Introducing

Graph processing

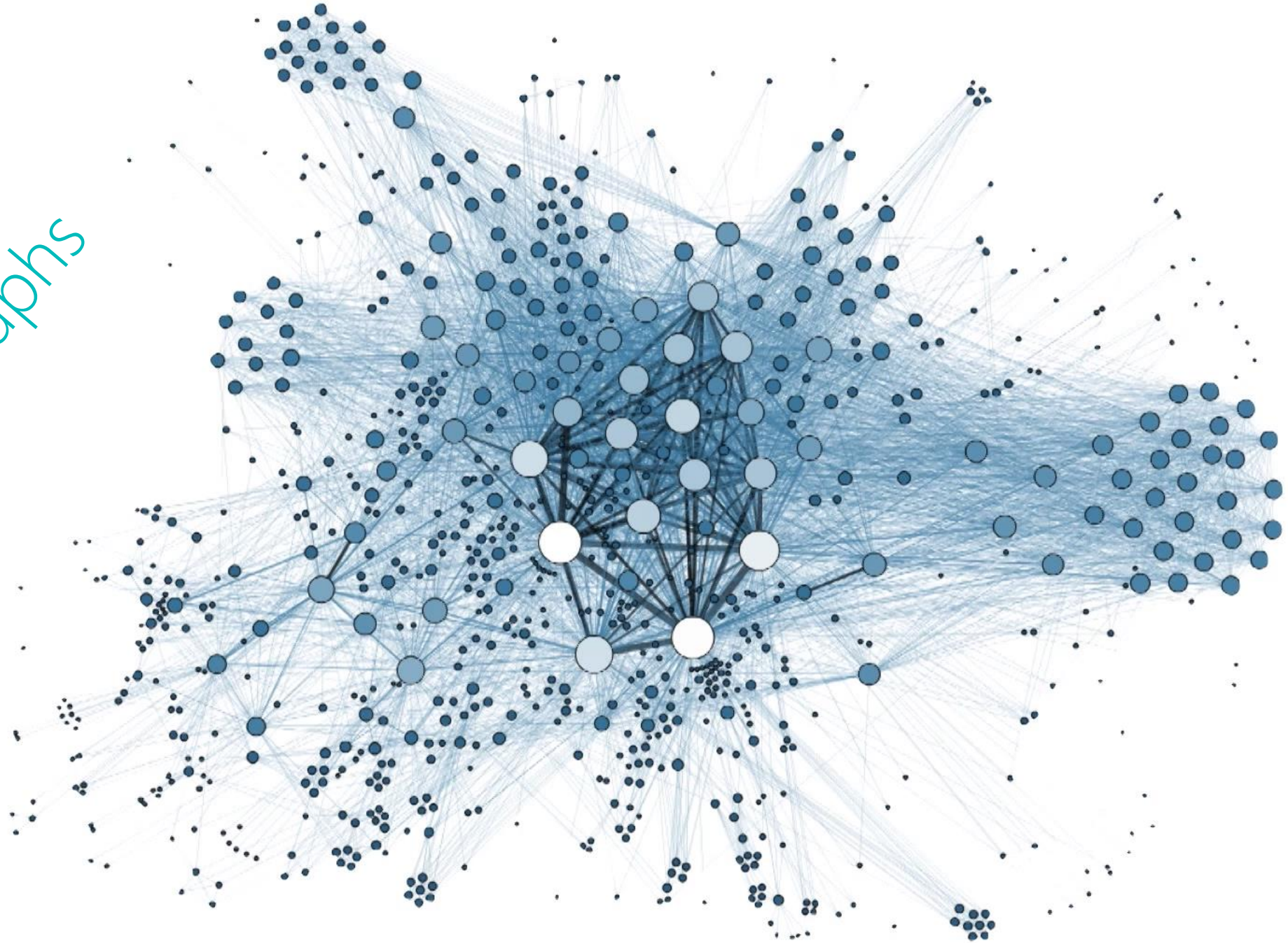
Author:

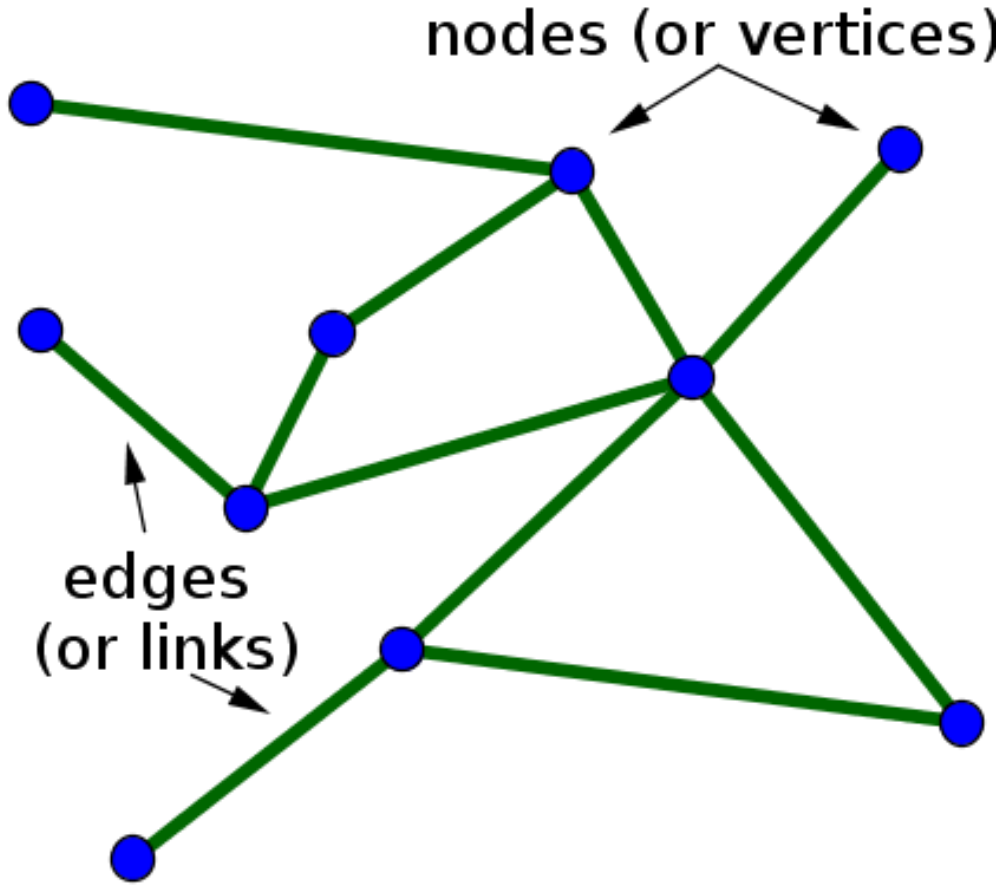
AmirHossein Mohammadi

MACHINE LEARNING

BIG DATA

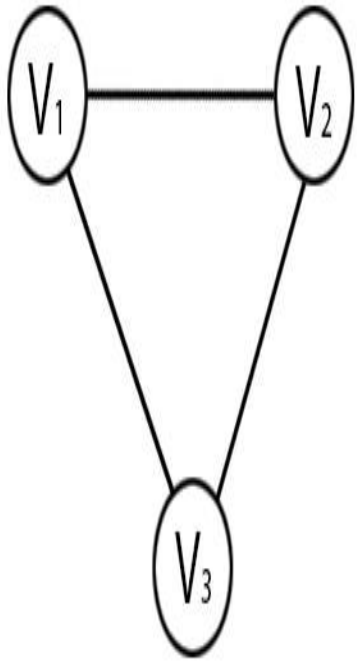
Graphs



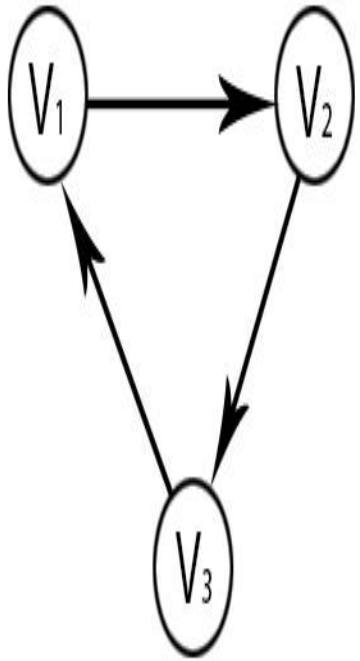


Nodes and Edges

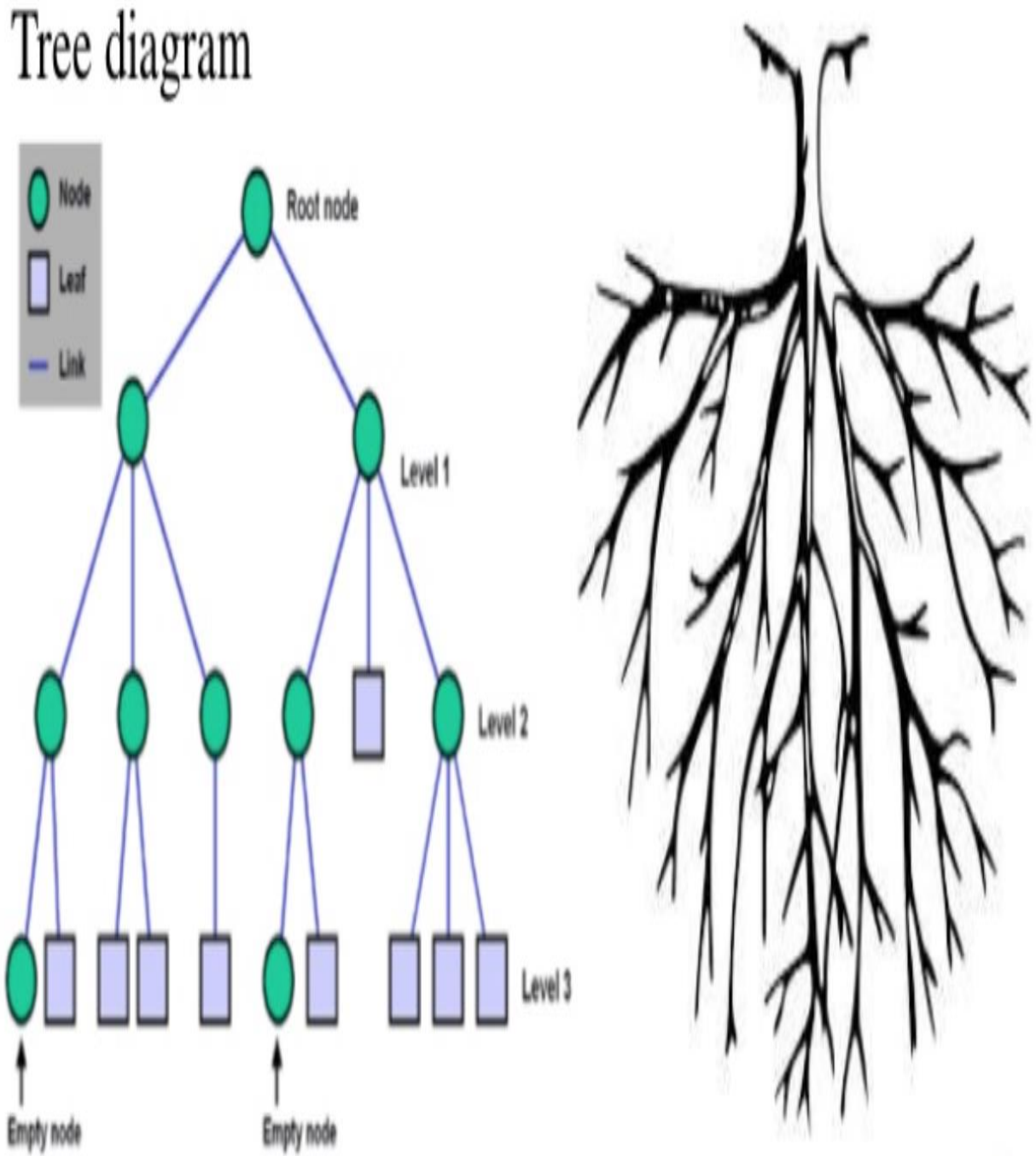
Undirected Graph



Directed Graph



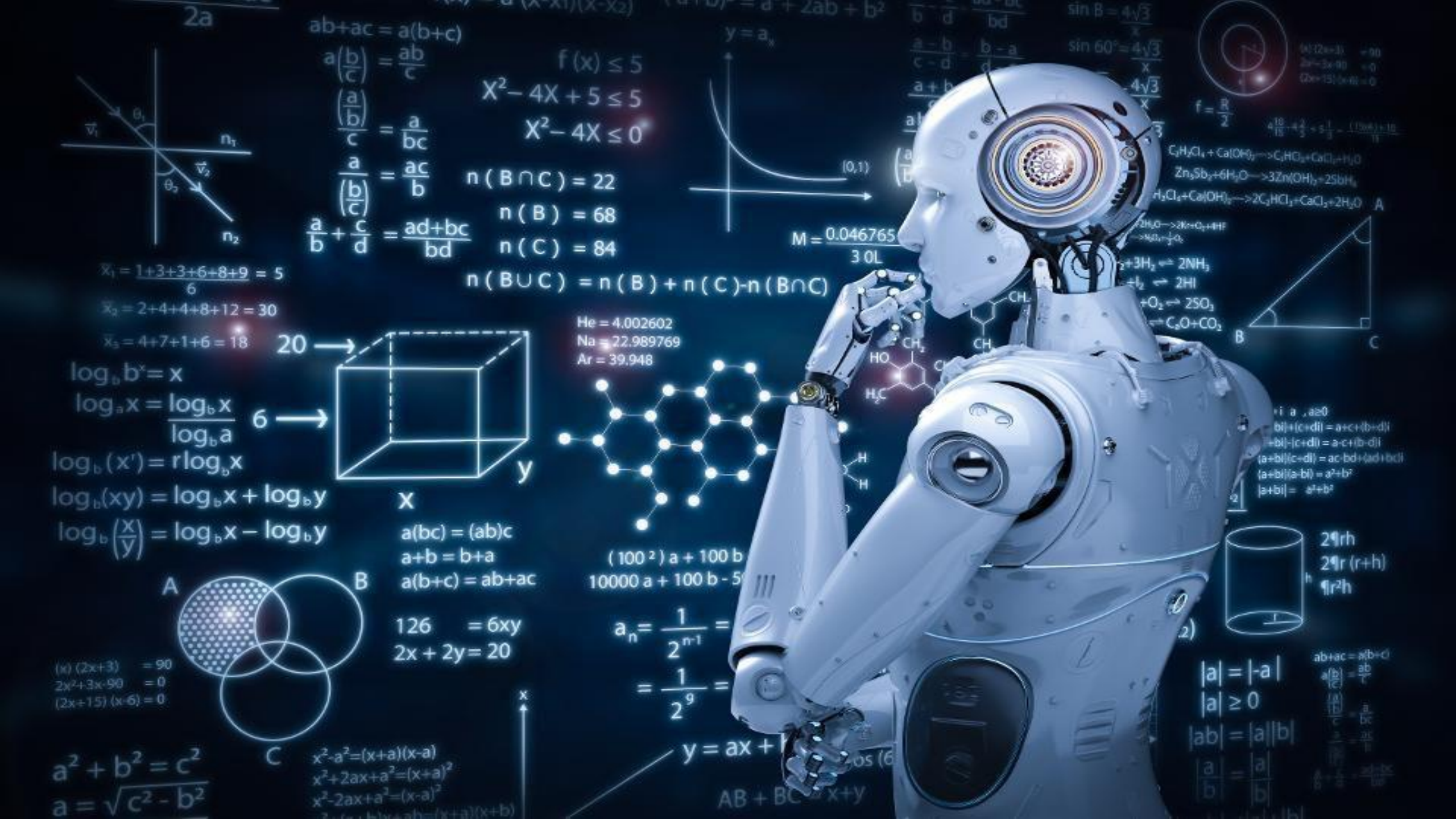
جهت دار و بدون جهت



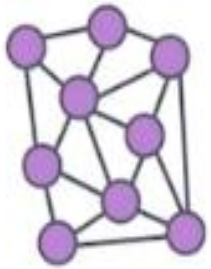
Tree

In **graph** theory, a **tree** is an undirected, connected and acyclic **graph**. In other words, a connected **graph** that does not contain even a single cycle is called a **tree**. A **tree** represents hierarchical structure in a graphical form. The elements of **trees** are called their nodes and the edges of the **tree** are called branches.

گراف یکی از بهترین ابزارها برای
مدل کردن مسائل دشوار و سخت

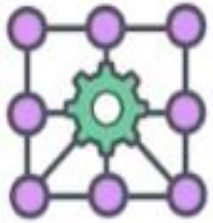

ab+ac = a(b+c)
f(x) ≤ 5
X² − 4X + 5 ≤ 5
X² − 4X ≤ 0
n(B∩C) = 22
n(B) = 68
n(C) = 84
n(B∪C) = n(B) + n(C) − n(B∩C)
20
6
y
x
He = 4.002602
Na = 22.989769
Ar = 39.948
log_b b^x = x
log_a x = log_b x / log_b a
log_b (x^r) = r log_b x
log_b (xy) = log_b x + log_b y
log_b (x/y) = log_b x − log_b y
a(bc) = (ab)c
a+b = b+a
a(b+c) = ab+ac
126 = 6xy
2x + 2y = 20
A
B
C
a² + b² = c²
a = √(c² − b²)
2¶rh
2¶r (r+h)
¶r²h
|a| = |−a|
|a| ≥ 0
|ab| = |a||b|

Graphs Enhance AI by Providing Context



Connected Feature Extraction

Context for Accuracy

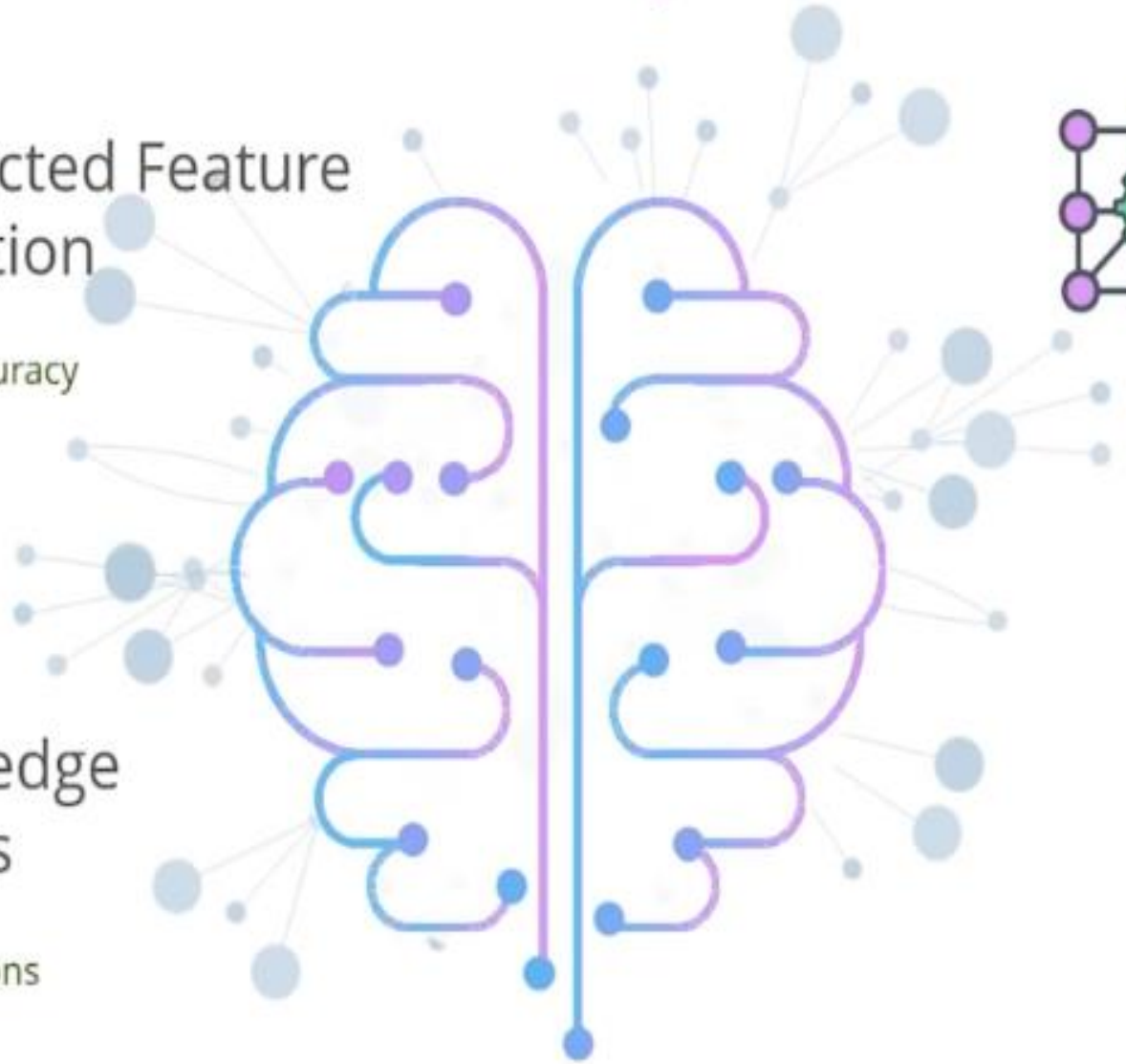


Graph Accelerated ML

Context for Efficiency

Knowledge Graphs

Context for Decisions



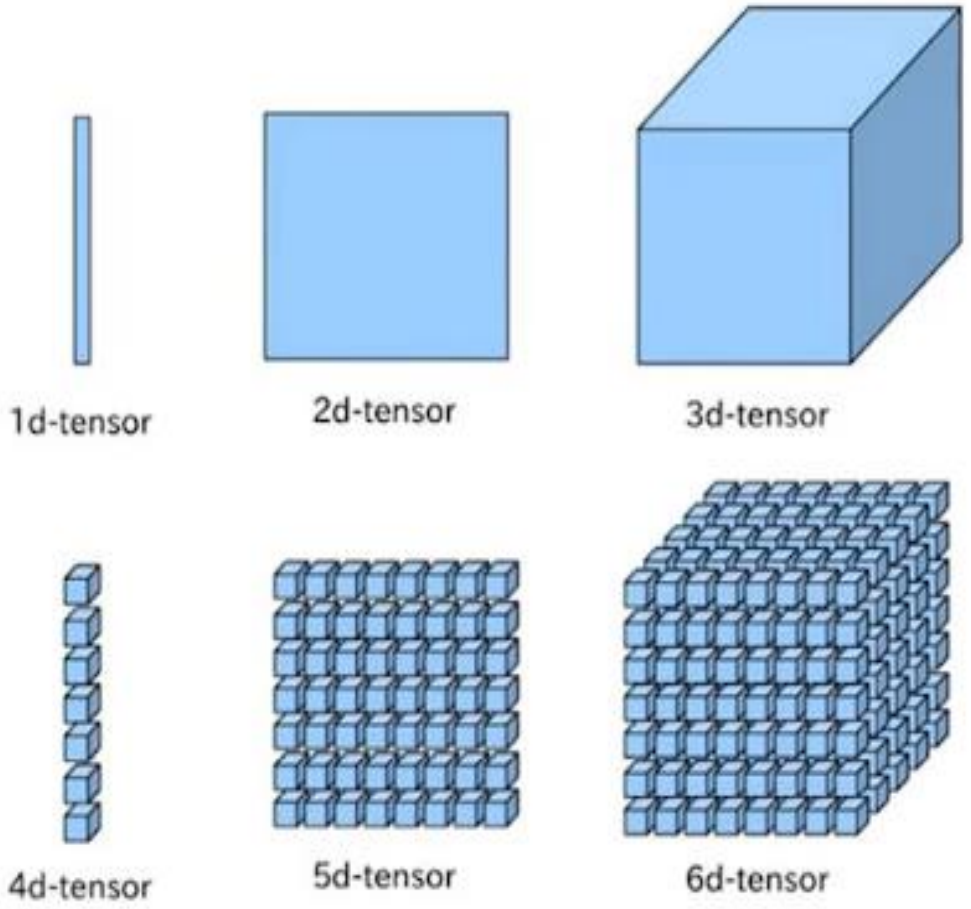
Graphs are Context for Accuracy

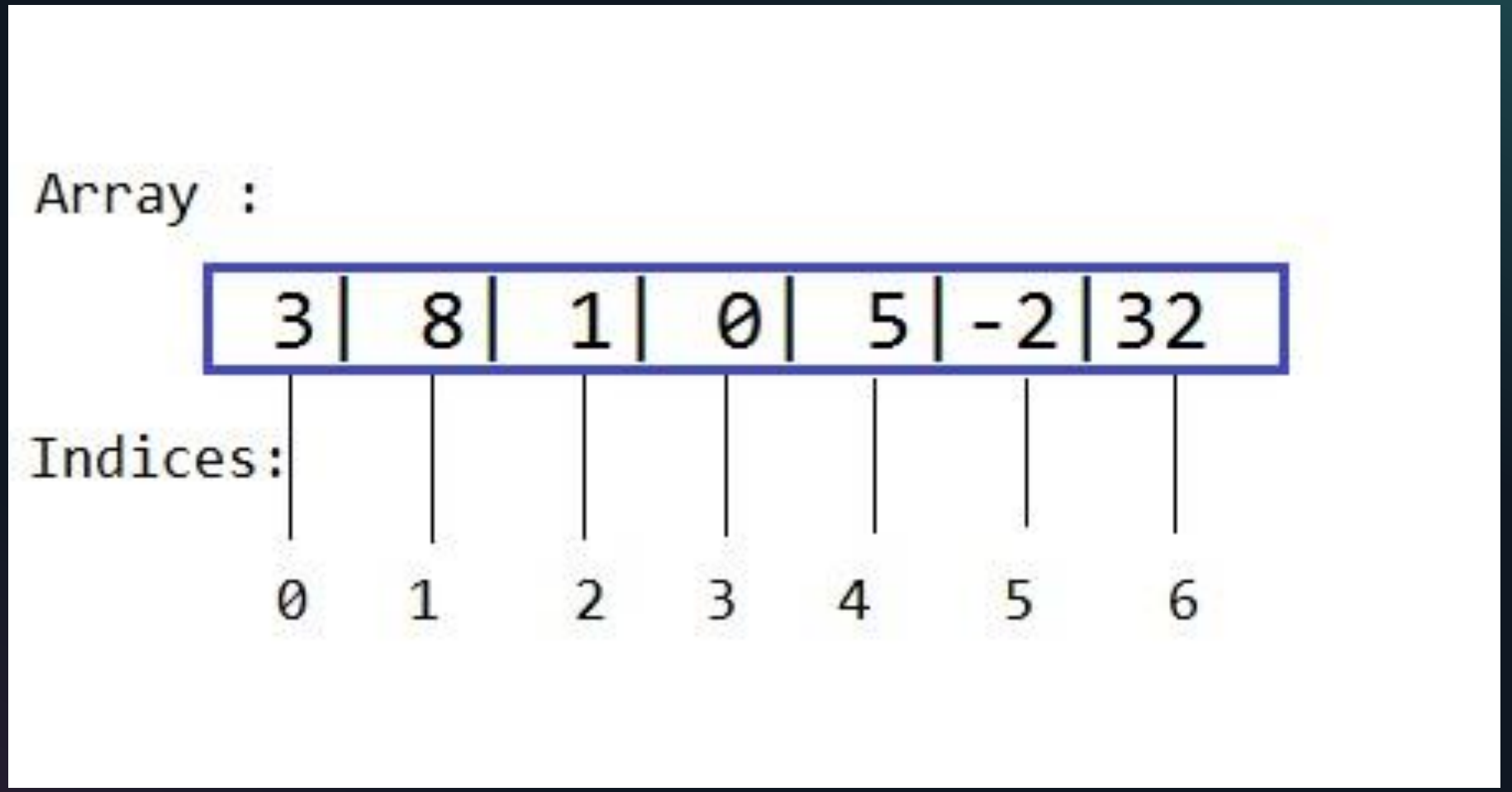
Connected Feature Extraction

Connected Feature Extraction

Sometimes it's who you know

- Relationships are often the strongest predictors of behavior
- Current machine learning methods rely on vectors, matrices, and tensors built from tables
- These methods simplify, or leave out entirely, predictive relationship and network data
- Graphs add highly predictive features to these models, adding accuracy without altering algorithms
- Graphs can infer relationships and add data where sparse

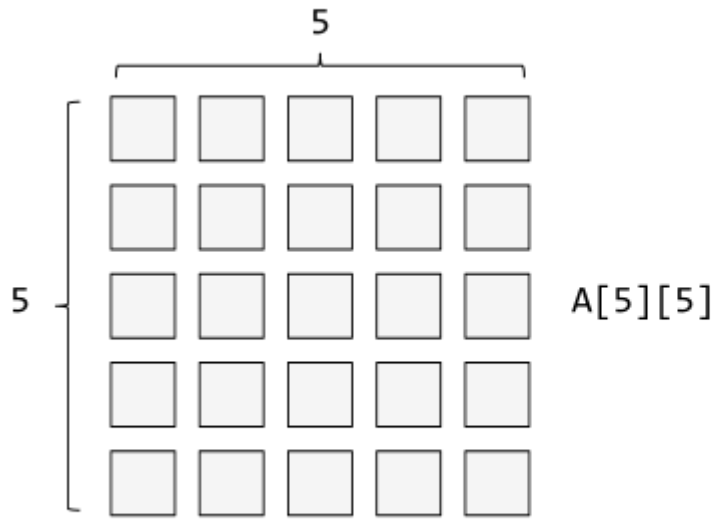




برای پیاده سازی این مفاهیم پیوسته با داده ها سر و کار داشتم

سوال اصلی اینه که آیا میتونستیم ما با این ساختار های داده ای روابط بین داده ها رو برای گرفتن تصمیمات استخراج کنیم؟

SQUARE MATRIX



Any two dimensional array can be interpreted as a matrix. An array A[n][n] is also a NxN matrix!

One degree of relationship

Graph vs array

عده ای اعتقاد دارند که گراف ها بیشتر برای شبکه های اجتمایی است که اگر اینطوری فکر نید که این به این معنی است که فقط بیگ دیتا در شبکه های اجتمایی هست

Graphs are Context for Efficiency

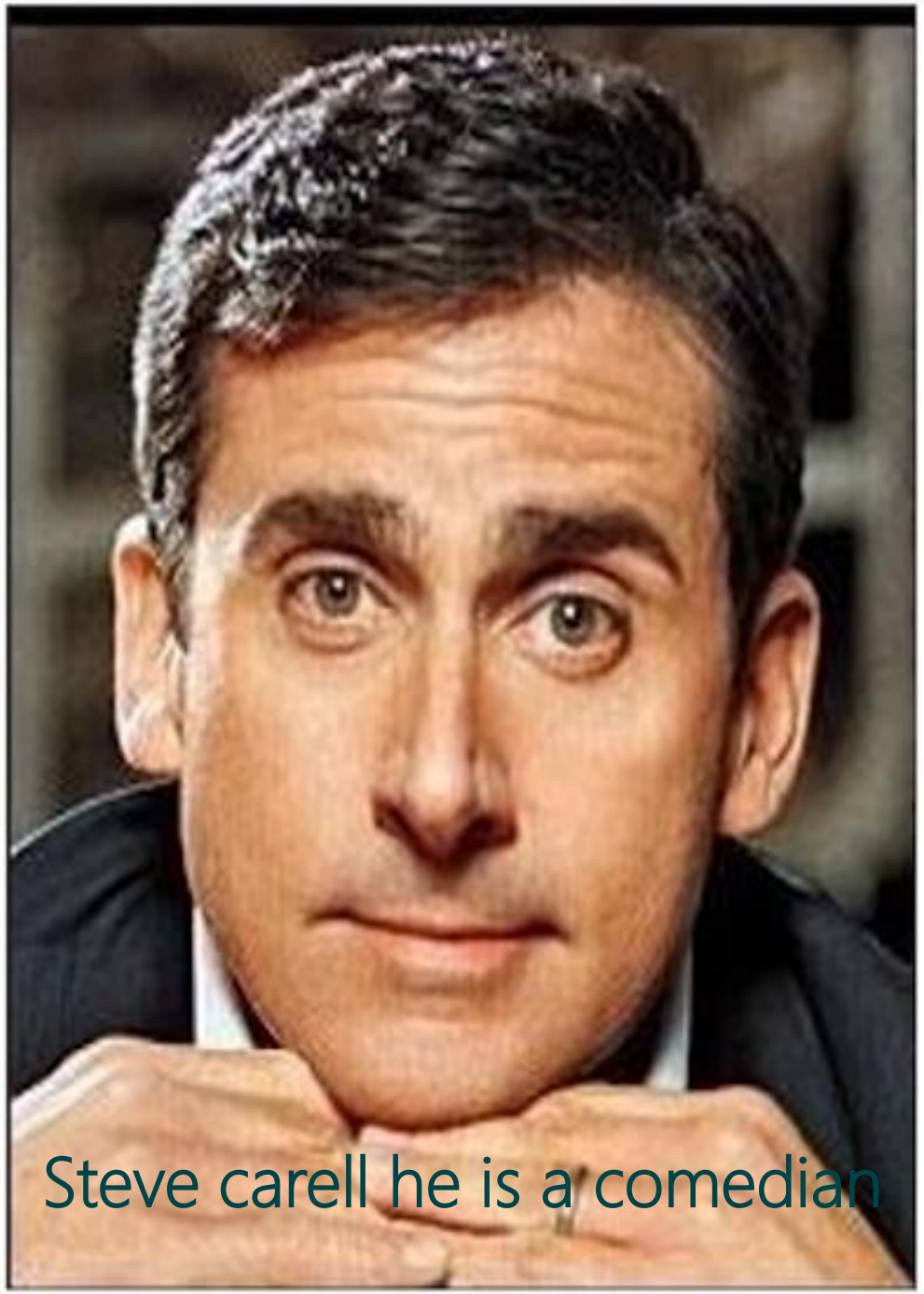
Graph Accelerated ML

Learinng....

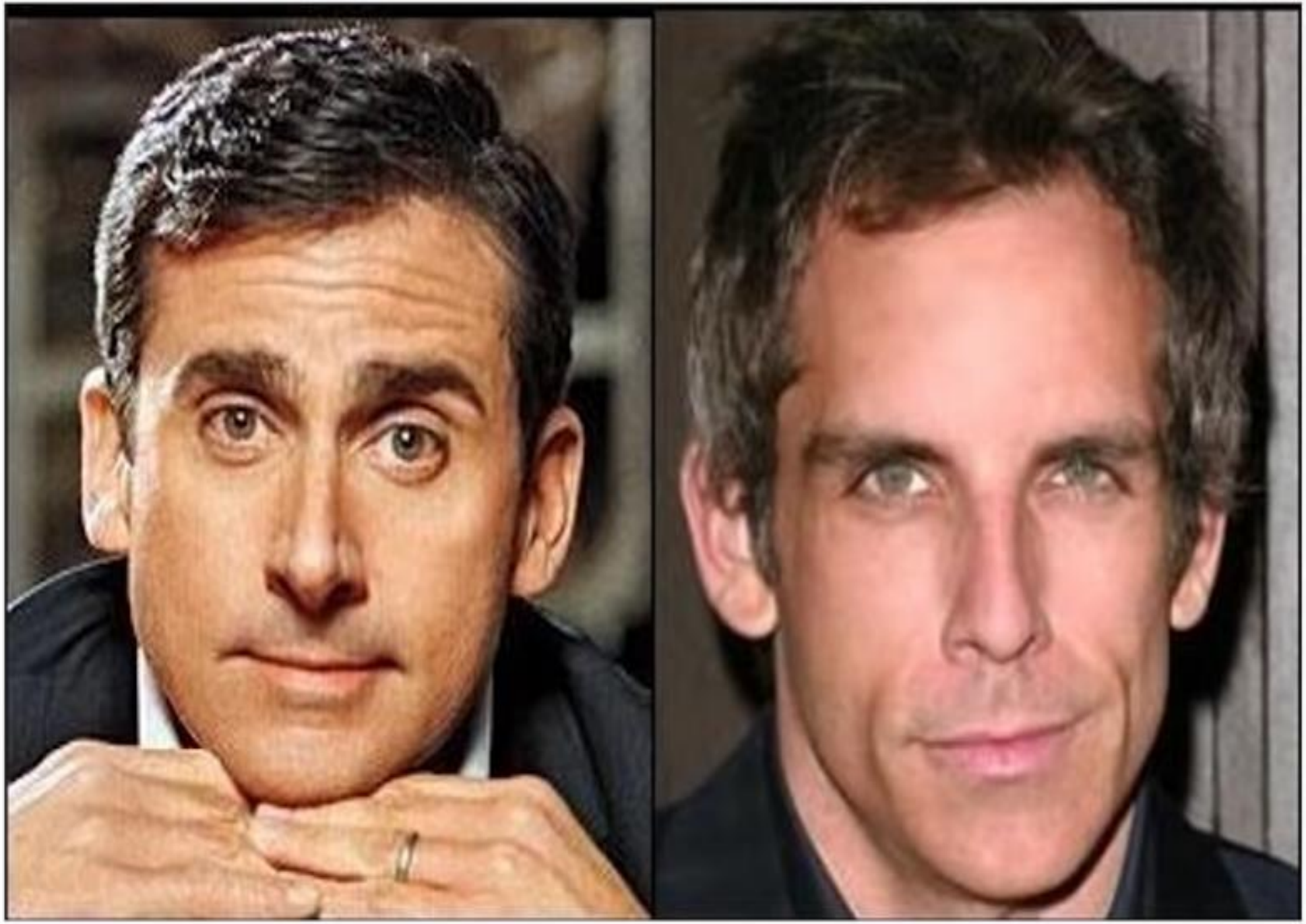
- بدست آوردن دانش و یا فهم از طریق مطالعه ،آموزش و یا تجربه
- بهبود عملکرد از طریق تجربه

Origin	Manufacturer	Color	Decade	Type	Example Type
Japan	Honda	Blue	1980	Economy	Positive
Japan	Toyota	Green	1970	Sports	Negative
Japan	Toyota	Blue	1990	Economy	Positive
USA	Chrysler	Red	1980	Economy	Negative
Japan	Honda	White	1980	Economy	Positive
Japan	Toyota	Green	1980	Economy	Positive
Japan	Honda	Red	1990	Economy	Negative


HAVE
FUN



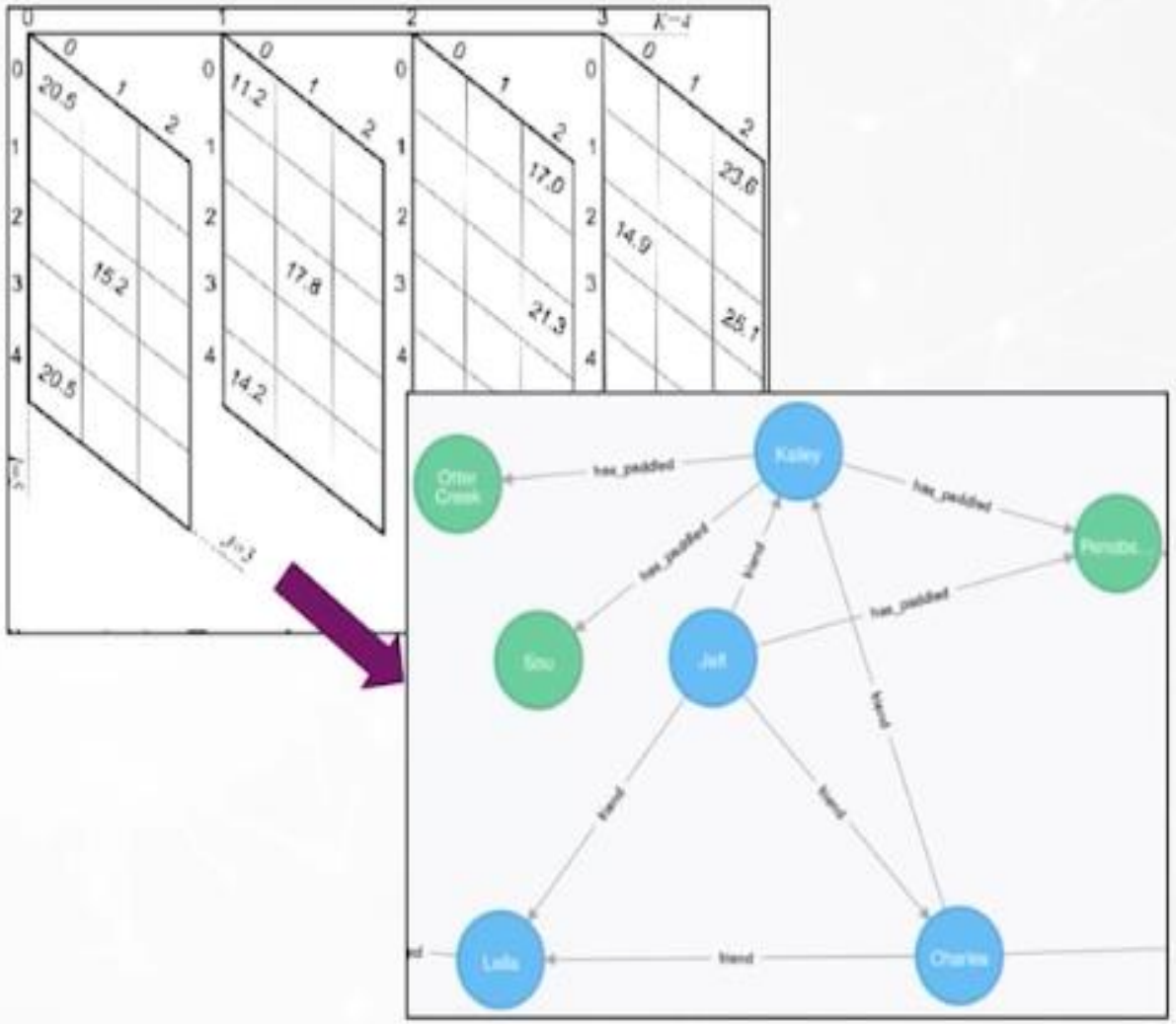
Steve carell he is a comedian



Model Optimization

Brute force is as inelegant as it sounds

- 56% of enterprise CIOs say iterative model training is the largest ML challenge[1]
- Renting more and more GPU time is not the answer – not every problem is "embarrassingly parallel"
- Table joins bog down data pipelines
- Sparse matrix compression methods are inefficient (more tables!)



Accelerate Your ML Process

Methods:

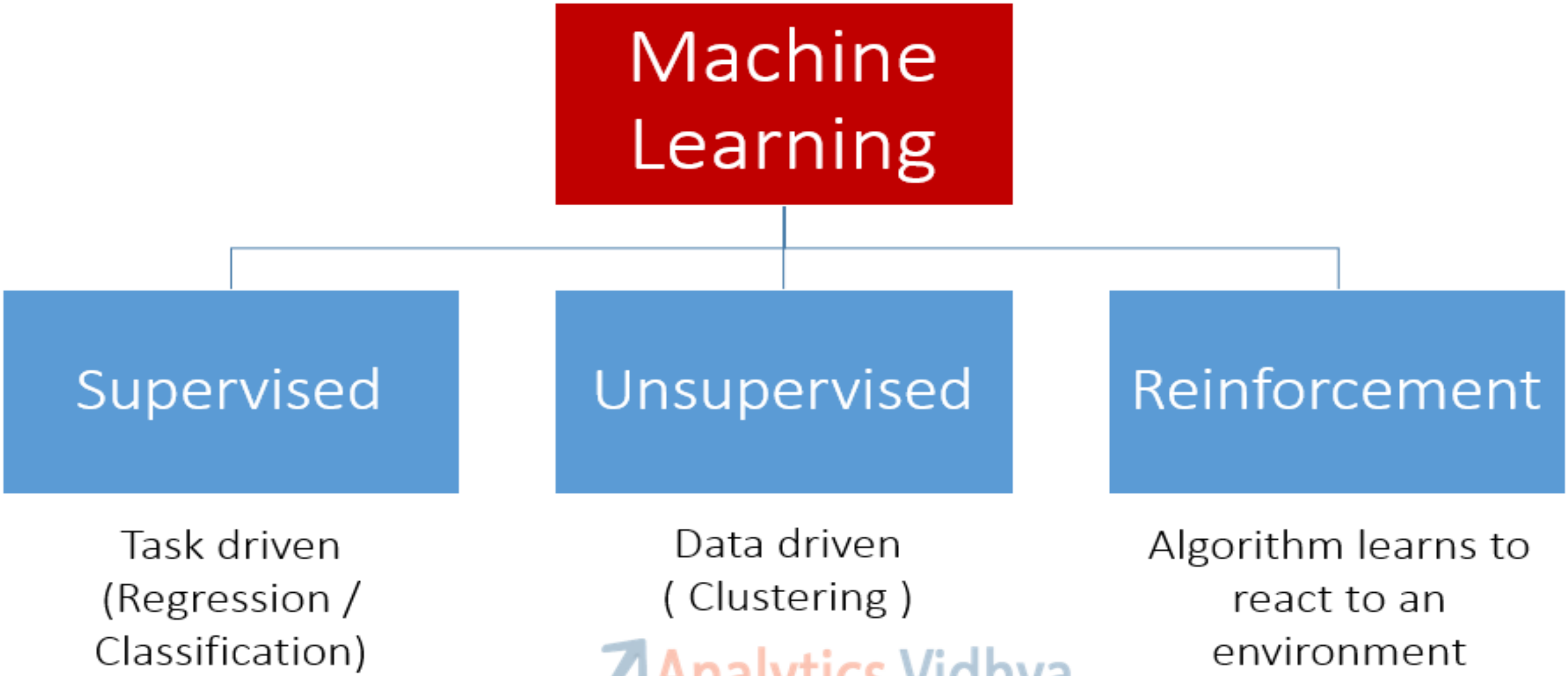
- Replace table joins with graph queries
- Replace sparse matrices and directional relationships with more efficient graph structures (i.e. collaborative filtering via Cypher query vs. matrix factorization)
- Use subgraph filtering to accelerate ML pipelines (Cypher queries, collaborative filtering, community detection, clustering, etc.)

Example: Recommendations

- Real time recommendations
- Customer Segmentation/KYC
- Churn analysis
- Dynamic pricing
- Promotions
- Patient modeling

Where are the graphs of machine learning?

Types of Machine Learning



Classification

Classification
/
رده بندی

وقتی که تابع هدف،به صورت گسسته باشید. روش هایی مانند درخت تصمیم از این نوع می باشد

INPUT

OUTPUT

Sky	Temp	Humid	Wind	Water	Forecast	C(x)
sunny	warm	normal	strong	warm	same	1
sunny	warm	high	strong	warm	same	1
rainy	cold	high	strong	warm	change	0
sunny	warm	high	strong	cool	change	1

Iris setosa

Iris versicolor

Iris virginica

petal length, petal width, sepal length, sepal width


Programming Time

GraphConnect 2017 → GraphConnect 2018

Pathfinding & Search

- Parallel Breadth First Search & DFS
- Shortest Path
- Single-Source Shortest Path
- All Pairs Shortest Path
- Minimum Spanning Tree

- A* Shortest Path
- Yen's K Shortest Path
- K-Spanning Tree (MST)

Centrality / Importance

- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- PageRank

- Harmonic Closeness Centrality
- Dangalchev Closeness Centrality
- Wasserman & Faust Closeness Centrality
- Approximate Betweenness Centrality
- Personalized PageRank

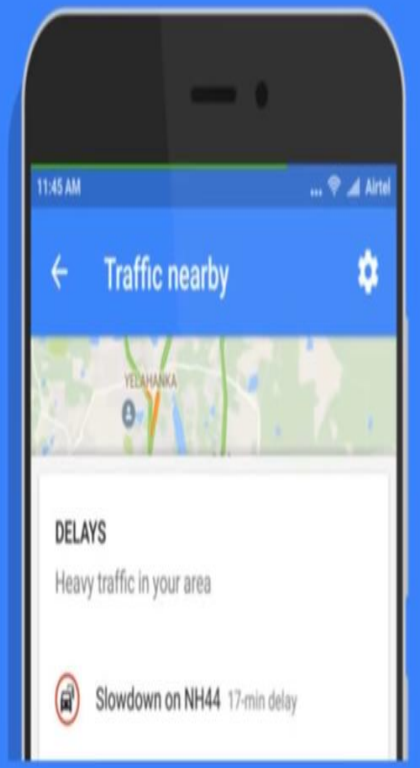
Community Detection

- Triangle Count
- Clustering Coefficients
- Connected Components (Union Find)
- Strongly Connected Components
- Label Propagation

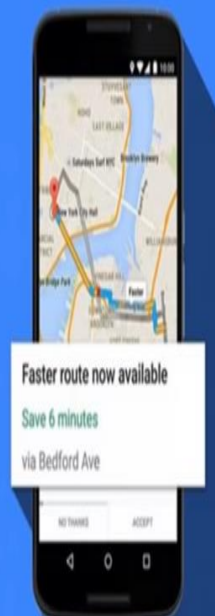
- Balanced Triad (identification)
- Louvain – Multi-Step


THE WORLD
POLITICAL
RUSSIA
CANADA
U.S.A.
UNITED STATES
MEXICO
BRAZIL
ARGENTINA
CHILE
PERU
BOLIVIA
PARAGUAY
URUGUAY
VENEZUELA
COLOMBIA
ECUADOR
CHINA
INDIA
KAZAKHSTAN
MONGOLIA
IRAN
AFGHANISTAN
PAKISTAN
SAUDI ARABIA
YEMEN
OMAN
IRAQ
TURKEY
SYRIA
JAPAN
INDONESIA
PHILIPPINES
MALAYSIA
THAILAND
VIETNAM
AUSTRALIA
NEW ZEALAND
PAPUA NEW GUINEA
FEDERATED STATES OF MICRONESIA
MARSHALL ISLANDS
PALAU
KIRIBATI
TUVALU
FIJI
SOLOMON ISLANDS
VANUATU
NORWAY
SWEDEN
FINLAND
ICELAND
GREENLAND
(KALAALLIT NUNAAT)
UNITED KINGDOM
IRELAND
FRANCE
SPAIN
PORTUGAL
GERMANY
POLAND
ITALY
GREECE
UKRAINE
BELARUS
ROMANIA
HUNGARY
DENMARK
NETHERLANDS
BELGIUM
SWITZERLAND
AUSTRIA
CZECH REPUBLIC
SLOVAKIA
SLOVENIA
CROATIA
BOSNIA AND HERZEGOVINA
SERBIA
BULGARIA
ESTONIA
LATVIA
LITHUANIA
MOLDOVA
GEORGIA
ARMENIA
AZERBAIJAN
UZBEKISTAN
TURKMENISTAN
KYRGYZSTAN
TAJIKISTAN
NORTH KOREA
SOUTH KOREA
TAIWAN
SRI LANKA
MYANMAR (BURMA)
LAOS
CAMBODIA
BRUNEI
NEPAL
BHUTAN
BANGLADESH
MALDIVES
ALGERIA
LIBYA
EGYPT
MOROCCO
TUNISIA
MAURITANIA
MALI
NIGER
CHAD
SUDAN
NIGERIA
ETHIOPIA
SOMALIA
KENYA
UGANDA
TANZANIA
DEMOCRATIC REPUBLIC OF THE CONGO
CENTRAL AFRICAN REPUBLIC
CAMEROON
GABON
CONGO
ANGOLA
ZAMBIA
ZIMBABWE
MOZAMBIQUE
MADAGASCAR
NAMIBIA
BOTSWANA
SOUTH AFRICA
LESOTHO
SWAZILAND
MAURITIUS
SEYCHELLES
COMOROS
NORTH ATLANTIC OCEAN
PACIFIC OCEAN
SOUTH ATLANTIC OCEAN
INDIAN OCEAN
SOUTHERN OCEAN
ARCTIC OCEAN
ANTARCTICA
BERING SEA
SEA OF OKHOTSK
SEA OF JAPAN (EAST SEA)
YELLOW SEA
EAST SEA
SOUTH CHINA SEA
PHILIPPINE SEA
CORAL SEA
TASMAN SEA
ARABIAN SEA
BAY OF BENGAL
CARIBBEAN SEA
MEDITERRANEAN SEA
BLACK SEA
CASPIAN SEA
RED SEA
GULF OF MEXICO
GULF OF GUINEA
BARENTS SEA
KARA SEA
LAPTEV SEA
GREENLAND SEA
NORWEGIAN SEA
BAFFIN BAY
LABRADOR SEA
HUDSON BAY
BEAUFORT SEA
CHUKCHI SEA
EAST SIBERIAN SEA
WEDDELL SEA
SARGASSO SEA
GOBI DESERT
SAHARA
ARABIAN DESERT
GREAT SANDY DESERT
GREAT VICTORIA DESERT
GIBSON DESERT
SIBERIA
VAN DER GRINTEN PROJECTION
EQUATORIAL SCALE 1: 20 000 000
0 400 800 1200 1600 km
0 200 400 600 800 1000 miles
Mumbai over 20m
Istanbul 10-20m
Toronto 5-10m
Birmingham 1-5m
Vigo 250,000 - 1,000,000
Cork 100,000 - 250,000
Torino Less than 100,000
National capitals are shown in CAPS.
Administered by or belonging to ( )
International boundary
State boundary
Disputed/undefined boundary
Major railway
Highest peak in continent
Highest peak in country
Other peak
Depth figure
International Date Line
mapsinternational.com

Google Maps



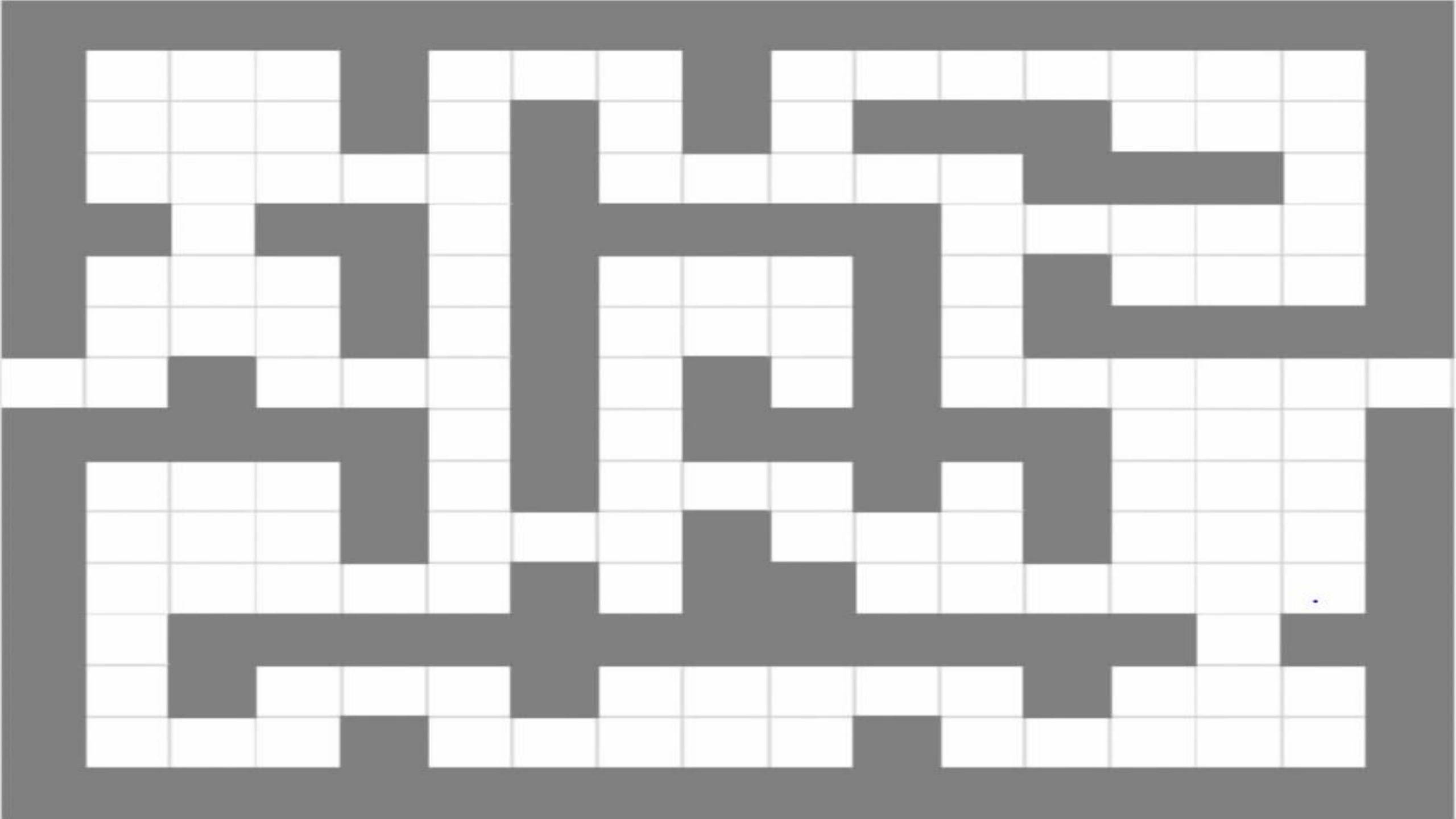
1

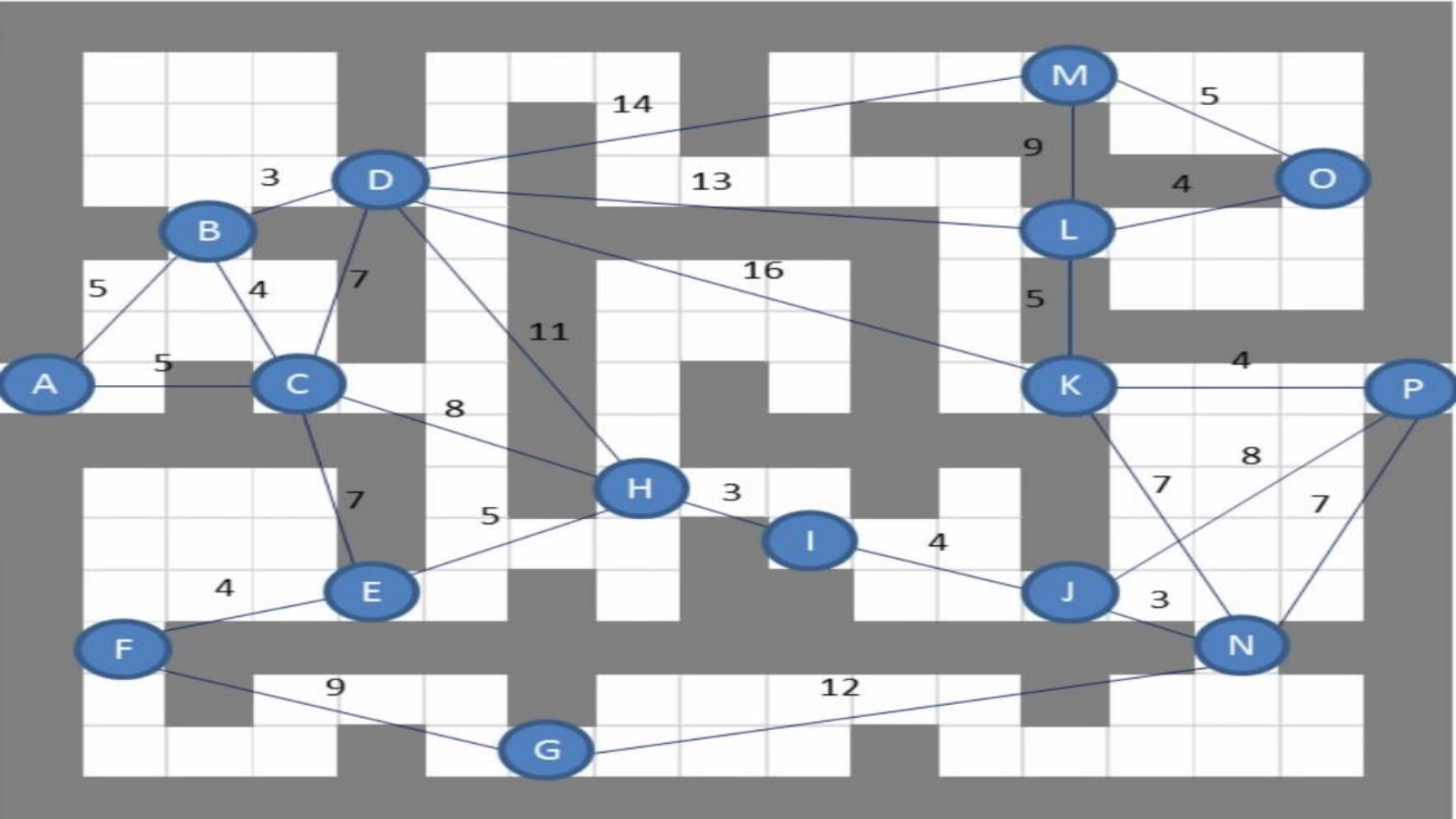


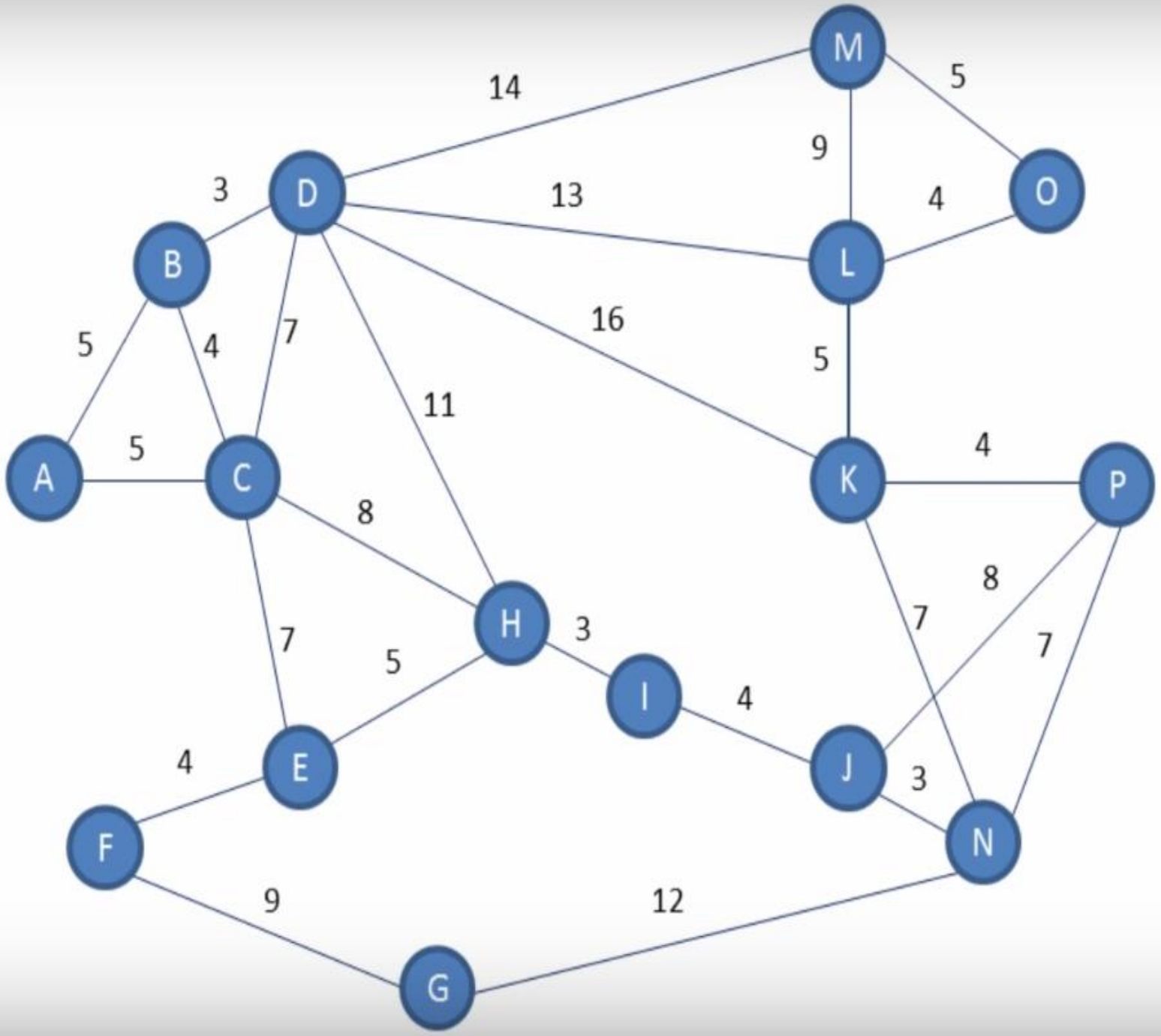
SUBSCRIBE

Google Maps

A* Pathfinding Algorithm

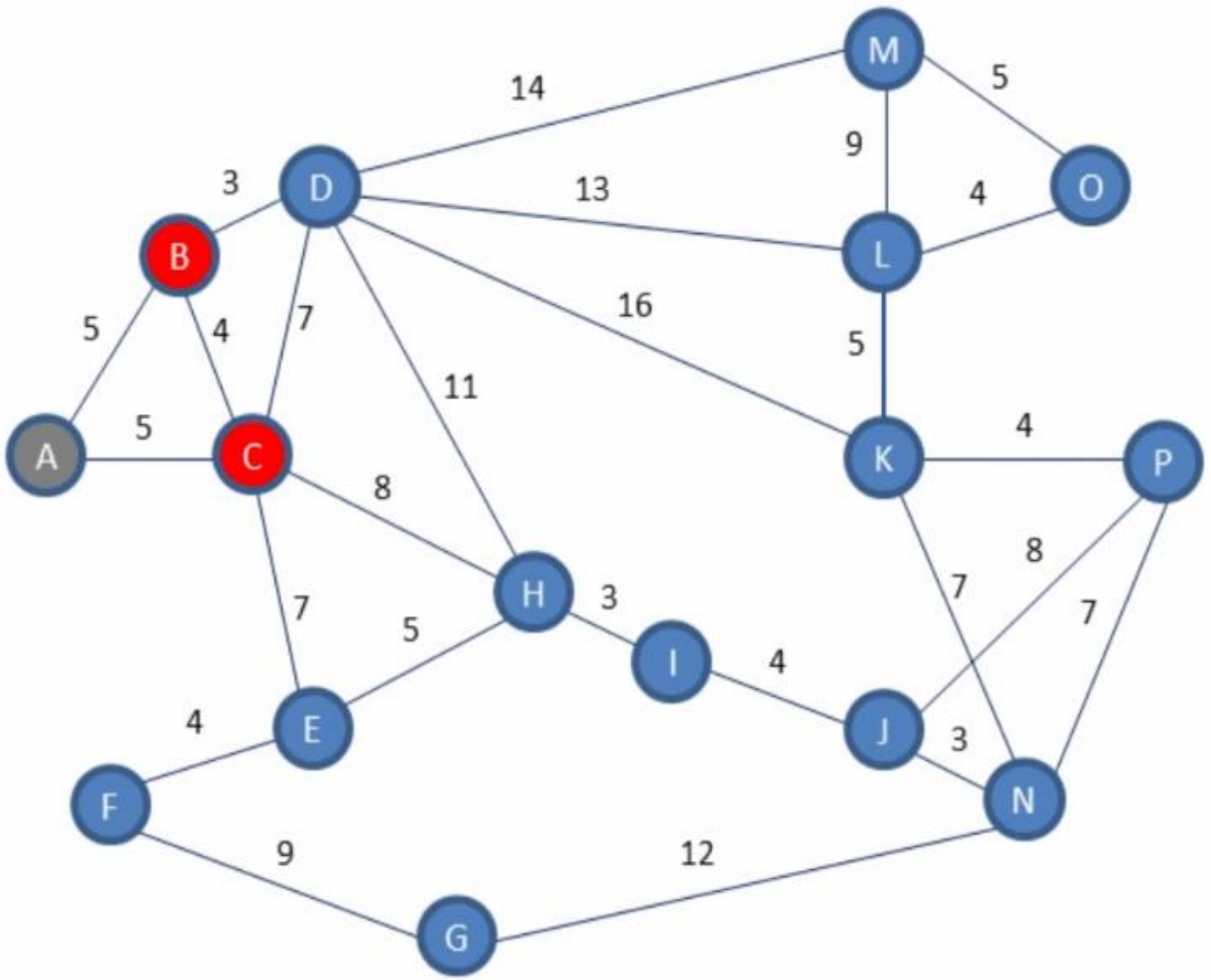



M
14
5
9
3
D
13
4
O
B
L
16
5
4
7
5
11
5
A
C
K
4
P
8
8
7
H
3
7
5
I
4
E
J
4
3
F
N
9
12
G


M
O
D
B
L
A
C
K
P
H
I
J
E
F
N
G
14
5
9
3
13
4
5
4
7
16
5
11
5
4
8
8
7
7
5
3
4
4
3
9
12

Open B C

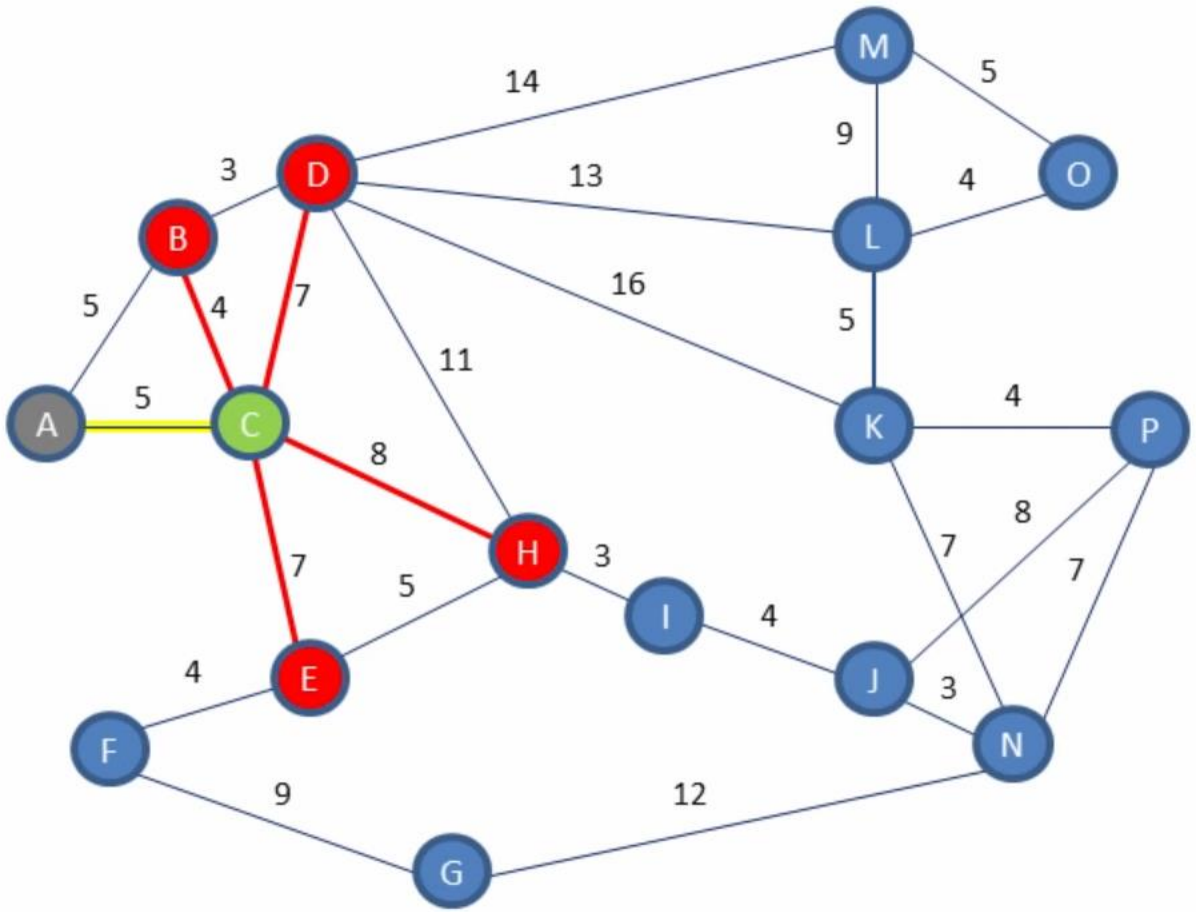
Closed A



Vertex	Distance from A (g)	Heuristic distance (h)	f = g + h	Previous vertex
A	0	16	16	
B	5	17	22	A
C	5	13	18	A
D		16		
E		16		
F		20		
G		17		
H		11		
I		10		
J		8		
K		4		
L		7		
M		10		
N		7		
O		5		
P		0		

Open B C D H E

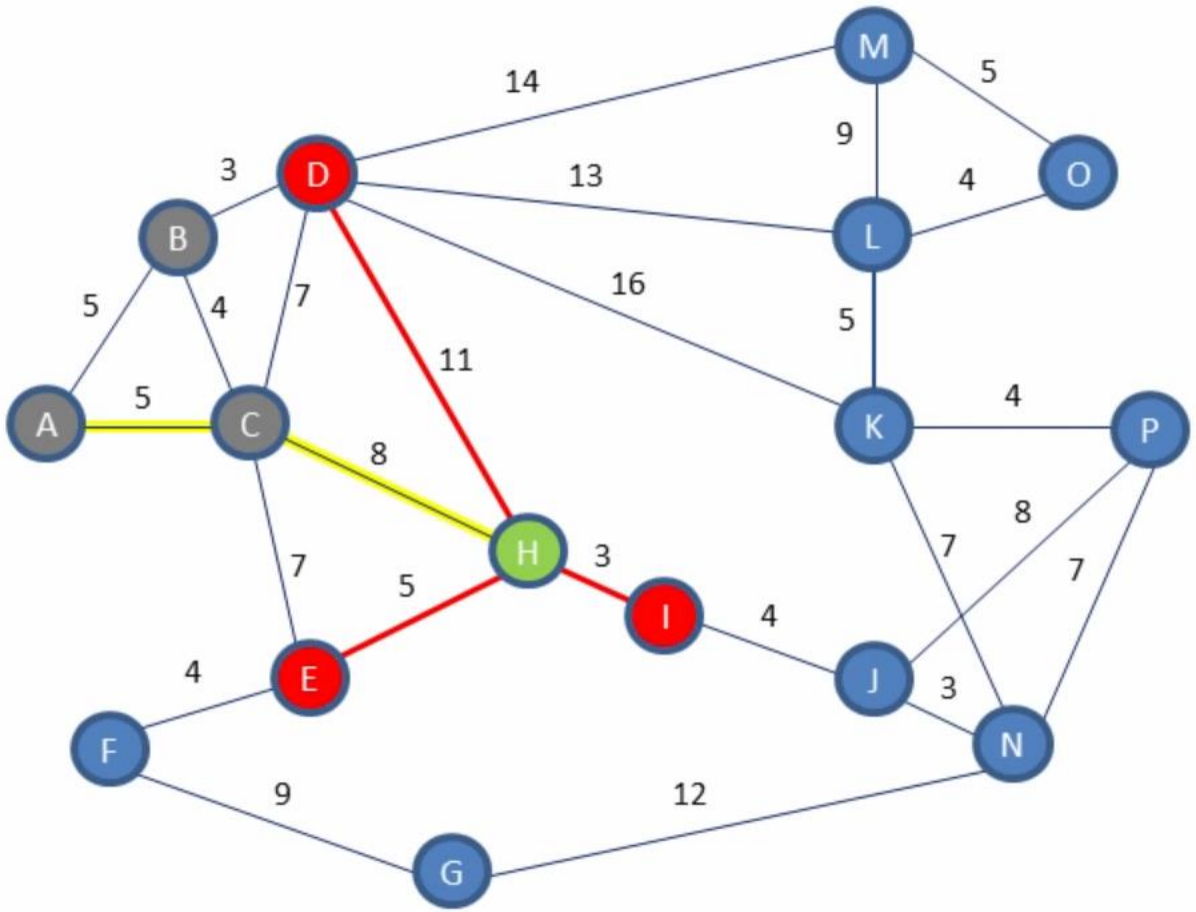
Closed A



Vertex	Distance from A (g)	Heuristic distance (h)	f = g + h	Previous vertex
A	0	16	16	
B	5 **9**	17	22 **26**	A **C**
C	5	13	18	A
D	**12**	16	**28**	**C**
E	**12**	16	**28**	**C**
F		20		
G		17		
H	**13**	11	**24**	**C**
I		10		
J		8		
K		4		
L		7		
M		10		
N		7		
O		5		
P		0		

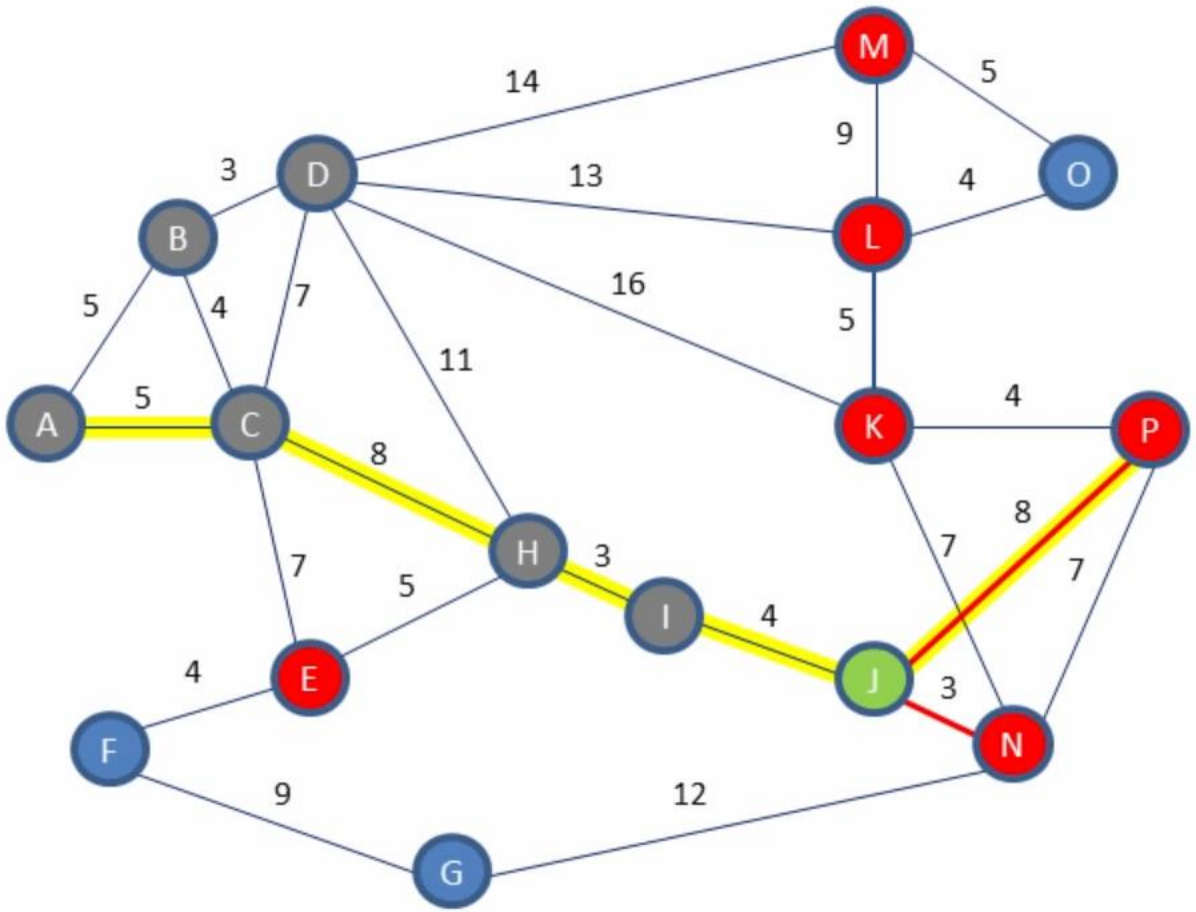
Open D H E I

Closed A C B



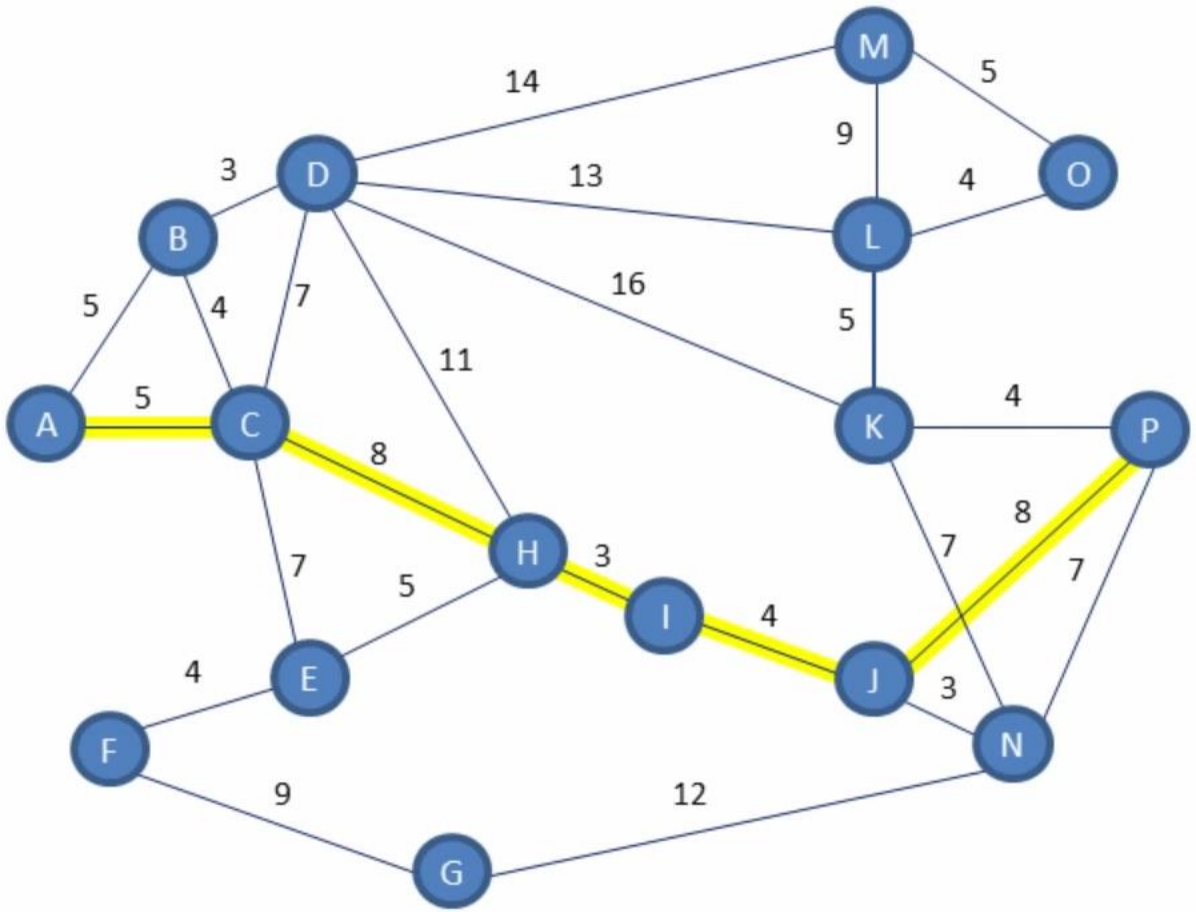
Vertex	Distance from A (g)	Heuristic distance (h)	f = g + h	Previous vertex
A	0	16	16	
B	5	17	22	A
C	5	13	18	A
D	8 **24**	16	24 **40**	B **H**
E	12 **18**	16	28 **34**	C **H**
F		20		
G		17		
H	13	11	24	C
I	**16**	10	**26**	**H**
J		8		
K		4		
L		7		
M		10		
N		7		
O		5		
P		0		

Open	E	M	L	K	J	P	N
Closed	A	C	B	H	D	I	

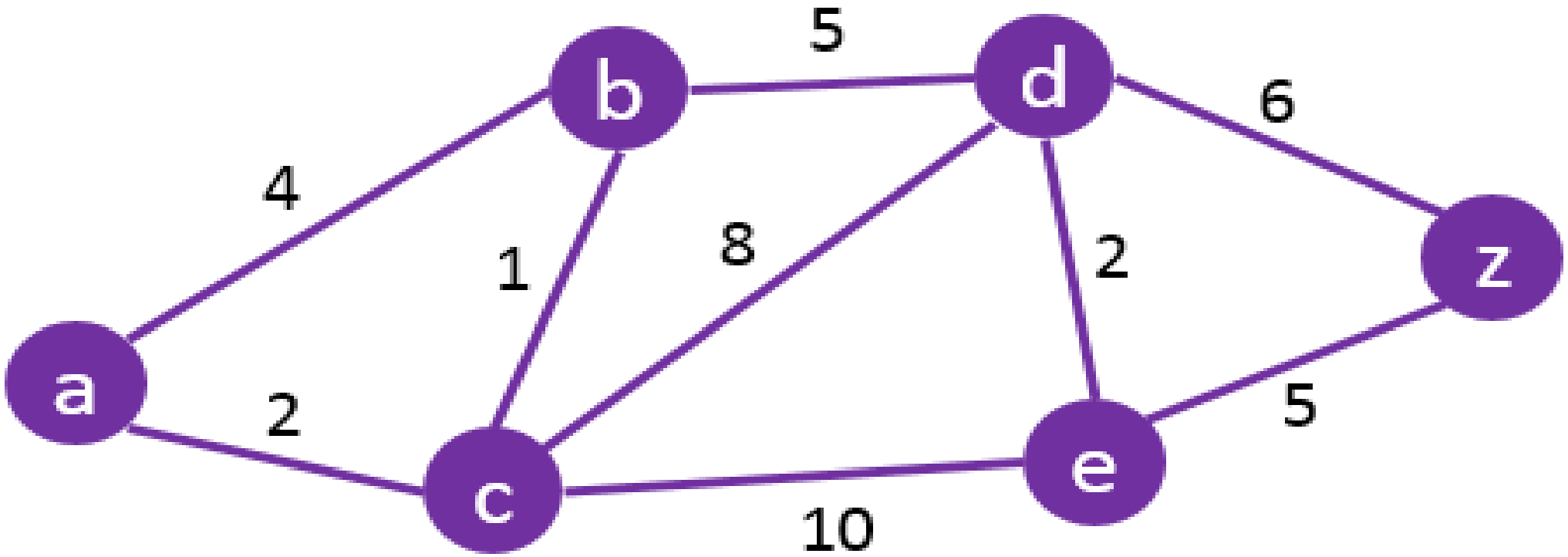


Vertex	Distance from A (g)	Heuristic distance (h)	f = g + h	Previous vertex
A	0	16	16	
B	5	17	22	A
C	5	13	18	A
D	8	16	24	B
E	12	16	28	C
F		20		
G		17		
H	13	11	24	C
I	16	10	26	H
J	20	8	28	I
K	24	4	28	D
L	21	7	28	D
M	22	10	32	D
N	**23**	7	**30**	**J**
O		5		
P	**28**	0	**28**	**J**

Open	E	M	L	K	J	P	N
Closed	A	C	B	H	D	I	



Vertex	Distance from A (g)	Heuristic distance (h)	f = g + h	Previous vertex
A	0	16	16	
B	5	17	22	A
C	5	13	18	A
D	8	16	24	B
E	12	16	28	C
F		20		
G		17		
H	13	11	24	C
I	16	10	26	H
J	20	8	28	I
K	24	4	28	D
L	21	7	28	D
M	22	10	32	D
N	23	7	30	J
O		5		
P	28	0	28	J

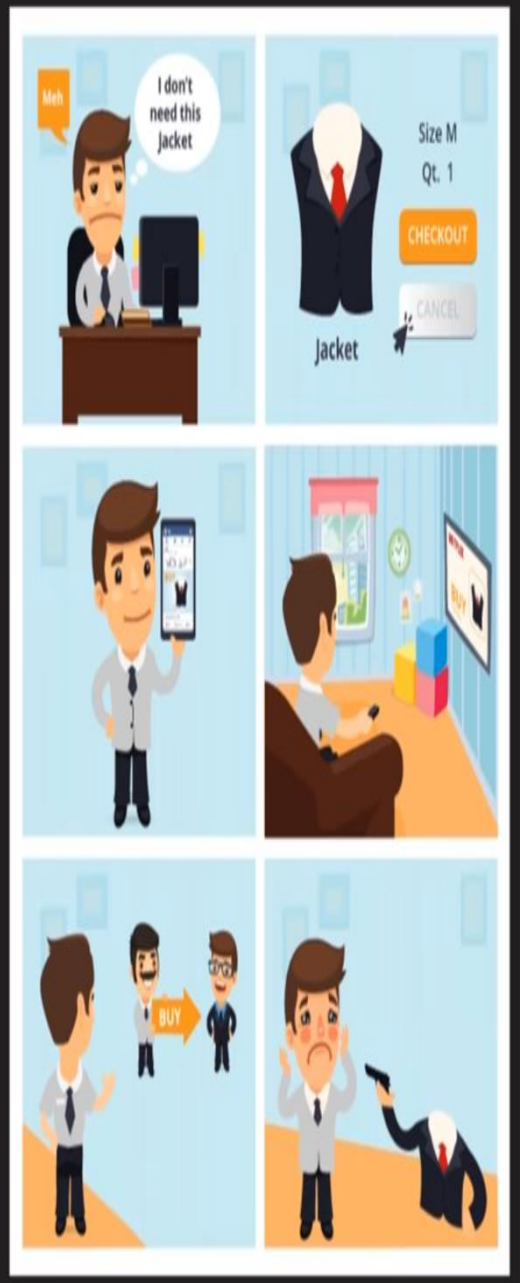


Dijkstra's Algorithm

What is the shortest path to travel from A to Z?

	GraphConnect 2017	GraphConnect 2018
Pathfinding & Search	• Parallel Breadth First Search & DFS • Shortest Path • Single-Source Shortest Path • All Pairs Shortest Path • Minimum Spanning Tree	• A* Shortest Path • Yen's K Shortest Path • K-Spanning Tree (MST)
Centrality / Importance	• Degree Centrality • Closeness Centrality • Betweenness Centrality • PageRank	• Harmonic Closeness Centrality • Dangalchev Closeness Centrality • Wasserman & Faust Closeness Centrality • Approximate Betweenness Centrality • Personalized PageRank
Community Detection	• Triangle Count • Clustering Coefficients • Connected Components (Union Find) • Strongly Connected Components • Label Propagation	• Balanced Triad (identification) • Louvain – Multi-Step

Ads Recommendation



2

amazon

NETFLIX

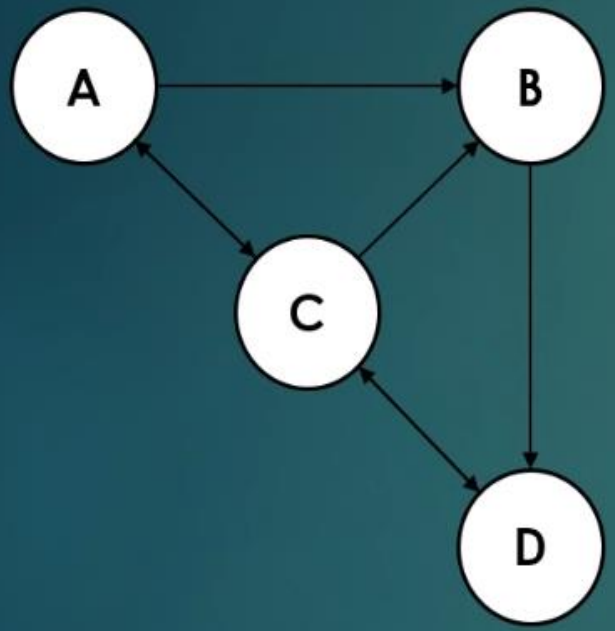
NETFLIX

Google

PageRank

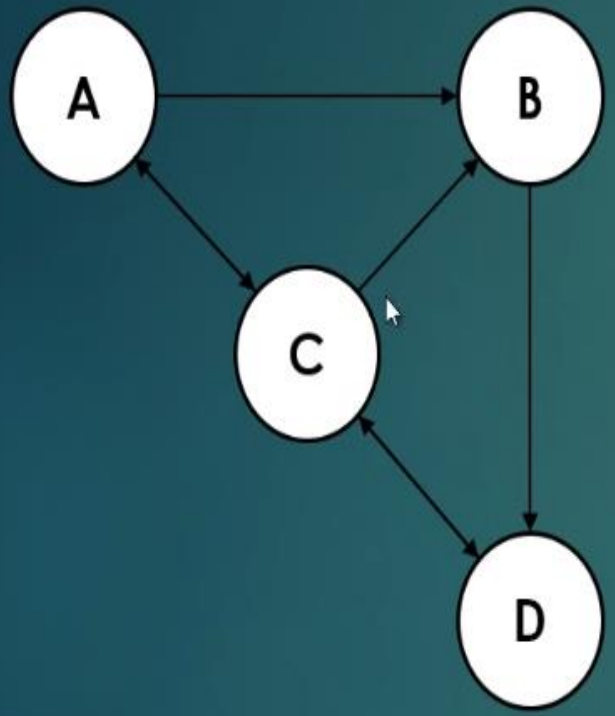
10

PageRank algorithm



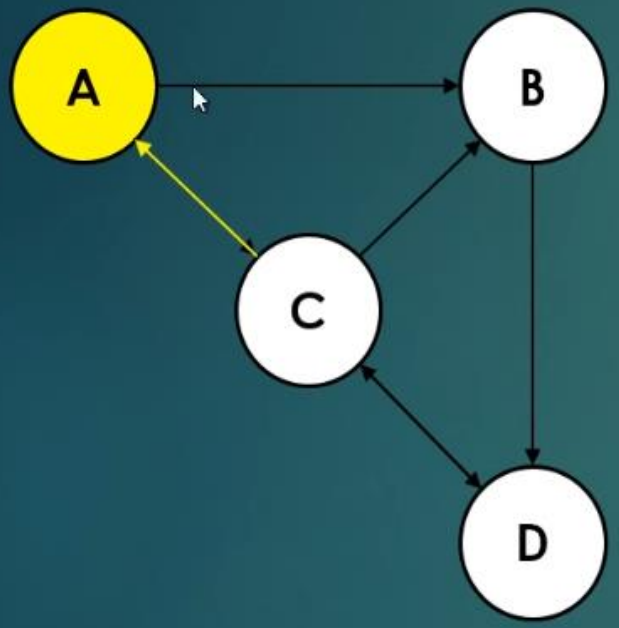
	Iteration 0	Iteration 1	Iteration 2	PageRank
A				
B				
C				
D				

PageRank algorithm



	Iteration 0	Iteration 1	Iteration 2	PageRank
A	1/4			
B	1/4			
C	1/4			
D	1/4			

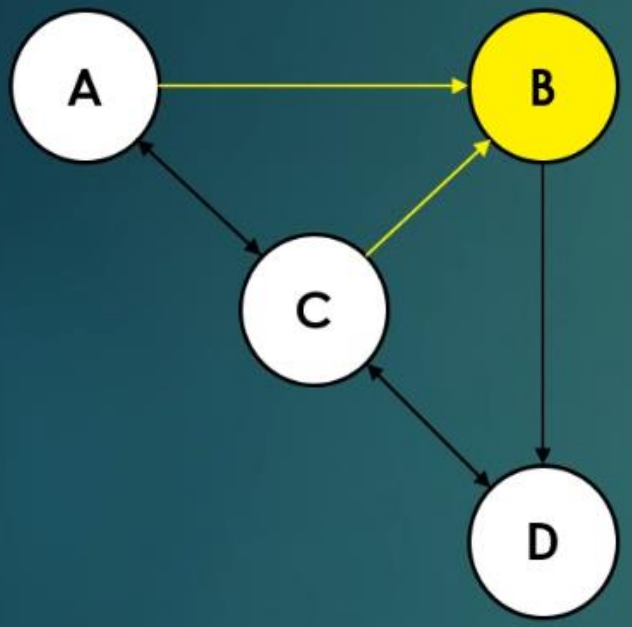
PageRank algorithm



	Iteration 0	Iteration 1	Iteration 2	PageRank
A	1/4	1/12		
B	1/4			
C	1/4			
D	1/4			

$$PR(A) = \frac{\frac{1}{4}}{3}$$

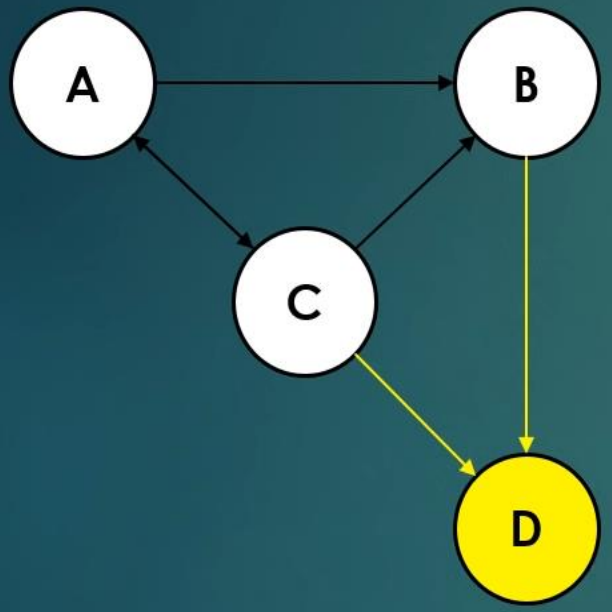
PageRank algorithm



	Iteration 0	Iteration 1	Iteration 2	PageRank
A	1/4	1/12		
B	1/4	2.5/12		
C	1/4			
D	1/4			

$$PR(B) = \frac{\frac{1}{4}}{2} + \frac{\frac{1}{4}}{3}$$

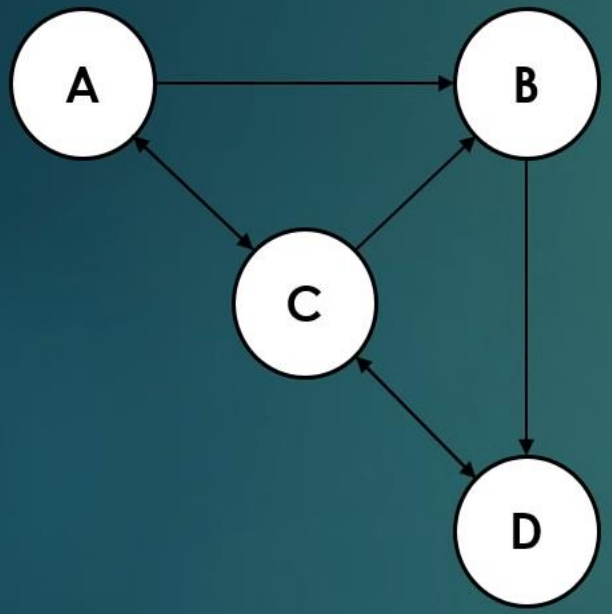
PageRank algorithm



	Iteration 0	Iteration 1	Iteration 2	PageRank
A	1/4	1/12	1.5/12	
B	1/4	2.5/12	2/12	
C	1/4	4.5/12	4.5/12	
D	1/4	4/12	4/12	

$$PR(D) = \frac{\frac{2.5}{12}}{1} + \frac{\frac{4.5}{12}}{3}$$

PageRank algorithm



	Iteration 0	Iteration 1	Iteration 2	PageRank
A	1/4	1/12	1.5/12	1
B	1/4	2.5/12	2/12	2
C	1/4	4.5/12	4.5/12	4
D	1/4	4/12	4/12	3

GraphConnect 2017 → GraphConnect 2018

Pathfinding & Search

- Parallel Breadth First Search & DFS
- Shortest Path
- Single-Source Shortest Path
- All Pairs Shortest Path
- Minimum Spanning Tree

- A* Shortest Path
- Yen's K Shortest Path
- K-Spanning Tree (MST)

Centrality / Importance

- Degree Centrality
- Closeness Centrality
- Betweenness Centrality
- PageRank

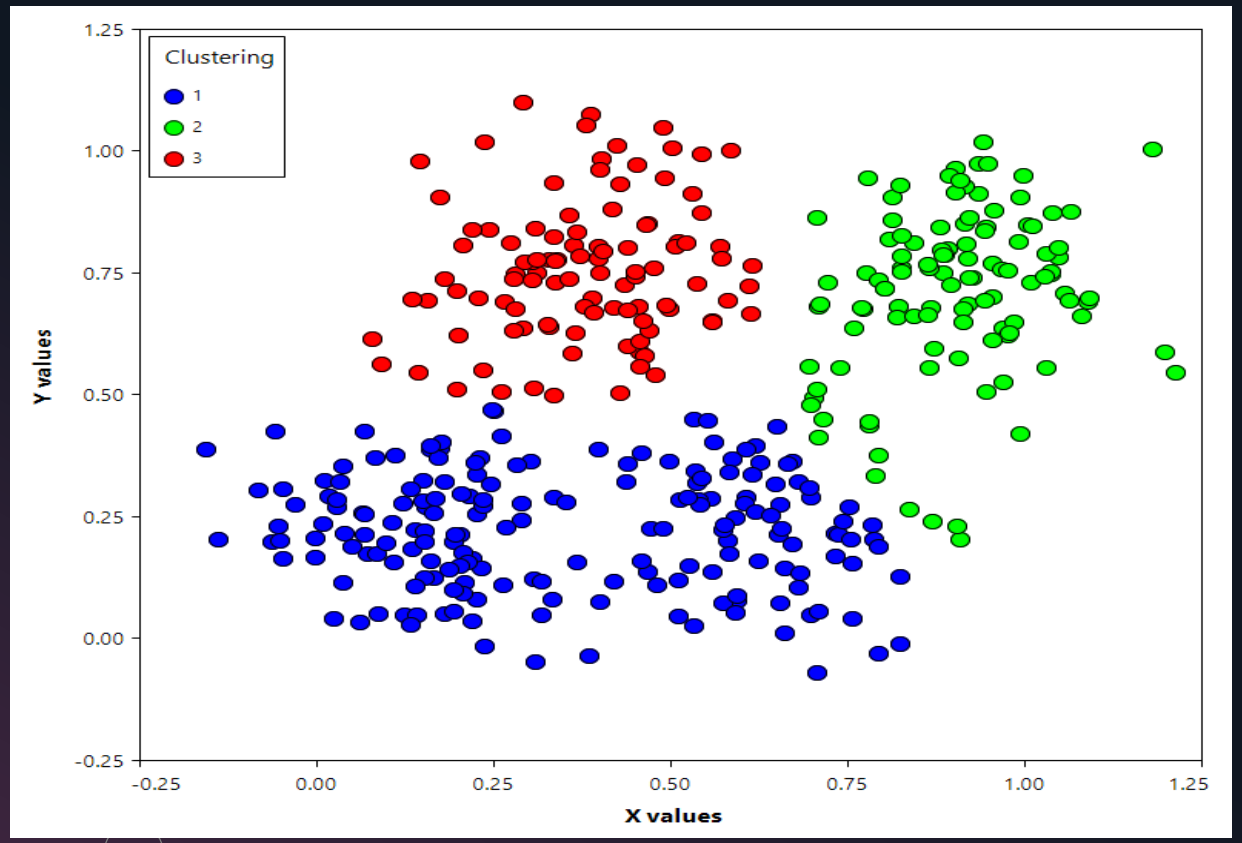
- Harmonic Closeness Centrality
- Dangalchev Closeness Centrality
- Wasserman & Faust Closeness Centrality
- Approximate Betweenness Centrality
- Personalized PageRank

Community Detection

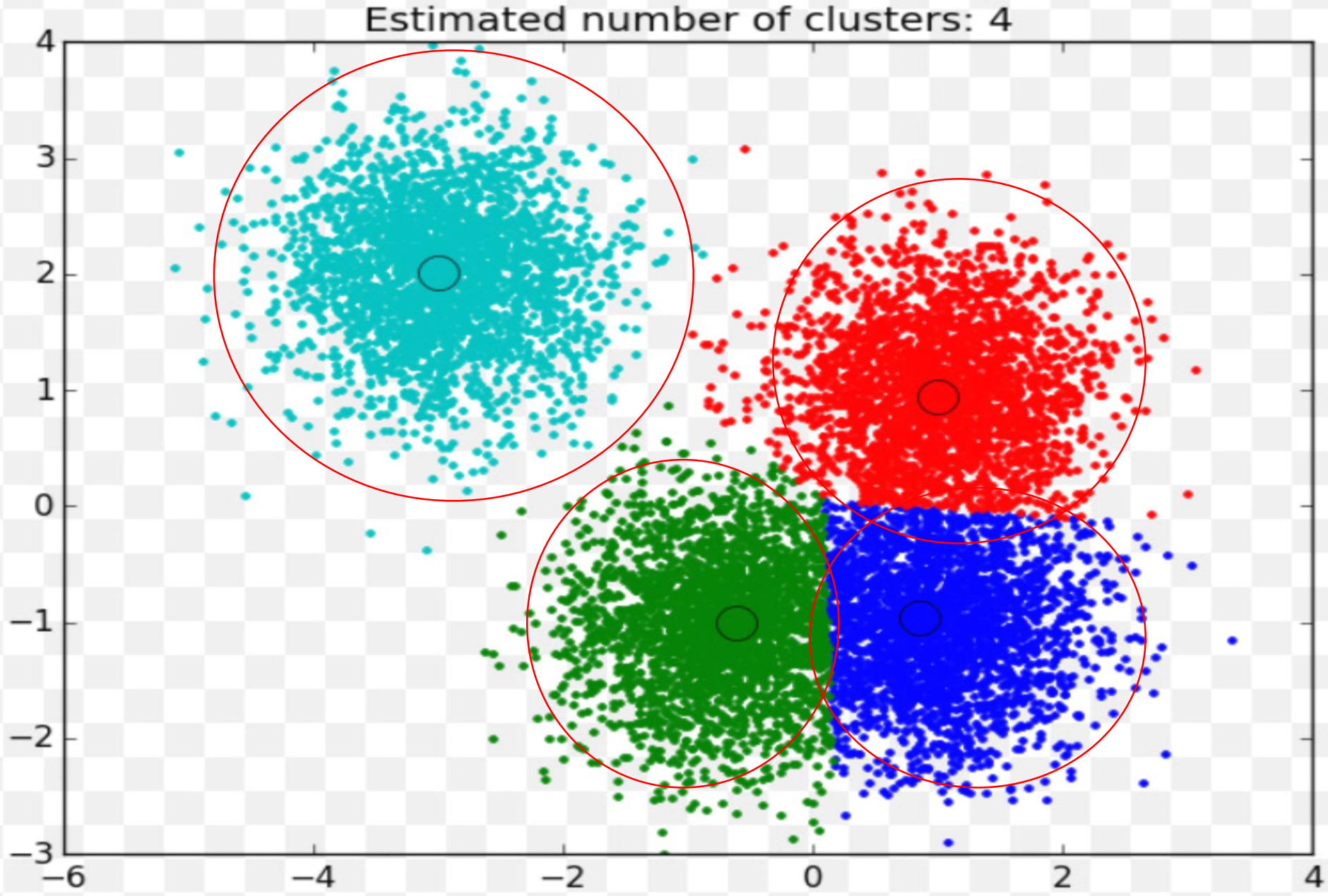
- Triangle Count
- Clustering Coefficients
- Connected Components (Union Find)
- Strongly Connected Components
- Label Propagation

- Balanced Triad (identification)
- Louvain – Multi-Step

روش یادگیری بدون نظارت

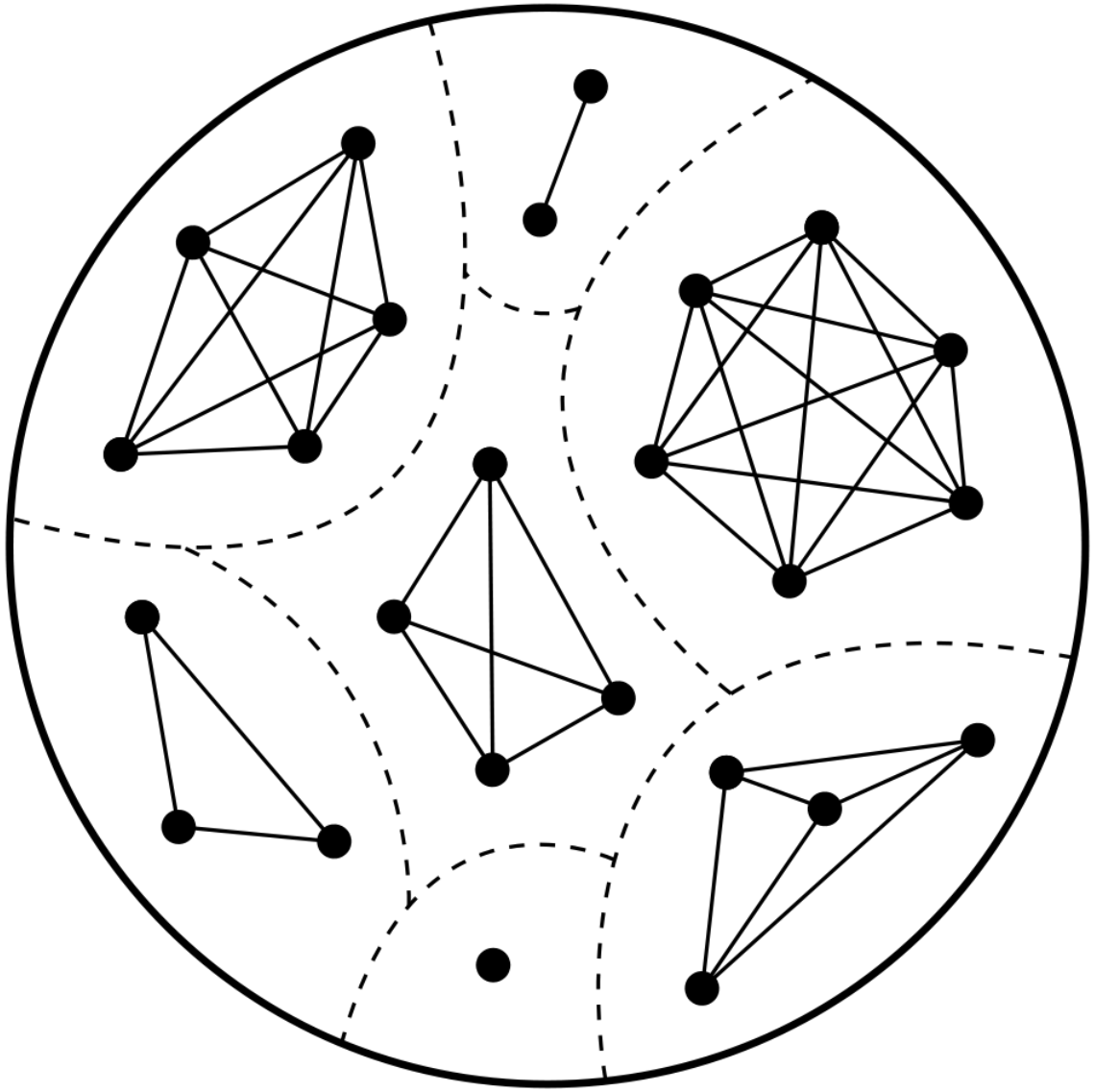


Clustering
/
خوشه بندی


Estimated number of clusters: 4
4
3
2
1
0
−1
−2
−3
−6
−4
−2
0
2
4

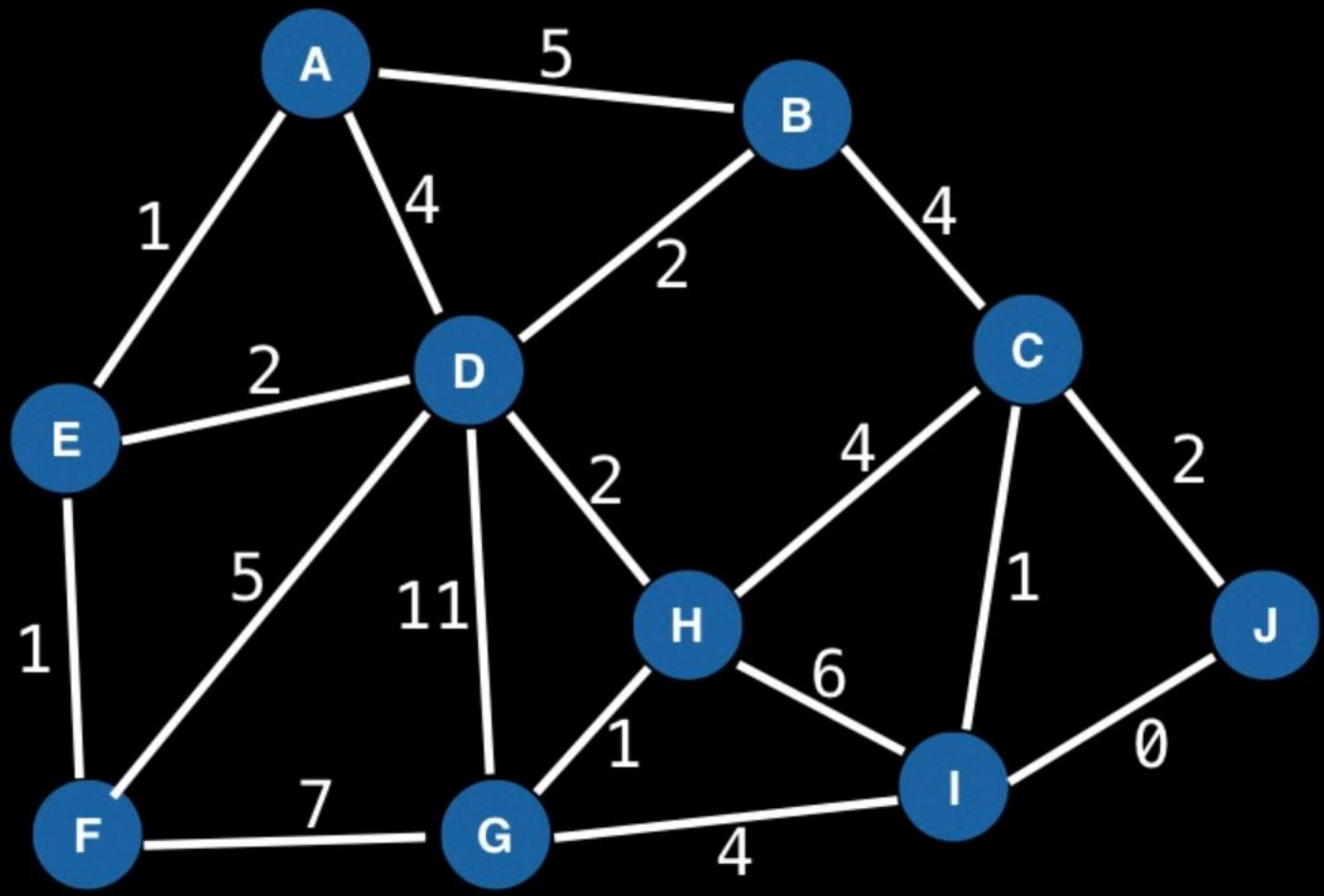
Union Find

Kruskal's Algorithm



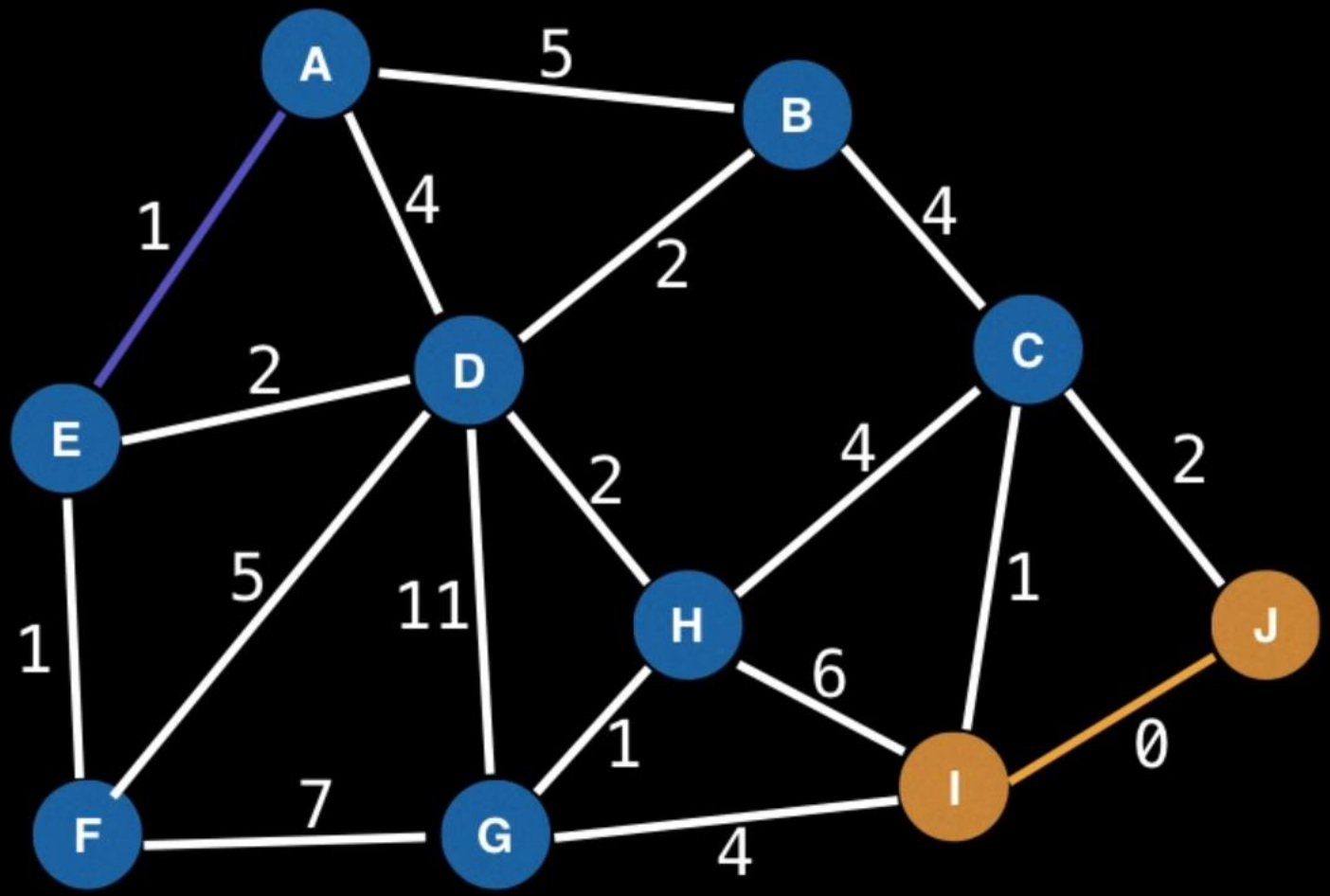
Union Find application: Kruskal's Minimum Spanning Tree

I to J = 0
A to E = 1
C to I = 1
E to F = 1
G to H = 1
B to D = 2
C to J = 2
D to E = 2
D to H = 2
A to D = 4
B to C = 4
C to H = 4
G to I = 4
A to B = 5
D to F = 5
H to I = 6
F to G = 7
D to G = 11



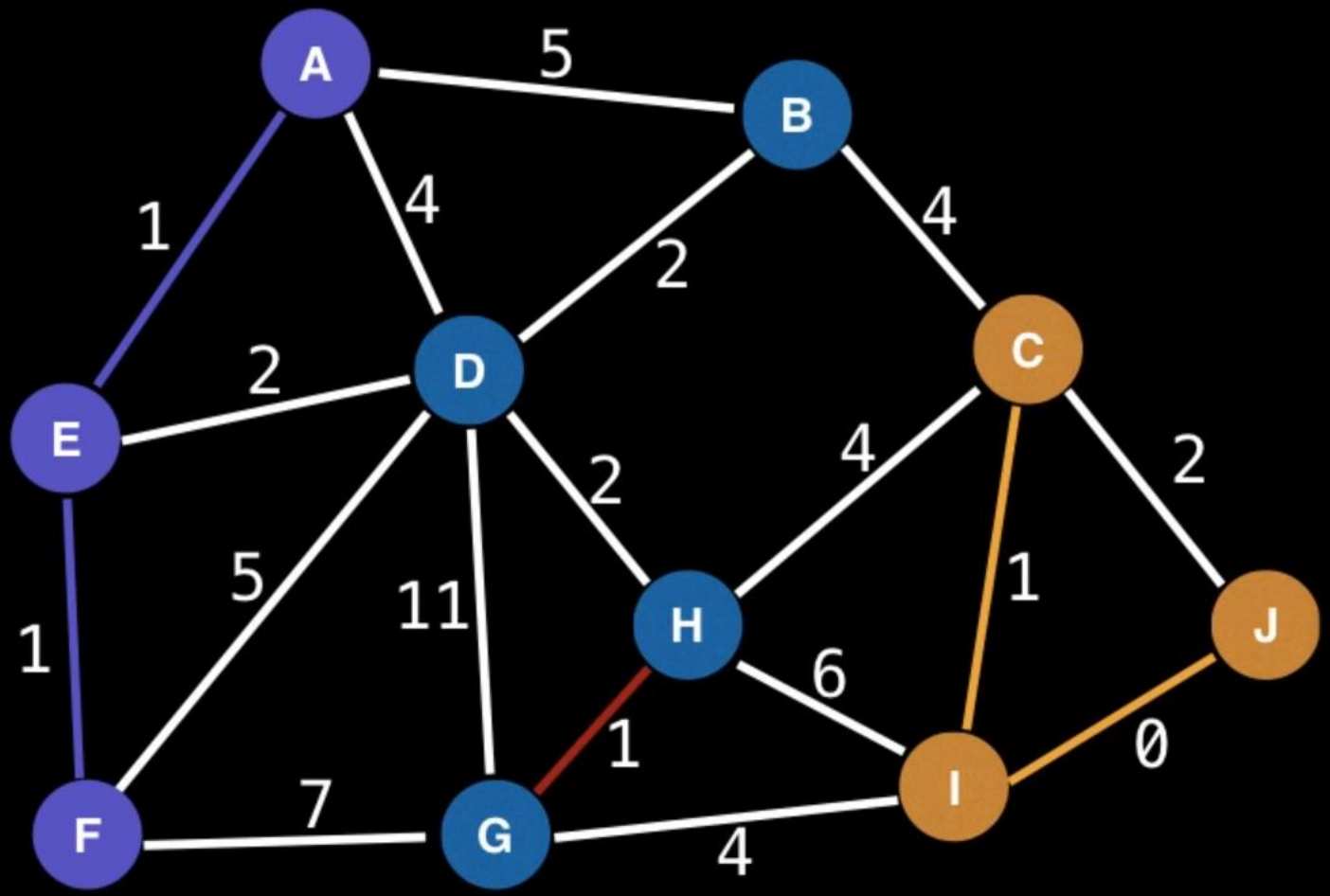
Union Find application: Kruskal's Minimum Spanning Tree

I to J = 0
A to E = 1
C to I = 1
E to F = 1
G to H = 1
B to D = 2
C to J = 2
D to E = 2
D to H = 2
A to D = 4
B to C = 4
C to H = 4
G to I = 4
A to B = 5
D to F = 5
H to I = 6
F to G = 7
D to G = 11



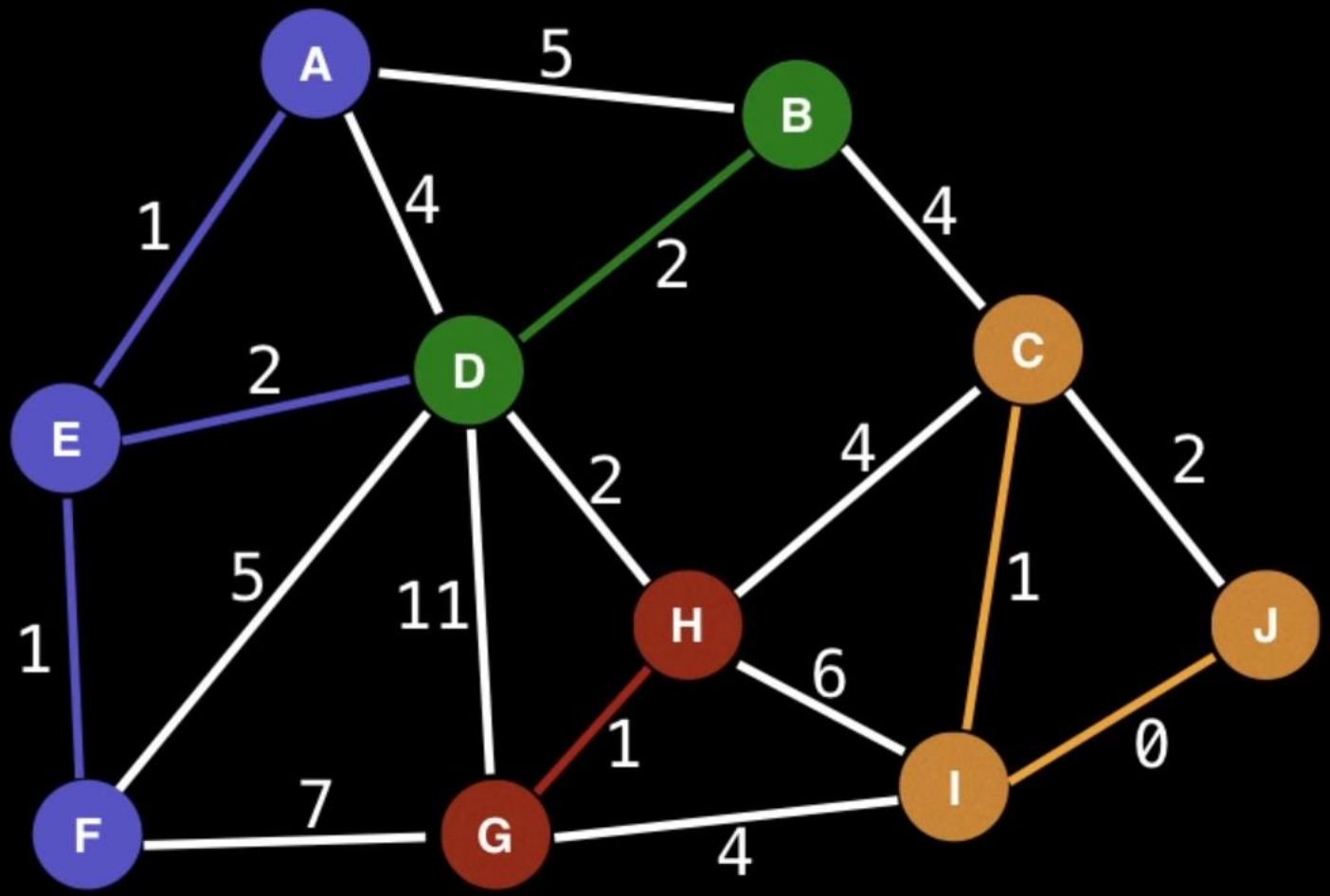
Union Find application: Kruskal's Minimum Spanning Tree

I to J = 0
A to E = 1
C to I = 1
E to F = 1
G to H = 1
B to D = 2
C to J = 2
D to E = 2
D to H = 2
A to D = 4
B to C = 4
C to H = 4
G to I = 4
A to B = 5
D to F = 5
H to I = 6
F to G = 7
D to G = 11



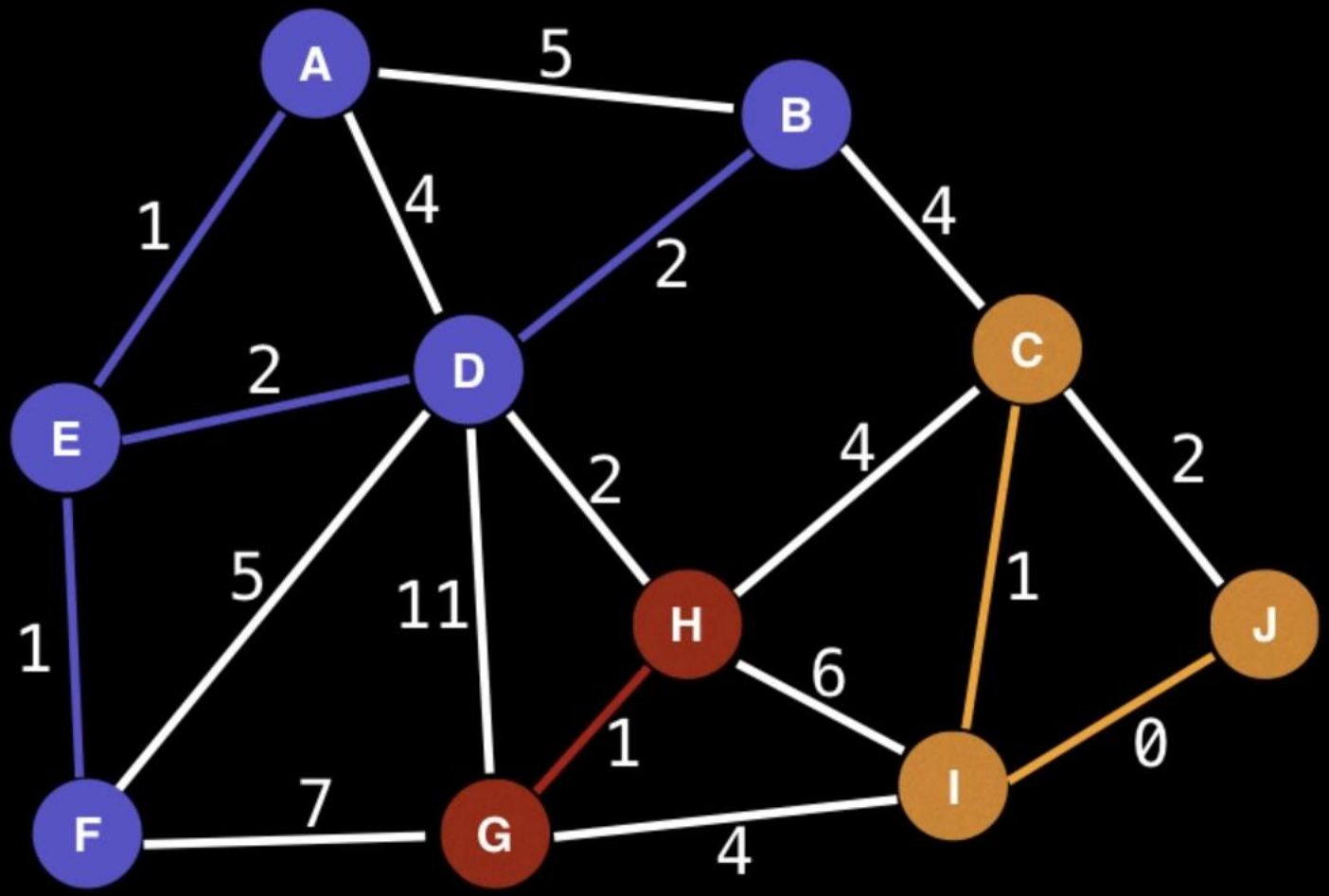
Union Find application: Kruskal's Minimum Spanning Tree

I to J = 0
A to E = 1
C to I = 1
E to F = 1
G to H = 1
B to D = 2
C to J = 2
D to E = 2
D to H = 2
A to D = 4
B to C = 4
C to H = 4
G to I = 4
A to B = 5
D to F = 5
H to I = 6
F to G = 7
D to G = 11



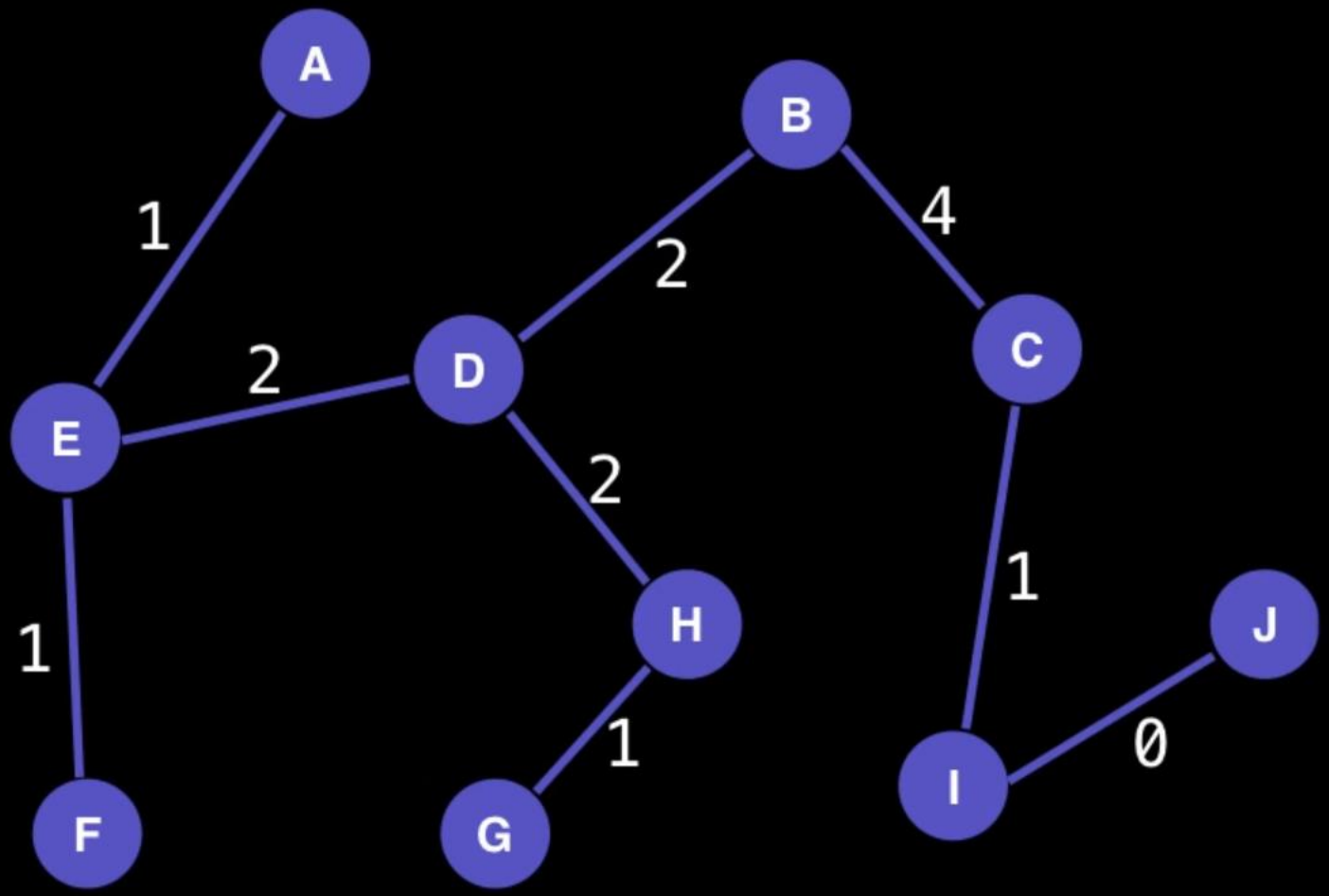
Union Find application: Kruskal's Minimum Spanning Tree

I to J = 0
A to E = 1
C to I = 1
E to F = 1
G to H = 1
B to D = 2
C to J = 2
D to E = 2
D to H = 2
A to D = 4
B to C = 4
C to H = 4
G to I = 4
A to B = 5
D to F = 5
H to I = 6
F to G = 7
D to G = 11



Union Find application: Kruskal's Minimum Spanning Tree

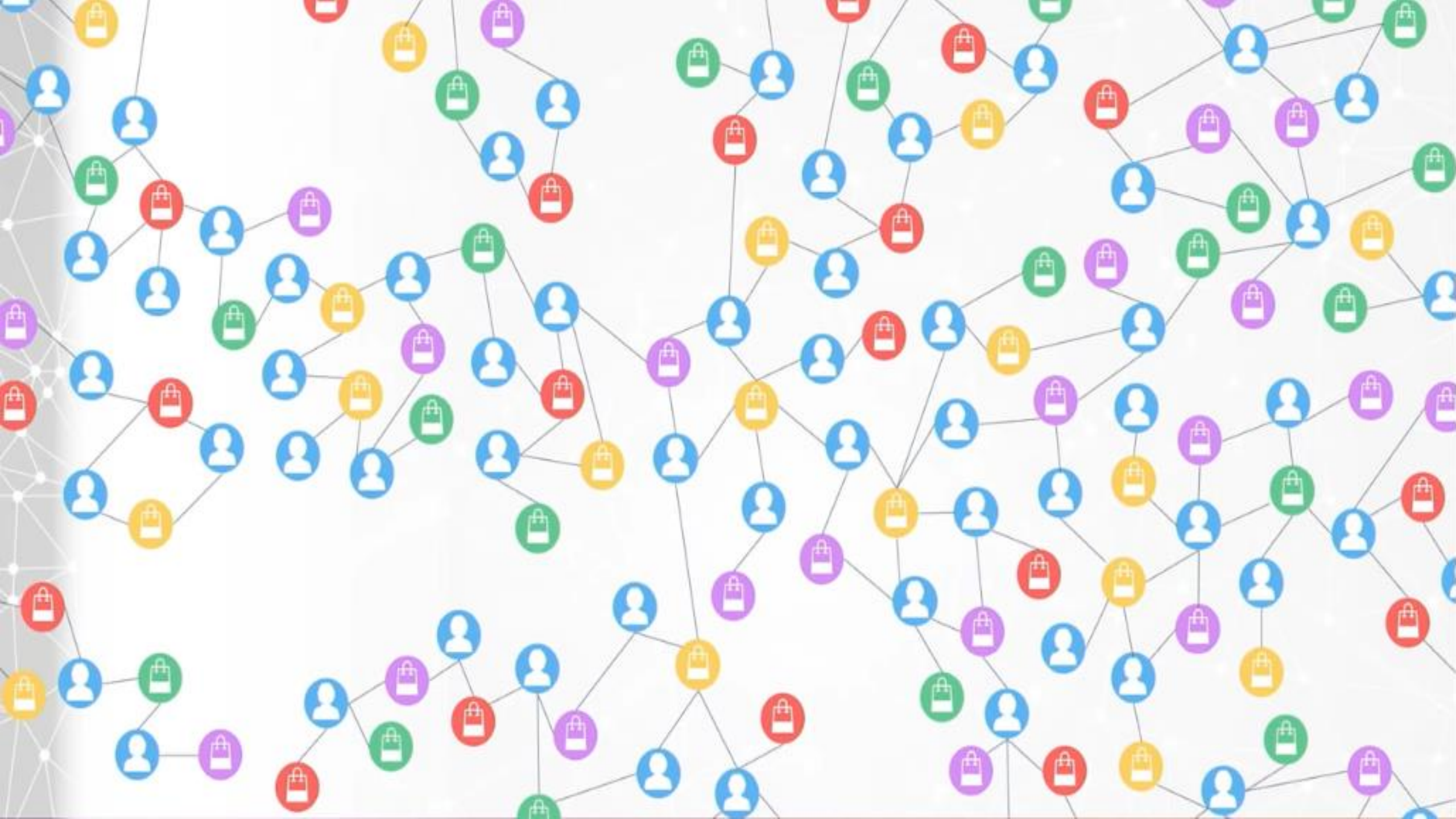
I to J = 0
A to E = 1
C to I = 1
E to F = 1
G to H = 1
B to D = 2
C to J = 2
D to E = 2
D to H = 2
A to D = 4
B to C = 4
C to H = 4
G to I = 4
A to B = 5
D to F = 5
H to I = 6
F to G = 7
D to G = 11



Graphs are Context for Decisions

Knowledge Graphs

LET'S SEE....



https://www.google.com/search?client=firefox-b-d&q=iran

Google

iran

All Images Maps News Videos More Settings Tools

About 1,180,000,000 results (0.57 seconds)

Top stories

L'Iran a bien arrêté l'anthropologue franco-iranienne Fariba Adelkhah

Le Monde

12 hours ago

Iran rejects suggestion its missile programme is negotiable

BBC

1 hour ago

At summit with Russia, Israel and US demanded Iran leave Lebanon, Iraq …

The Times of Israel

2 hours ago

More for iran

Iran

Country in the Middle East

Iran, also called Persia, and officially the Islamic Republic of Iran, is a country in Western Asia. With 82 million inhabitants, Iran is the world's 18th most populous country. Its territory spans 1,648,195 km², making it the second largest country in the Middle East and the 17th largest in the world. Wikipedia

https://www.apnews.com/Iran

TEHRAN, **Iran** (AP) — **Iran's** president says his country is ready to negotiate with the United States if Washington lifts its economic sanctions.

Iran - latest news, breaking stories and comment - The Independent

https://www.independent.co.uk/topic/Iran

All the latest breaking news on **Iran**. Browse The Independent's complete collection of articles and commentary on **Iran**.

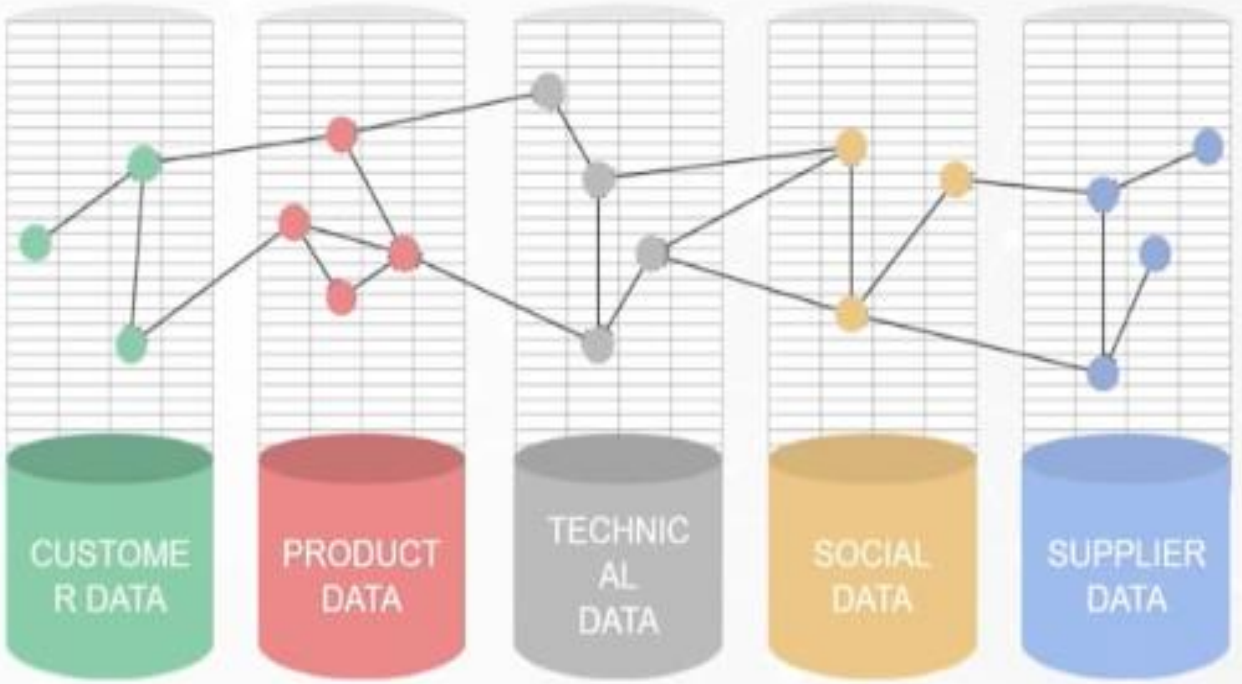
Searches related to iran

iran **facts**	iran **people**
iran **headlines**	iran **war**
iran **breaking news**	iran **culture**
iran **news**	**current situation in** iran

Knowledge Graphs

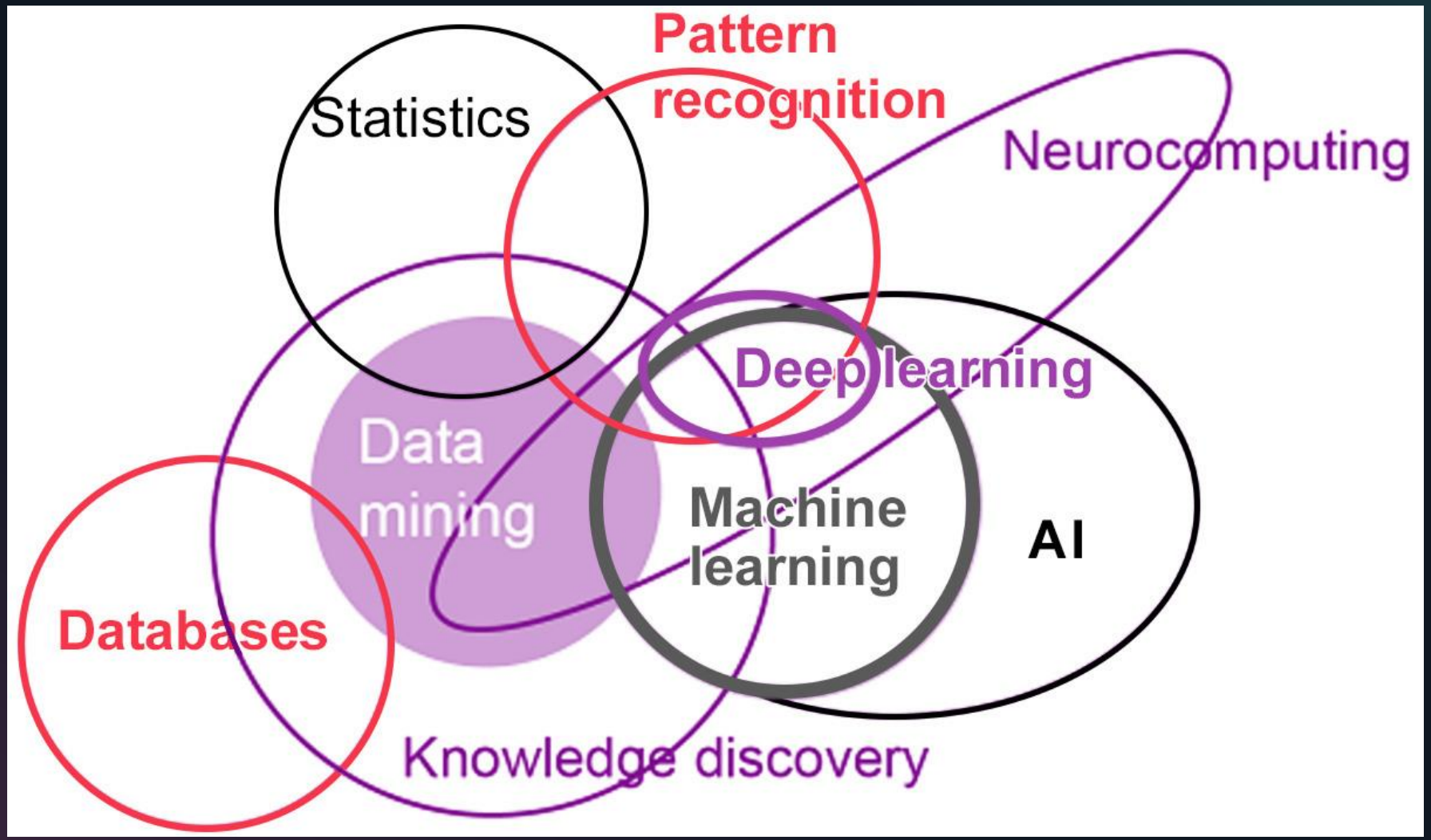
Context doesn't fit cleanly in an equation

- A connected, dynamic, and understandable repository of different data types
- Link siloed or external data sources in an intelligent way
- Key to understand your unique, enterprise language
- **Knowledge Base ≠ Knowledge Graph**



neo4j

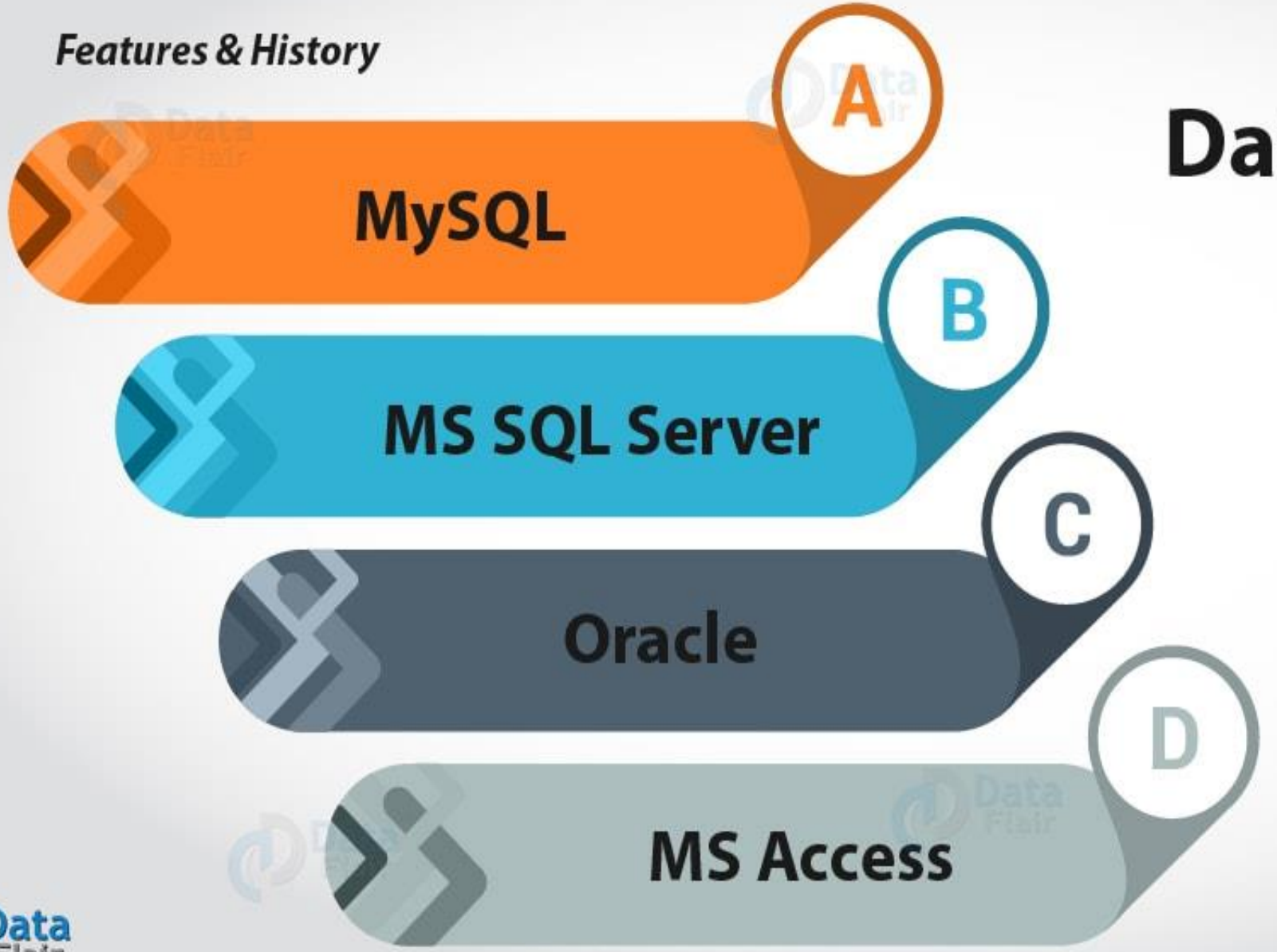

DON'T
FORGET


Statistics
Pattern recognition
Neurocomputing
Deep learning
Data mining
Machine learning
AI
Databases
Knowledge discovery

BIG DATA

Features & History

SQL RDBMS Database System



employeeNumb	lastName	firstName	extension	email	officeCode	reportsTo	jobTitle
1002	Murphy	Diane	x5800	dmurphy@classicmodelcars.com	1	NULL	President
1056	Patterson	Mary	x4611	mpatterso@classicmodelcars.com	1	1002	VP Sales
1076	Firrelli	Jeff	x9273	jfirrelli@classicmodelcars.com	1	1002	VP Marketing
1088	Patterson	William	x4871	wpatterson@classicmodelcars.com	6	1056	Sales Manager (APAC)
1102	Bondur	Gerard	x5408	gbondur@classicmodelcars.com	4	1056	Sale Manager (EMEA)
1143	Bow	Anthony	x5428	abow@classicmodelcars.com	1	1056	Sales Manager (NA)
1165	Jennings	Leslie	x3291	ljennings@classicmodelcars.com	1	1143	Sales Rep
1166	Thompson	Leslie	x4065	lthompson@classicmodelcars.com	1	1143	Sales Rep
1188	Firrelli	Julie	x2173	jfirrelli@classicmodelcars.com	2	1143	Sales Rep
1216	Patterson	Steve	x4334	spatterson@classicmodelcars.com	2	1143	Sales Rep
1286	Tseng	Foon Yue	x2248	ftseng@classicmodelcars.com	3	1143	Sales Rep
1323	Vanauf	George	x4102	gvanauf@classicmodelcars.com	3	1143	Sales Rep
1337	Bondur	Loui	x6493	lbondur@classicmodelcars.com	4	1102	Sales Rep
1370	Hernandez	Gerard	x2028	ghernande@classicmodelcars.com	4	1102	Sales Rep
1401	Castillo	Pamela	x2759	pcastillo@classicmodelcars.com	4	1102	Sales Rep
1501	Bott	Larry	x2311	lbott@classicmodelcars.com	7	1102	Sales Rep

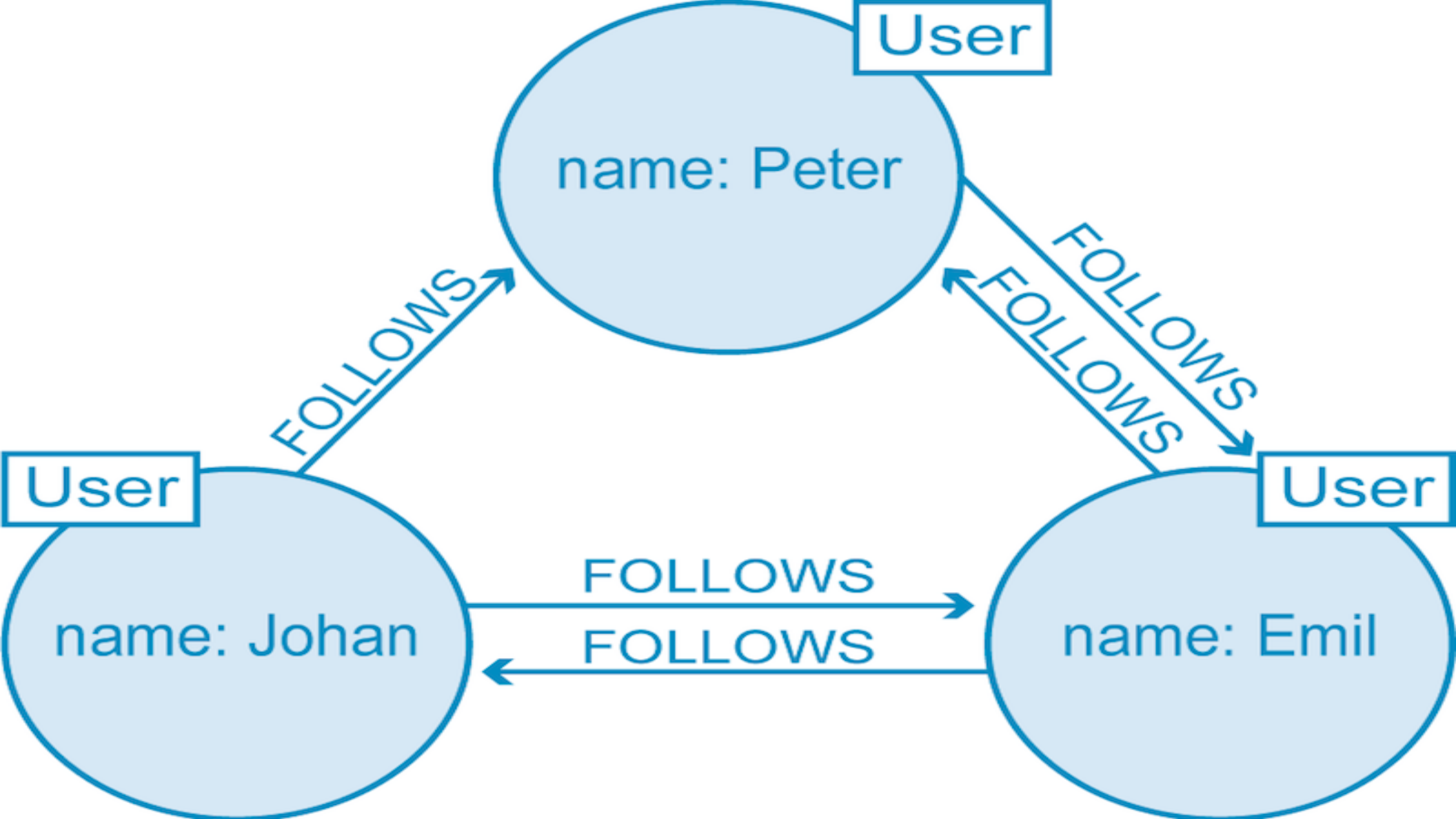
✓جرا باید برای پیدا کردن یک رابطه بین داده ها کلی join داشته باشیم؟

✓چرا باید با همه داده ها کار کنیم ؟

✓زمان چه می شود؟

✓آیا در نهایت به جواب می رسیم؟

✓جدول ها قابل این را ندارند که داده های ساخت نیافته را ذخیره کنند؟


User
name: Peter
FOLLOWS
FOLLOWS
FOLLOWS
FOLLOWS
User
name: Johan
User
name: Emil
FOLLOWS
FOLLOWS

Graph Database

&

Cypher

Typical Complex SQL Join

```
(SELECT T.directReportees AS directReportees, sum(T.count) AS count
FROM (
SELECT manager.pid AS directReportees, 0 AS count
  FROM person_reportee manager
  WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
UNION
  SELECT manager.pid AS directReportees, count(manager.directly_manages) AS count
FROM person_reportee manager
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
UNION
SELECT manager.pid AS directReportees, count(reportee.directly_manages) AS count
FROM person_reportee manager
JOIN person_reportee reportee
ON manager.directly_manages = reportee.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
UNION
SELECT manager.pid AS directReportees, count(L2Reportees.directly_manages) AS count
FROM person_reportee manager
JOIN person_reportee L1Reportees
ON manager.directly_manages = L1Reportees.pid
JOIN person_reportee L2Reportees
ON L1Reportees.directly_manages = L2Reportees.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
) AS T
GROUP BY directReportees)
UNION
(SELECT T.directReportees AS directReportees, sum(T.count) AS count
FROM (
SELECT manager.directly_manages AS directReportees, 0 AS count
FROM person_reportee manager
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
UNION
SELECT reportee.pid AS directReportees, count(reportee.directly_manages) AS count
FROM person_reportee manager
JOIN person_reportee reportee
ON manager.directly_manages = reportee.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
UNION
SELECT depth1Reportees.pid AS directReportees,
count(depth2Reportees.directly_manages) AS count
FROM person_reportee manager
JOIN person_reportee L1Reportees
ON manager.directly_manages = L1Reportees.pid
JOIN person_reportee L2Reportees
ON L1Reportees.directly_manages = L2Reportees.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
) AS T
GROUP BY directReportees)
UNION
(SELECT T.directReportees AS directReportees, sum(T.count) AS count
  FROM(
  SELECT reportee.directly_manages AS directReportees, 0 AS count
FROM person_reportee manager
JOIN person_reportee reportee
ON manager.directly_manages = reportee.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
UNION
SELECT L2Reportees.pid AS directReportees, count(L2Reportees.directly_manages)
AS count
FROM person_reportee manager
JOIN person_reportee L1Reportees
ON manager.directly_manages = L1Reportees.pid
JOIN person_reportee L2Reportees
ON L1Reportees.directly_manages = L2Reportees.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
GROUP BY directReportees
) AS T
GROUP BY directReportees)
UNION
(SELECT L2Reportees.directly_manages AS directReportees, 0 AS count
FROM person_reportee manager
JOIN person_reportee L1Reportees
ON manager.directly_manages = L1Reportees.pid
JOIN person_reportee L2Reportees
ON L1Reportees.directly_manages = L2Reportees.pid
WHERE manager.pid = (SELECT id FROM person WHERE name = "fName lName")
)
```

The Same Query using Cypher

```
MATCH (boss)-[:MANAGES*0..3]->(sub),
      (sub)-[:MANAGES*1..3]->(report)
WHERE boss.name = "John Doe"
RETURN sub.name AS Subordinate,
  count(report) AS Total
```

NEO4j USE CASES

Real Time Recommendations

Master Data Management

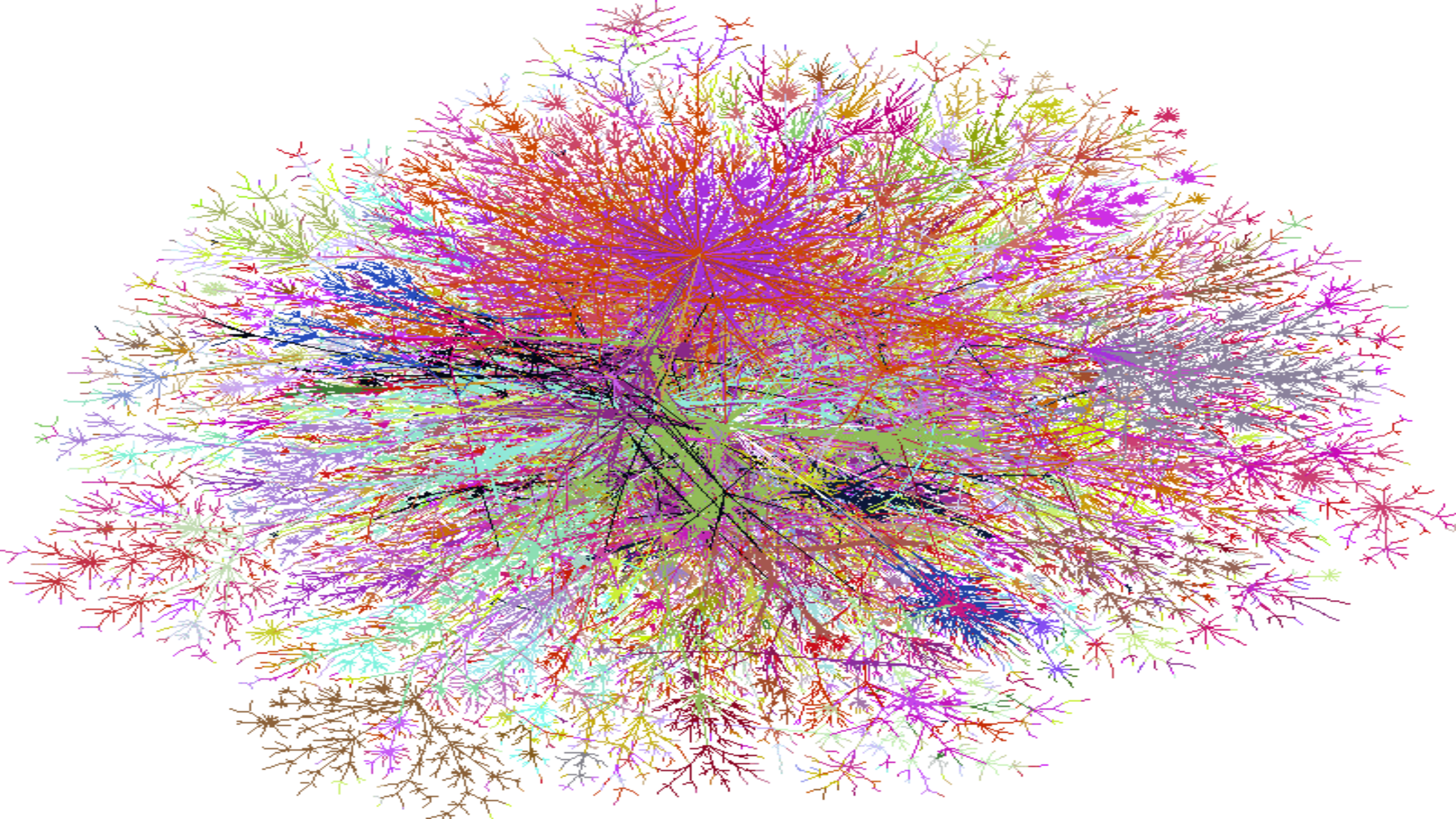
Fraud Detection

Graph Based Search

Network & IT-Operations

Identity & Access Management


Challenges



What is Big Data?

edureka!

"Big data is the term for a collection of data sets so large and complex that it becomes difficult to process using on-hand database management tools or traditional data processing applications"

Volume

Processing increasing huge data sets

Variety

Processing different types of data

Velocity

Data is being generated at an alarming rate

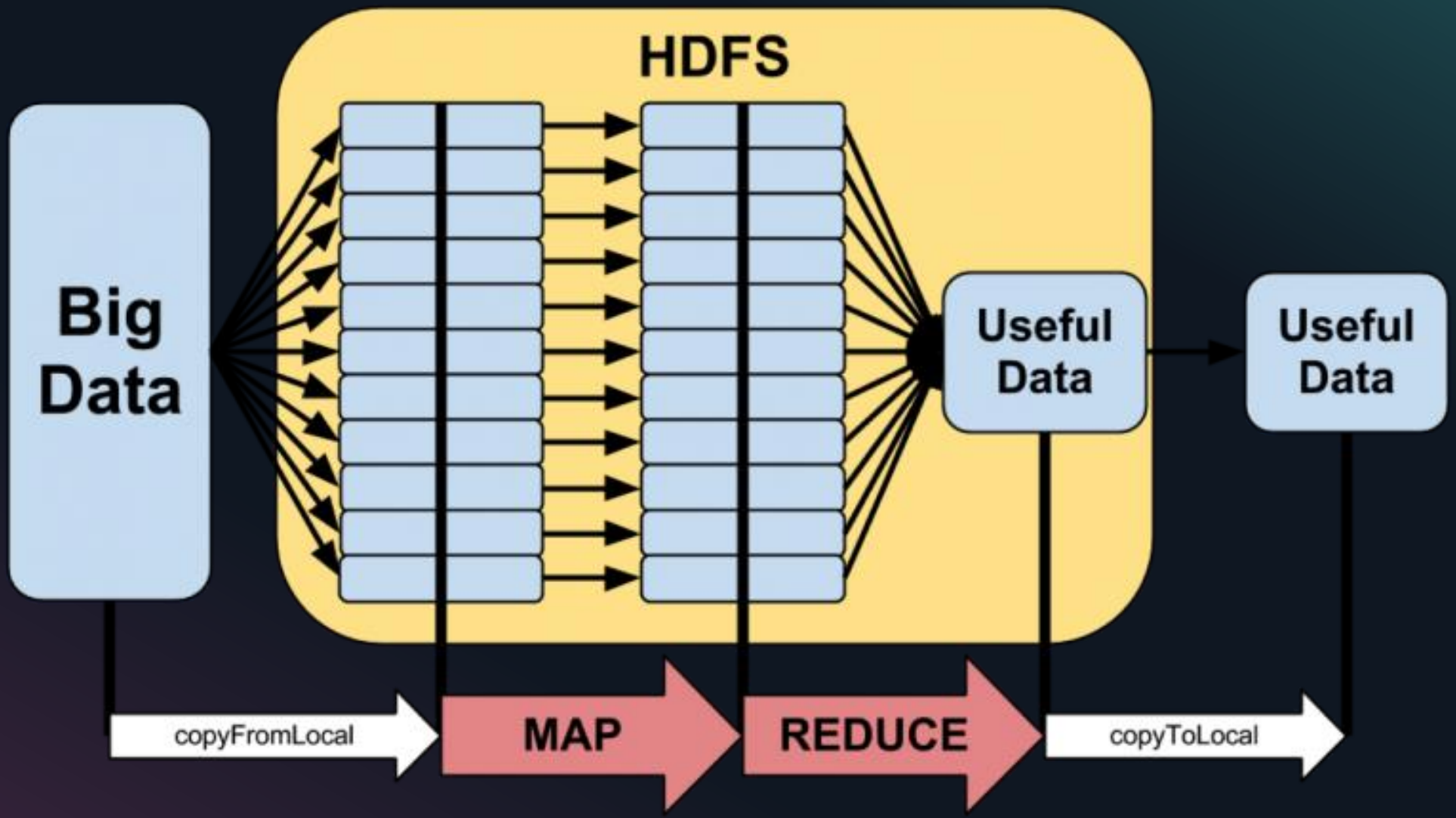
Value

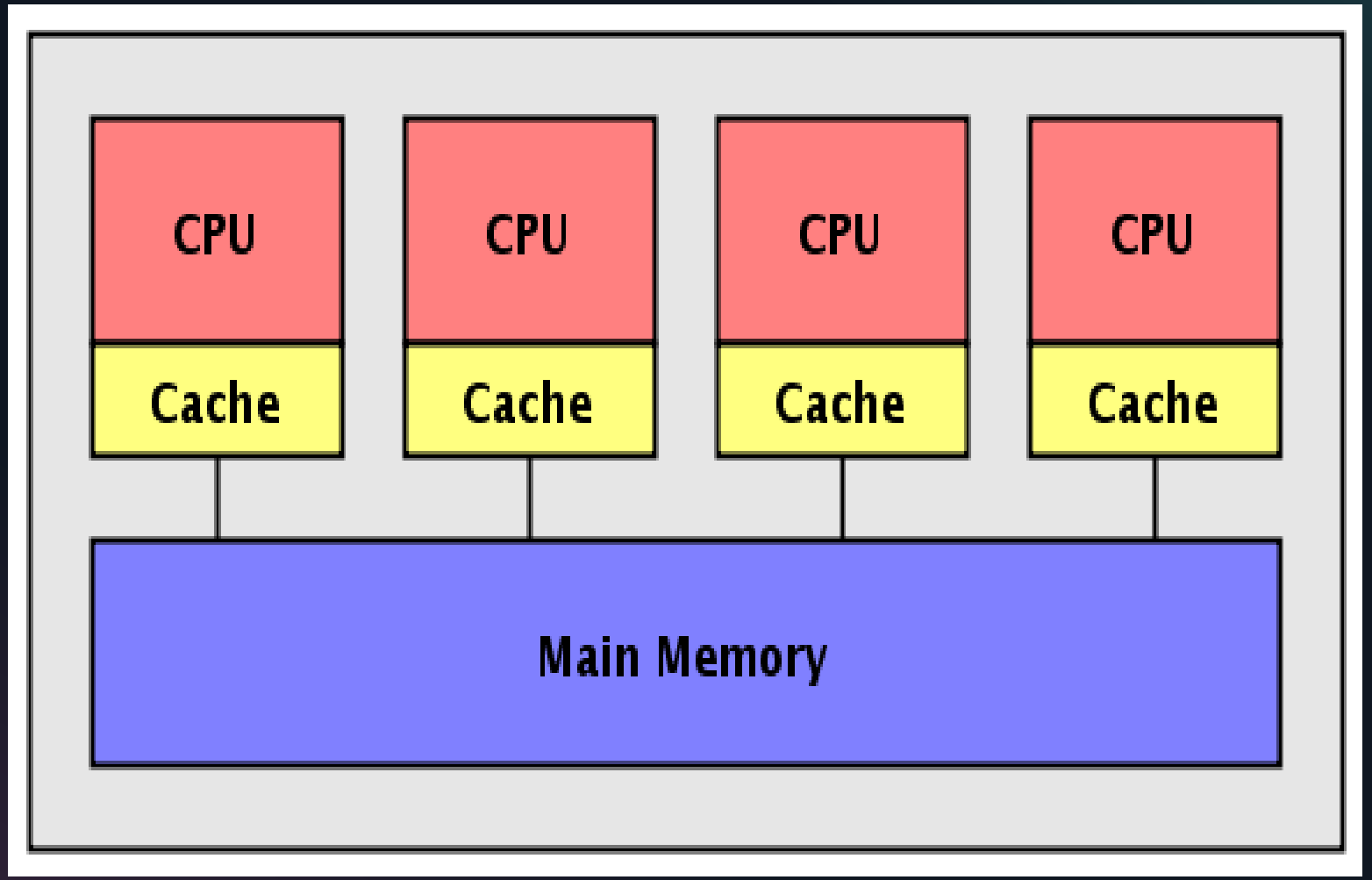
Finding correct meaning out of the data

Veracity

Min	Max	Mean	SD
4.3	?	5.84	0.83
2.0	4.4	3.05	[illegible]
15000	7.9	1.20	0.43
0.1	2.5	?	0.76

Uncertainty and inconsistencies in the data


HDFS
Big Data
Useful Data
Useful Data
copyFromLocal
MAP
REDUCE
copyToLocal


CPU
CPU
CPU
CPU
Cache
Cache
Cache
Cache
Main Memory

Distributed graph processing with Pregel and ArangoDB

Batch Graph Processing

- Sometimes graph query language can not cover all use-cases
- We want execute algorithms on **entire** graph at **once**

Reuse general highly parallel distributed processing systems

- E.g. Google's MapReduce or Apache Hadoop
- It can be difficult to adapt graph algorithms to a general model

Specialized systems, i.e. graph focused computing model

- Google's Pregel / Apache Giraph
- Apache Spark with **GraphX** module
- Apache Flink with **Gelly** module

Large Graphs are increasingly common in many applications:

- Social networks form graphs
  - Facebook's Friends Graph
  - Twitter's follower network
- The Web can be modeled as graph
- This data needs to be stored and processed

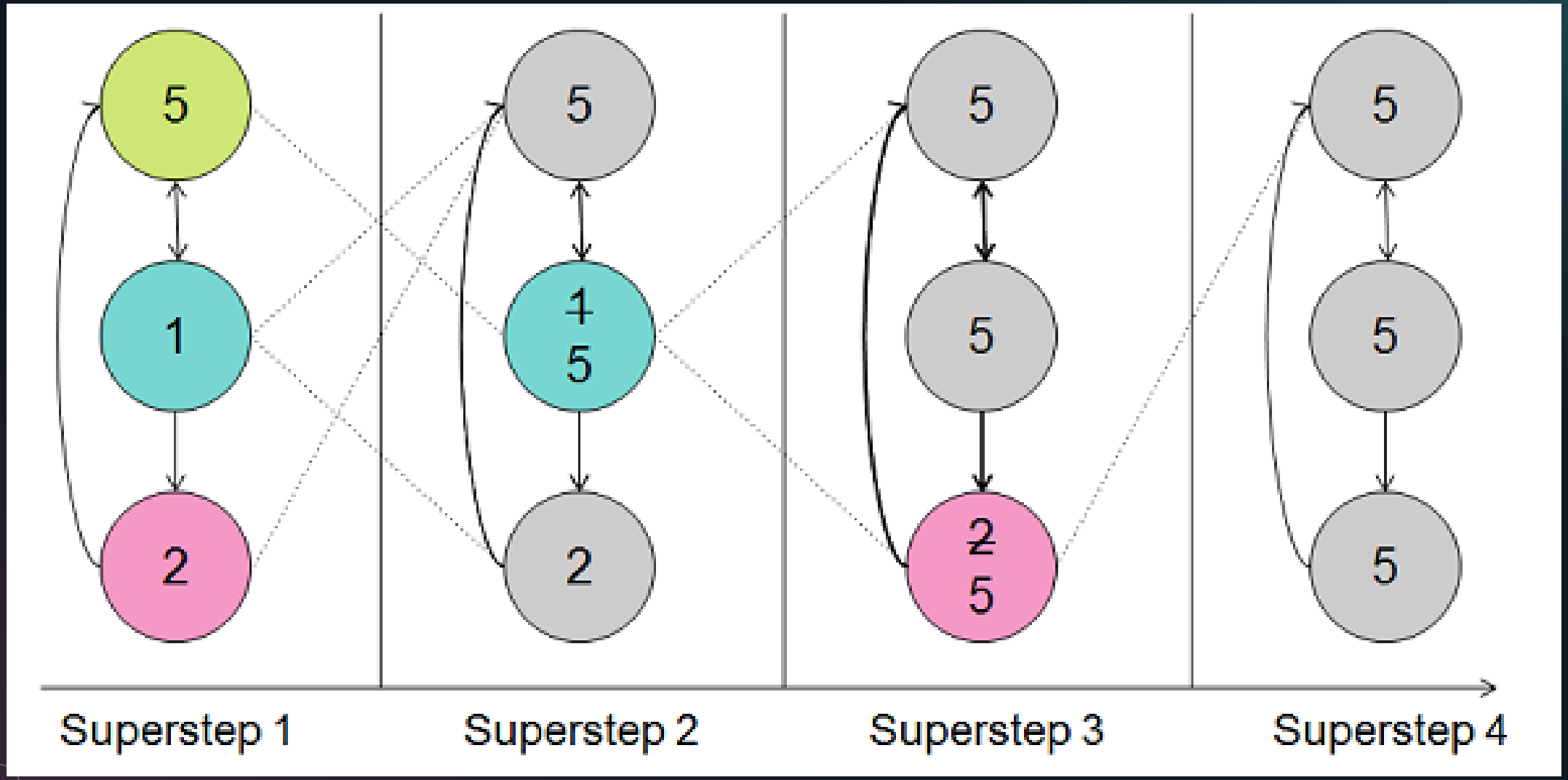
Network	Nodes	Edges
Facebook	2.06 billion (users)	hundreds of billions
Twitter	328 million (users)	65 billion
Web Pages[1]	1.7 billion	64 billion (hyperlinks)

What is Pregel ?

- A distributed message passing system specialized for large-scale graph computations
- Exposes a specialized API, where vertices and Edges are first class citizens
  - As opposed to more general systems like MapReduce / Hadoop or even Apache Spark without GraphX

Historical Background: Bulk Synchronous Parallel Computers

- **BSP** is an abstract model of a parallel computer by Leslie Valiant
  - A number of machines which can independently perform computations
  - Machines can exchange messages over a network
  - The computation is divided into **supersteps**, between steps is a global synchronized barrier.
- All messages sent during one superstep are *guaranteed* to be received


5
1
2
5
1
5
2
5
2
5
5
5
5
Superstep 1
Superstep 2
Superstep 3
Superstep 4