Introduction to Network Analysis

Filipe Teixeira

GHENT UNIVERSITY

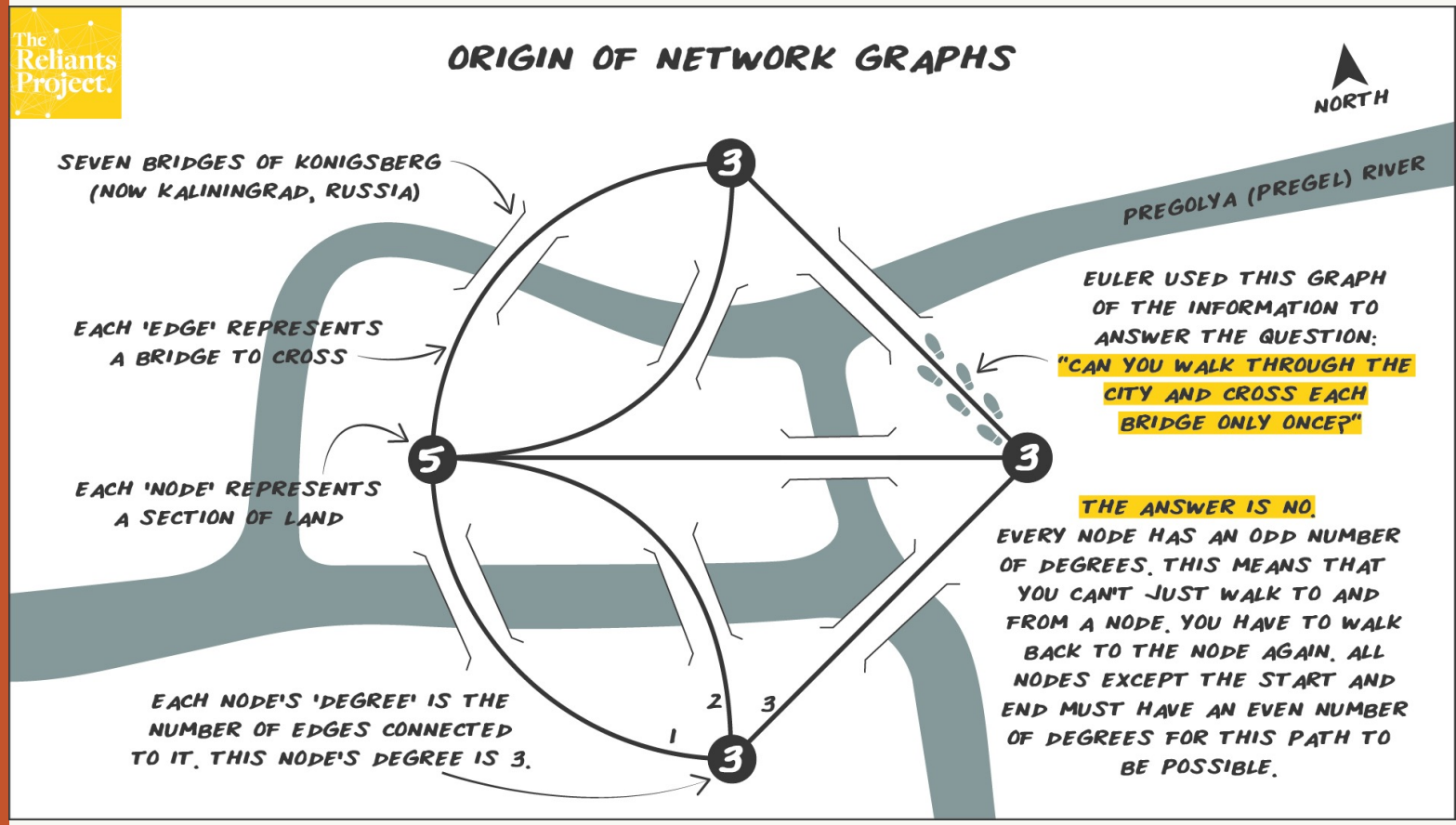
FLAMES

“In God we trust, all others must bring data.”

W. Edwards Deming

GRAPH THEORY

THE BRIDGES OF KÖNIGSBERG



Networks are everywhere, from the Internet, to social networks, and the genetic networks that determine our biological existence. (...)

From the origins of the six degrees of separation to explaining why networks are robust to random failures, (...) how viruses like Ebola and H1N1 spread, and why it is that our friends have more friends than we do.

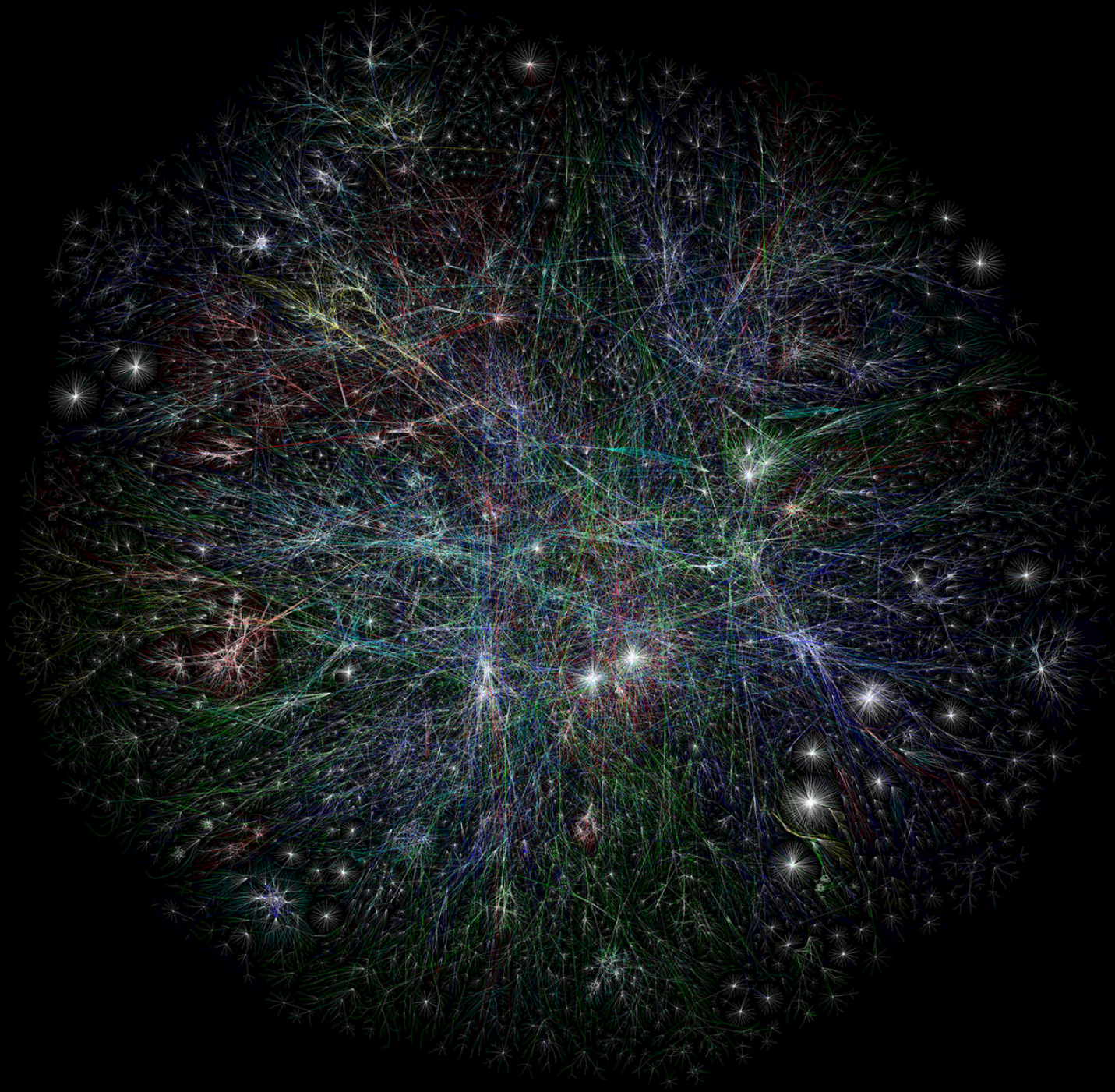
- Albert-Lásló Barabási

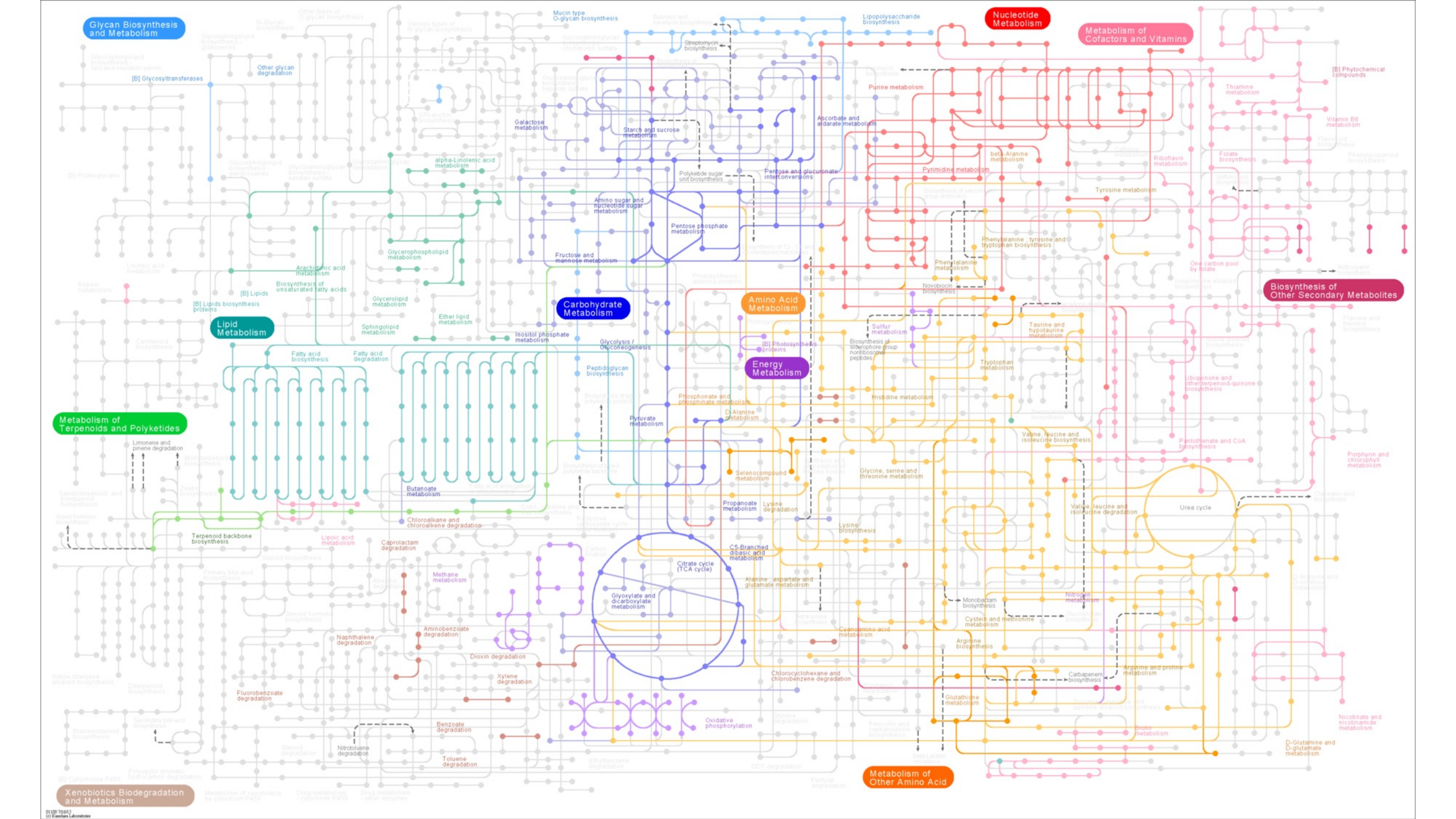
NETWORKS...
NETWORKS EVERYWHERE

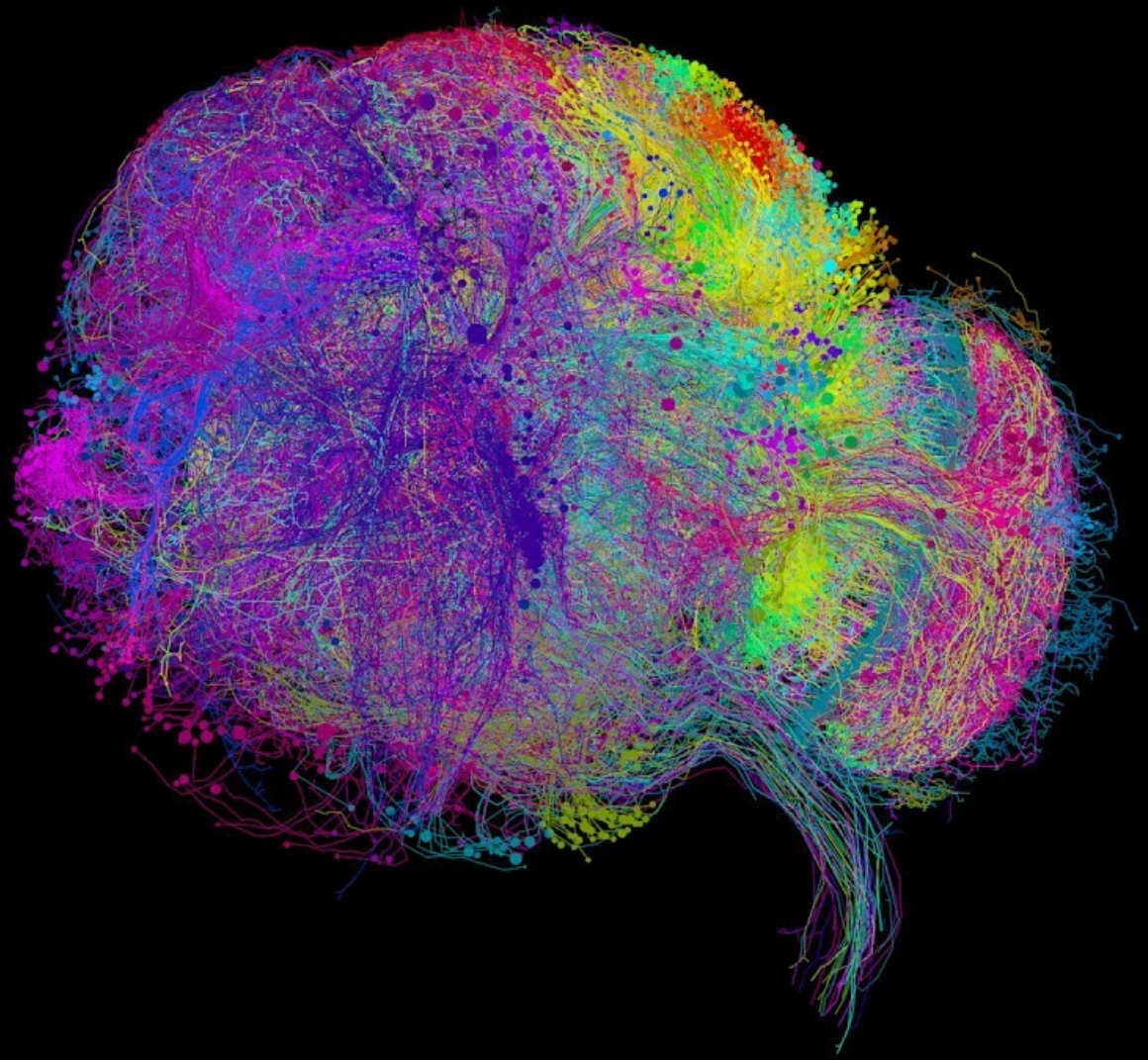
NETWORKS...
NETWORKS EVERYWHERE

Next: Cliché royalty free video showing traffic, people using Facebook and drone shots of cities at night.

credit: Martin Grandjean




Glycan Biosynthesis and Metabolism
Nucleotide Metabolism
Metabolism of Cofactors and Vitamins
Carbohydrate Metabolism
Amino Acid Metabolism
Lipid Metabolism
Energy Metabolism
Biosynthesis of Other Secondary Metabolites
Metabolism of Terpenoids and Polyketides
Xenobiotics Biodegradation and Metabolism
Metabolism of Other Amino Acid
Mucin type O-glycan biosynthesis
Lipopolysaccharide biosynthesis
Streptomycin biosynthesis
Other glycan degradation
[B] Glycosyltransferases
Galactose metabolism
Starch and sucrose metabolism
Ascorbate and aldarate metabolism
Purine metabolism
[B] Phytochemical compounds
Thiamine metabolism
Vitamin B6 metabolism
alpha-Linolenic acid metabolism
Polyketide sugar unit biosynthesis
Pentose and glucuronate interconversions
beta-Alanine metabolism
Pyrimidine metabolism
Riboflavin metabolism
Folate biosynthesis
Amino sugar and nucleotide sugar metabolism
Pentose phosphate metabolism
Tyrosine metabolism
Glycerophospholipid metabolism
Fructose and mannose metabolism
Phenylalanine, tyrosine and tryptophan biosynthesis
Phenylalanine metabolism
Arachidonic acid metabolism
Biosynthesis of unsaturated fatty acids
[B] Lipids
Novobiocin biosynthesis
One carbon pool by folate
[B] Lipids biosynthesis proteins
Glycerolipid metabolism
Ether lipid metabolism
Sphingolipid metabolism
Inositol phosphate metabolism
Glycolysis / Gluconeogenesis
Sulfur metabolism
Taurine and hypotaurine metabolism
[B] Photosynthesis proteins
Biosynthesis of siderophore group nonribosomal peptides
Fatty acid biosynthesis
Fatty acid degradation
Peptidoglycan biosynthesis
Tryptophan metabolism
Ubiquinone and other terpenoid-quinone biosynthesis
Phosphonate and phosphinate metabolism
Histidine metabolism
D-Alanine metabolism
Pyruvate metabolism
Limonene and pinene degradation
Valine, leucine and isoleucine biosynthesis
Pantothenate and CoA biosynthesis
Porphyrin and chlorophyll metabolism
Selenocompound metabolism
Glycine, serine and threonine metabolism
Butanoate metabolism
Propanoate metabolism
Lysine degradation
Valine, leucine and isoleucine degradation
Urea cycle
Chloroalkane and chloroalkene degradation
Lysine biosynthesis
Terpenoid backbone biosynthesis
Lipoic acid metabolism
Caprolactam degradation
C5-Branched dibasic acid metabolism
Citrate cycle (TCA cycle)
Alanine, aspartate and glutamate metabolism
Methane metabolism
Glyoxylate and dicarboxylate metabolism
Monobactam biosynthesis
Nitrogen metabolism
Cysteine and methionine metabolism
Aminobenzoate degradation
Cyanoamino acid metabolism
Naphthalene degradation
Arginine biosynthesis
Dioxin degradation
Xylene degradation
Chlorocyclohexane and chlorobenzene degradation
Carbapenem biosynthesis
Arginine and proline metabolism
Fluorobenzoate degradation
Glutathione metabolism
Oxidative phosphorylation
Nicotinate and nicotinamide metabolism
Benzoate degradation
Biotin metabolism
Nitrotoluene degradation
D-Glutamine and D-glutamate metabolism
Toluene degradation



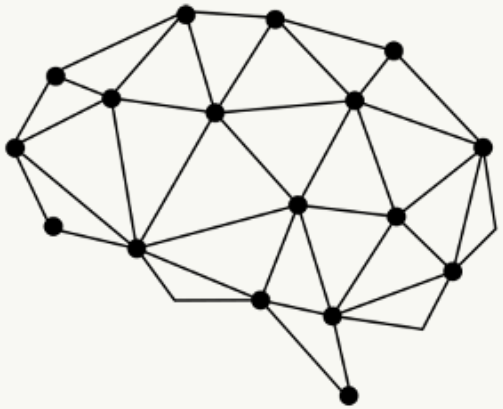
NOW WHAT?

NAILS...
NAILS EVERYWHERE

"If all you have is a hammer, everything looks like a nail"

- Abraham Maslow

- Network Analysis is being broadly overused.
- At the same time not used enough...



Cambridge Analytica

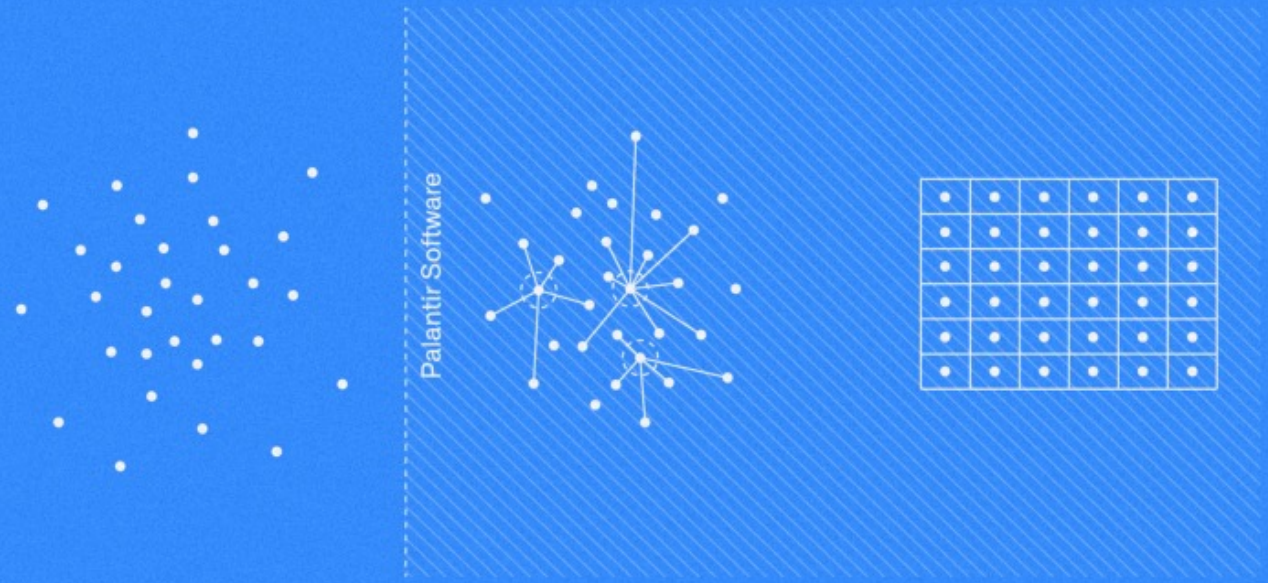
Palantir

PALANTIR EXPLAINED SERIES | ISSUE № 1

Palantir is Not a Data Company →

Palantir Technologies | Security + Functionality | palantir.com

Specific Client Data



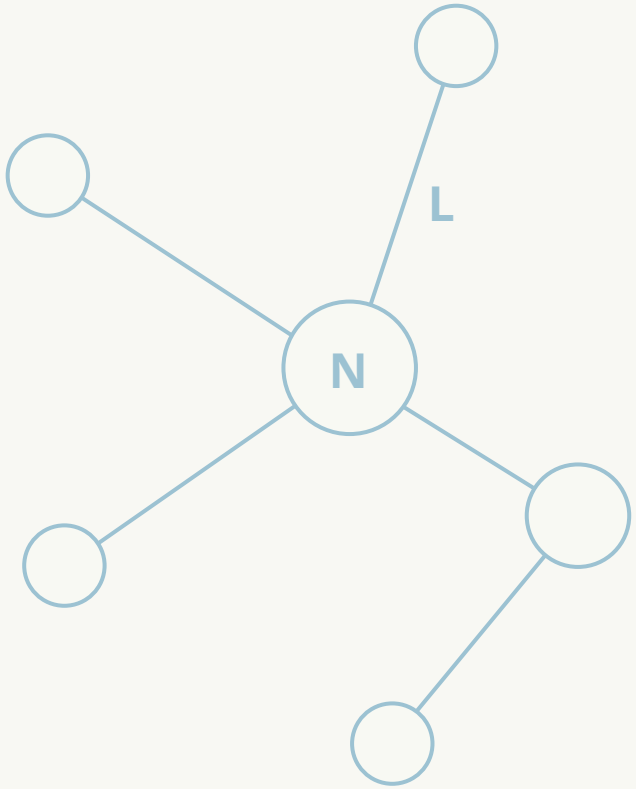
01
Fragmentation

02
Connection

03
Organization
& Security

IT’S A KIND OF MAGIC

Network Statistics

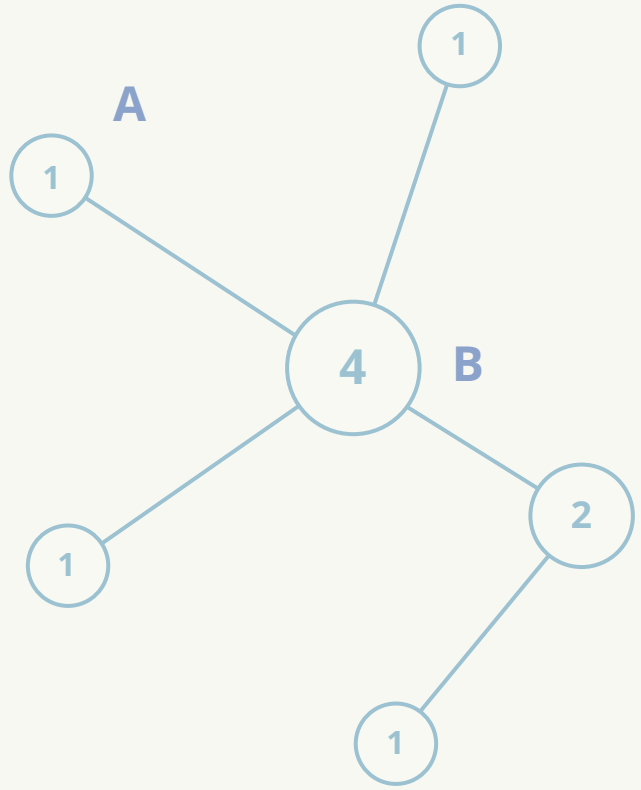


Components (N): nodes, vertices

Interactions (L): links, edges

System (N,L): network, graph

Network Statistics

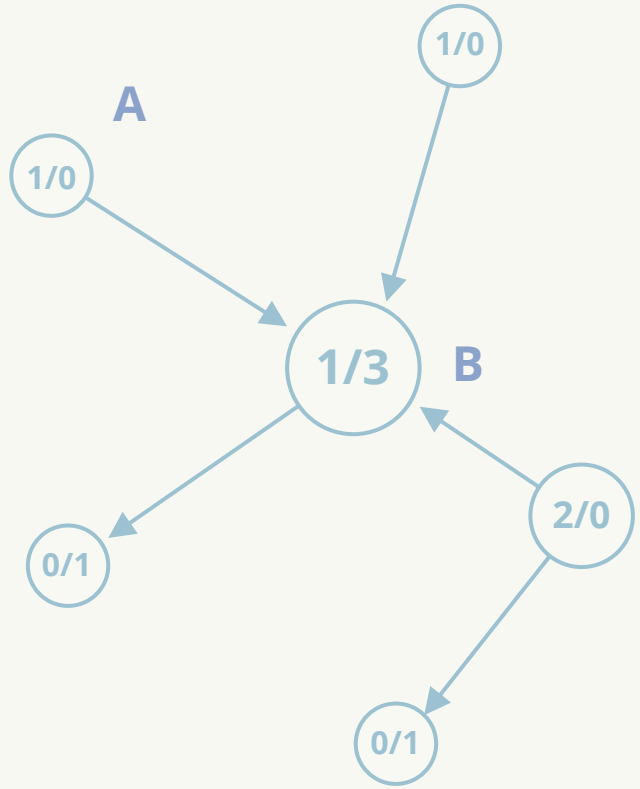


Degree

(undirected)

$k_A = 1$ $k_B = 4$

Network Statistics



Degree

(directed)

$k_{A\_out} = 1$

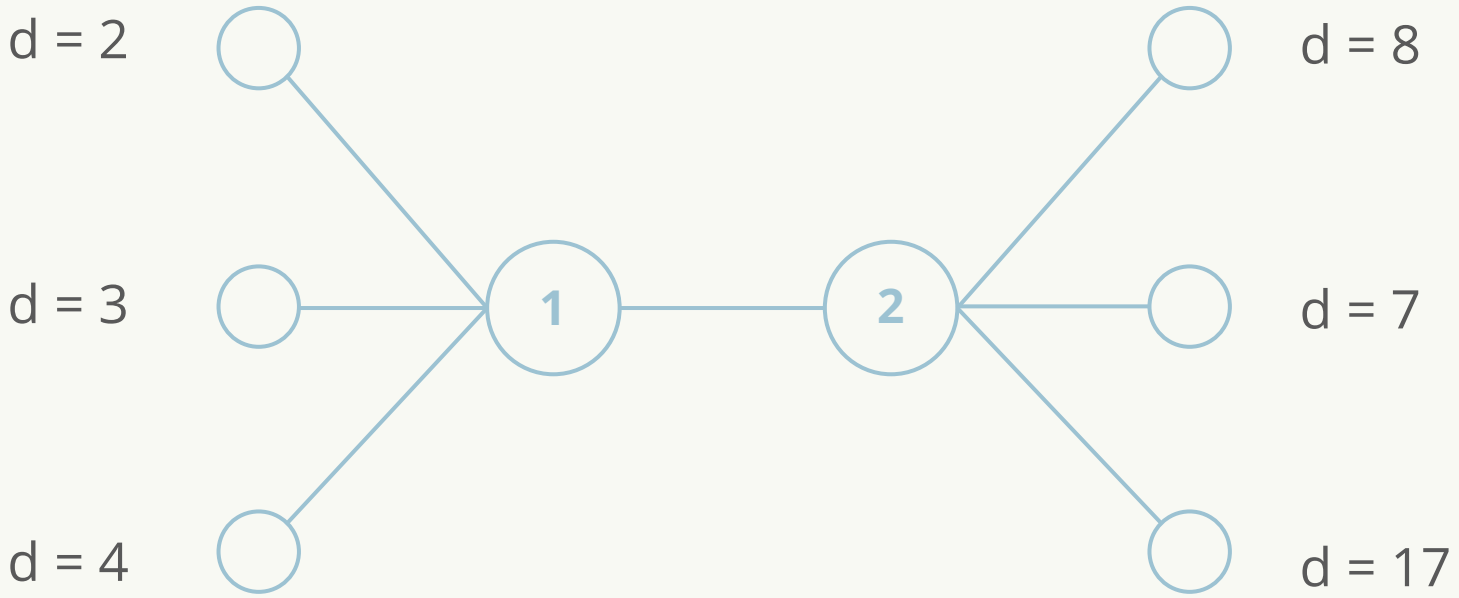
$k_{A\_in} = 0$

$k_{B\_out} = 1$

$k_{B\_in} = 3$

More statistics

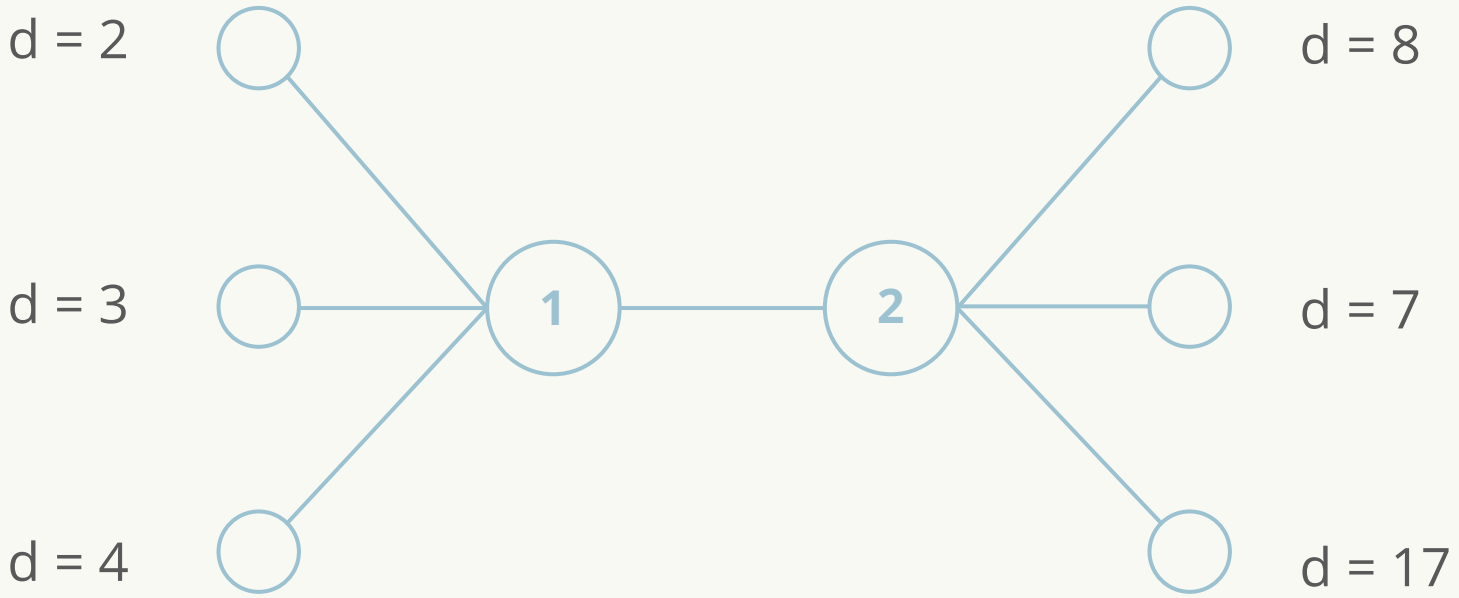
Degree is a local measure



N_1 and N_2 have the same degree

More statistics

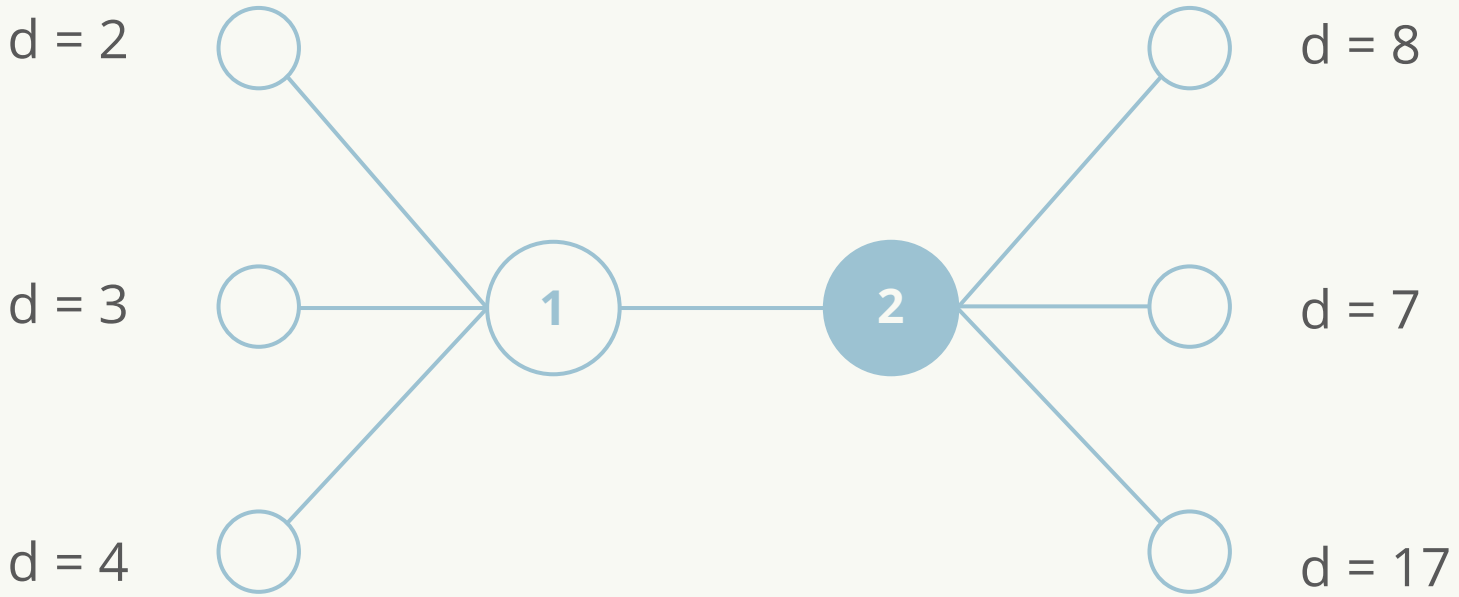
Eigenvector centrality



How many links do a node’ connections (neighbors) have

More statistics

Eigenvector centrality



N_2 will rank better than N_1

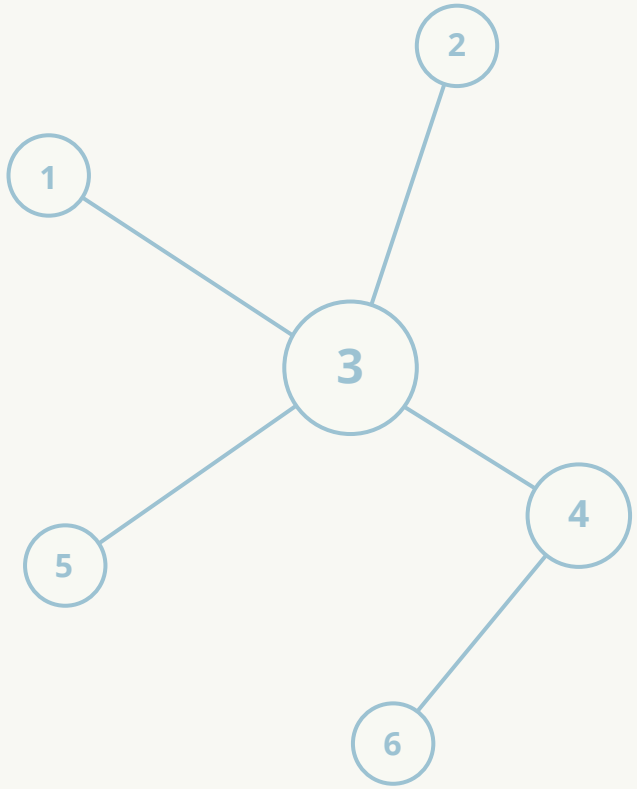
PATHOLOGY

(not the attempt of a bad pun)

Paths

The distance (**shortest path, geodesic path**) between two nodes is defined as the number of edges along the shortest path connecting them.

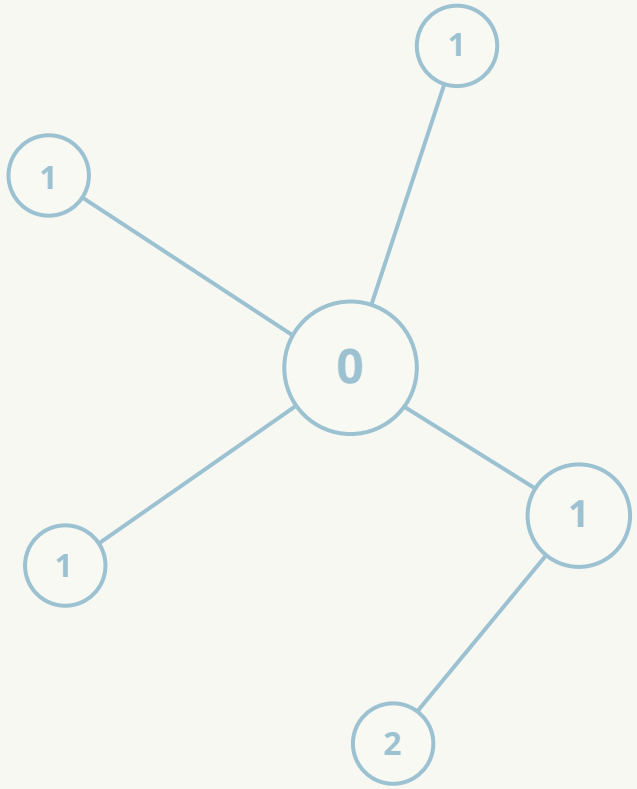
*If the two nodes are disconnected, the distance is infinity.



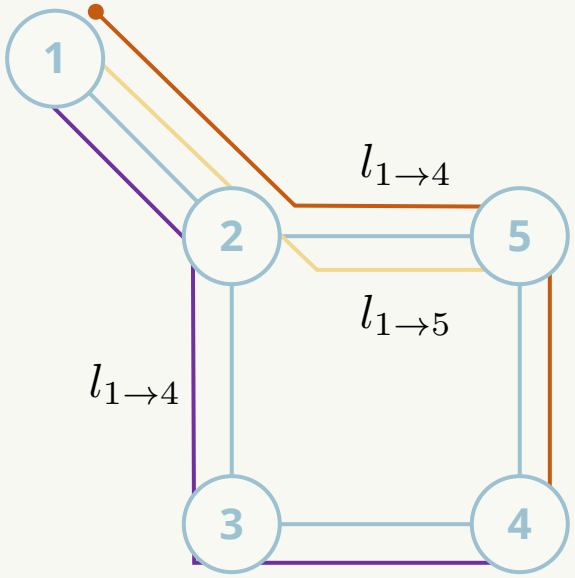
Paths

The distance (**shortest path, geodesic path**) between two nodes is defined as the number of edges along the shortest path connecting them.

*If the two nodes are disconnected, the distance is infinity.



Shortest Path



$$l_{1\to4} = 3$$

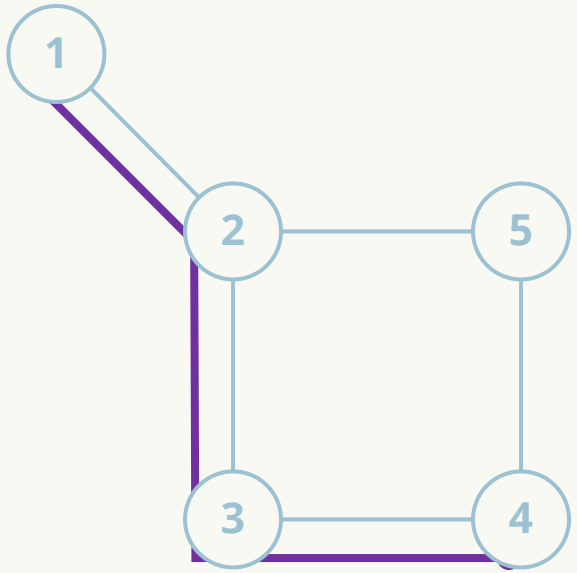
$$l_{1\to5} = 2$$

The path with the shortest length between two nodes (distance).

Paths

Paths

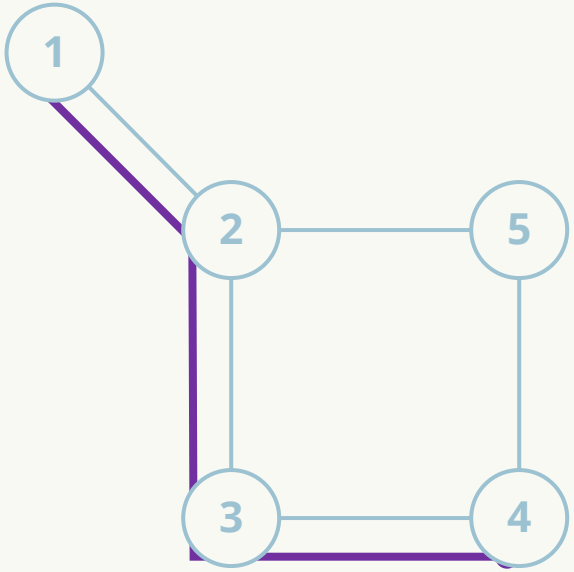
Diameter



$$l_{1\to 4} = 3$$

The longest shortest path in a graph

Average Path Length

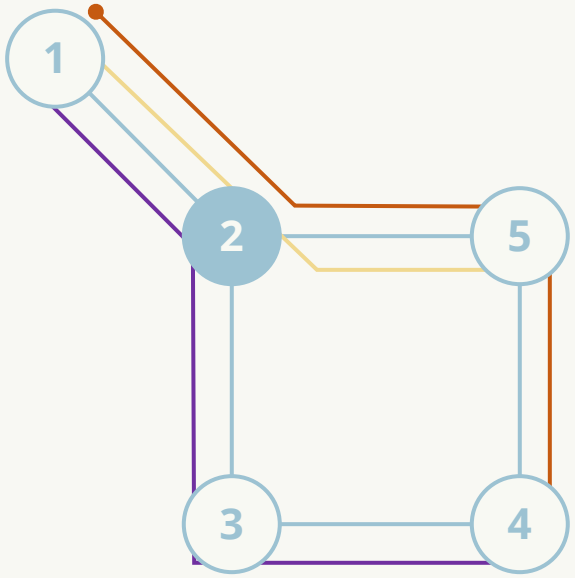


$$(l_{1\to 2} + l_{1\to 3} + l_{1\to 4} + \\ + l_{1\to 5} + l_{2\to 3} + l_{2\to 4} + \\ + l_{2\to 5} + l_{3\to 4} + l_{3\to 5} + \\ + l_{4\to 5})\,/10 = 1.6$$

The average of the shortest paths for all pairs of nodes.

Paths (sort of)

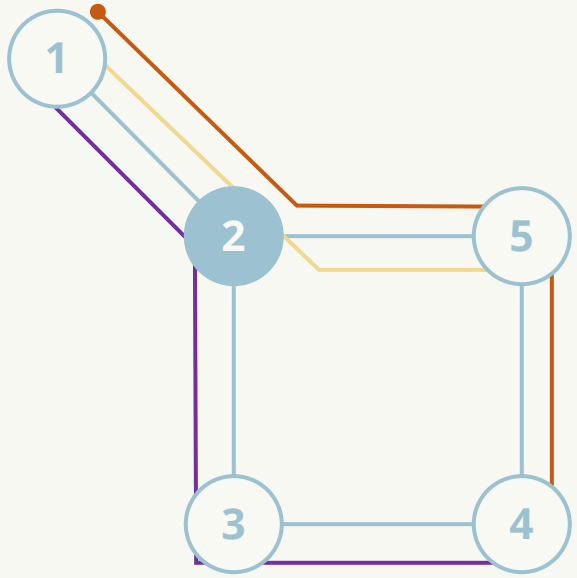
Closeness centrality



Shortest distance from graph nodes

(or average path length)

Betweenness centrality



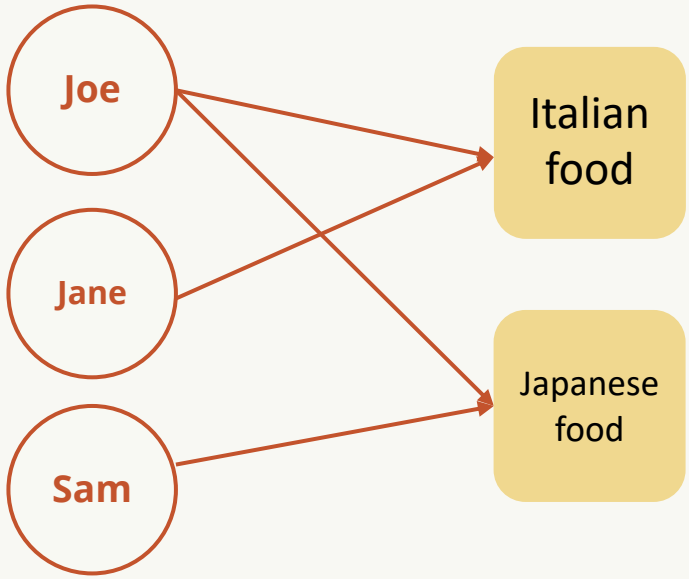
percentage of shortest paths where
our studied node lies in them

(information flow)

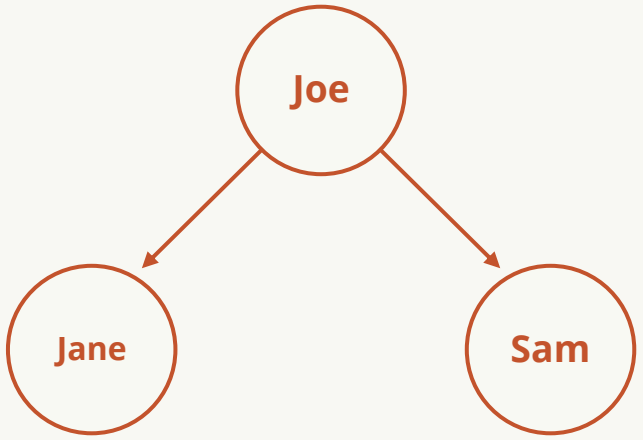
Paths (sort of)

THERE’S MORE

One possible network

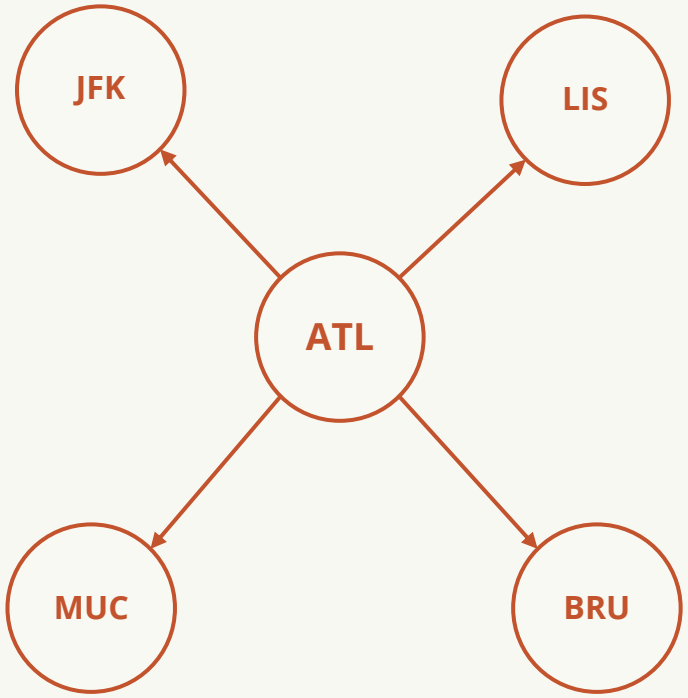


Another possibility

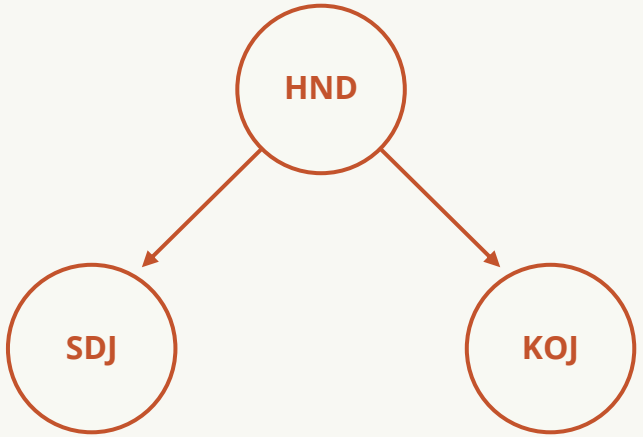


More than one!!!

Hub and spoke



Point-to-point



More than one!!!

RESOURCES

https://project-awesome.org/briatte/awesome-network-analysis

R(rrrr)

(ok this one is a bad pun)

R universe

Questions???

Type in the chat box or turn on your microphone

RStudio