRank%20(PR)%20is%20an%20algorithm,the%20importance%20of%20website%20pages.

PageRank Algorithm

– Assume a small universe of four web pages: **A**, **B**, **C**, and **D**. Links are ignored from a page to itself. Multiple outbound links are treated as a single link from one page to another. PageRank is initialized to the same value for all pages.

– In the original form of PageRank, the sum of PageRank over all pages was the total number of pages on the web at that time, so each page in this example would have an initial value of 1. However, later versions of PageRank and the remainder of this section assume a probability distribution between 0 and 1. Hence, the initial value for each page in this example is 0.25.

https://en.wikipedia.org/wiki/PageRank#:~:text=PageRank%20(PR)%20is%20an%20algorithm,the%20importance%20of%20website%20pages.

PageRank Algorithm

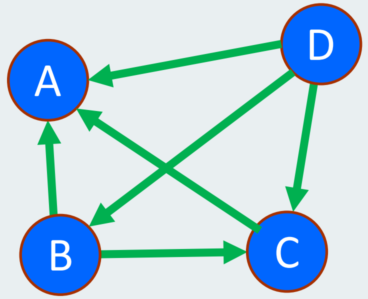
– The PageRank transferred from a given page to the targets of its outbound links upon the next iteration is divided equally among all outbound links.

– If the only links in the system were from pages **B**, **C**, and **D** to **A**, each link would transfer 0.25 PageRank to **A** upon the next iteration, for a total of 0.75.

$$PR(A) = PR(B) + PR(C) + PR(D)$$

https://en.wikipedia.org/wiki/PageRank#:~:text=PageRank%20(PR)%20is%20an%20algorithm,the%20importance%20of%20website%20pages.

PageRank Algorithm



– Suppose instead that page **B** had a link to pages **C** and **A**, page **C** had a link to page **A**, and page **D** had links to all three pages. Thus, upon the first iteration, page **B** would transfer half of its existing value (0.125) to page **A** and the other half (0.125) to page **C**. Page **C** would transfer all of its existing value (0.25) to the only page it links to, **A**. Since **D** had three outbound links, it would transfer one third of its existing value, or approximately 0.083, to **A**. At the completion of this iteration, page **A** will have a PageRank of approximately 0.458.

$$PR(A) = \frac{PR(B)}{2} + \frac{PR(C)}{1} + \frac{PR(D)}{3}$$

https://en.wikipedia.org/wiki/PageRank#:~:text=PageRank%20(PR)%20is%20an%20algorithm,the%20importance%20of%20website%20pages.

PageRank Algorithm

- In other words, the PageRank conferred by an outbound link is equal to the document's own PageRank score divided by the number of outbound links $\boldsymbol{L}$.

$$PR(A) = \frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)}$$

- In the general case, the PageRank value for any page $\boldsymbol{u}$ can be expressed as:

$$PR(u) = \sum_{v \in \mathrm{B}_u} \frac{PR(v)}{L(v)}$$

- i.e. the PageRank value for a page $\boldsymbol{u}$ is dependent on the PageRank values for each page $\boldsymbol{v}$ contained in the set $\boldsymbol{B_u}$ (the set containing all pages linking to page $\boldsymbol{u}$), divided by the number $L(v)$ of links from page $\boldsymbol{v}$.

https://en.wikipedia.org/wiki/PageRank#:~:text=PageRank%20(PR)%20is%20an%20algorithm,the%20importance%20of%20website%20pages.

PageRank Algorithm

– The PageRank theory holds that an imaginary surfer who is randomly clicking on links will eventually stop clicking. The probability, at any step, that the person will continue following links is a damping factor d. The probability that they instead jump to any random page is $1 - d$.

– Various studies have tested different damping factors, but it is generally assumed that the damping factor will be set around 0.85.

https://en.wikipedia.org/wiki/PageRank#:~:text=PageRank%20(PR)%20is%20an%20algorithm,the%20importance%20of%20website%20pages.

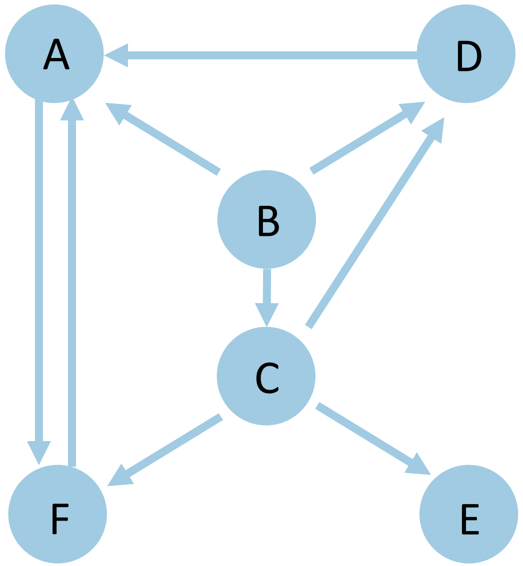
PageRank Algorithm

– The damping factor is subtracted from 1 (and in some algorithm variations, the result is divided by the number of documents (N) in the collection. This term is then added to the product of the damping factor and the sum of the incoming PageRank scores. That is,

$$PR(A) = \frac{1-d}{N} + d\left(\frac{PR(B)}{L(B)} + \frac{PR(C)}{L(C)} + \frac{PR(D)}{L(D)} + \cdots\right)$$

https://en.wikipedia.org/wiki/PageRank#:~:text=PageRank%20(PR)%20is%20an%20algorithm,the%20importance%20of%20website%20pages.

PageRank Algorithm



Adjacency matrix

$$G = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}$$

$\boldsymbol{C_j}$: ***column sum***

$$a_{ij} = \begin{cases} \frac{g_{ij}}{c_j}, & if\ C_j \neq 0 \\ 0, & if\ C_j = 0 \end{cases}$$

 https://www.youtube.com/watch?v=dZ1hvXQofyo

PageRank Algorithm

Type 1 Probability Matrix	Type 2 Probability Matrix
Convert from adjacency matrix	Random walk

$$A_1 = \begin{bmatrix} 0 & 0 & 0 & 1/1 & 0 & 1/1 \\ 1/2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 0 \\ 0 & 1/2 & 1/3 & 0 & 0 & 0 \\ 0 & 0 & 1/3 & 0 & 0 & 0 \\ 1/2 & 0 & 1/3 & 0 & 0 & 0 \end{bmatrix}$$

$$A_2 = \begin{bmatrix} 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \\ 1/6 & 1/6 & 1/6 & 1/6 & 1/6 & 1/6 \end{bmatrix}$$

Basic concept: $A = pA_1 + (1-p)A_2$

Actual formula: $a_{ij} = \begin{cases} \frac{pg_{ij}}{C_j} + \frac{(1-p)}{n} & if\ C_j \neq 0 \\ \frac{1}{n} & if\ C_j = 0 \end{cases}$

Transition Probability Matrix

- A is known as the Markov matrix
  - All elements range from 0 to 1, and the sum of each column is equal to 1.
- If z is the initial probability of each node, Az is the probability after 1 transition, A^2 is the probability after transitions, … and so on.

Transition Probability Matrix

- $A^k z$ converges to the node rank if k is large enough.
- $A^{k+1} z = A^k z$ when k is large ➔ $Ax = x$ with $x = A^k z$ when k is large ➔ x is the Google PageRank.
- Most elements of A are equal to $\frac{1-p}{n}$.
  - If $n = 4 \times 10^9$ and $p = 0.85$, then $\frac{1-p}{n} = 3.75 \times 10^{-11}$.

Compute the PageRank

- **Eigenvector method**
  - $Ax = x$ ➔ x is the eigenvector corresponding to eigenvalue 1.
  - **Fact 1**: A always has 1 as its eigenvalue of maximum magnitude
- **Power method**
  - Repeat $x = Ax$ until x converges
  - The only possible approach for a large n
  - **Fact 2**: $A^k z$ is not affected by z as k increases

Centrality and Behavior

– Centrality usually represents the “leadership” of the nodes (actors); however, the leadership of all nodes could be divided into different levels, for instance, global and local leaders.
– Sometimes, we want to choose a leader who can connect to other people from various sources.
– As a result, we will adopt different centrality measurements in various scenarios, depending on the targeted characteristics.

Paper Reading

The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles

R. Guimerà*, S. Mossa†, A. Turtschi‡, and L. A. N. Amaral*§

*Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208; †European Synchrotron Radiation Facility, B.P. 220, F-38043 Grenoble, Cedex, France; and ‡Avanade Deutschland Gmbh, Campus Kronberg 7, D-61476, Kronberg, Germany

Edited by Kenneth W. Wachter, University of California, Berkeley, CA, and approved April 5, 2005 (received for review October 27, 2004)

We analyze the global structure of the worldwide air transportation network, a critical infrastructure with an enormous impact on local, national, and international economies. We find that the worldwide air transportation network is a scale-free small-world network. In contrast to the prediction of scale-free network models, however, we find that the most connected cities are not necessarily the most central, resulting in anomalous values of the centrality. We demonstrate that these anomalies arise because of the multicommunity structure of the network. We identify the communities in the air transportation network and show that the community structure cannot be explained solely based on geographical constraints and that geopolitical considerations have to be taken into account. We identify each city's global role based on its pattern of intercommunity and intracommunity connections, which enables us to obtain scale-specific representations of the network.

complex networks | betweenness centrality | critical infrastructures

Like other critical infrastructures, the air transportation network has enormous impact on local, national, and international economies. It is thus natural that airports and national airline companies are often times associated with the image a country or region wants to project (1–4).

The air transportation system is also responsible, indirectly, for the propagation of diseases such as influenza and, recently, severe acute respiratory syndrome (SARS). The air transportation network thus plays for certain diseases a role that is analogous to that of the web of human sexual contacts (5) for the propagation of AIDS and other sexually transmitted infections (6, 7).

The worldwide air transportation network is responsible for the

Much research has been conducted on the definition of models and algorithms that enable one to solve problems of optimal network design (9, 10). However, a worldwide, "system" level analysis of the structure of the air transportation network is still lacking. However, just as one cannot fully understand the complex dynamics of ecosystems by looking at simple food chains (11) or the complex behavior in cells by studying isolated biochemical pathways (12, 13), one cannot fully understand the dynamics of the air transportation system without a "holistic" perspective. Modern "network analysis" (14–18) provides an ideal framework within which to pursue such a study.

We analyze here the worldwide air transportation network. We build a network of 3,883 locales, villages, towns, and cities with at least one airport and establish links between pairs of locales that are connected by nonstop passenger flights. We find that the worldwide air transportation network is a small-world network (19) for which (*i*) the number of nonstop connections from a given city and (*ii*) the number of shortest paths going through a given city have distributions that are scale-free. In contrast to the prediction of scale-free network models, we find that the most-connected cities are not necessarily the most "central," that is, the cities through which most shortest paths go. We show that this surprising result can be explained by the existence of several distinct "communities" within the air transportation network. We identify these communities by using algorithms recently developed for the study of complex networks and show that the structure of the communities cannot be explained solely based on geographical constraints and that geopolitical considerations also must be taken into account. The existence of communities leads us to the definition of each city's global role, based on its pattern of intercommunity and intracommunity connections.

Guimera, R., Mossa, S., Turtschi, A., & Amaral, L. N. (2005). The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. Proceedings of the National Academy of Sciences, 102(22), 7794-7799.

Questions:

1. What is the objective of this paper?
2. How did the authors formulate the transportation network and quantify the transportation characteristics with nodal indicators?
3. What are the findings of this study?
4. If you want to achieve the same objective, how do you formulate the network?

References

- Wasserman, S., & Faust, K. (1994). Social network analysis: Methods and applications.
- Tsvetovat, M., & Kouznetsov, A. (2011). Social Network Analysis for Startups: Finding connections on the social web. "O'Reilly Media, Inc.".
- Hanneman, R. A., & Riddle, M. (2005). Introduction to social network methods.
- https://en.wikipedia.org/wiki/Social_network

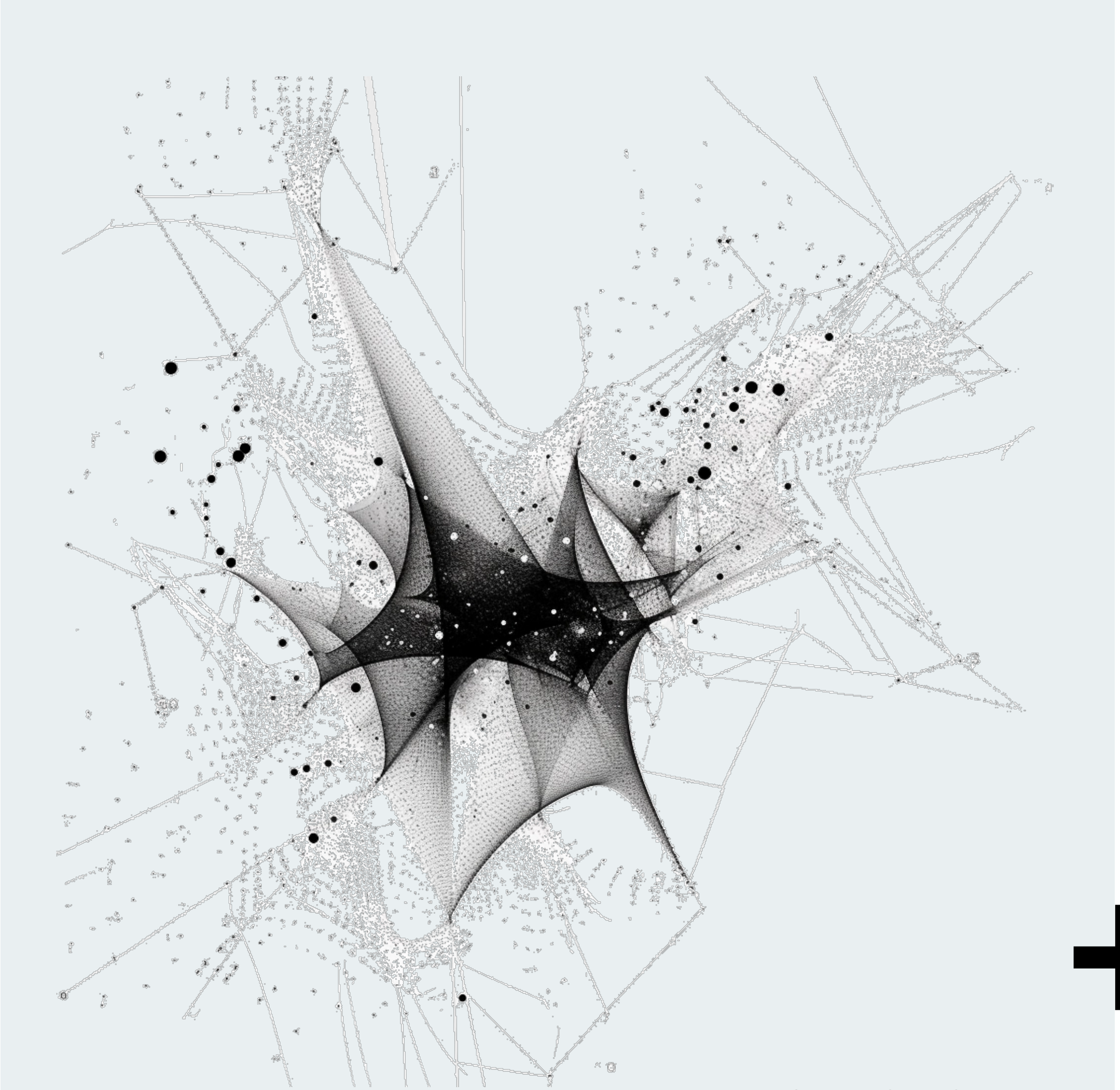


Photo credit: midjourney

The End

Thank you for your attention!

Email: chchan@ntnu.edu.tw
Website: toodou.github.io

Photo credit: midjourney