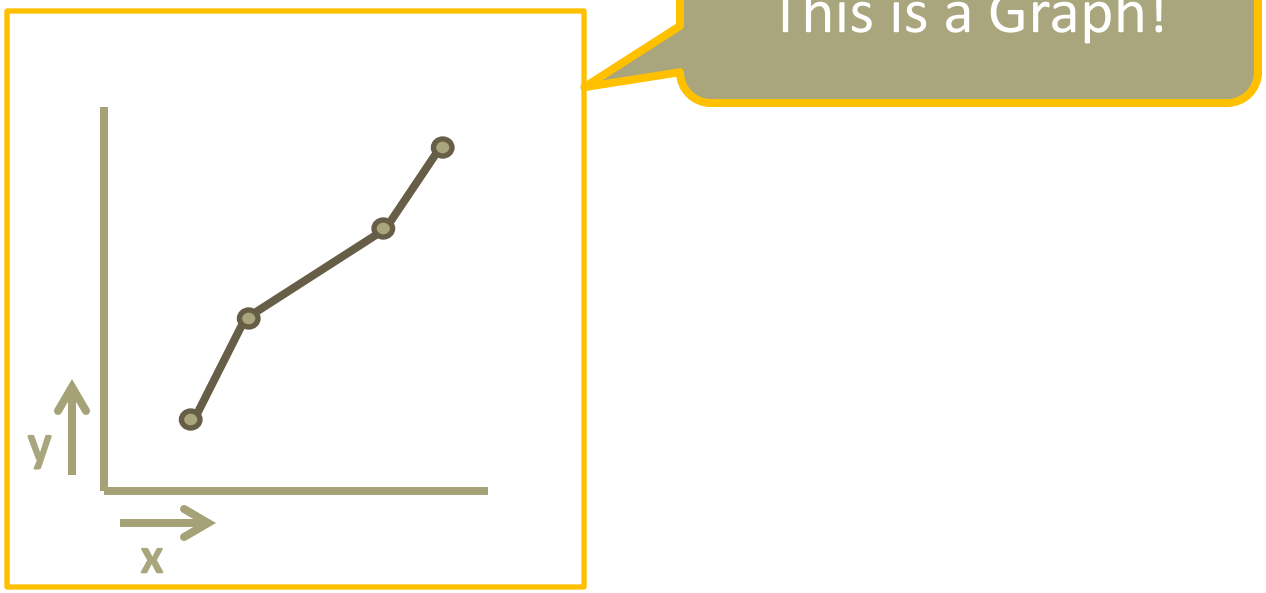
Introduction to Graph Data

Ernst Henle
ErnstHe@UW.edu
Skype: ernst.predixion

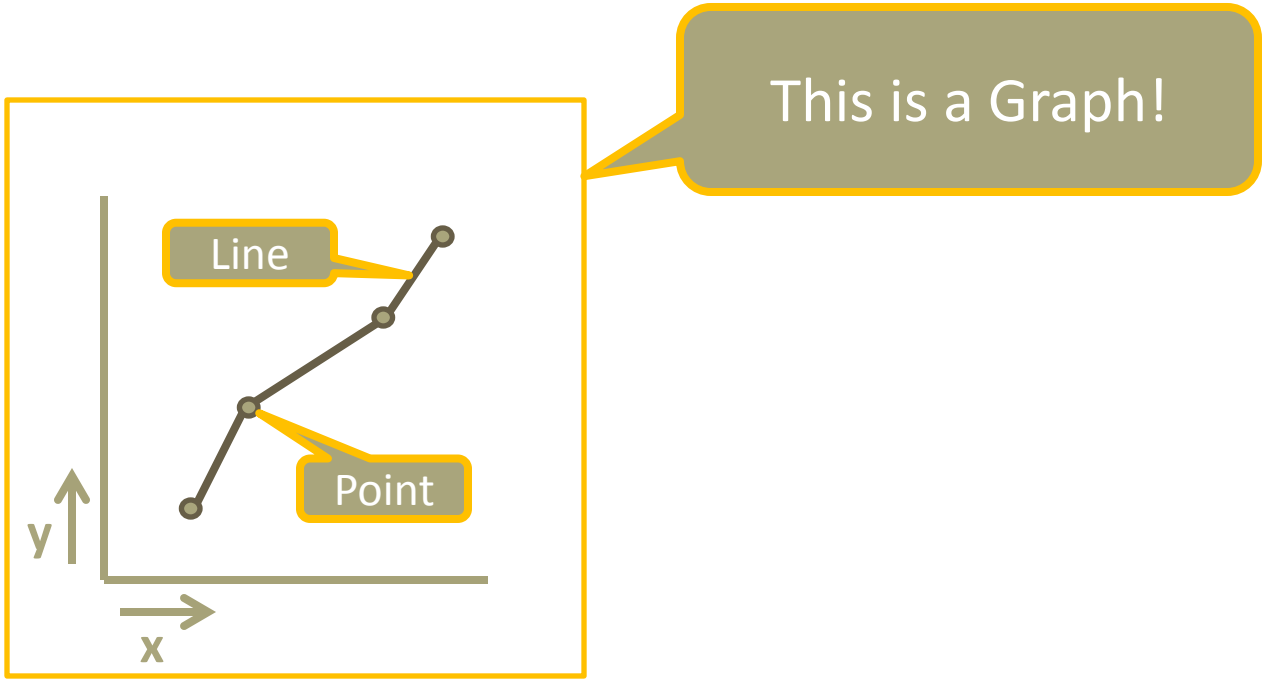
Graph Data: What is it? (0)

What is a Graph?

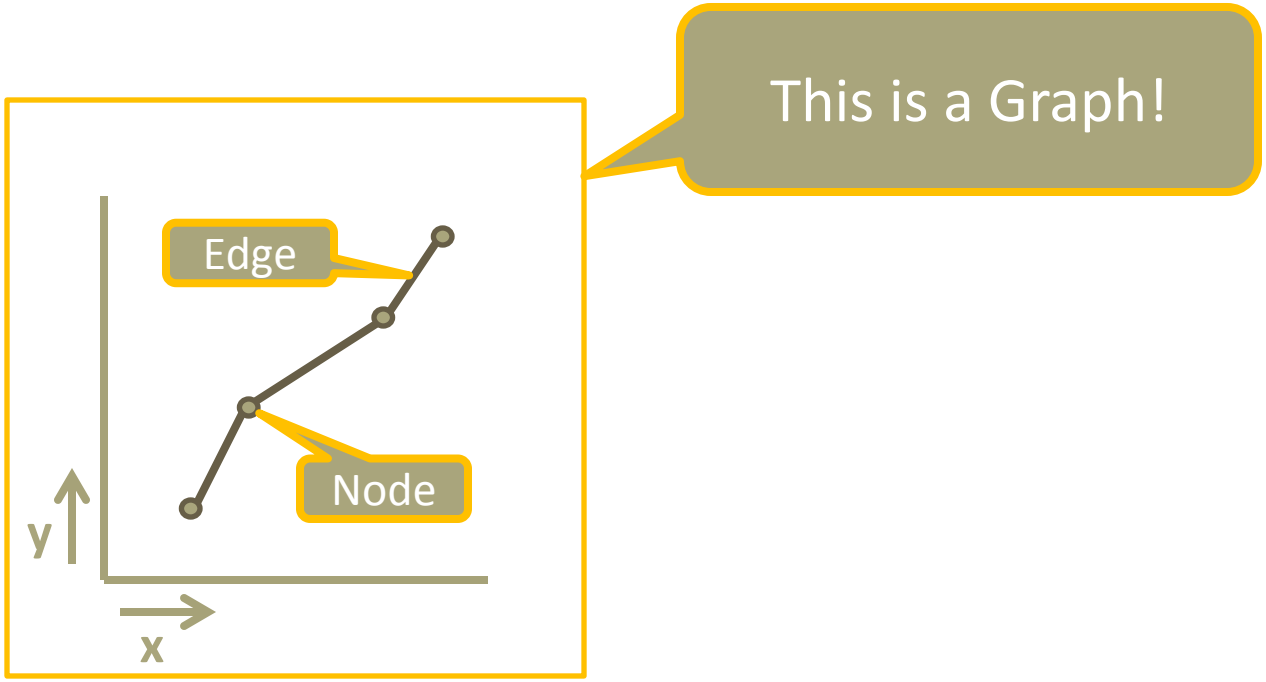
Graph Data: What is it? (1)



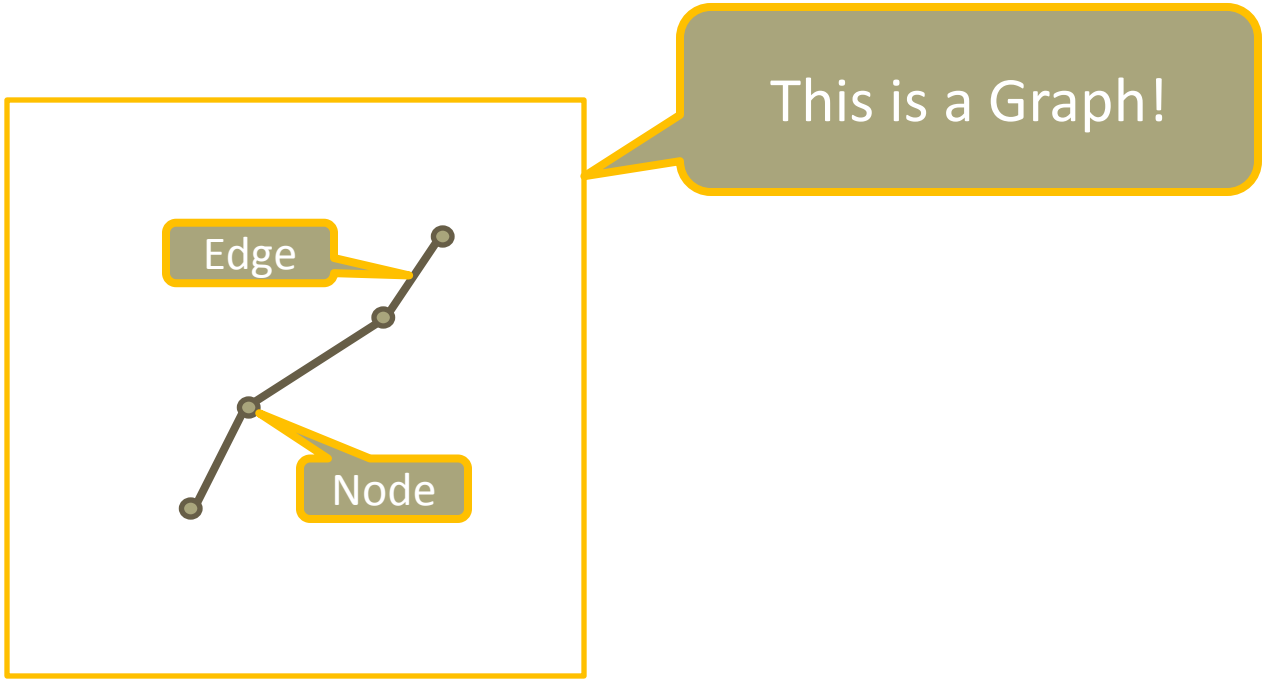
Graph Data: What is it? (2)



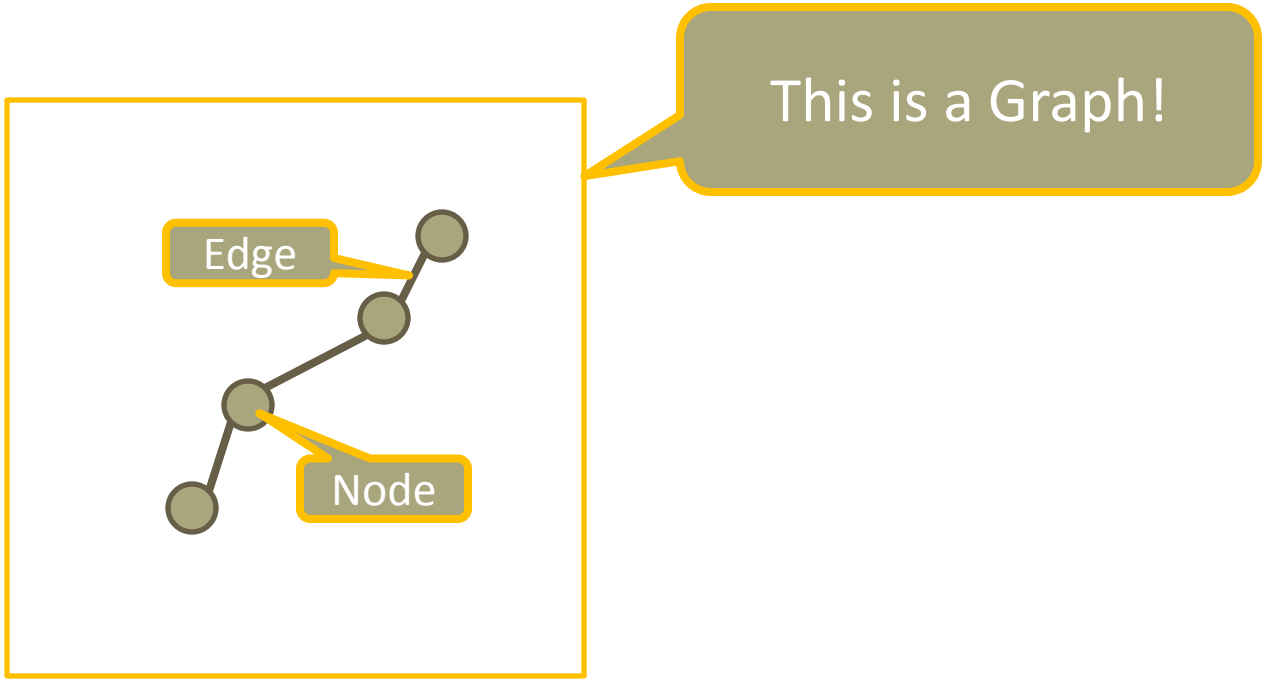
Graph Data: What is it? (3)



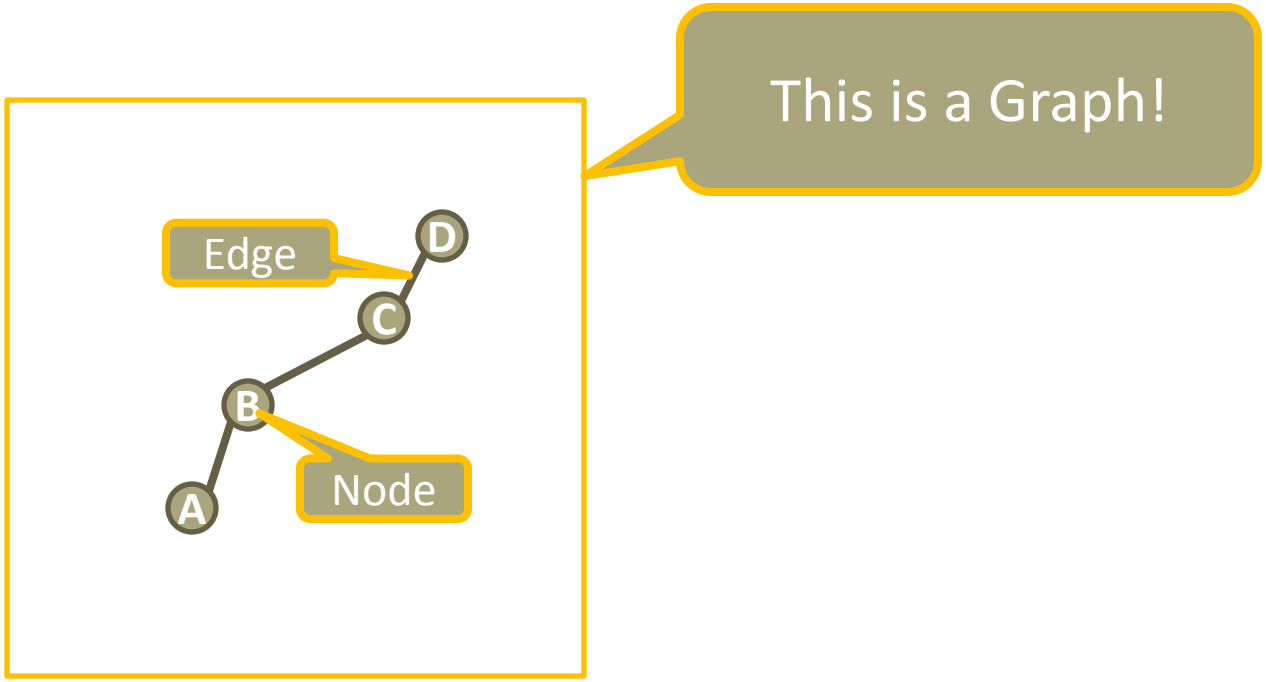
Graph Data: What is it? (4)



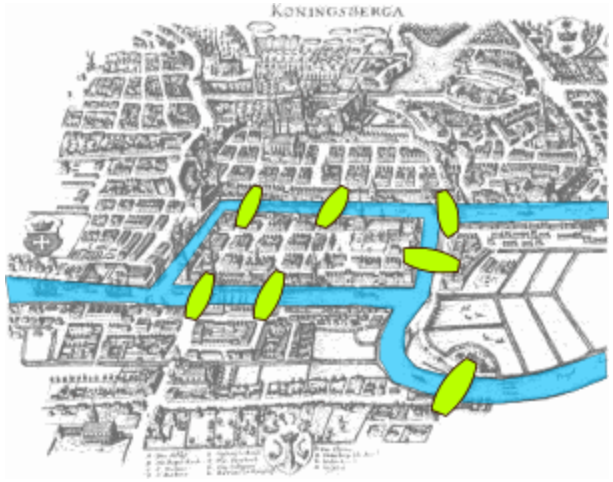
Graph Data: What is it? (5)



Graph Data: What is it? (6)



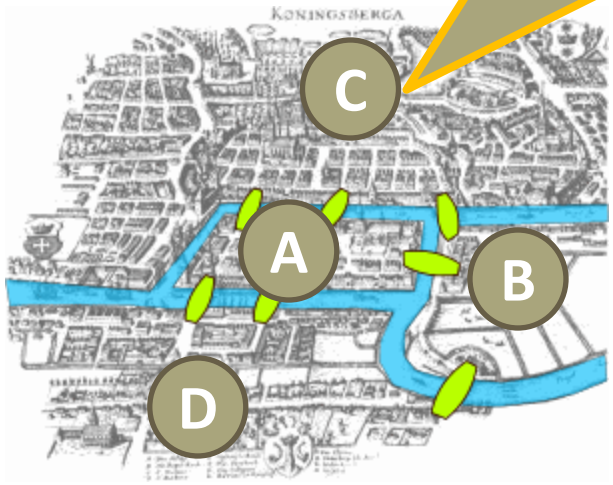
Graph Data: Euler's Seven Bridges of Königsberg (0)



"find a walk through the city that would cross each bridge once and only once"
http://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

Graph Data: Euler's Seven Bridges of Königsberg (1)

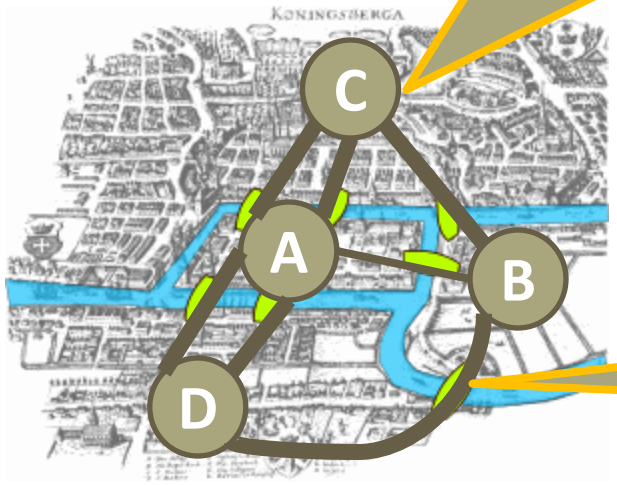
Each separated piece of land is a node



"find a walk through the city that would cross each bridge once and only once"
http://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

Graph Data: Euler's Seven Bridges of Königsberg (2)

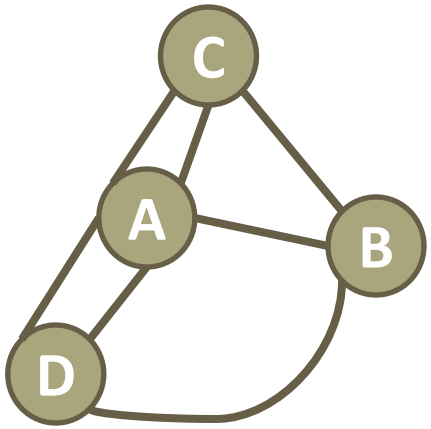
Each separated piece of land is a node



Each bridge is an edge

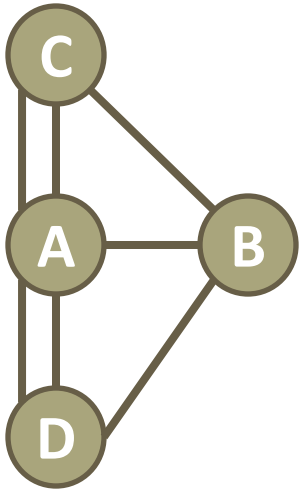
"find a walk through the city that would cross each bridge once and only once"
http://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

Graph Data: Euler's Seven Bridges of Königsberg (3)



"find a walk through the city that would cross each bridge once and only once"
http://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

Graph Data: Euler's Seven Bridges of Königsberg (4)

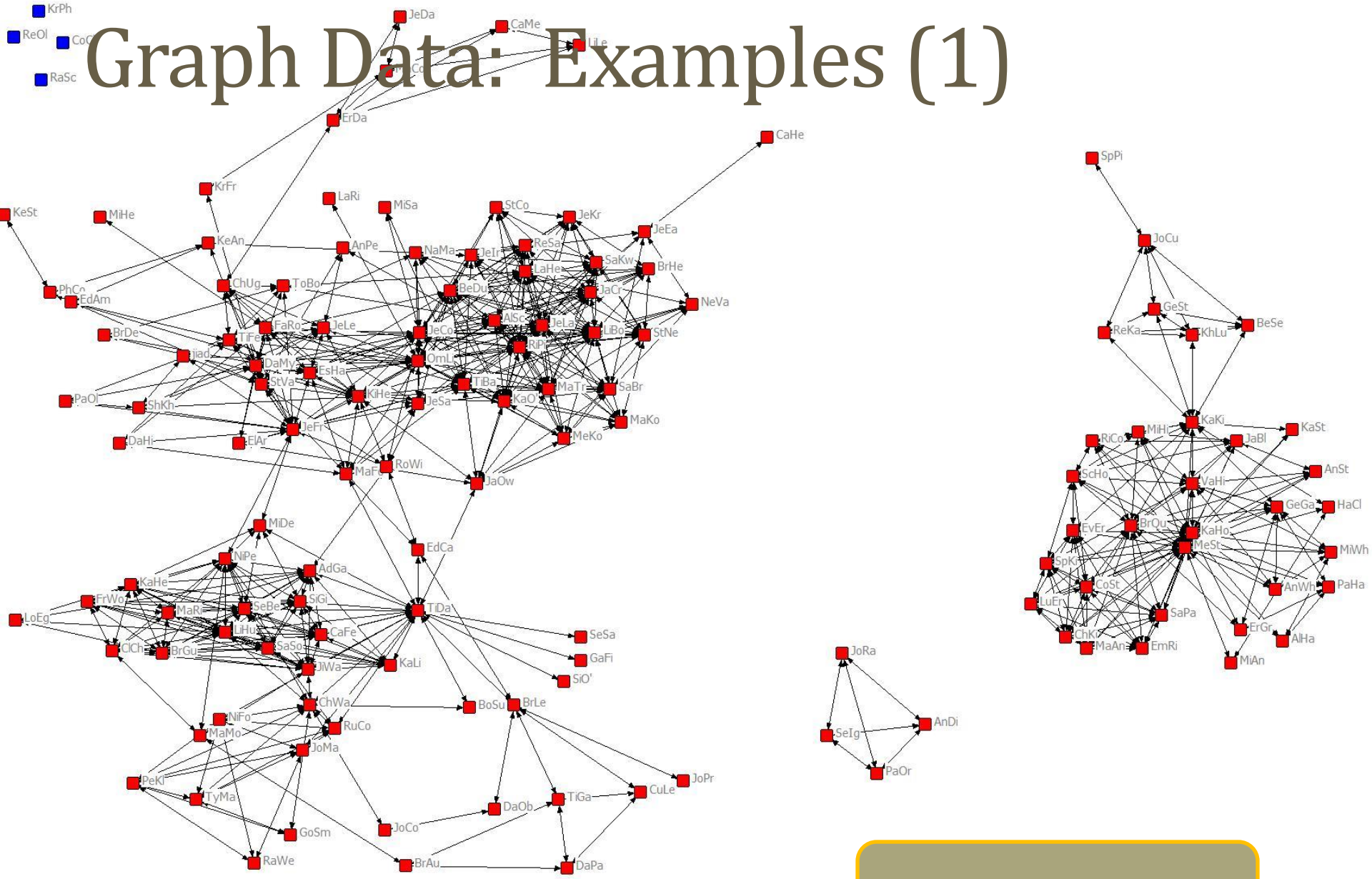


"find a walk through the city that would cross each bridge once and only once"
http://en.wikipedia.org/wiki/Seven_Bridges_of_K%C3%B6nigsberg

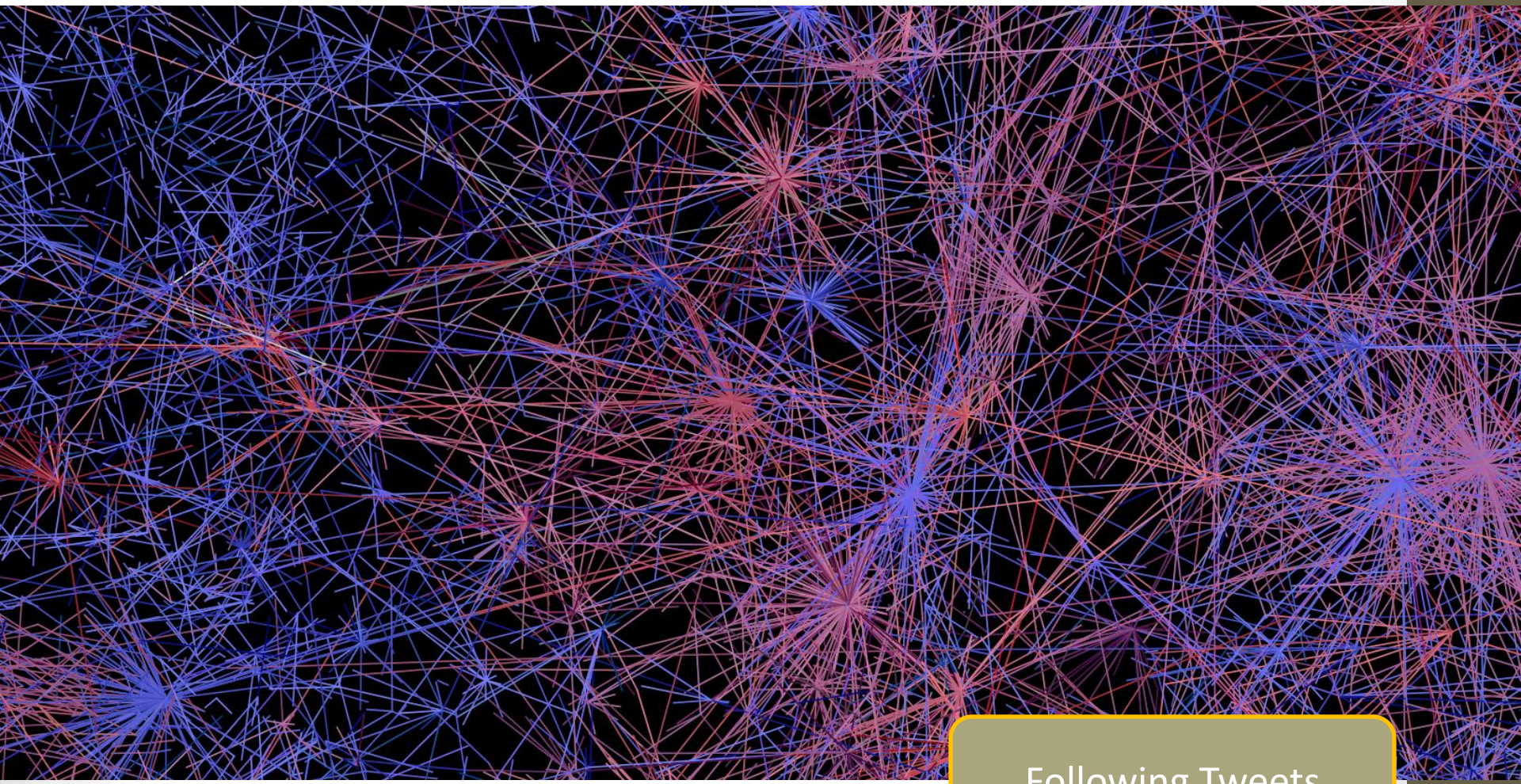
Graph Data: Examples (0)

- Graphs represent many things like Associations and Networks.
- Graphs represent the world wide web
- Graphs represent Facebook networks
- Graphs represent metabolic pathways

Graph Data: Examples (1)

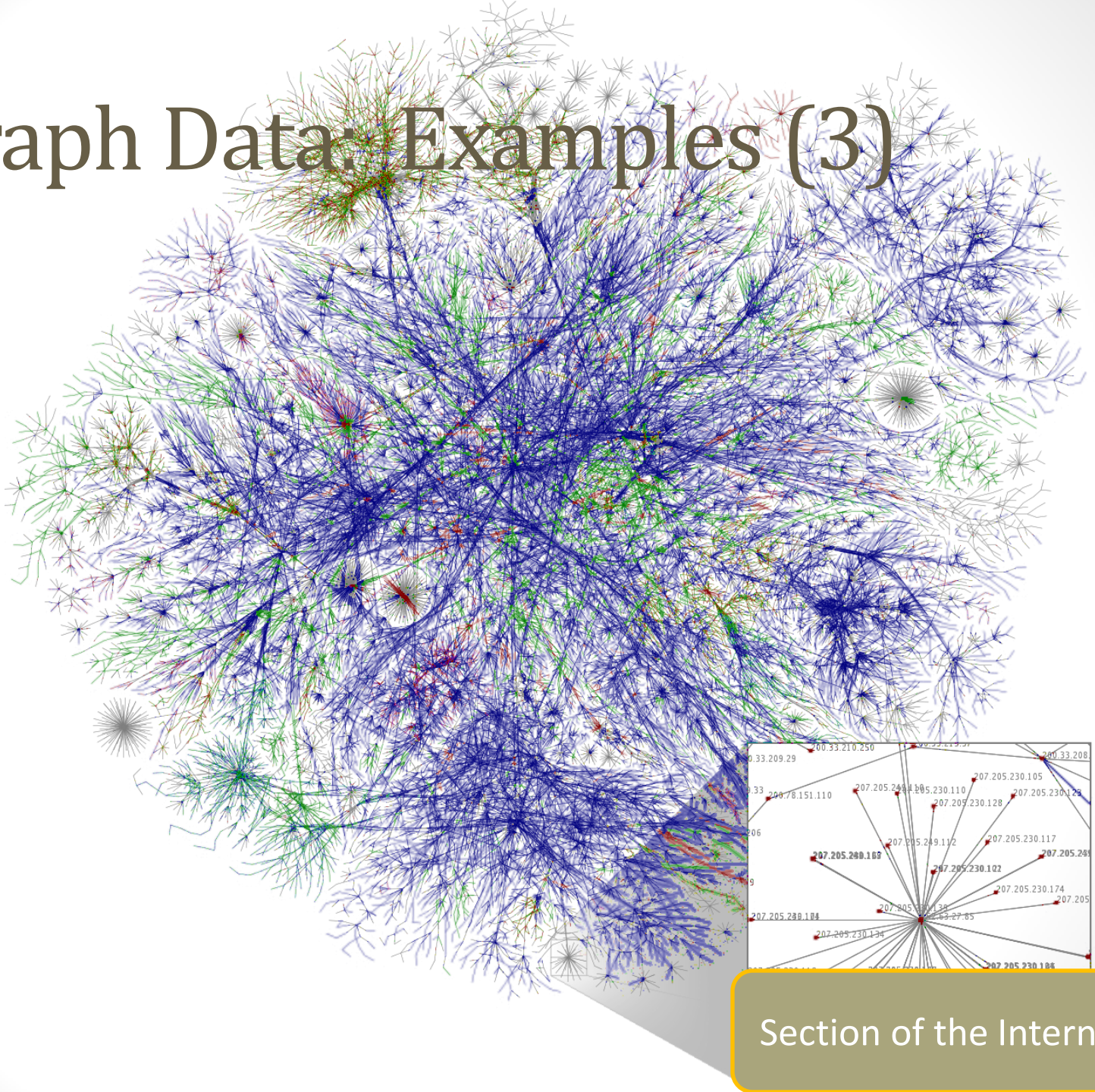


Graph Data: Examples (2)

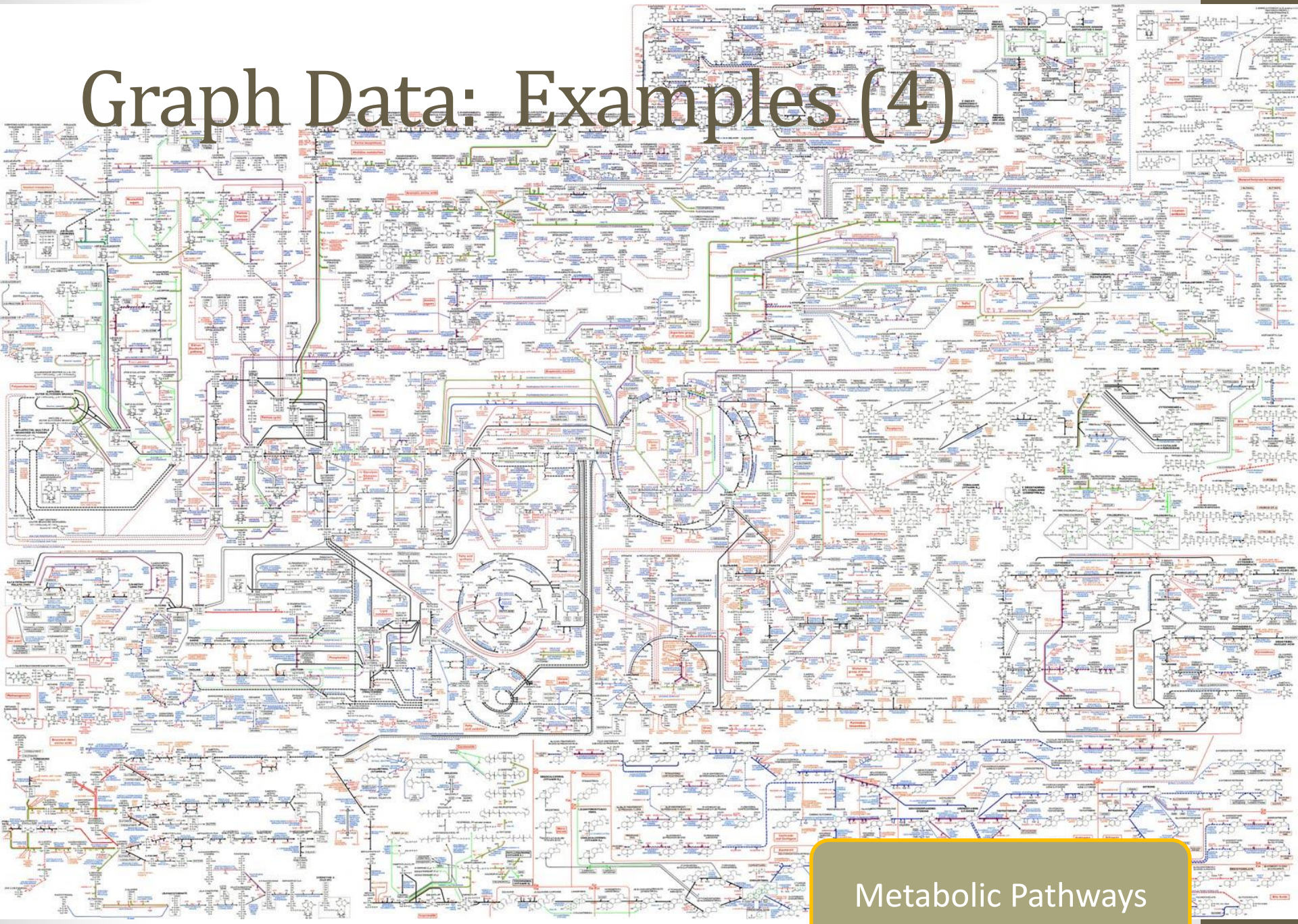


Following Tweets

Graph Data: Examples (3)

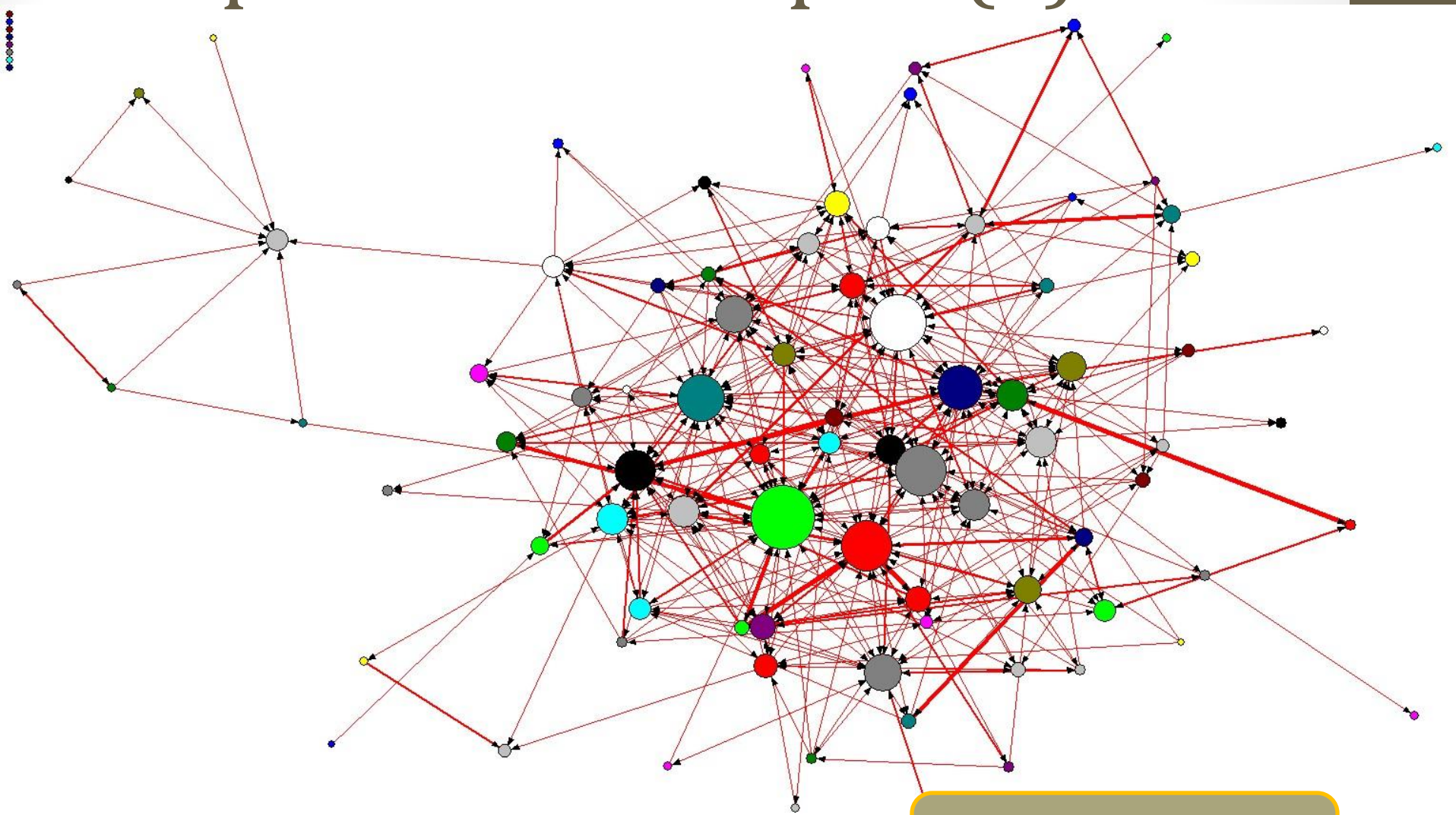


Graph Data: Examples (4)



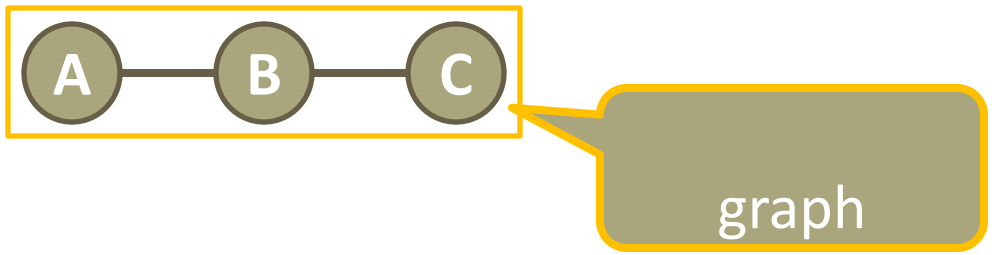
Metabolic Pathways

Graph Data: Examples (5)

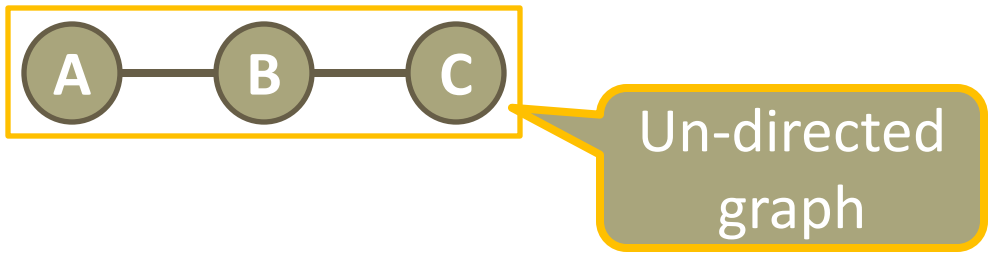


Ranked Web Pages

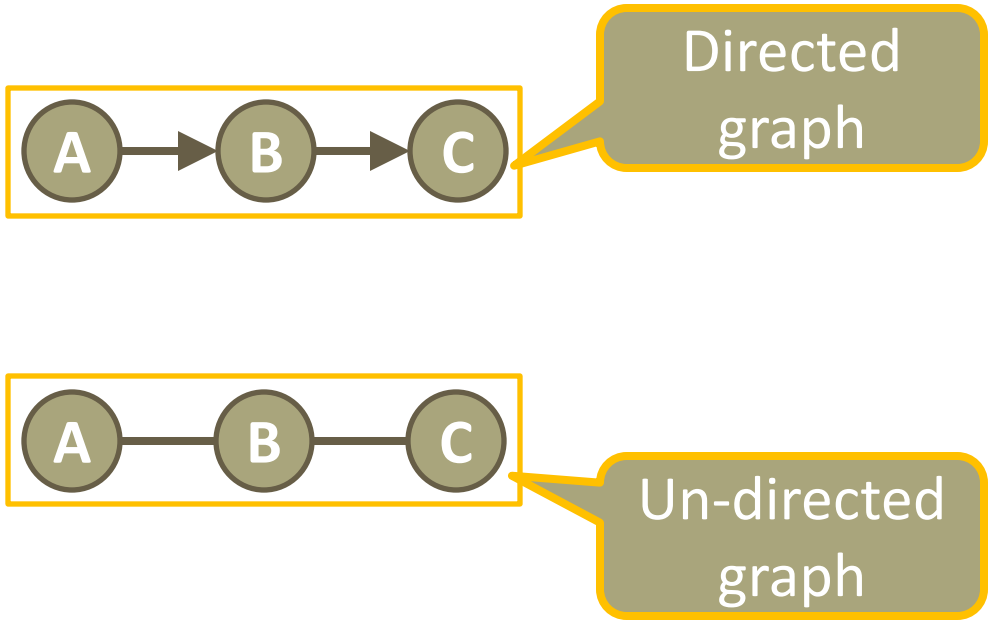
Graph Data: Directed and Un-directed Graphs (0)



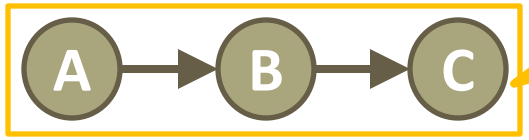
Graph Data: Directed and Un-directed Graphs (1)



Graph Data: Directed and Un-directed Graphs (2)

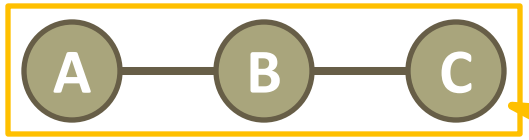


Graph Data: Directed and Un-directed Graphs (3)



Directed graph

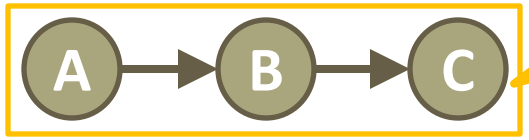
"The link goes from A to B"



Un-directed graph

"The link is between A and B"

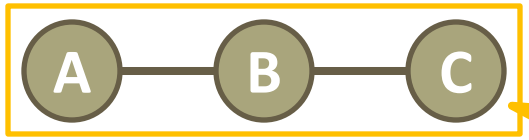
Graph Data: Directed and Un-directed Graphs (4)



Directed graph

"The link goes from A to B"

Example of directed graphs: World Wide Web

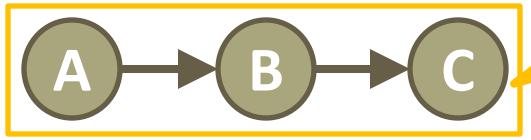


Un-directed graph

"The link is between A and B"

Examples of un-directed graphs: Facebook, LinkedIn

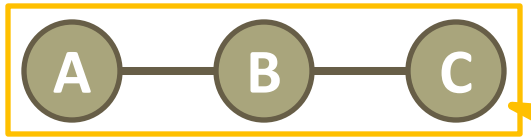
Graph Data: Directed and Un-directed Graphs (5)



Directed graph

"The link goes from A to B"

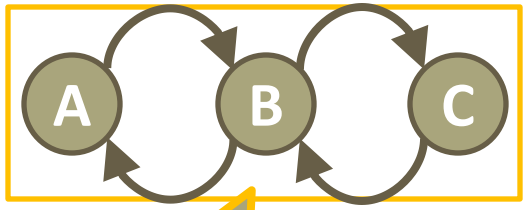
Example of directed graphs: World Wide Web



Un-directed graph

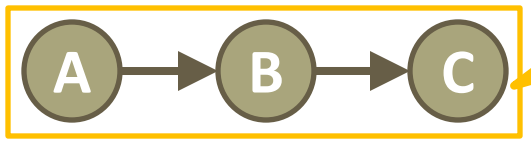
"The link is between A and B"

Examples of un-directed graphs: Facebook, LinkedIn



graph with bi-directional links

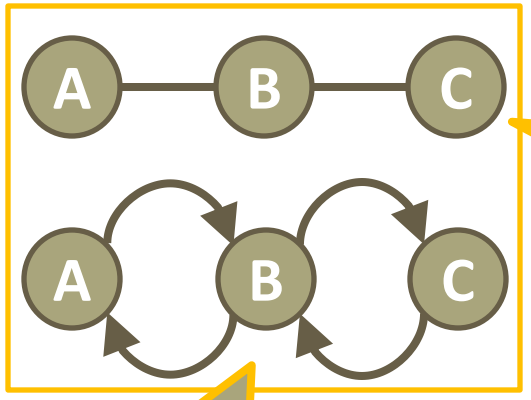
Graph Data: Directed and Un-directed Graphs (6)



Directed graph

"The link goes from A to B"

Example of directed graphs: World Wide Web



Un-directed graph

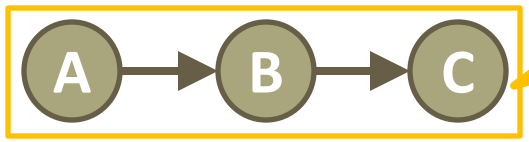
"The link is between A and B"

Examples of un-directed graphs: Facebook, LinkedIn

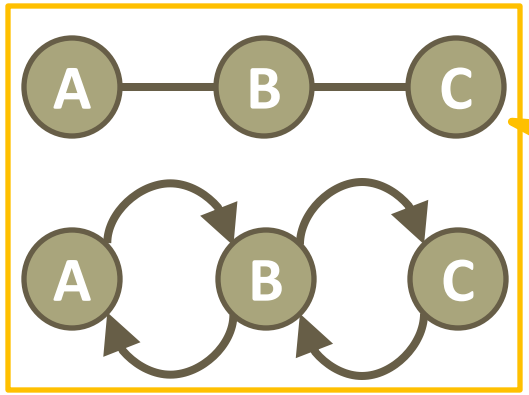
graph with bi-directional links

A graph with bi-directional links is like an undirected graph

Graph Data: Formalize as Rectangular Data (0)



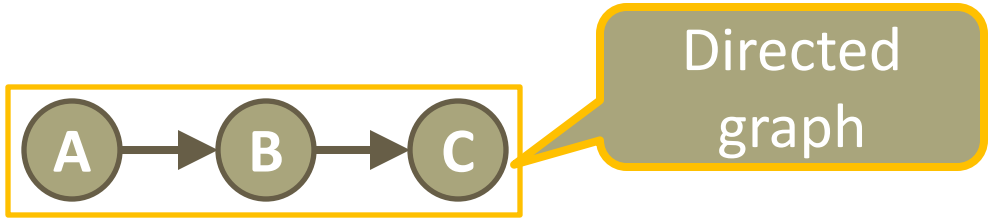
Directed graph



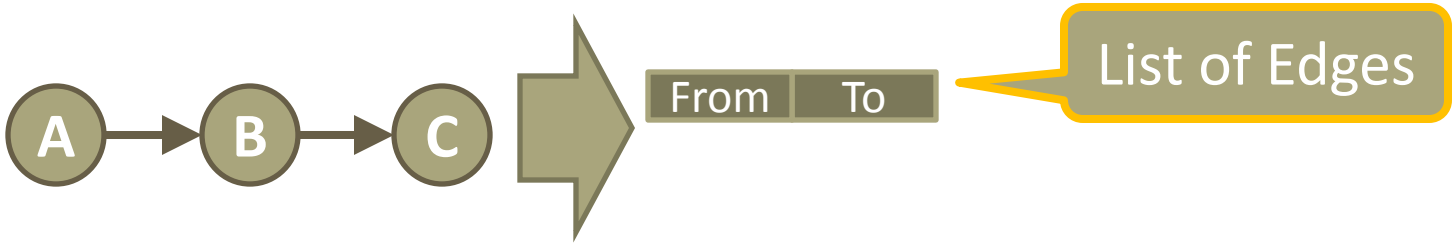
Un-directed graph

- How can we perform operations on graphs?
- How can we formalize graphs as rectangular data?

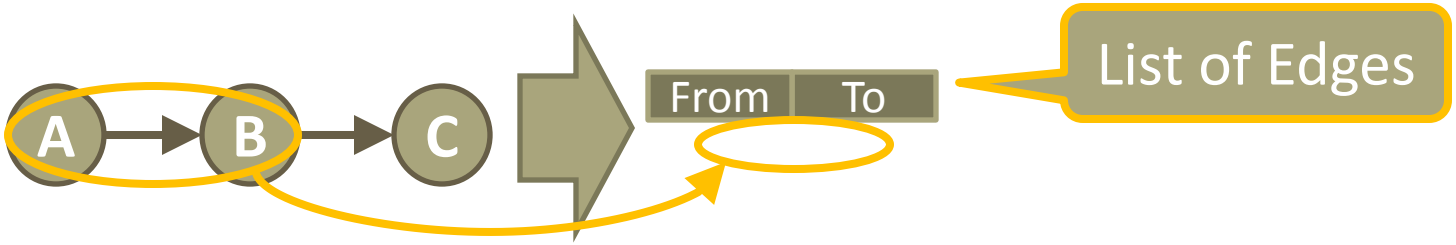
Graph Data: Formalize as Rectangular Data (1)



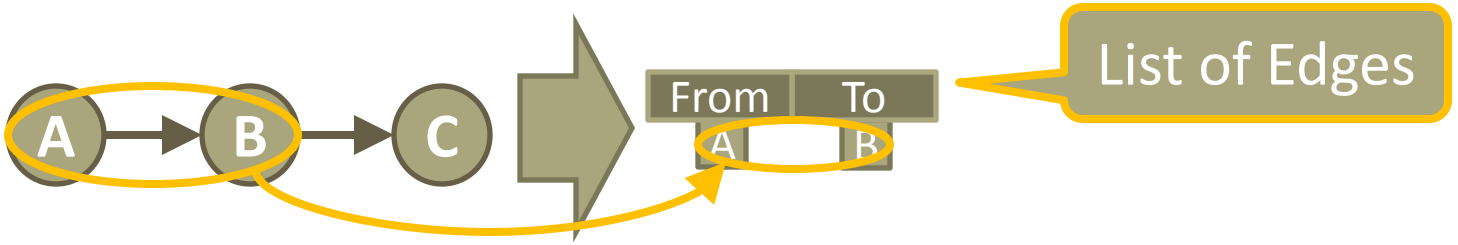
Graph Data: Formalize as Rectangular Data (2)



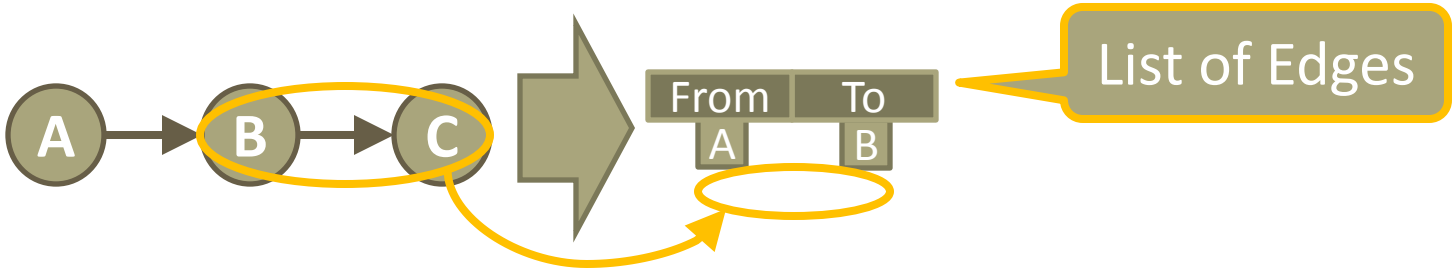
Graph Data: Formalize as Rectangular Data (3)



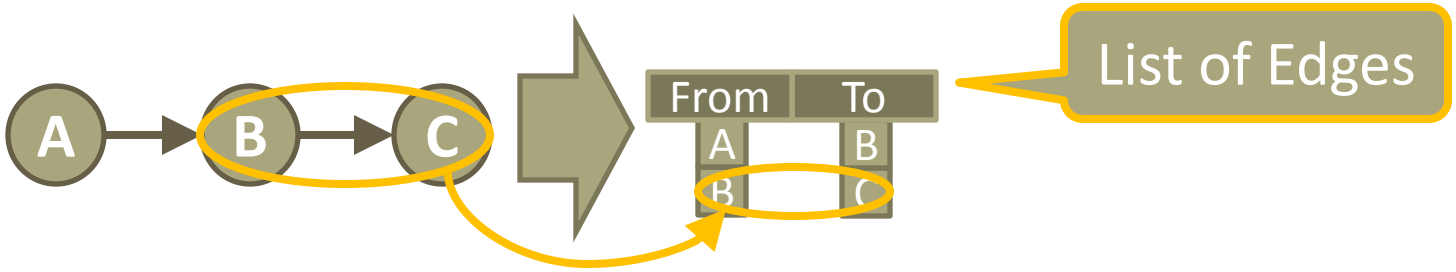
Graph Data: Formalize as Rectangular Data (4)



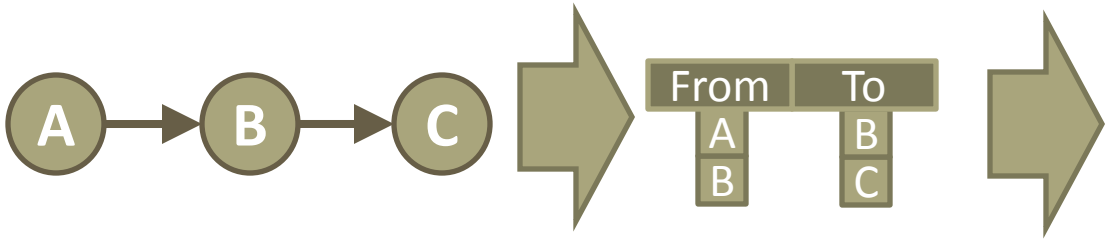
Graph Data: Formalize as Rectangular Data (5)



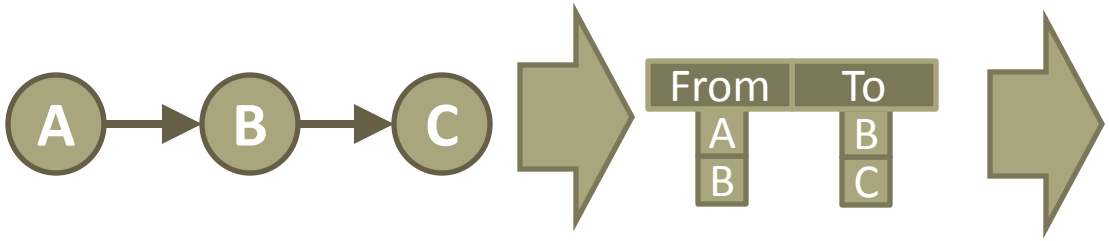
Graph Data: Formalize as Rectangular Data (6)



Graph Data: Formalize as Rectangular Data (7)



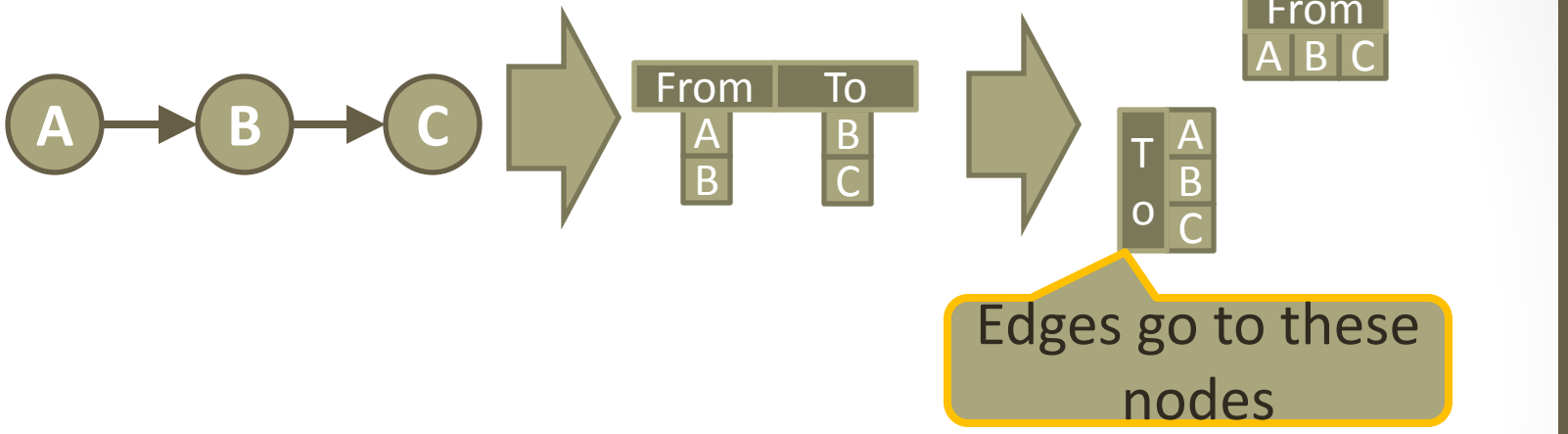
Graph Data: Formalize as Rectangular Data (8)



Edges start from these nodes

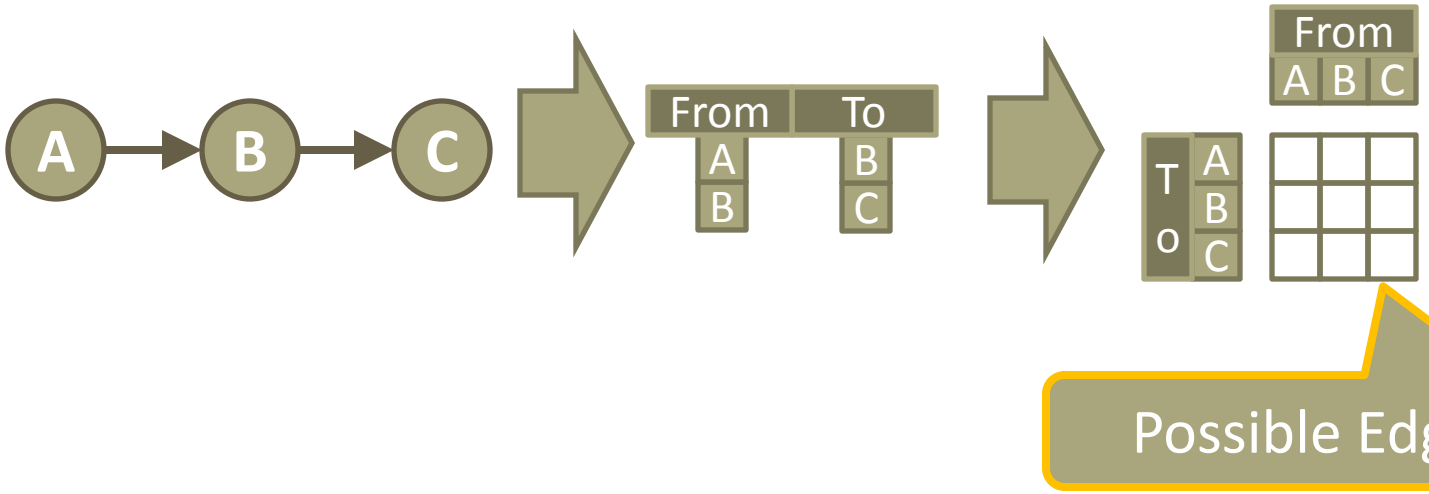
From
A
B
C

Graph Data: Formalize as Rectangular Data (9)



Every node could emanate and receive links.

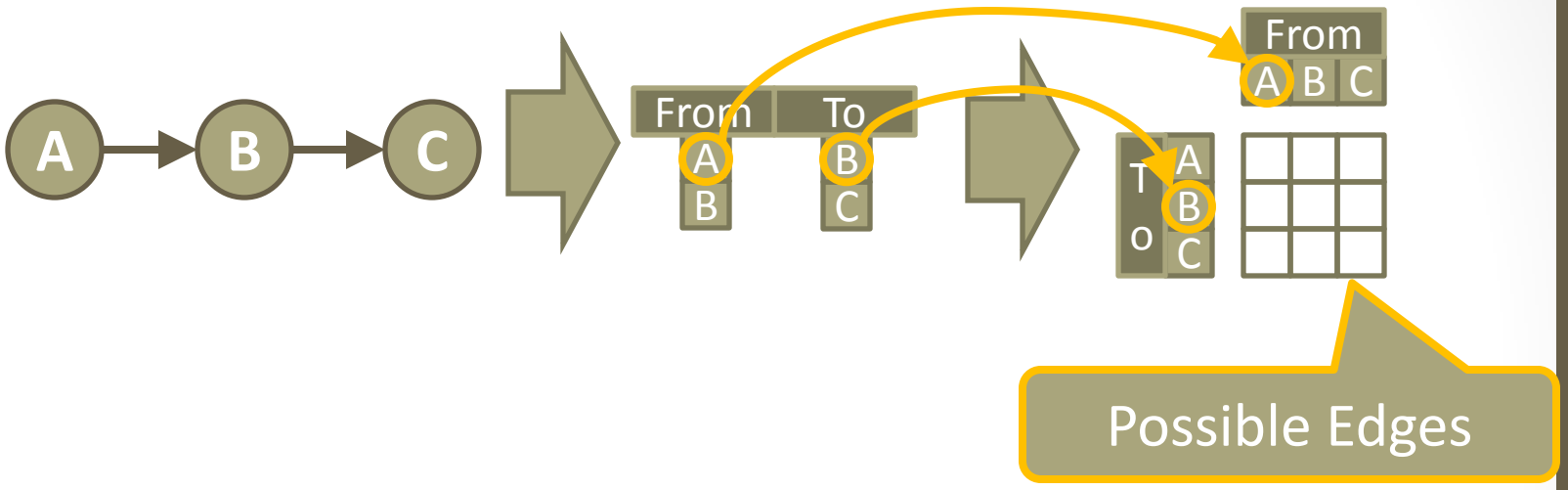
Graph Data: Formalize as Rectangular Data (10)



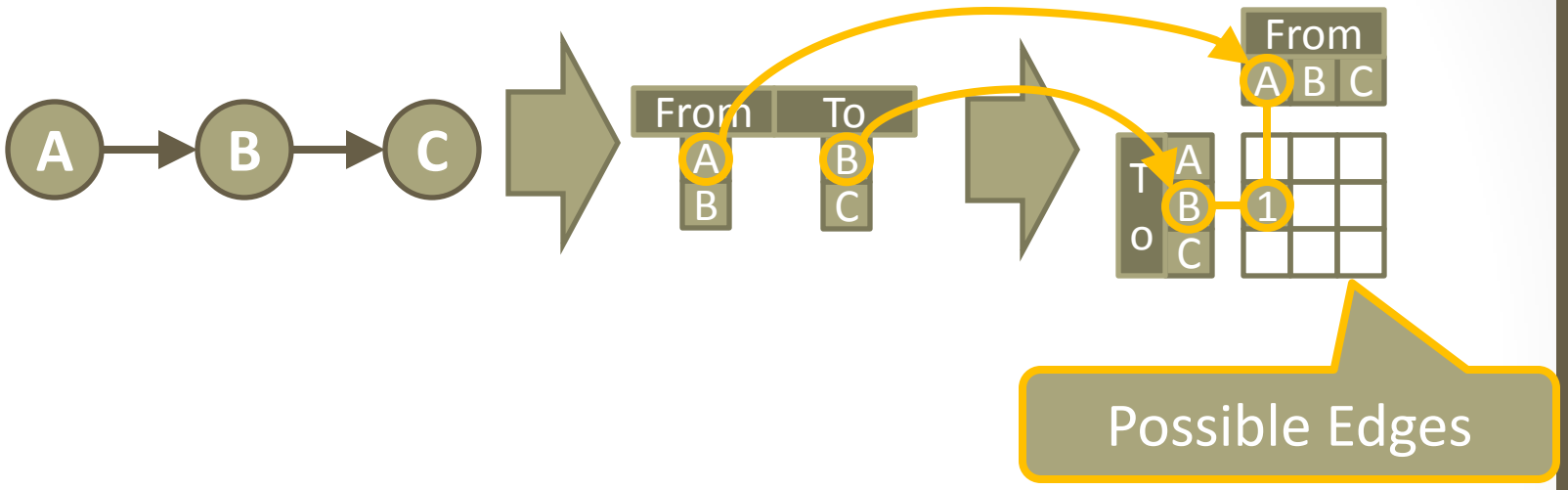
Possible Edges

Every node could emanate and receive links. Therefore the matrix is square.

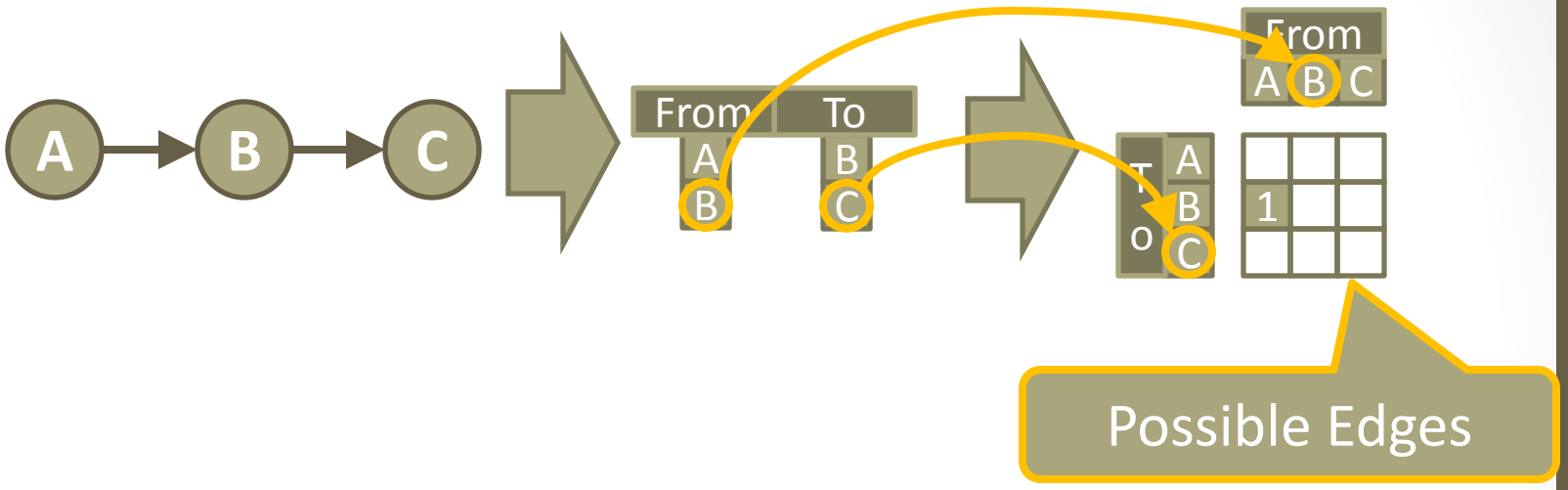
Graph Data: Formalize as Rectangular Data (11)



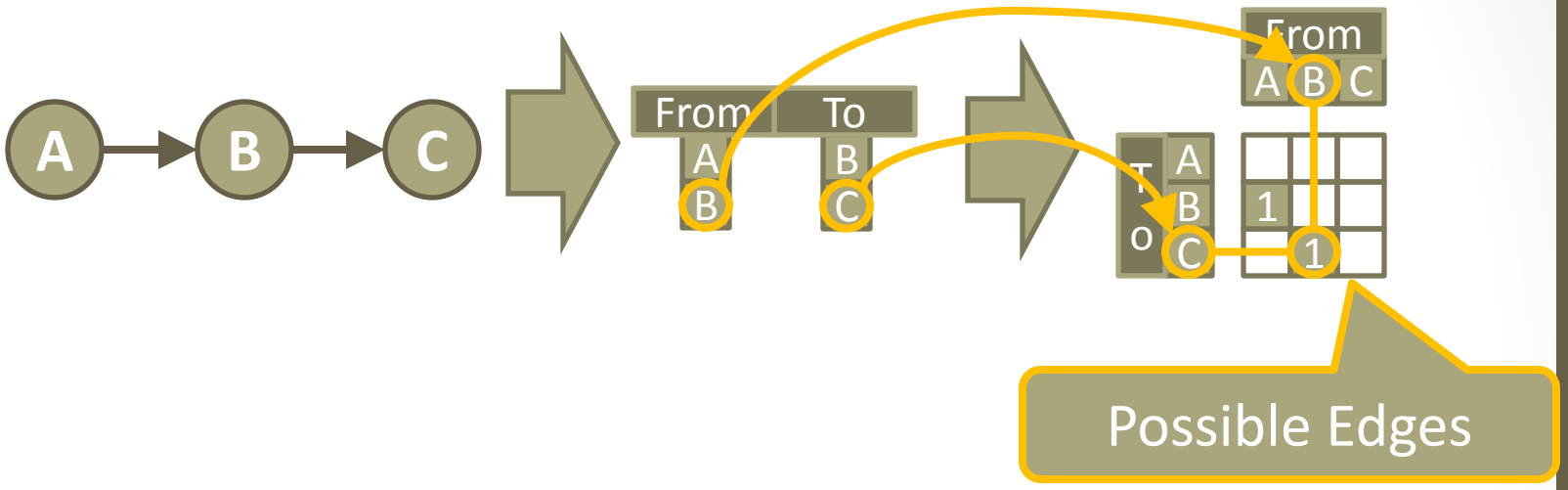
Graph Data: Formalize as Rectangular Data (12)



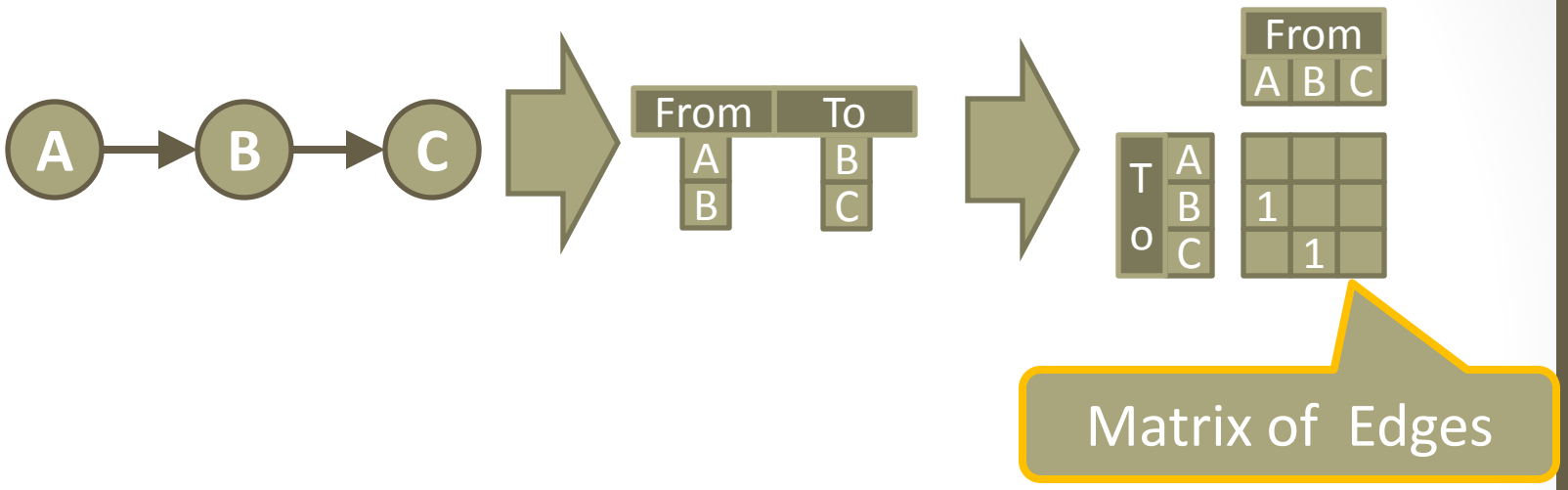
Graph Data: Formalize as Rectangular Data (13)



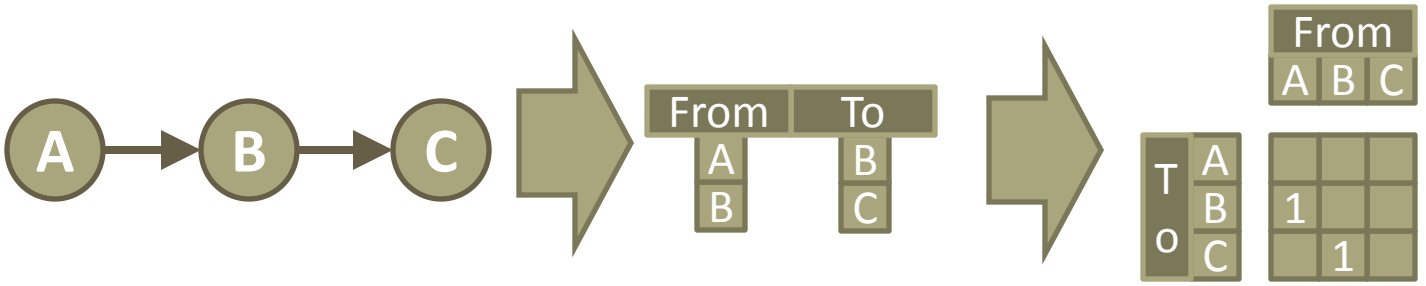
Graph Data: Formalize as Rectangular Data (14)

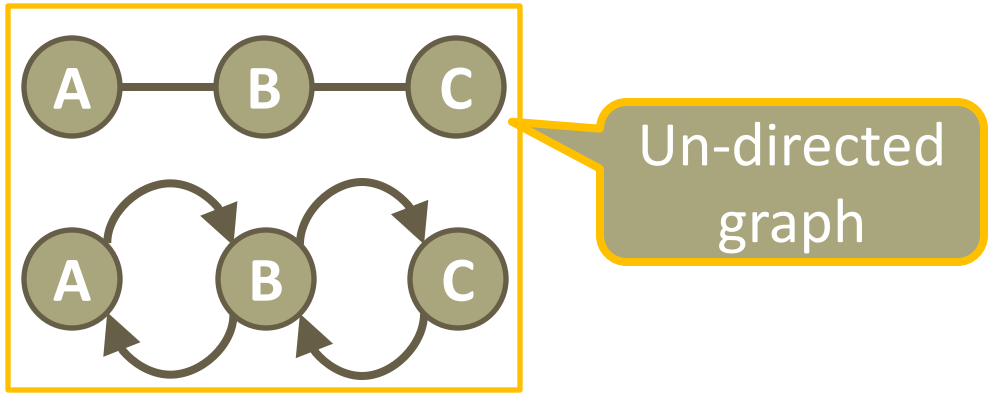


Graph Data: Formalize as Rectangular Data (15)

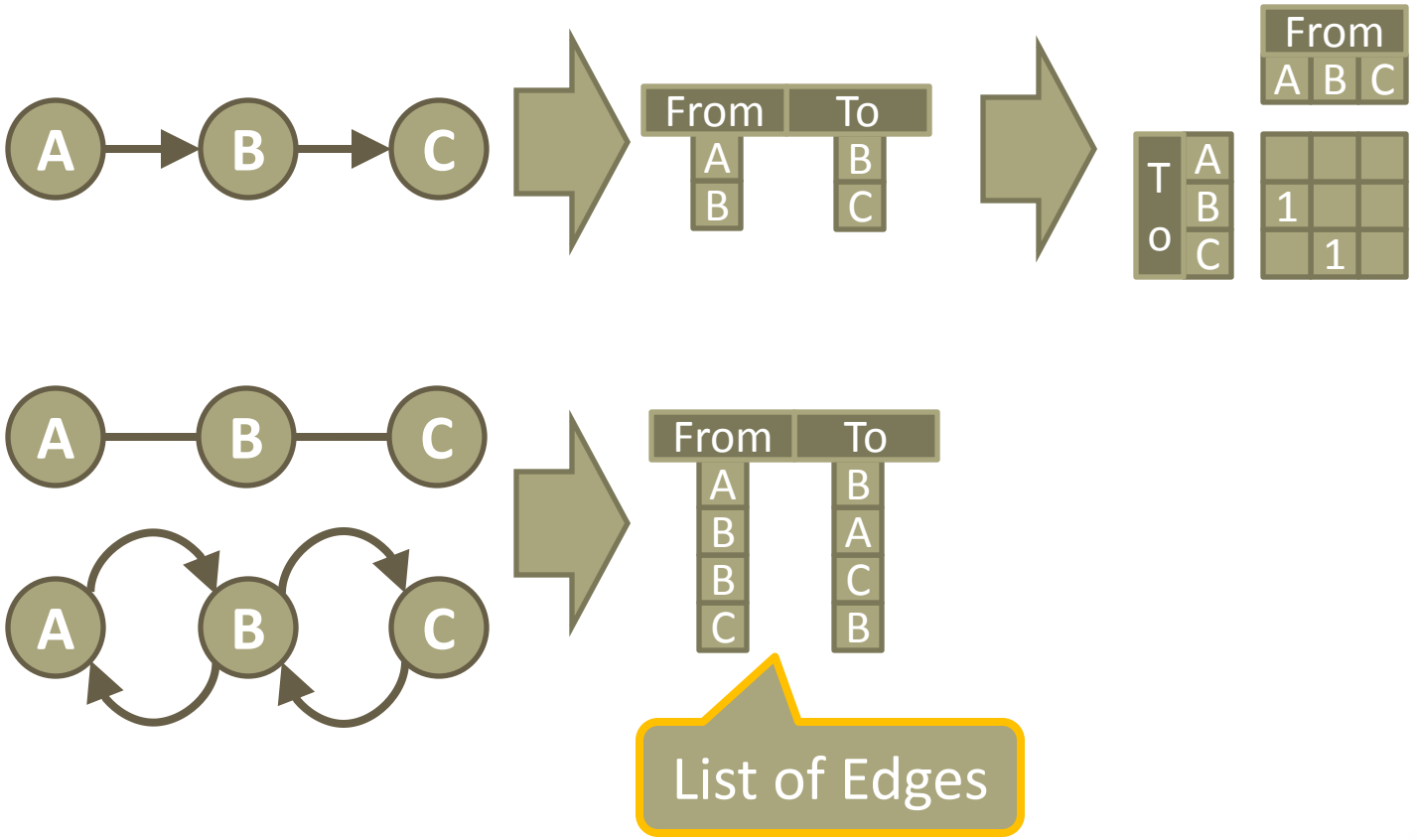


Graph Data: Formalize as Rectangular Data (16)

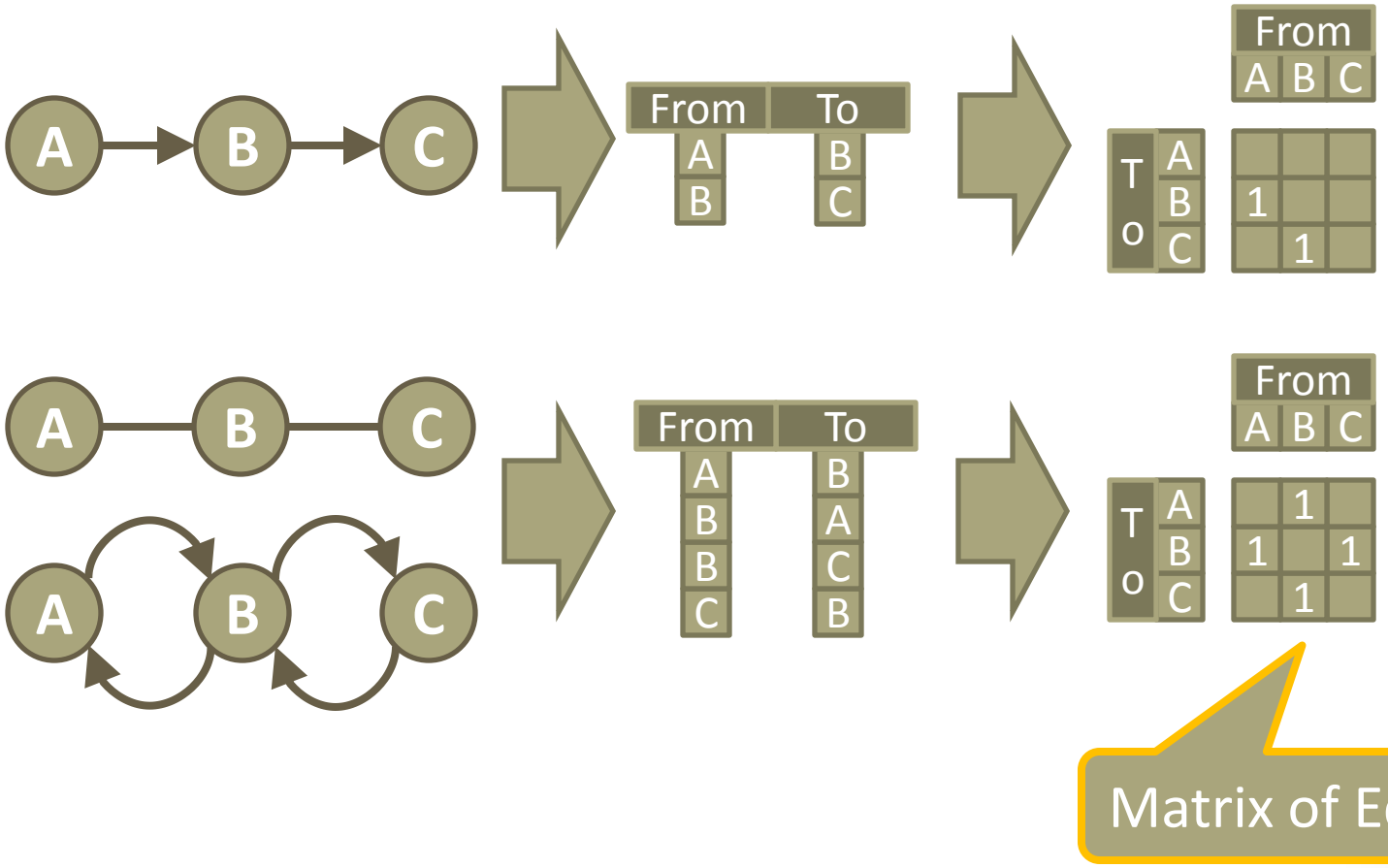




Graph Data: Formalize as Rectangular Data (17)

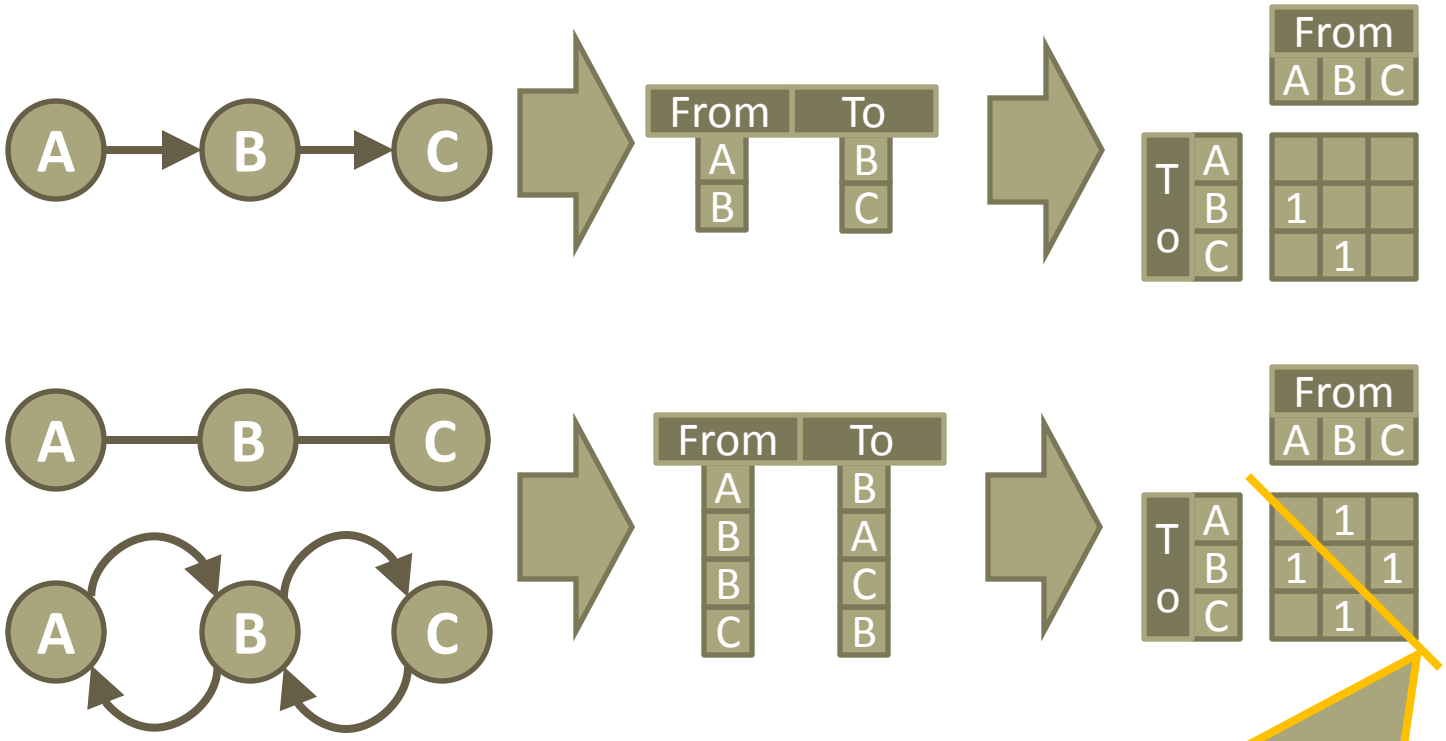


Graph Data: Formalize as Rectangular Data (18)



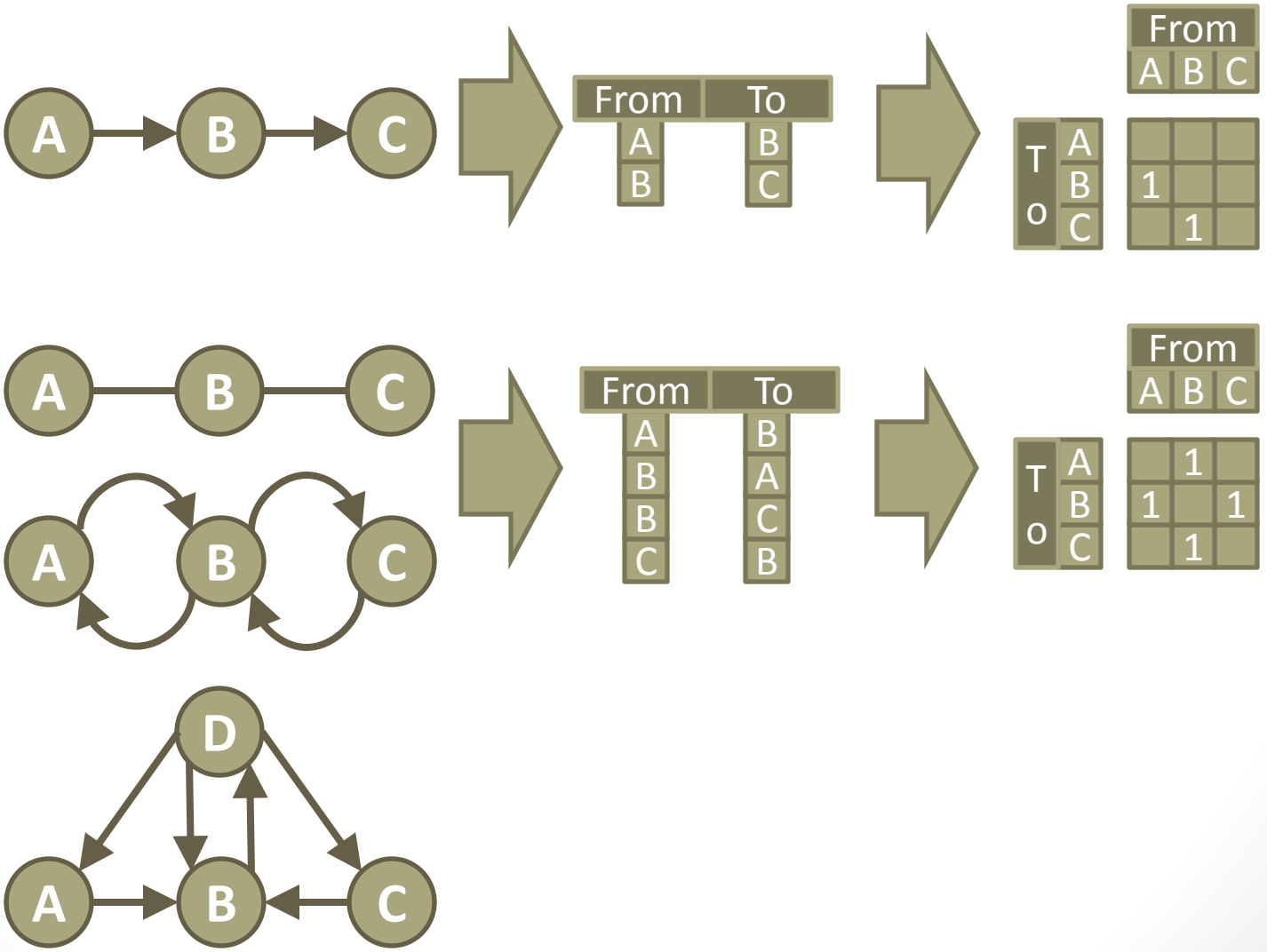
Matrix of Edges

Graph Data: Formalize as Rectangular Data (19)

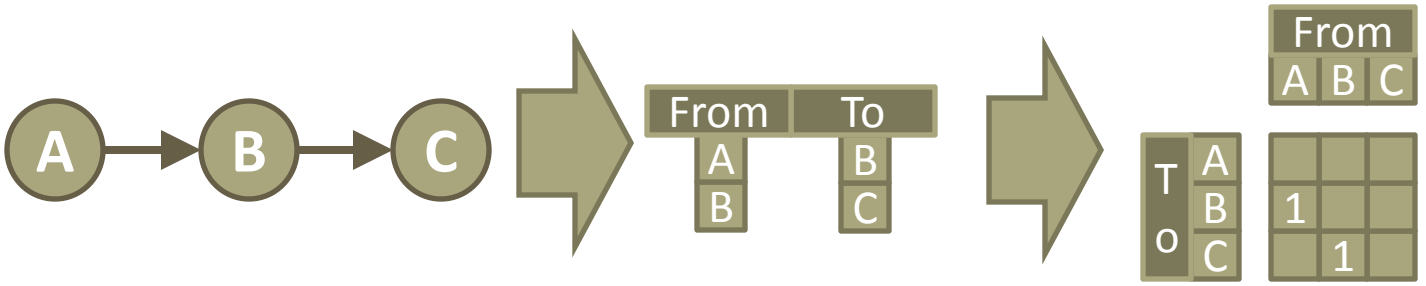


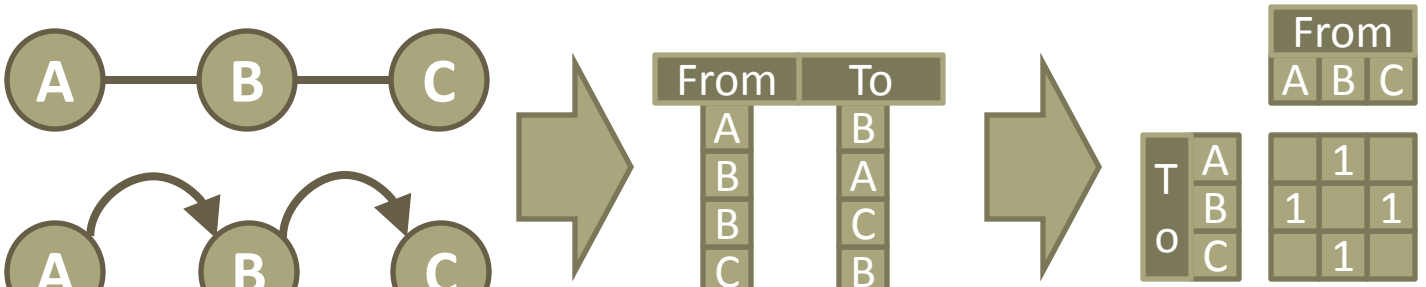
Note: undirected or bi-directional graphs Have symmetric matrices

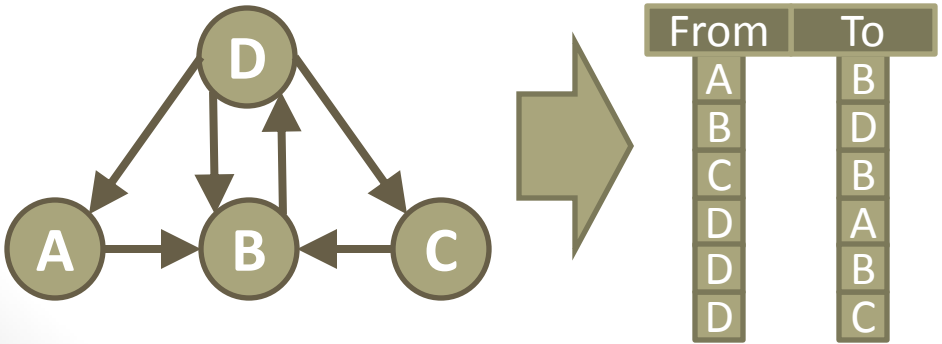
Graph Data: Formalize as Rectangular Data (20)



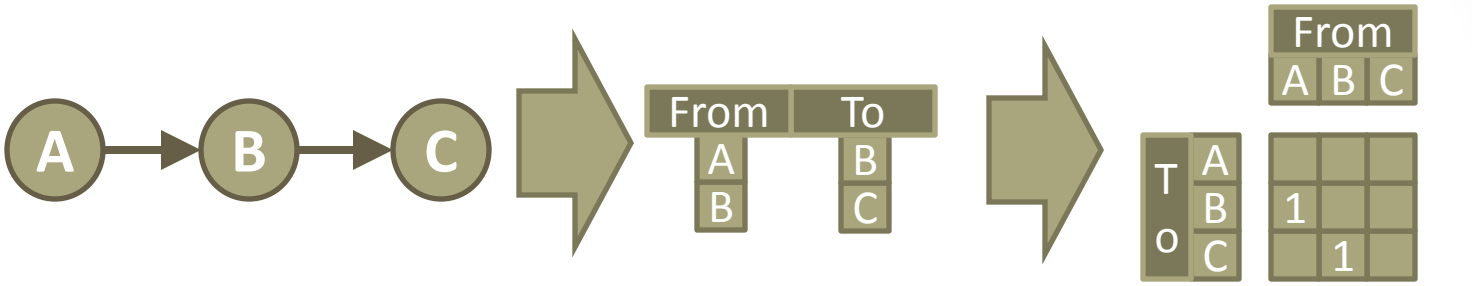
Graph Data: Formalize as Rectangular Data (21)

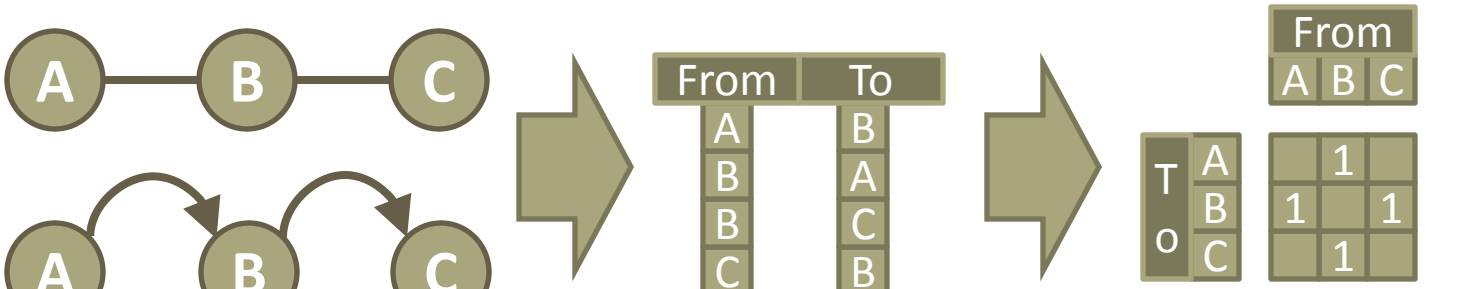


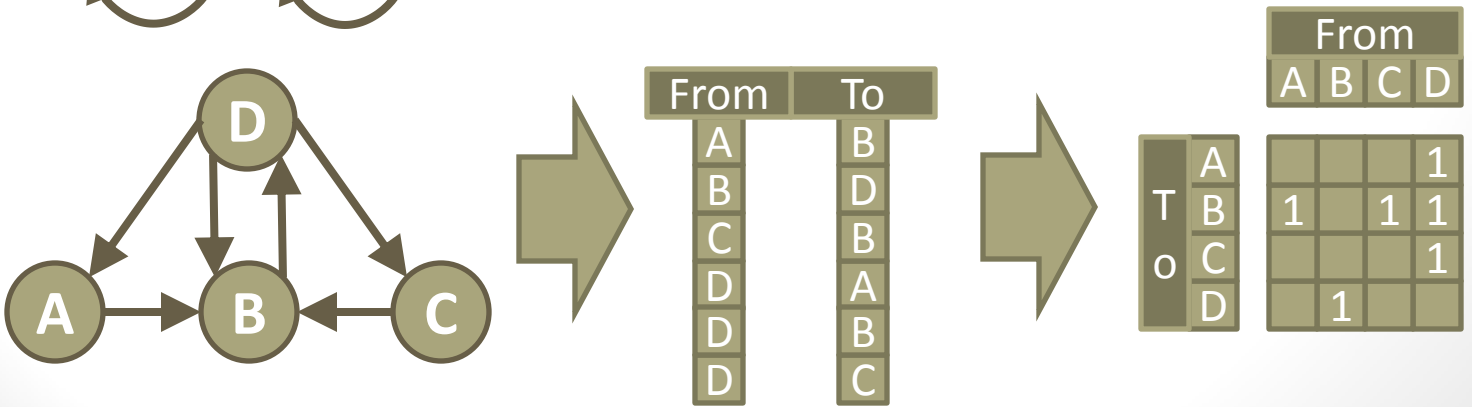




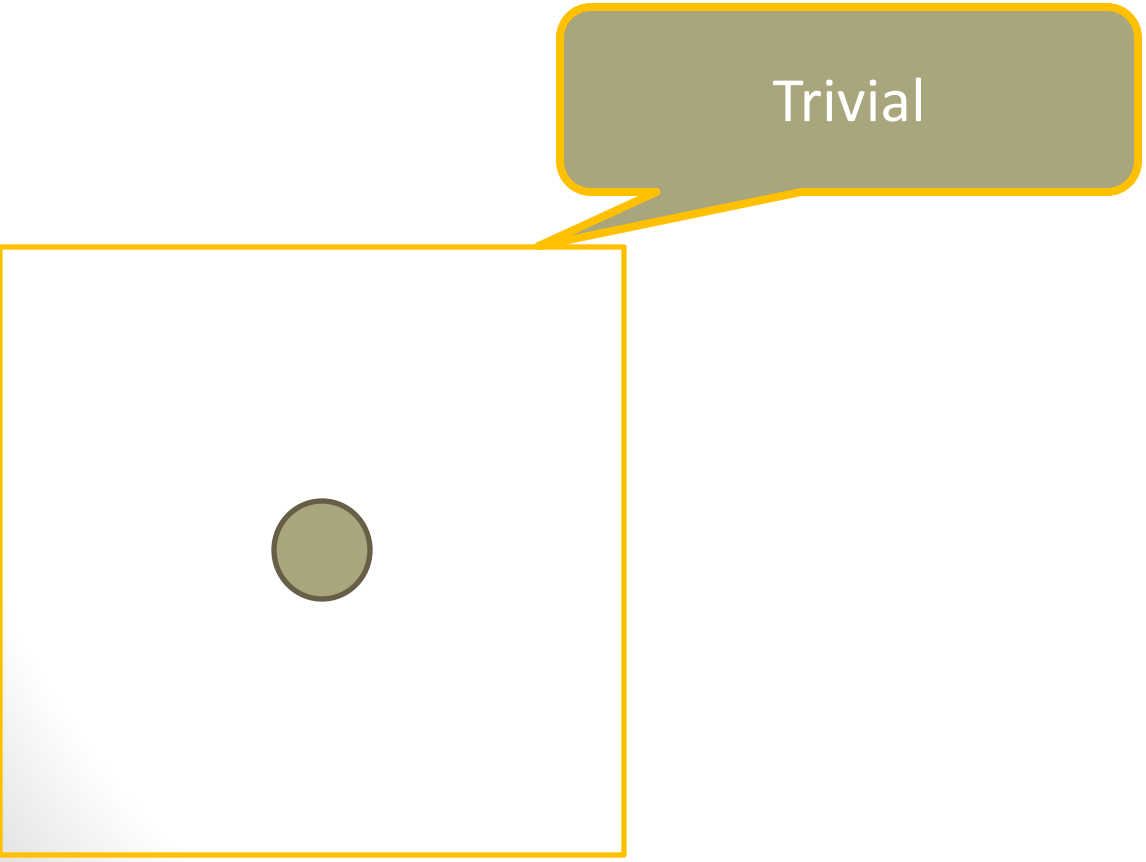
Graph Data: Formalize as Rectangular Data (22)



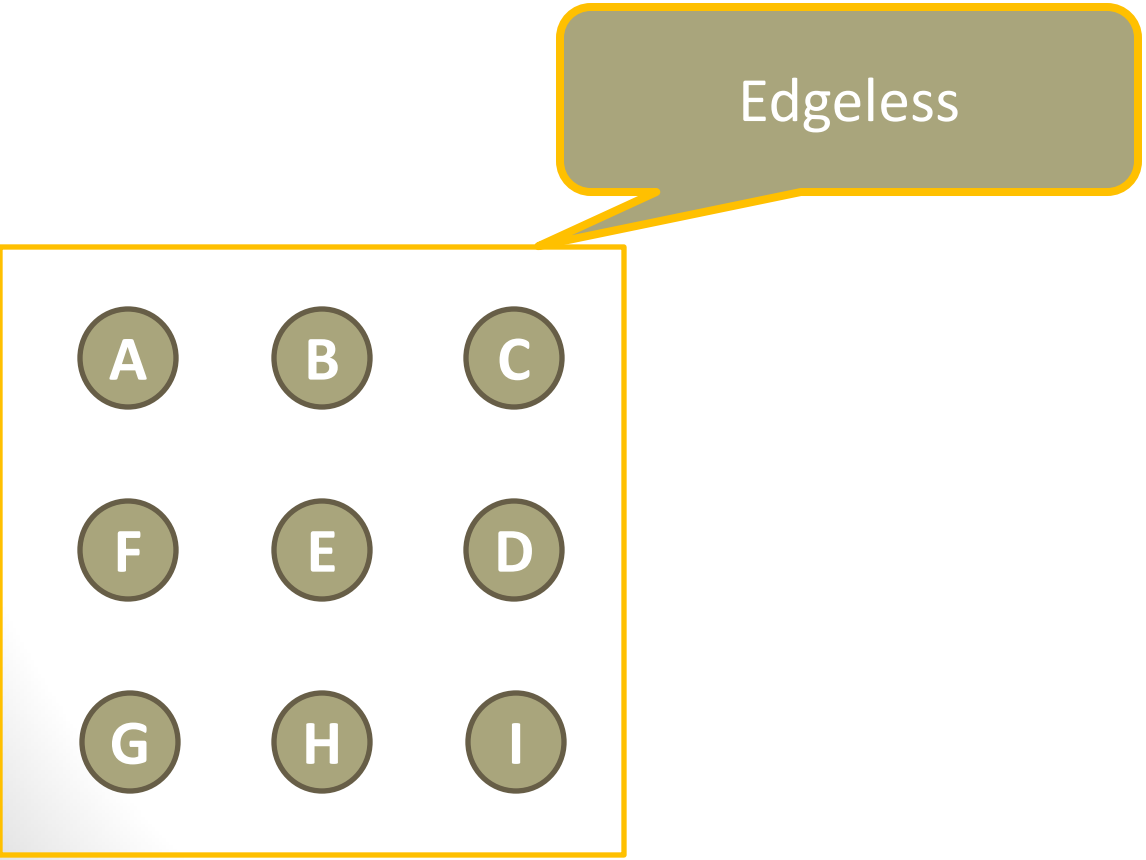




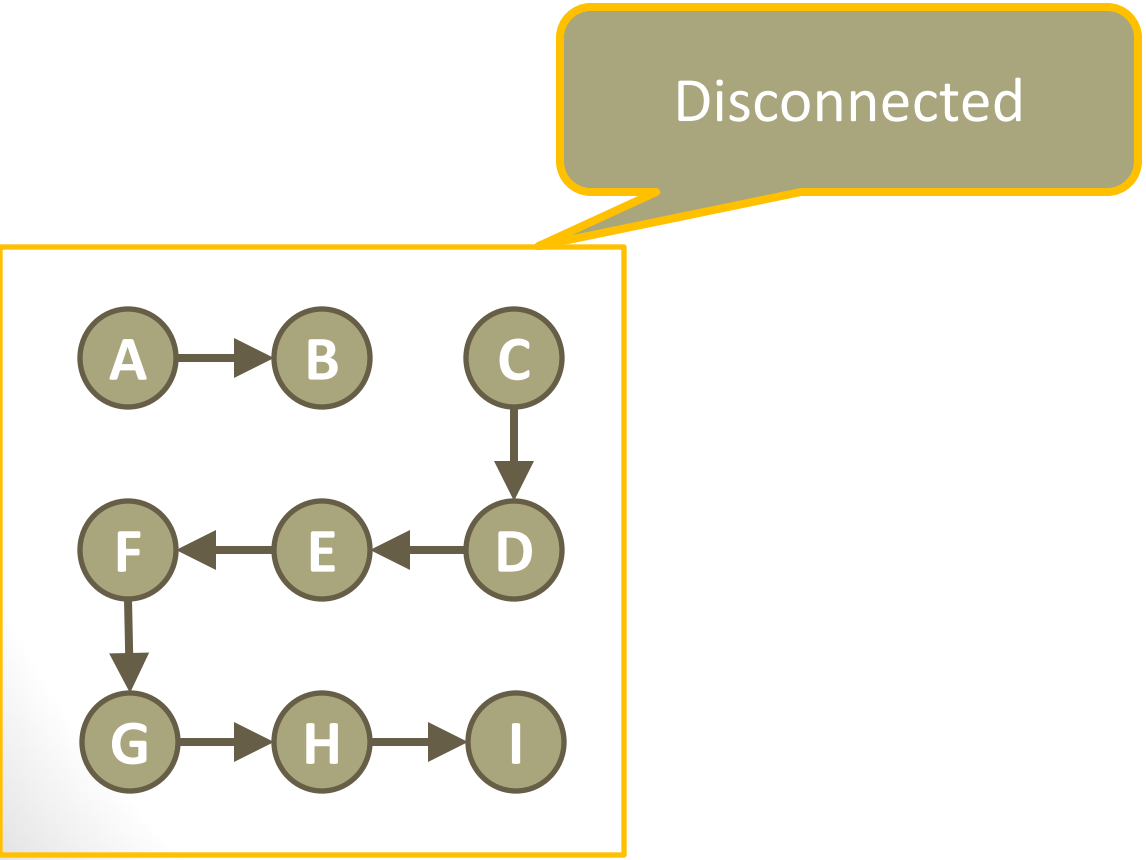
Graph Data: Connectedness and Density (0)



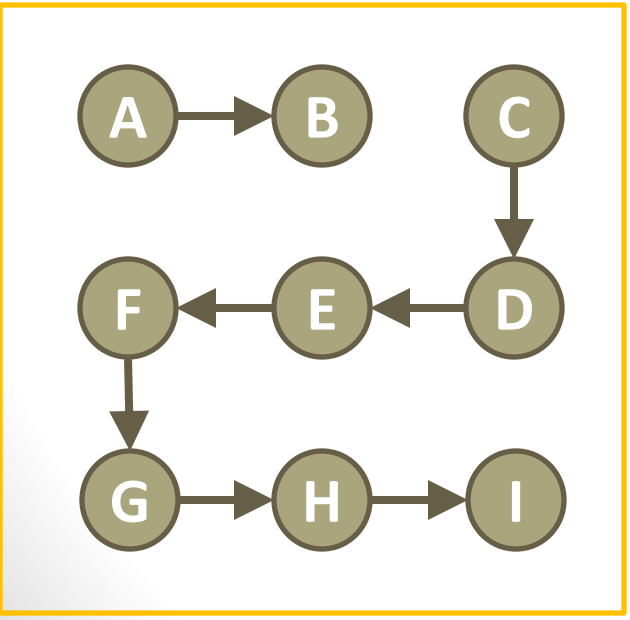
Graph Data: Connectedness and Density (1)

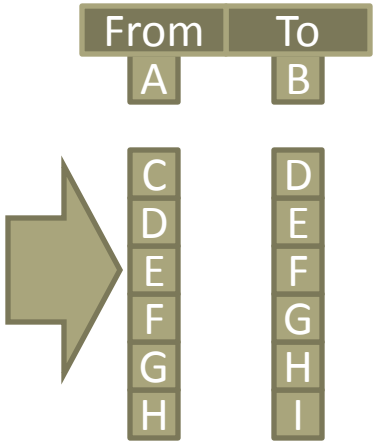


Graph Data: Connectedness and Density (2)

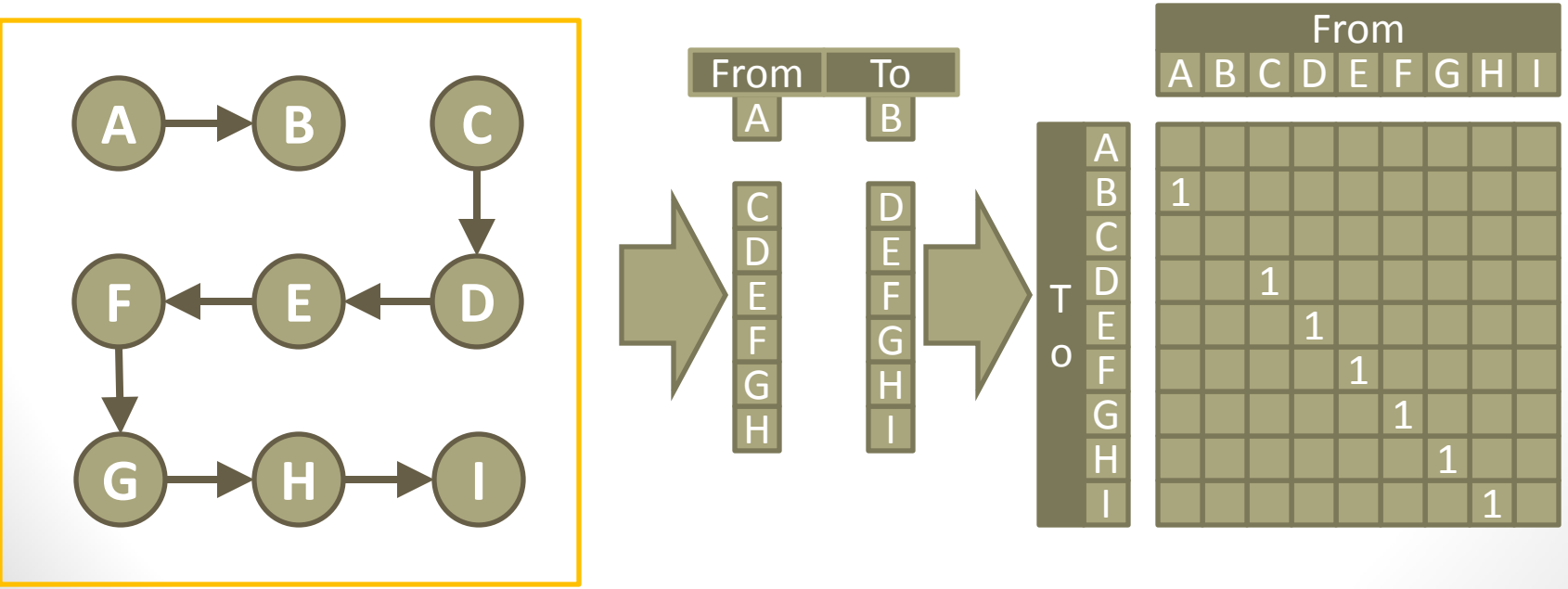


Graph Data: Connectedness and Density (3)

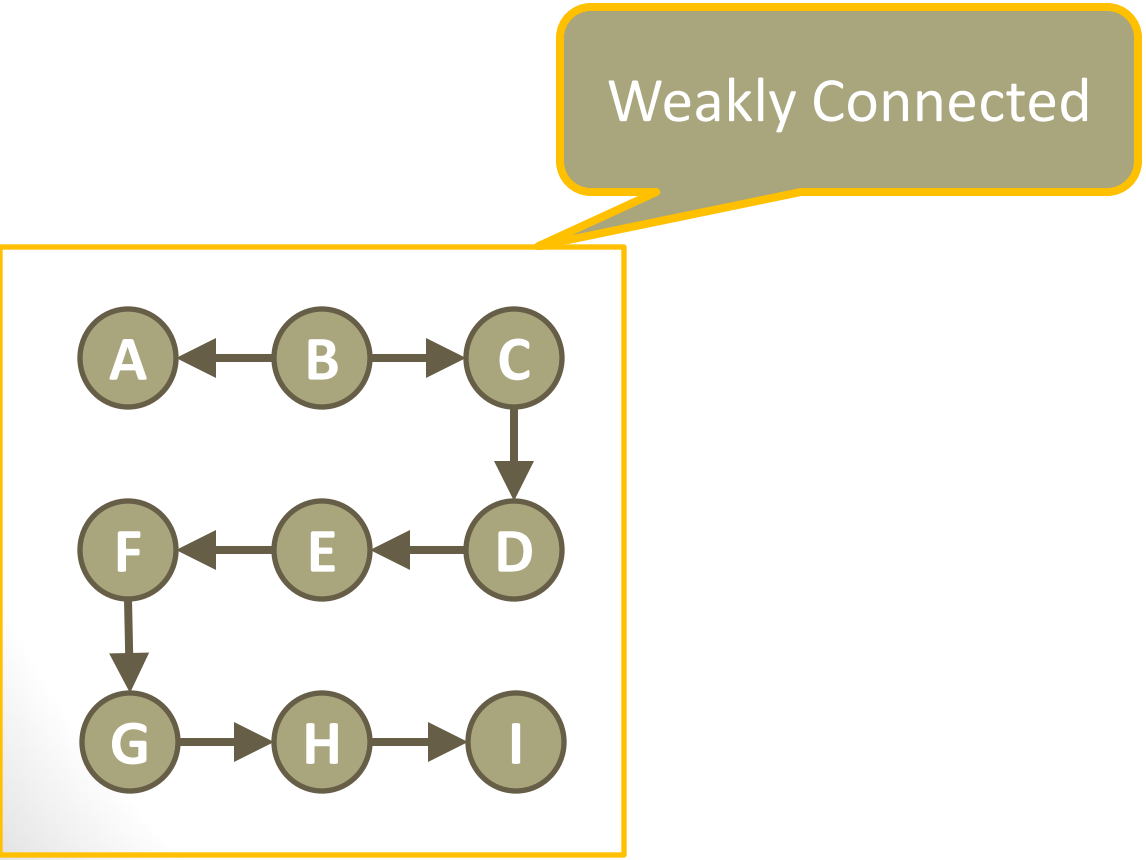




Graph Data: Connectedness and Density (4)

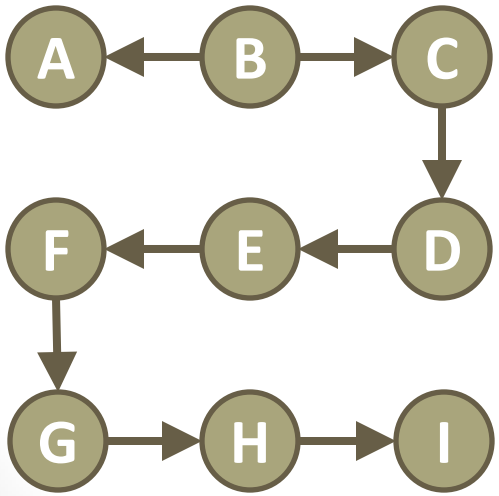


Graph Data: Connectedness and Density (5)



Graph Data: Connectedness and Density (6)

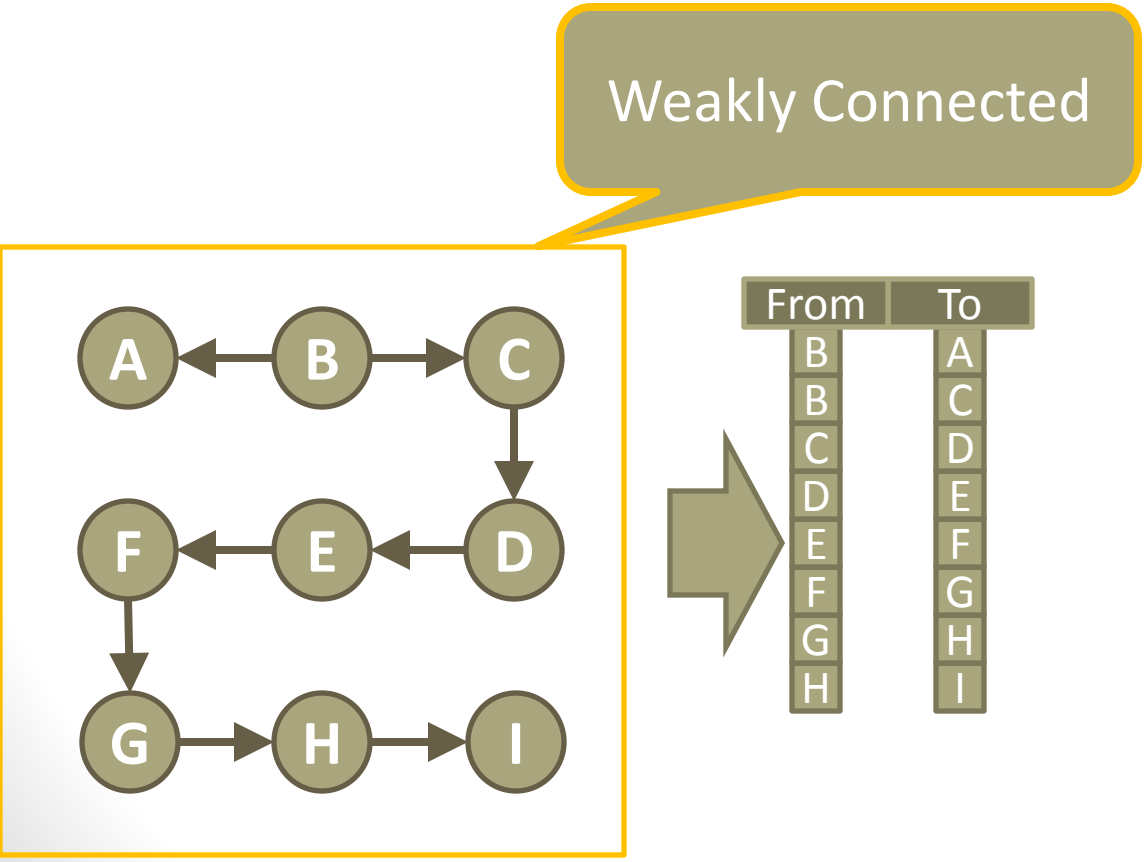
Weakly Connected



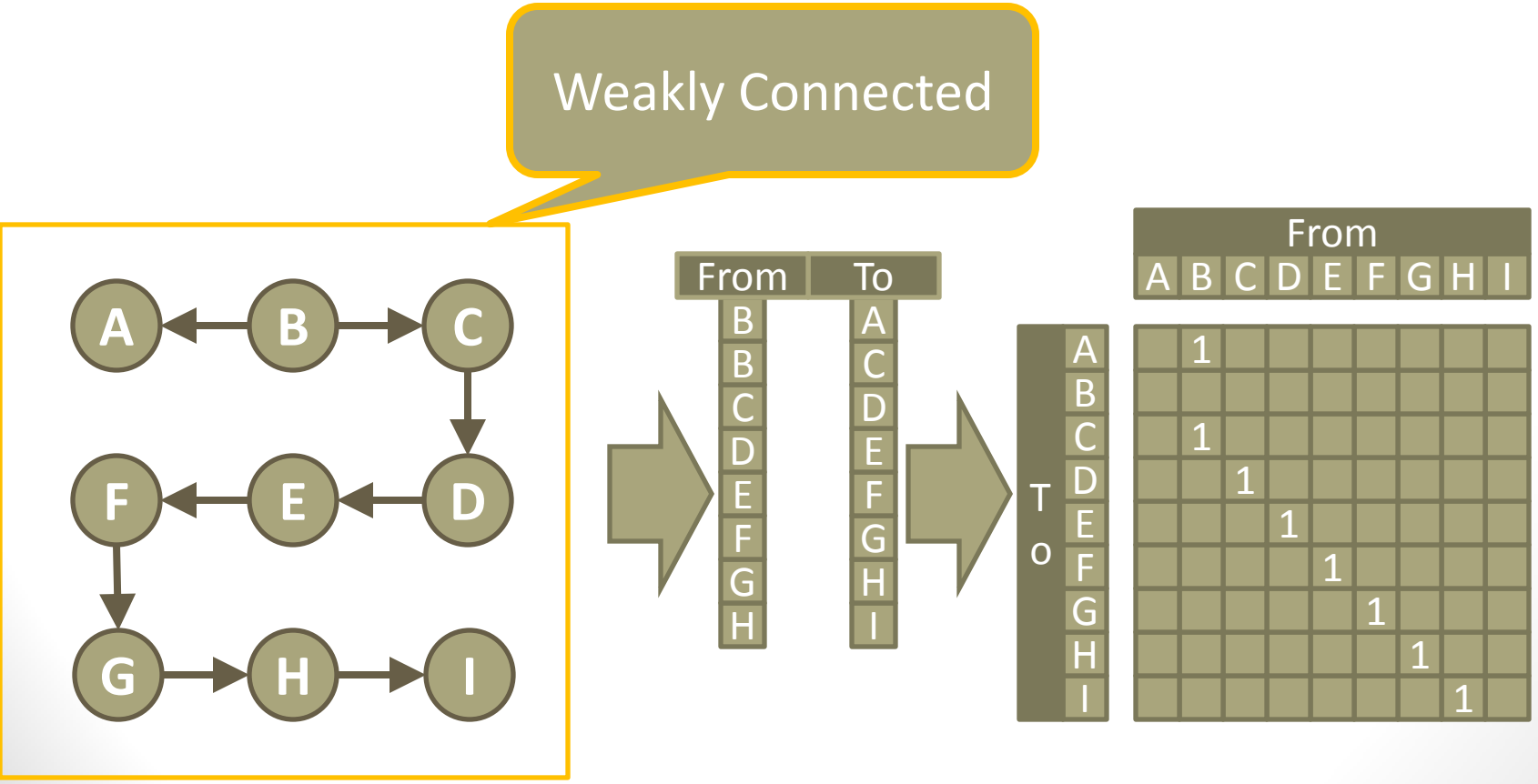
Weakly Connected:
For every pair of vertices, a path exists along the graph's edges regardless of the edge's direction. There exists at least one pair of vertices (u, v), where a path does not exist along the graph's directed edges from u to v an does not exist from v to u.

In this case: A path does not exist from A to C and does not exist from C to A.

Graph Data: Connectedness and Density (7)

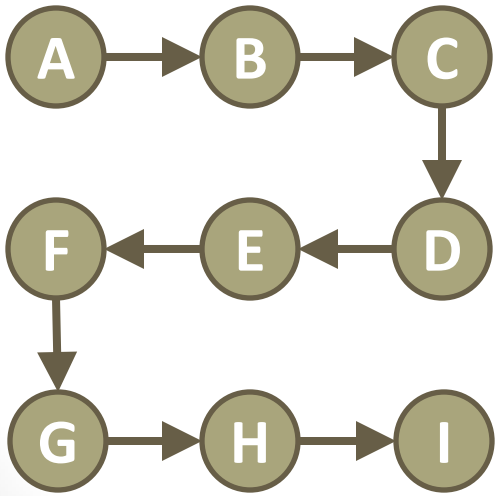


Graph Data: Connectedness and Density (8)



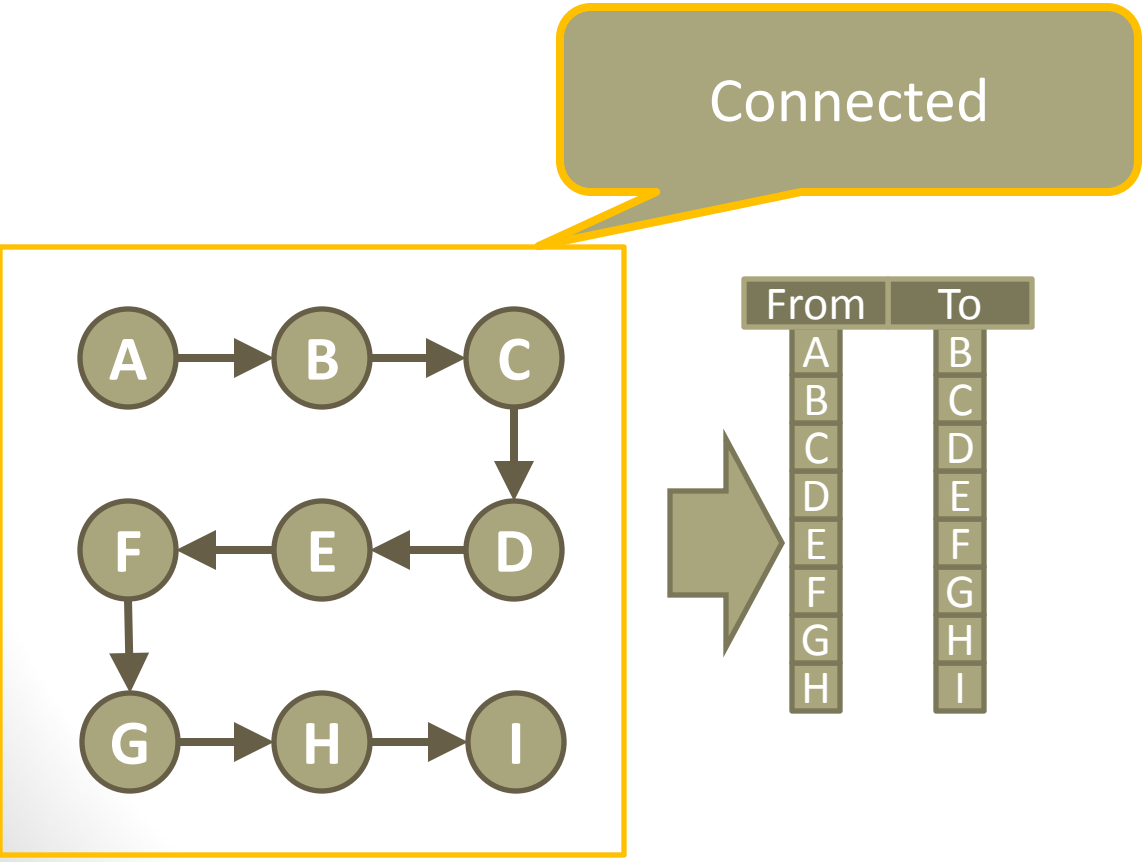
Graph Data: Connectedness and Density (9)

Connected

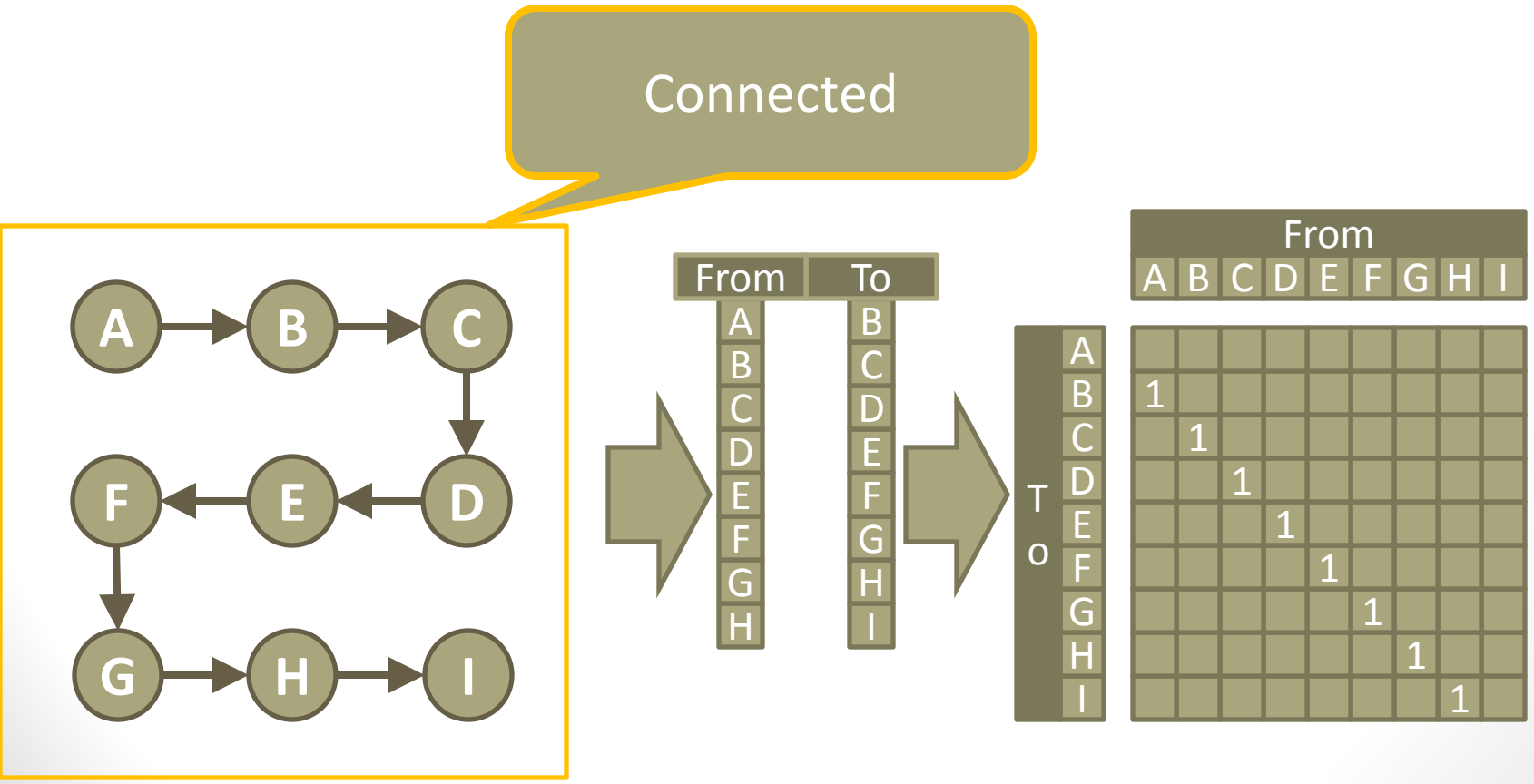


Connected:
For every pair of vertices (u, v), a path exists along the graph's directed edges from u to v or from v to u. There exists at least one pair of vertices (u, v), where a path does not exist along the graph's directed edges from either u to v or from v to u.

Graph Data: Connectedness and Density (10)

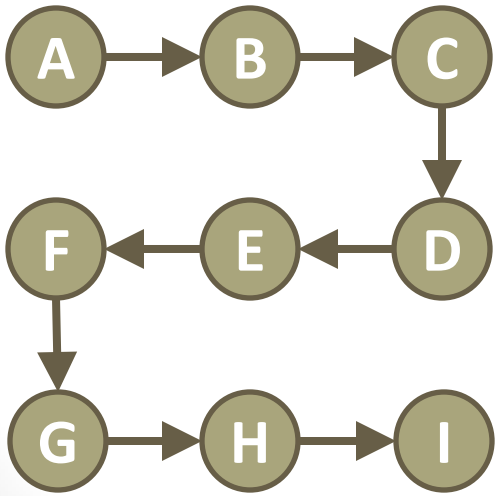


Graph Data: Connectedness and Density (11)



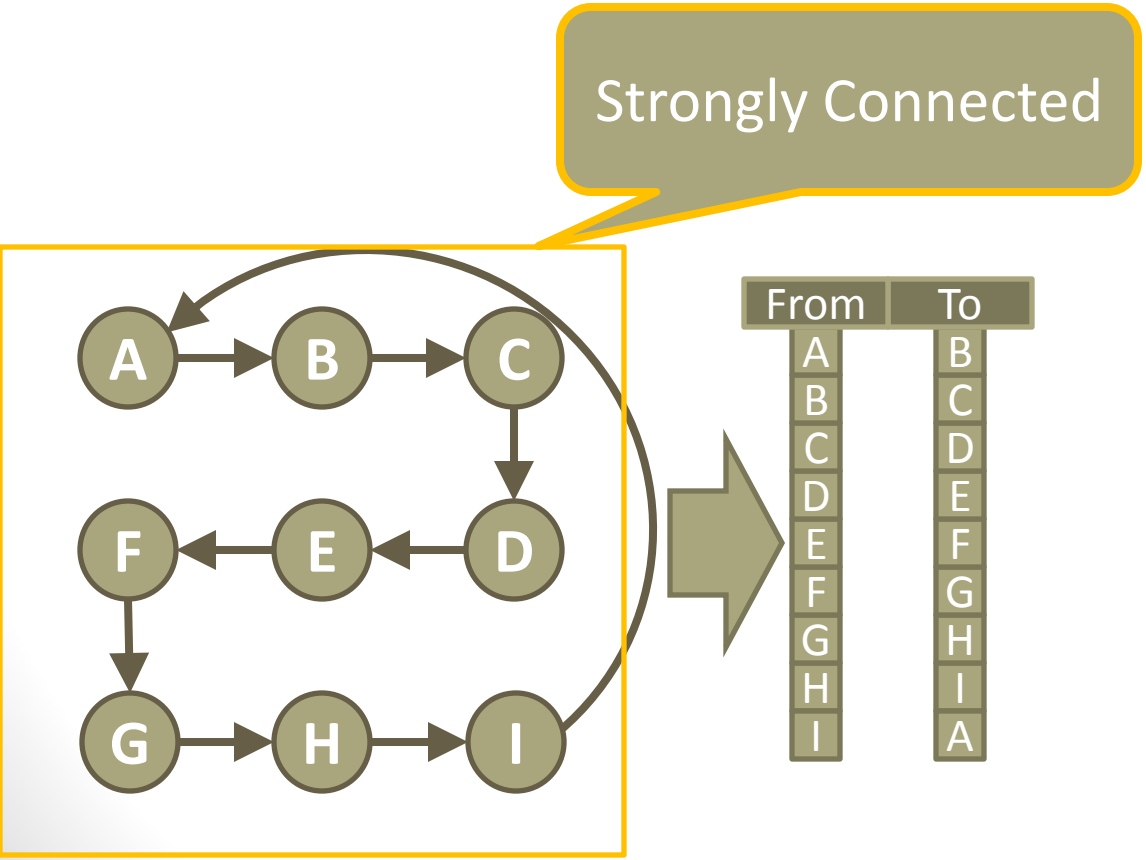
Graph Data: Connectedness and Density (12)

Connected

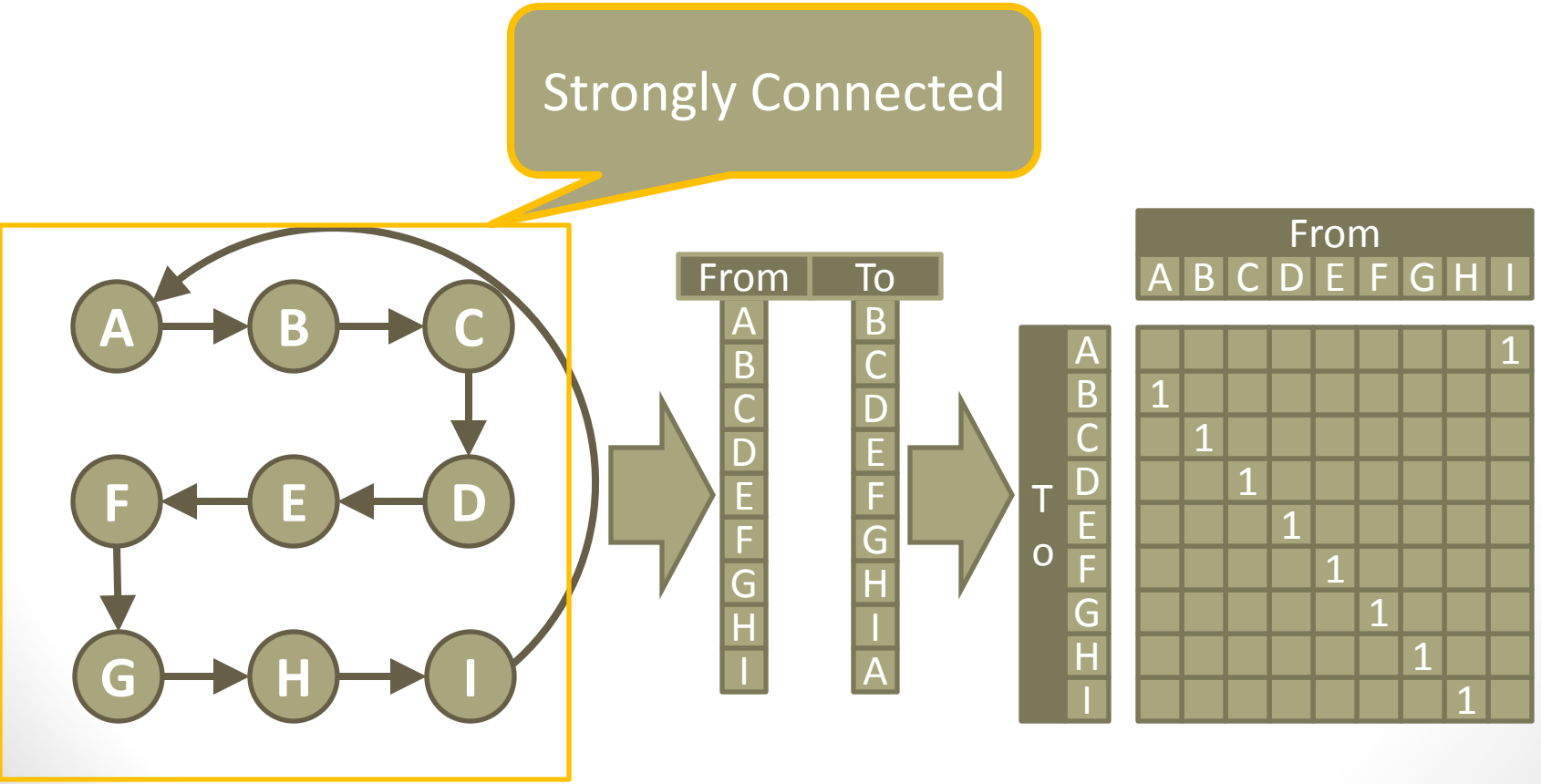


Connected:
A path exists along the graph's directed edges from every node to every node

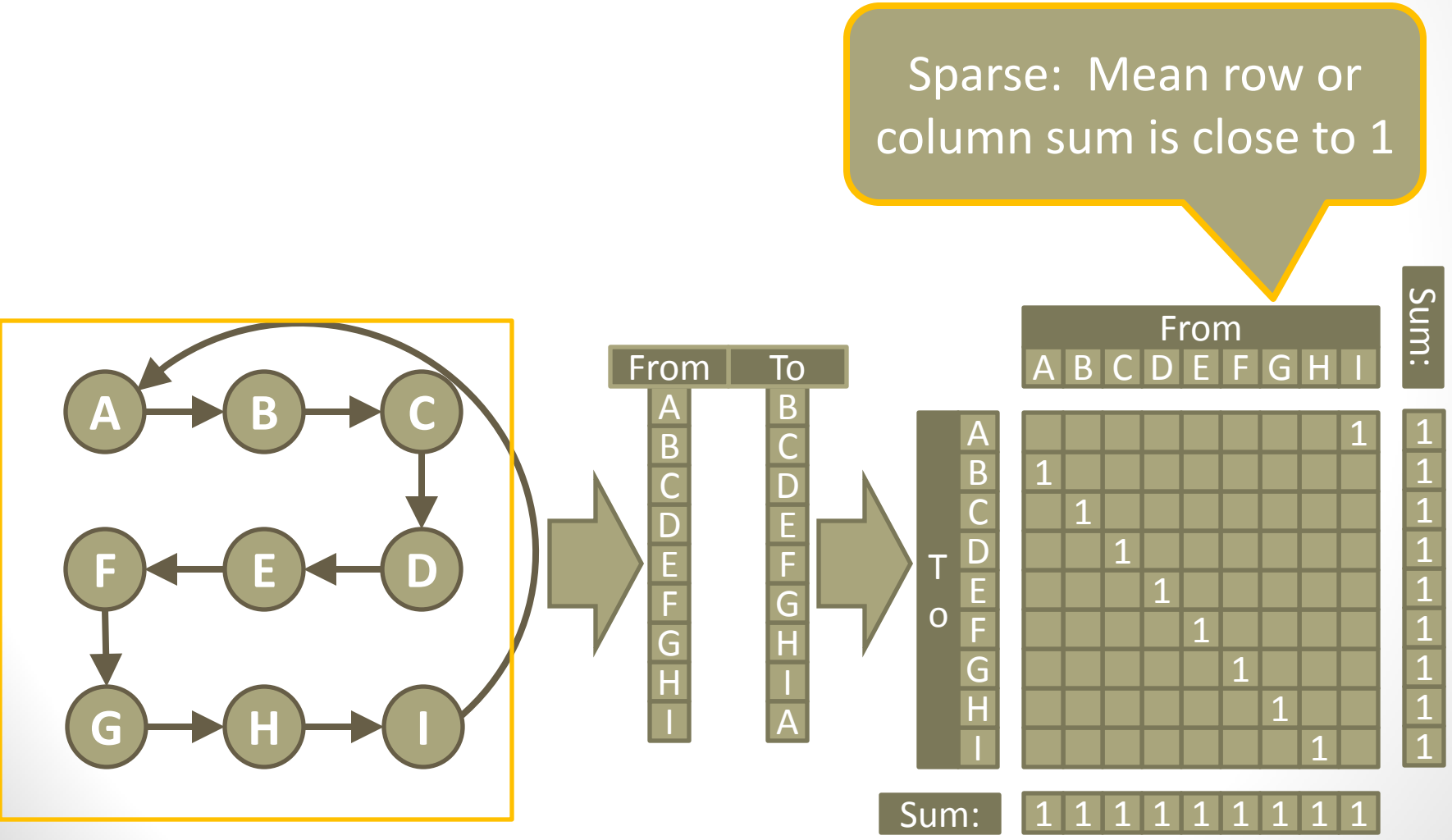
Graph Data: Connectedness and Density (13)



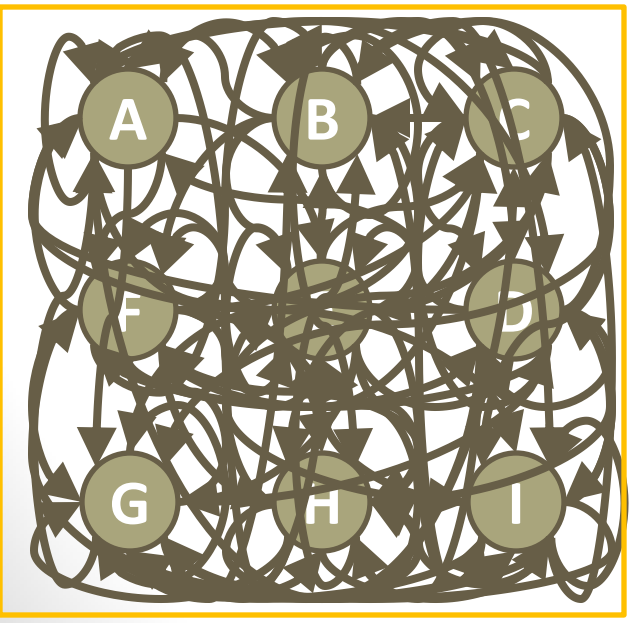
Graph Data: Connectedness and Density (14)



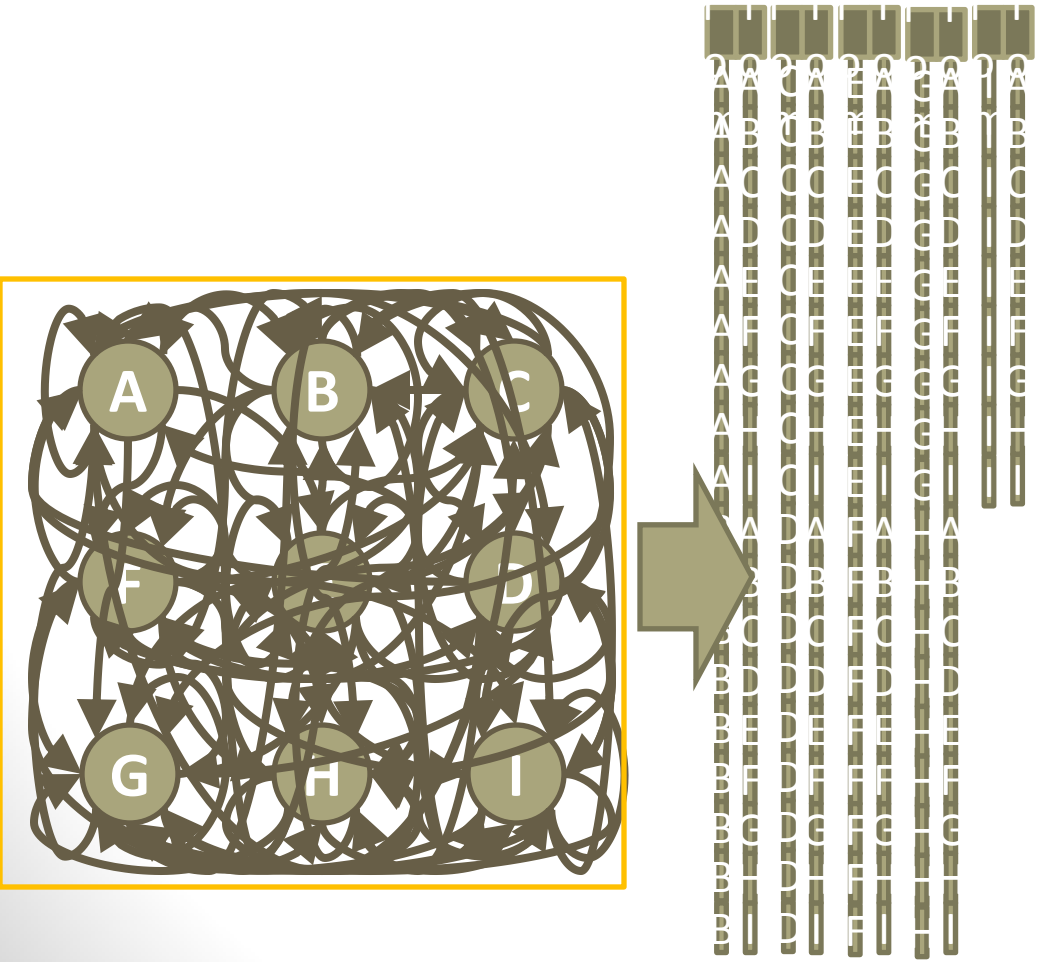
Graph Data: Connectedness and Density (15)



Graph Data: Connectedness and Density (16)



Graph Data: Connectedness and Density (17)



Graph Data: Connectedness and Density (18)

From	To
A	A
A	B
A	C
A	D
A	E
A	F
A	G
A	H
A	I
B	A
B	B
B	C
B	D
B	E
B	F
B	G
B	H
B	I

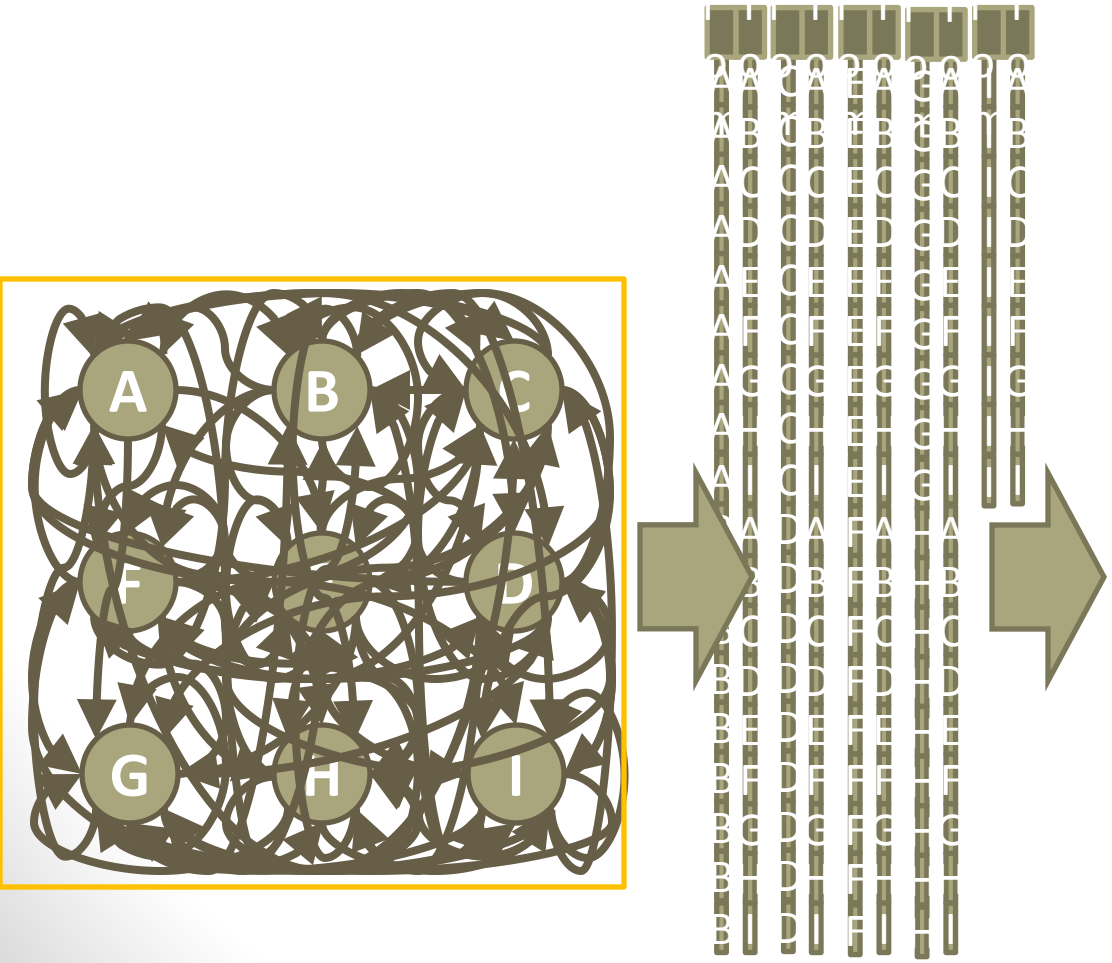
From	To
C	A
C	B
C	C
C	D
C	E
C	F
C	G
C	H
C	I
D	A
D	B
D	C
D	D
D	E
D	F
D	G
D	H
D	I

From	To
E	A
E	B
E	C
E	D
E	E
E	F
E	G
E	H
E	I
F	A
F	B
F	C
F	D
F	E
F	F
F	G
F	H
F	I

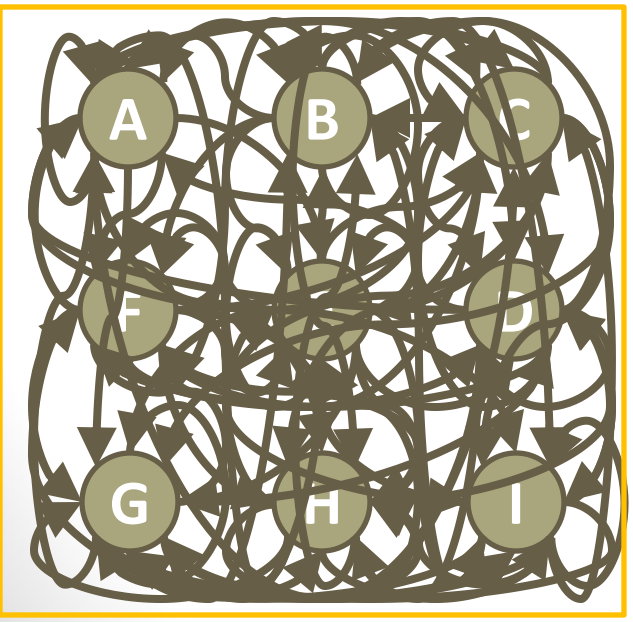
From	To
G	A
G	B
G	C
G	D
G	E
G	F
G	G
G	H
G	I
H	A
H	B
H	C
H	D
H	E
H	F
H	G
H	H
H	I

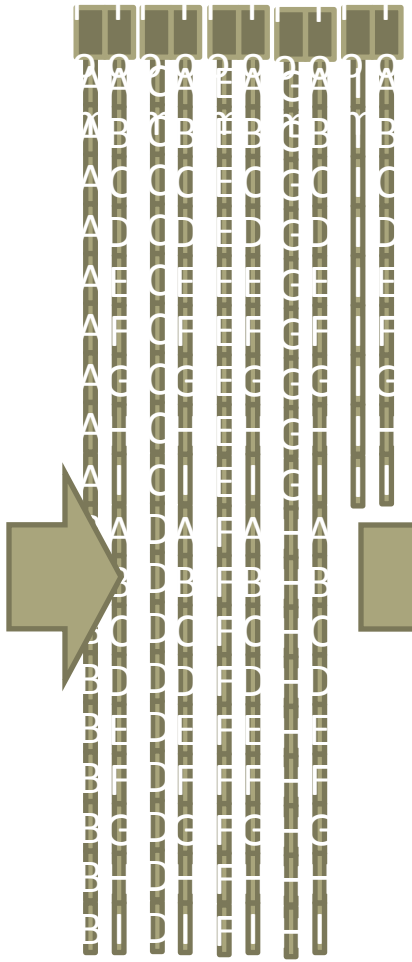
From	To
I	A
I	B
I	C
I	D
I	E
I	F
I	G
I	H
I	I

Graph Data: Connectedness and Density (19)



Graph Data: Connectedness and Density (20)

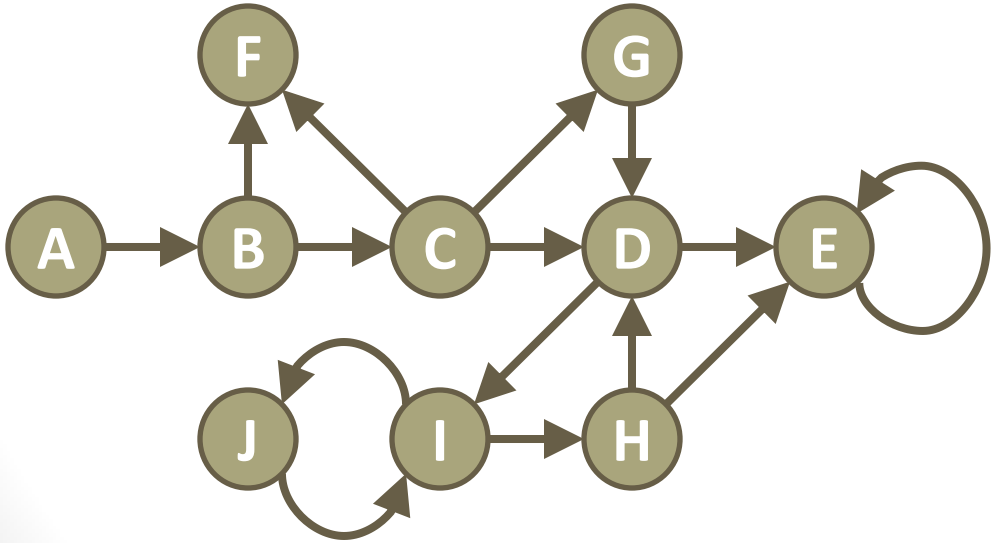




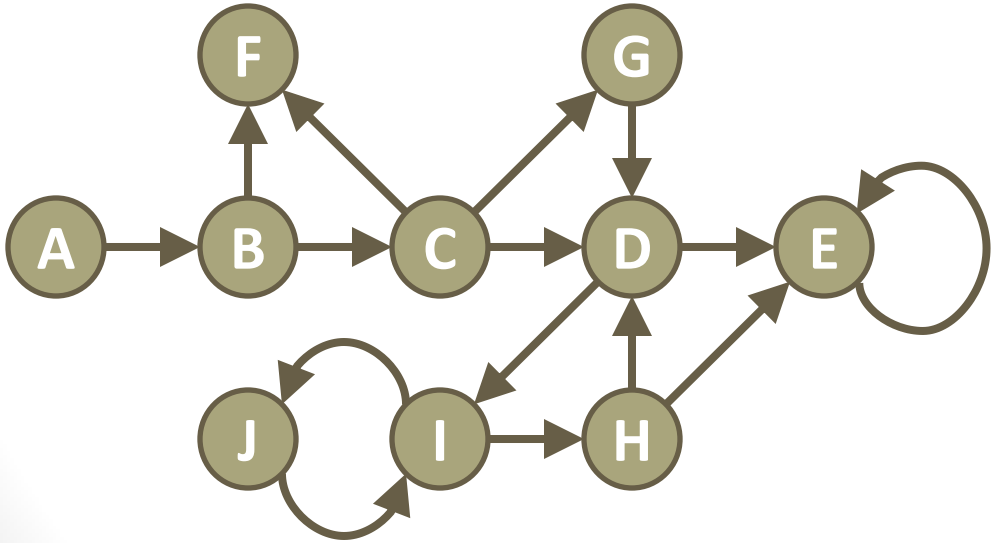
Dense: Mean row or column sum is close to number of nodes

To \ From	A	B	C	D	E	F	G	H	I	Sum:
A	1	1	1	1	1	1	1	1	1	9
B	1	1	1	1	1	1	1	1	1	9
C	1	1	1	1	1	1	1	1	1	9
D	1	1	1	1	1	1	1	1	1	9
E	1	1	1	1	1	1	1	1	1	9
F	1	1	1	1	1	1	1	1	1	9
G	1	1	1	1	1	1	1	1	1	9
H	1	1	1	1	1	1	1	1	1	9
I	1	1	1	1	1	1	1	1	1	9
Sum:	9	9	9	9	9	9	9	9	9	

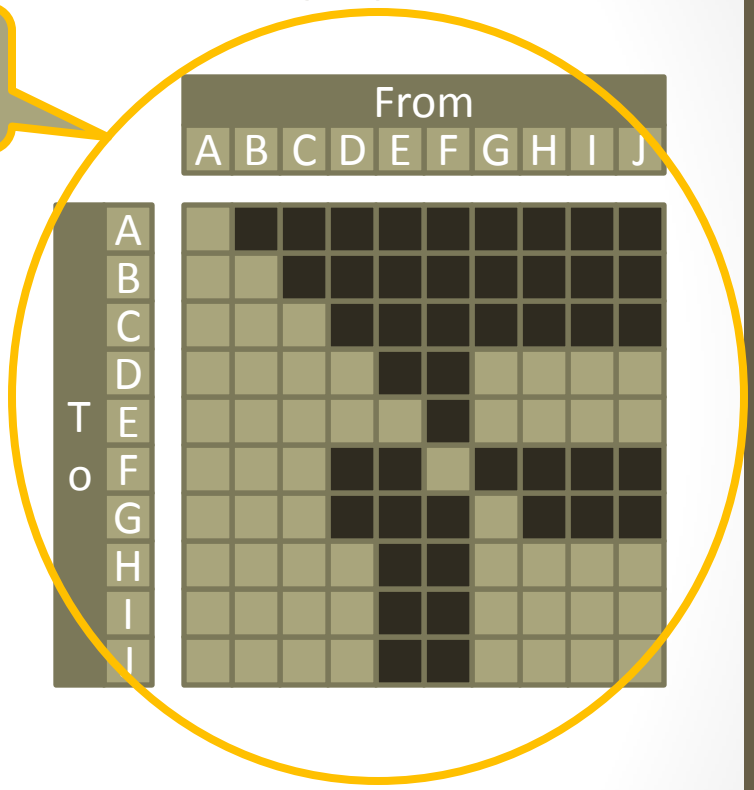
Graph Data: Allowed Paths, Distance, and Diameter (0)



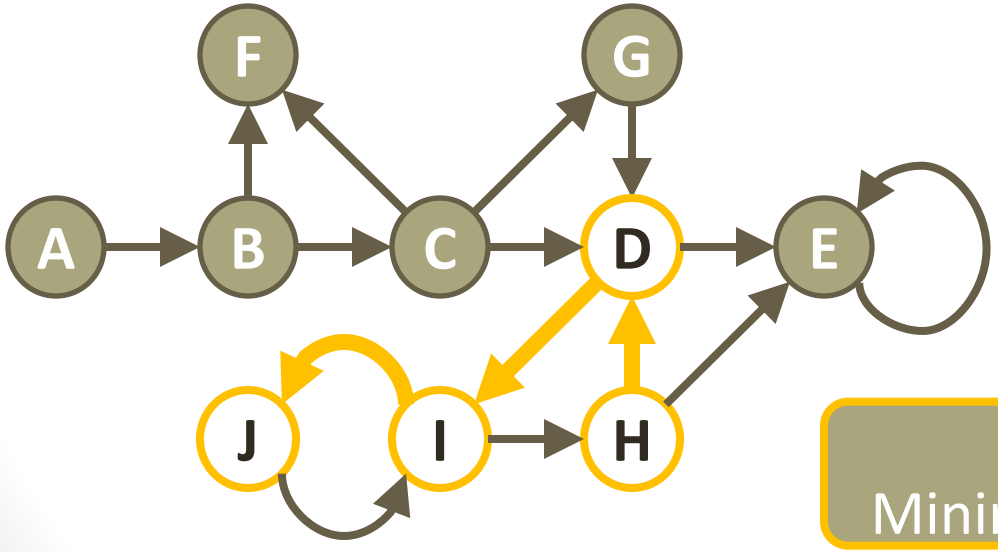
Graph Data: Allowed Paths, Distance, and Diameter (1)

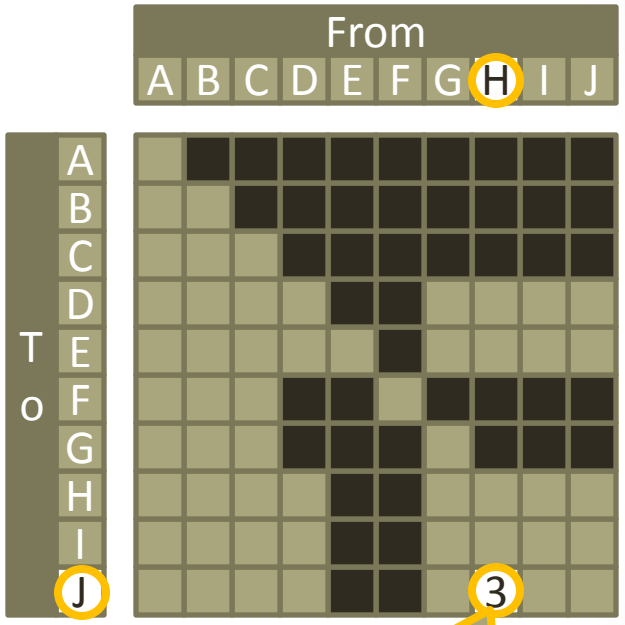


Allowed Paths



Graph Data: Allowed Paths, Distance, and Diameter (2)

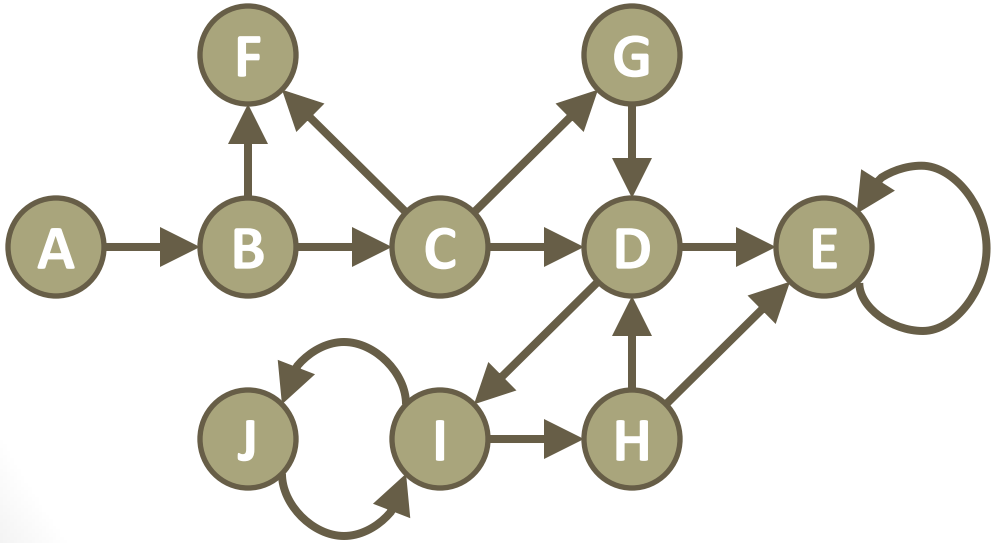




Path Distance:
Minimum Distance from H to J

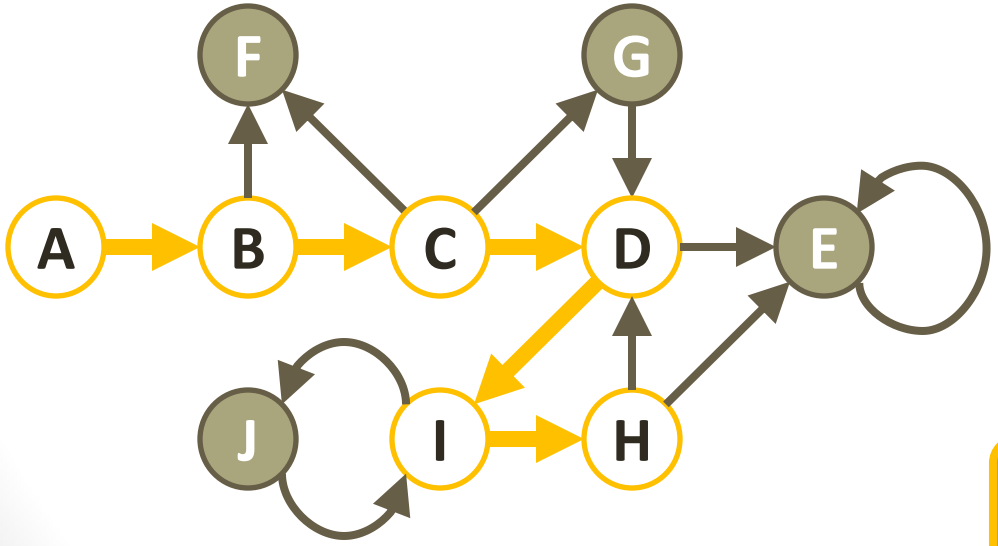
Graph Data: Allowed Paths, Distance, and Diameter (3)

Distance Matrix:
Minimum distance between any two vertices along a directed edge



To \ From	A	B	C	D	E	F	G	H	I	J
A	0									
B	1	0								
C	2	1	0							
D	3	2	1	0			1	1	2	3
E	4	3	2	1	0		2	1	2	3
F	2	1	1			0				
G	4	2	1				0			
H	5	4	3	2			3	0	1	2
I	4	3	2	1			2	2	0	1
J	5	4	3	2			3	3	1	0

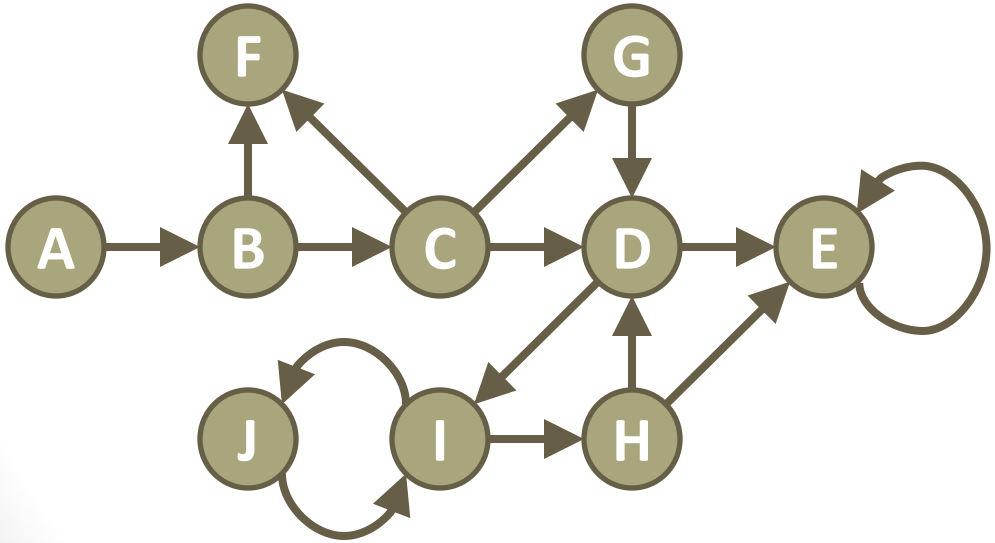
Graph Data: Allowed Paths, Distance, and Diameter (4)



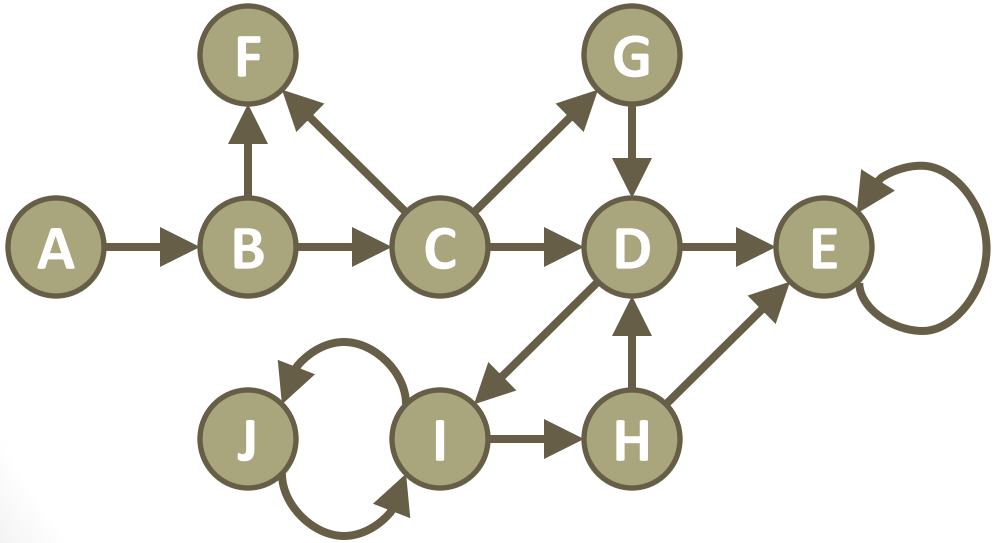
To \ From	A	B	C	D	E	F	G	H	I	J
A	0									
B	1	0								
C	2	1	0							
D	3	2	1	0			1	1	2	3
E	4	3	2	1	0		2	1	2	3
F	2	1	1			0				
G	4	2	1				0			
H	5	4	3	2			3	0	1	2
I	4	3	2	1			2	2	0	1
J	5	4	3	2			3	3	1	0

Diameter:
Max Path Distance

Graph Data: Order and Size (0)



Graph Data: Order and Size (1)

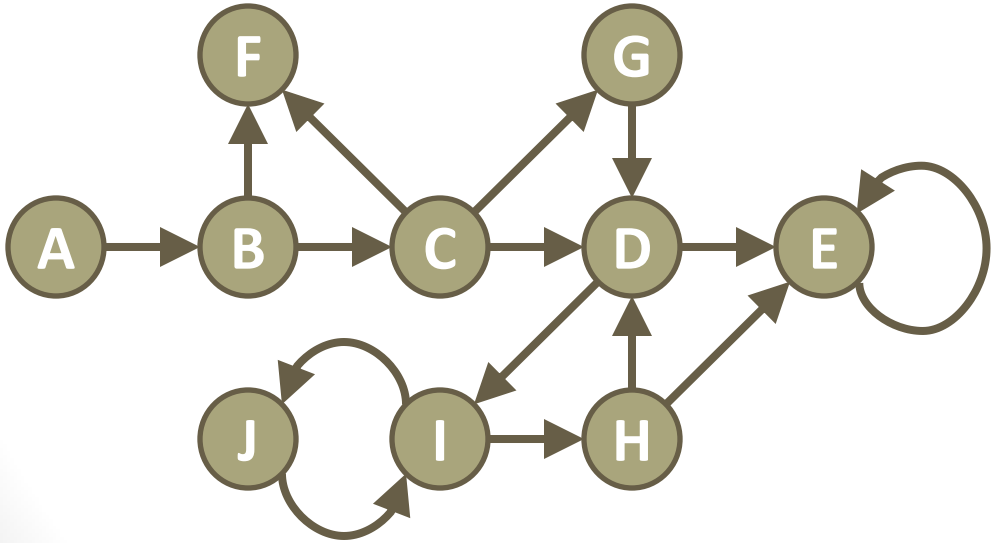


Matrix of Edges

To \ From	A	B	C	D	E	F	G	H	I	J
A										
B	1									
C		1								
D			1				1	1		
E				1	1			1		
F		1	1							
G			1							
H									1	
I				1						1
J									1	

Graph Data: Order and Size (2)

Order of Graph: 10

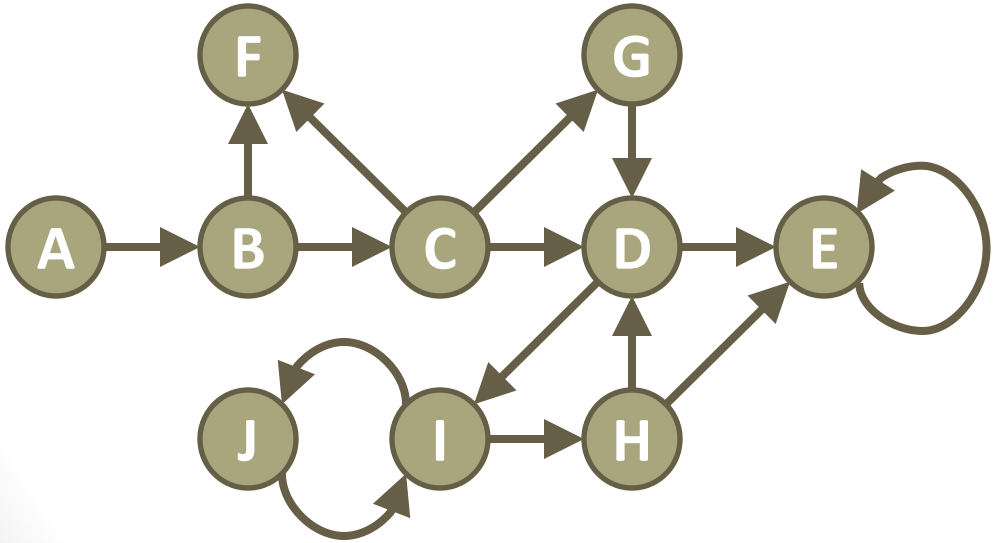


To \ From	A	B	C	D	E	F	G	H	I	J
A										
B	1									
C		1								
D			1				1	1		
E				1	1			1		
F		1	1							
G			1							
H									1	
I				1						1
J									1	

Order of Graph: Number of Nodes

Graph Data: Order and Size (3)

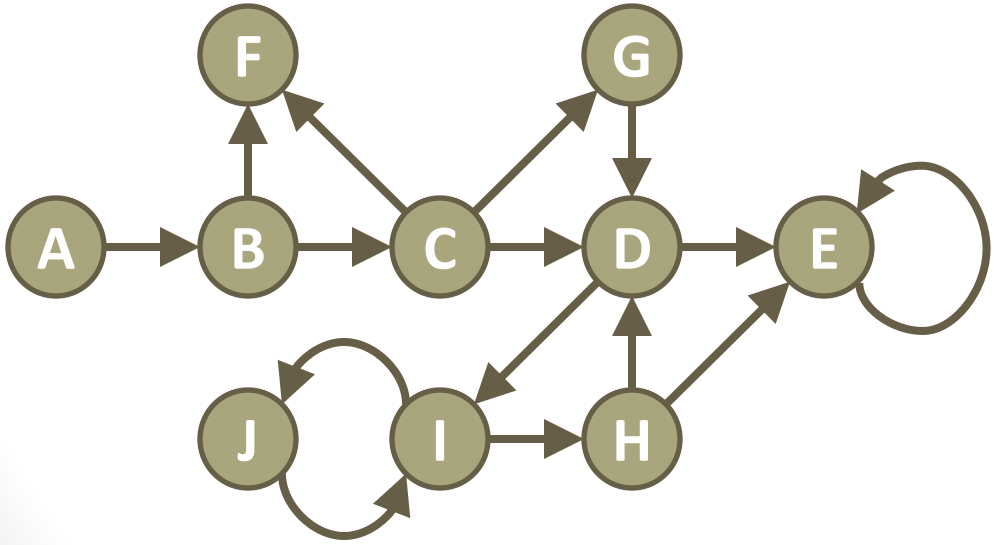
Order of Graph: 10
Size of Graph: 15



To \ From	A	B	C	D	E	F	G	H	I	J
A										
B	1									
C		1								
D			1				1	1		
E				1	1			1		
F		1	1							
G			1							
H									1	
I				1						1
J									1	

Size of Graph:
Number of Edges

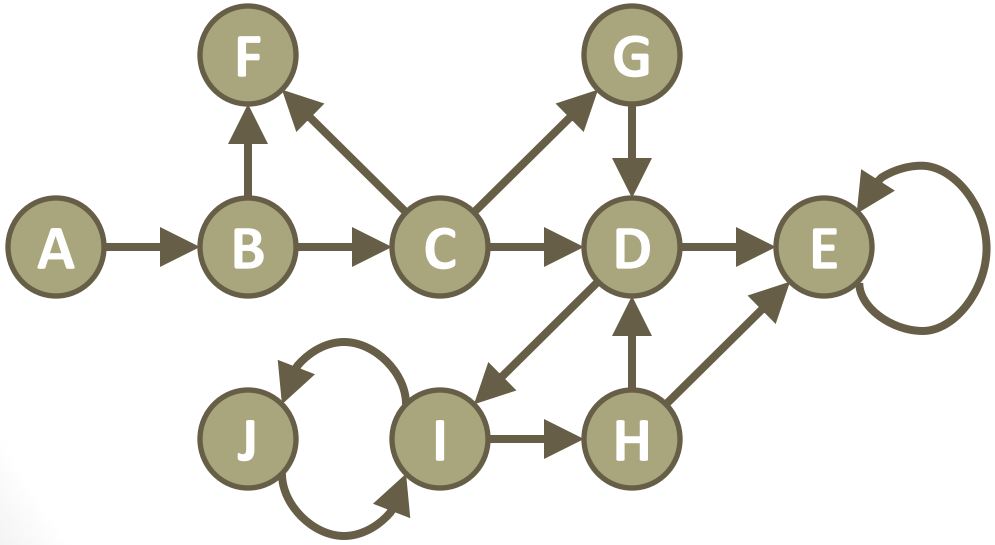
Graph Data: Popularity (0)



To \ From	A	B	C	D	E	F	G	H	I	J
A										
B	1									
C		1								
D			1				1	1		
E				1	1			1		
F		1	1							
G			1							
H									1	
I				1						1
J									1	

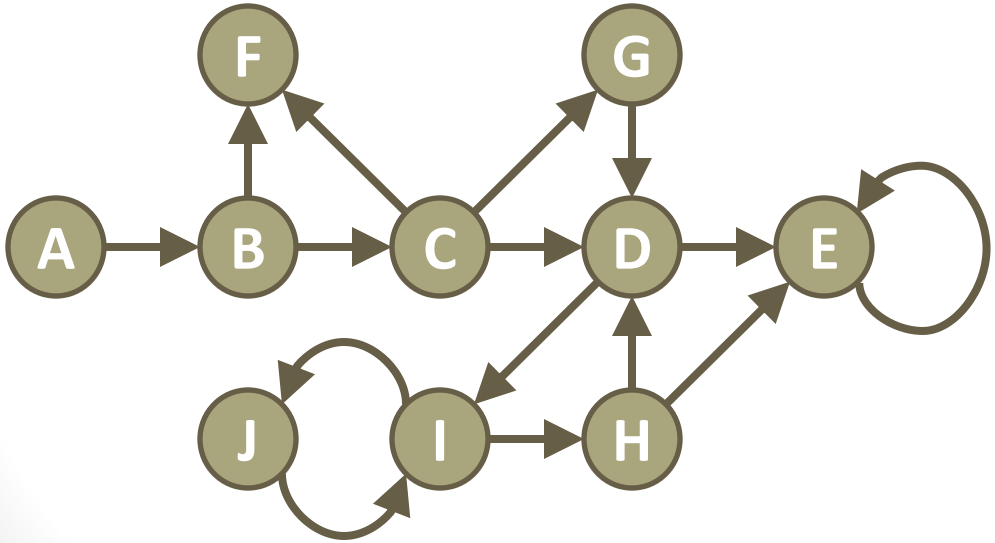
Graph Data: Popularity (1)

In-degree of a node: Popularity



To \ From	A	B	C	D	E	F	G	H	I	J	Sum:
A											0
B	1										1
C		1									1
D			1				1	1			3
E				1	1			1			3
F		1	1								2
G			1								1
H									1		1
I				1						1	2
J									1		1

Graph Data: Popularity (2)



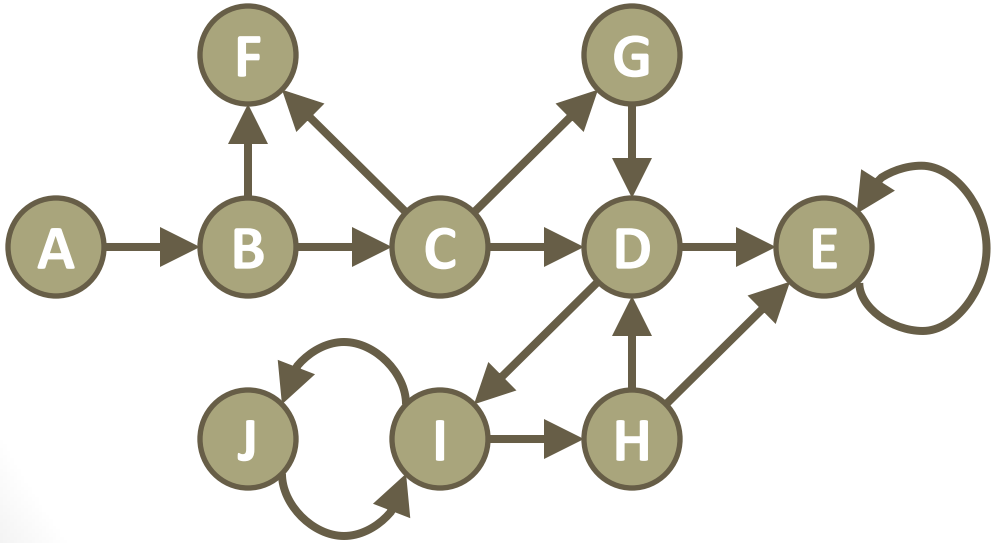
In-degree of a node: Popularity

To \ From	A	B	C	D	E	F	G	H	I	J	Sum:
A											0
B	1										1
C		1									1
D			1				1	1			3
E				1	1			1			3
F		1	1								2
G			1								1
H									1		1
I				1						1	2
J									1		1
Sum:	1	2	3	2	1	0	1	2	2	1	

Out-degree of a node: Gregariousness

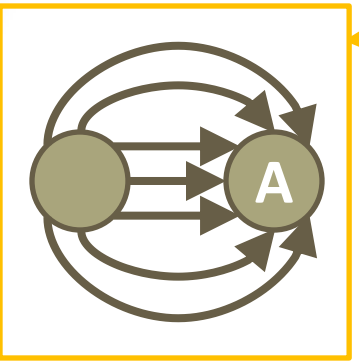
Graph Data: Popularity (3)

But, not every link is equally important. Otherwise, you could create a website with a thousand links that all point to your site.

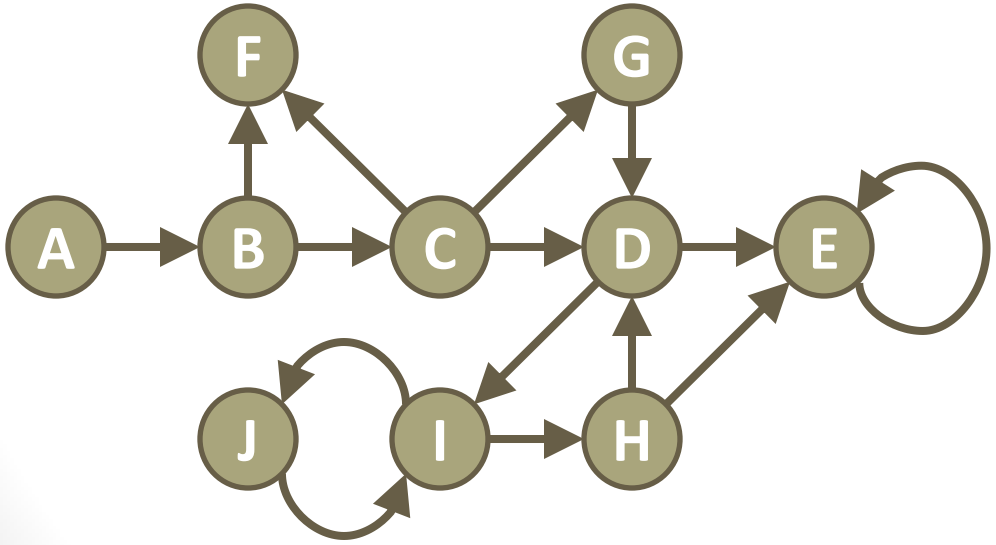


To \ From	A	B	C	D	E	F	G	H	I	J	Sum:
A											0
B	1										1
C		1									1
D			1				1	1			3
E				1	1			1			3
F		1	1								2
G			1								1
H									1		1
I				1						1	2
J									1		1
Sum:	1	2	3	2	1	0	1	2	2	1	

Graph Data: Popularity (4)



One very gregarious node could make another node very popular



To \ From	A	B	C	D	E	F	G	H	I	J	Sum:
A											0
B	1										1
C		1									1
D			1				1	1			3
E				1	1			1			3
F		1	1								2
G			1								1
H									1		1
I				1						1	2
J									1		1
Sum:	1	2	3	2	1	0	1	2	2	1	

Graph Data: Links (0)

- Graph Analytics:
  - http://en.wikipedia.org/wiki/Graph_theory
- Graph Diameter
  - http://people.hofstra.edu/geotrans/eng/methods/diameter1.html
- Distance
  - http://en.wikipedia.org/wiki/Distance_(graph_theory)
- Resistance
  - http://en.wikipedia.org/wiki/Resistance_distance
- Centrality
  - http://en.wikipedia.org/wiki/Betweenness#Betweenness_centrality

Graph Data: Links (1)

- http://en.wikipedia.org/wiki/Graph_(mathematics)
- http://en.wikipedia.org/wiki/Connectivity_(graph_theory)

Introduction to Graph Data