

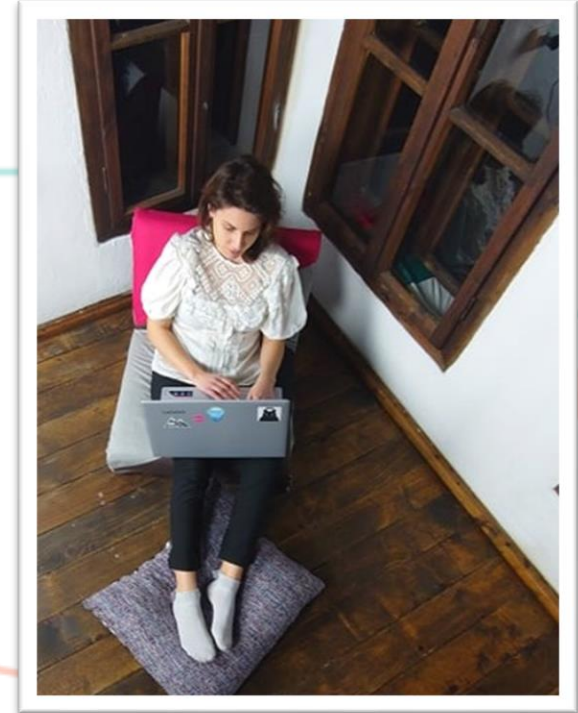
The background features a network of nodes and lines. A teal line with circular nodes runs horizontally across the top. An orange line with circular nodes runs horizontally across the middle. A blue line with circular nodes runs horizontally across the bottom. On the right side, there is a complex network of grey nodes connected by thin grey lines, with some nodes highlighted in teal, orange, and blue to match the lines.

# Introduction to Graph Databases

Knowledge bases course  
Master program of AI  
Sofia University, Bulgaria

# About me

- Melania Berbatova
- Master of AI'2020
- Freelance data scientist and
- Previously – ML engineer @Sciant and @ReceiptBank



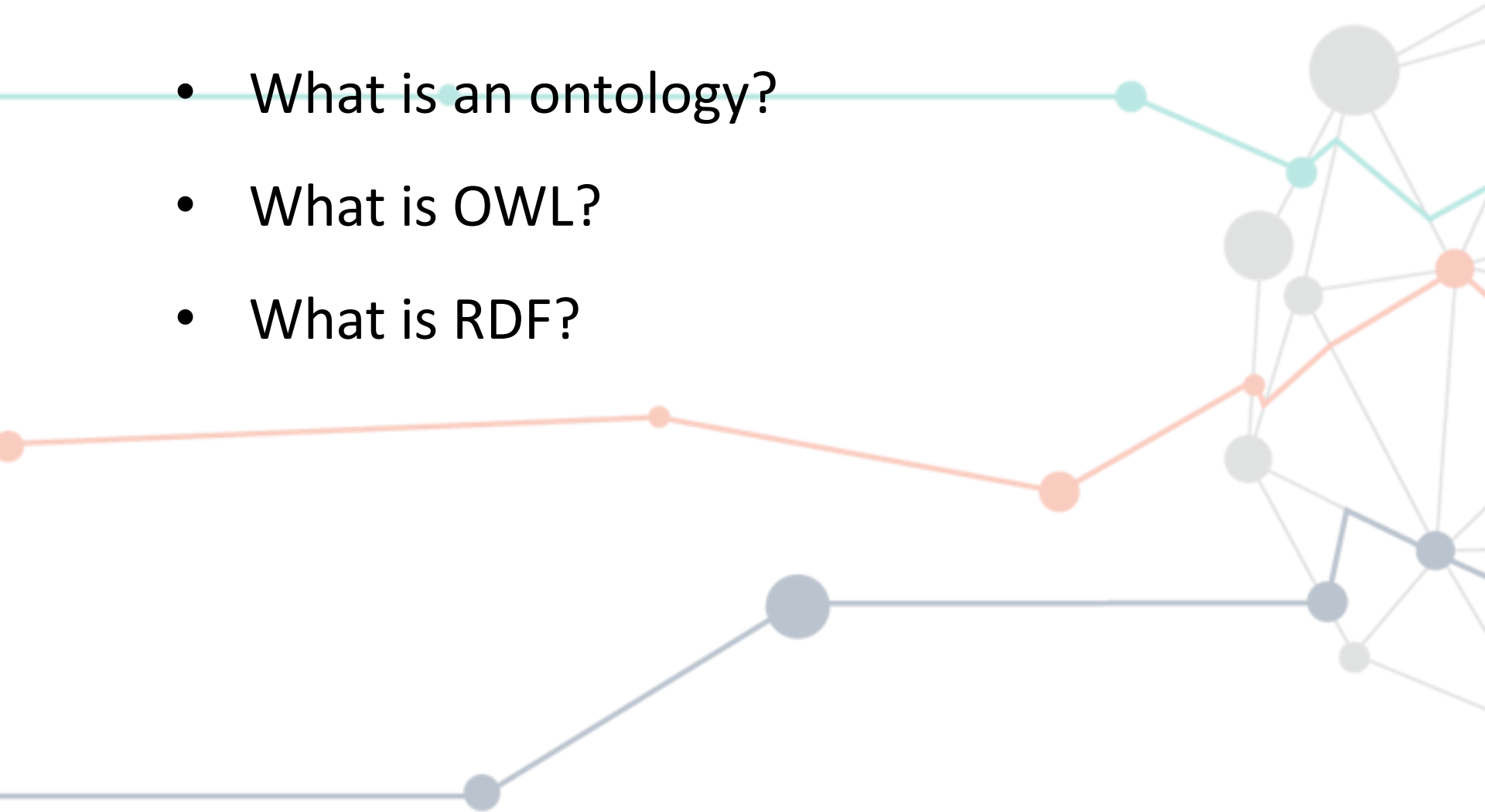
- Email me: [melania.berbatova@gmail.com](mailto:melania.berbatova@gmail.com)
- GitHub: <http://github.com/melaniab>

About you?



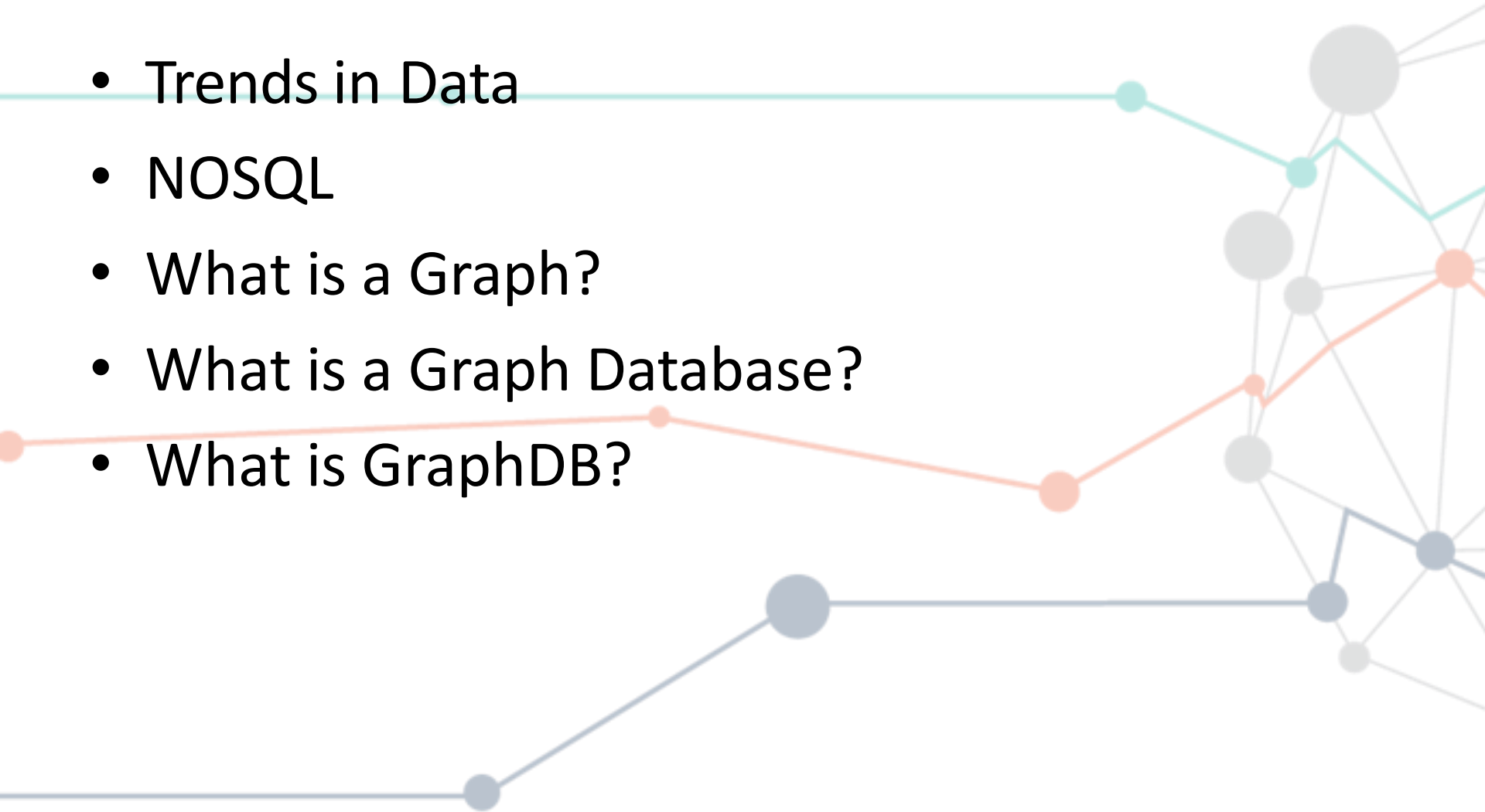
# From previous semester

- What is an ontology?
- What is OWL?
- What is RDF?



# Agenda

- Trends in Data
- NOSQL
- What is a Graph?
- What is a Graph Database?
- What is GraphDB?



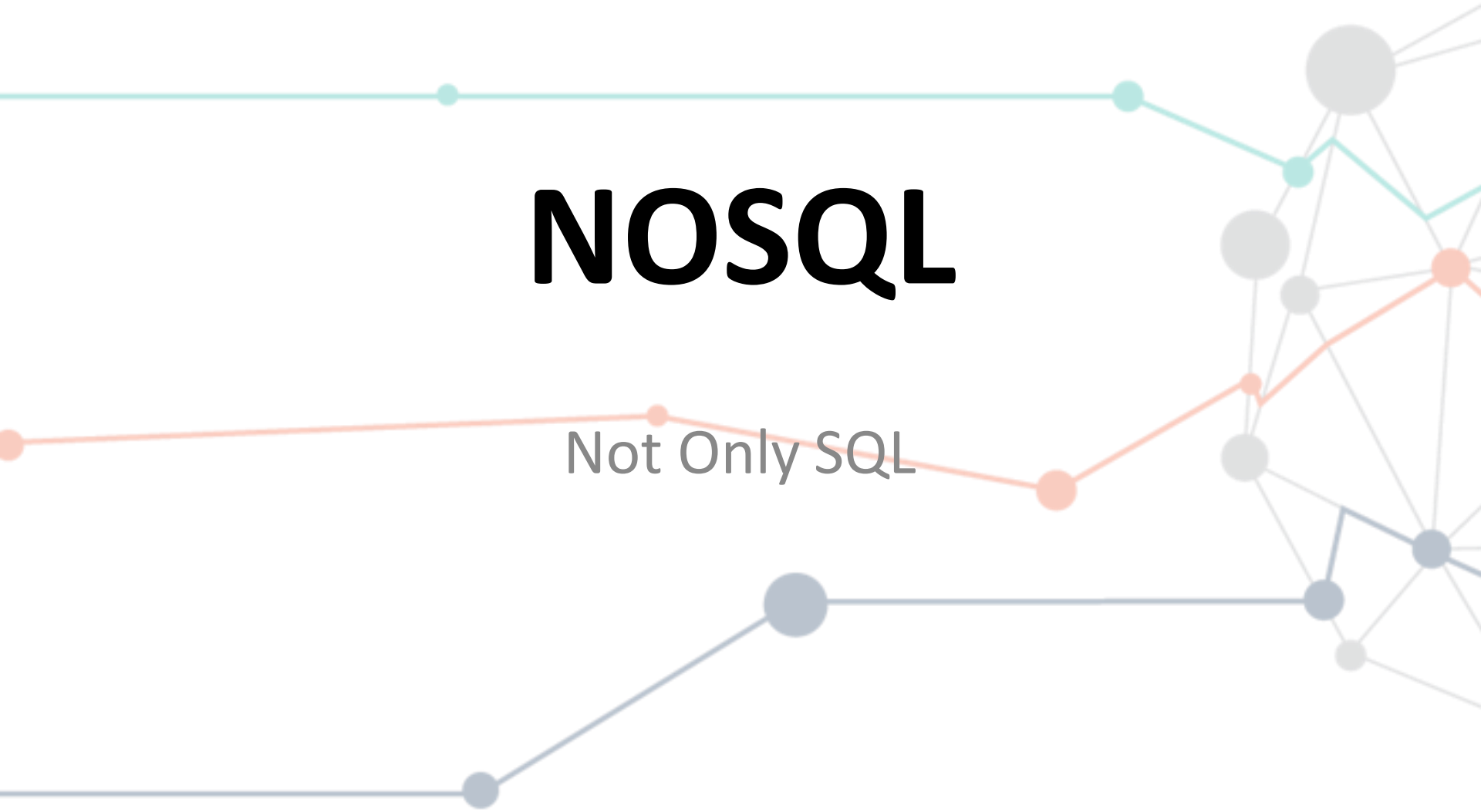
# Data is more Semi-Structured:

- If you tried to collect all the data of every movie ever made, how would you model it?
- Actors, Characters, Locations, Dates, Costs, Ratings, Showings, Ticket Sales, etc.



# NOSQL

Not Only SQL



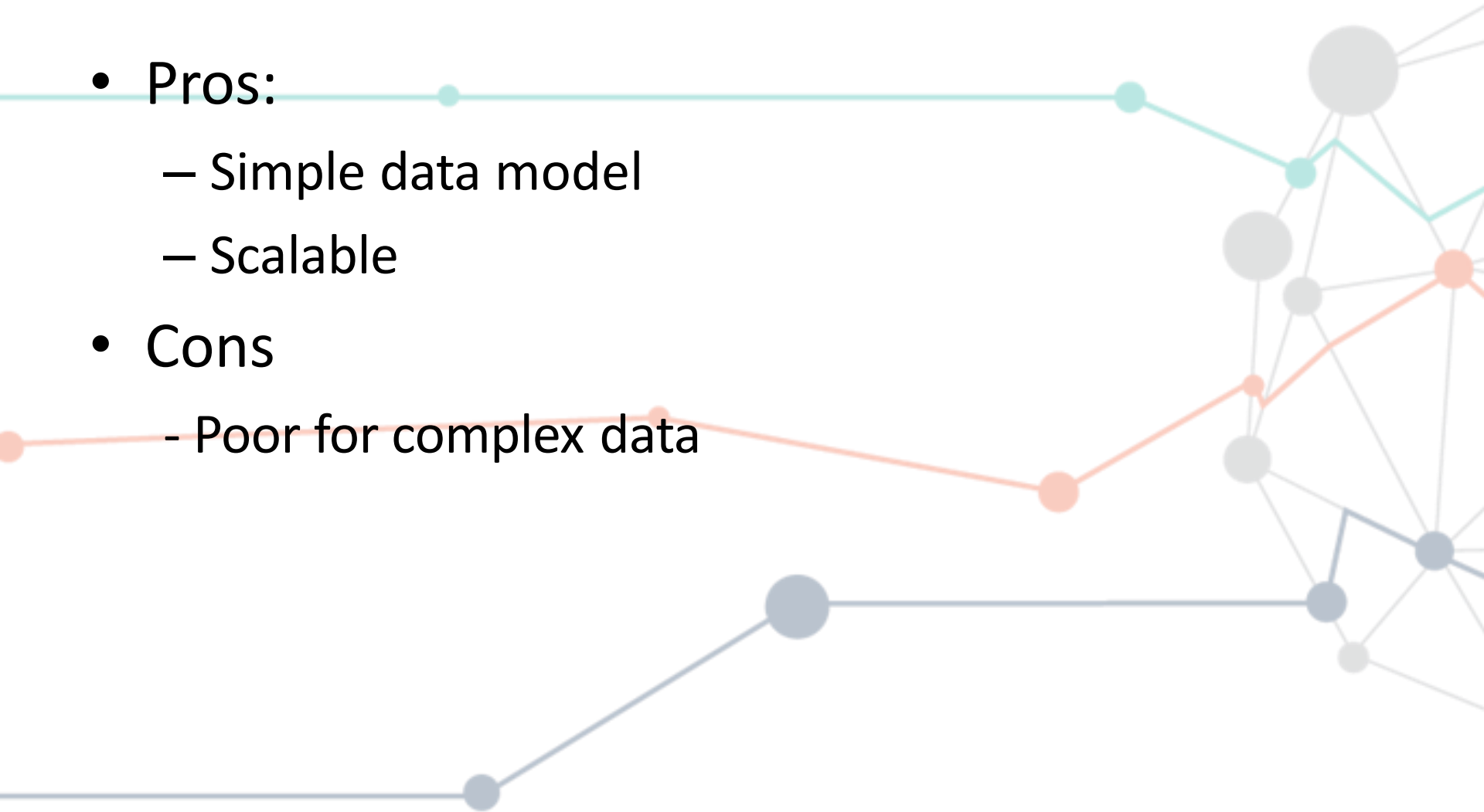
# NoSQL





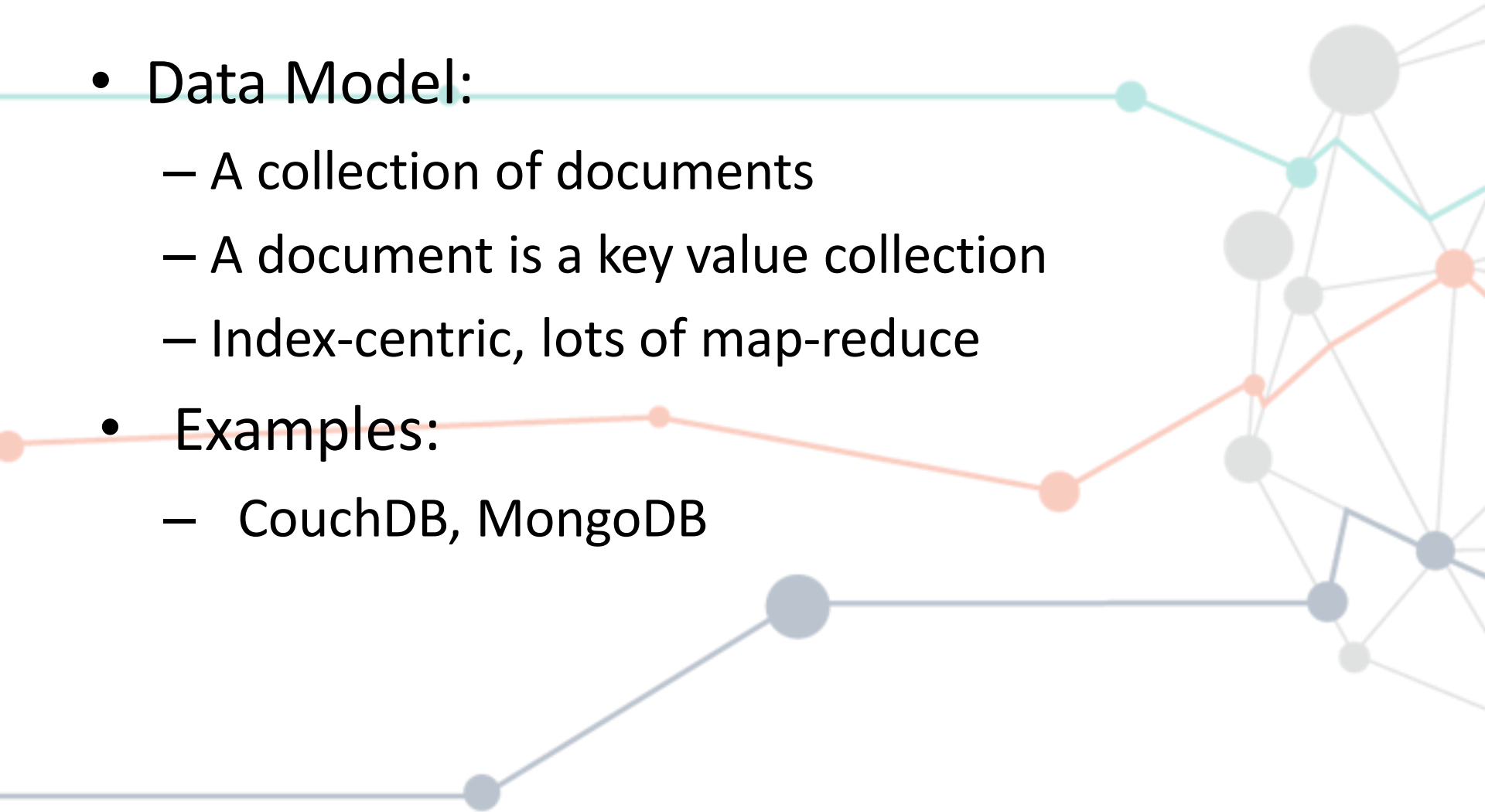
# Key Value Stores: Pros and Cons

- Pros:
  - Simple data model
  - Scalable
- Cons
  - Poor for complex data



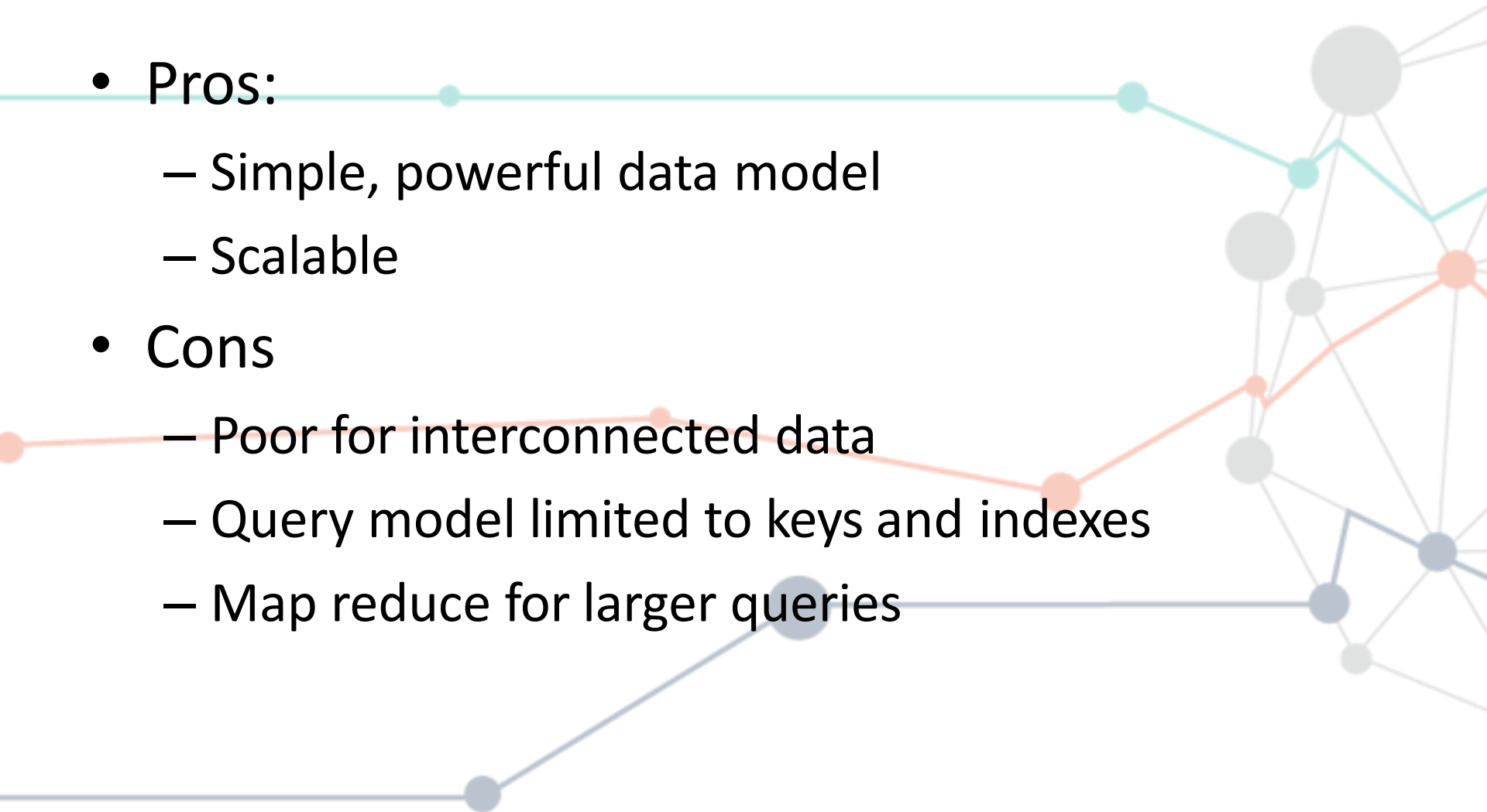
# Document Databases

- Data Model:
  - A collection of documents
  - A document is a key value collection
  - Index-centric, lots of map-reduce
- Examples:
  - CouchDB, MongoDB



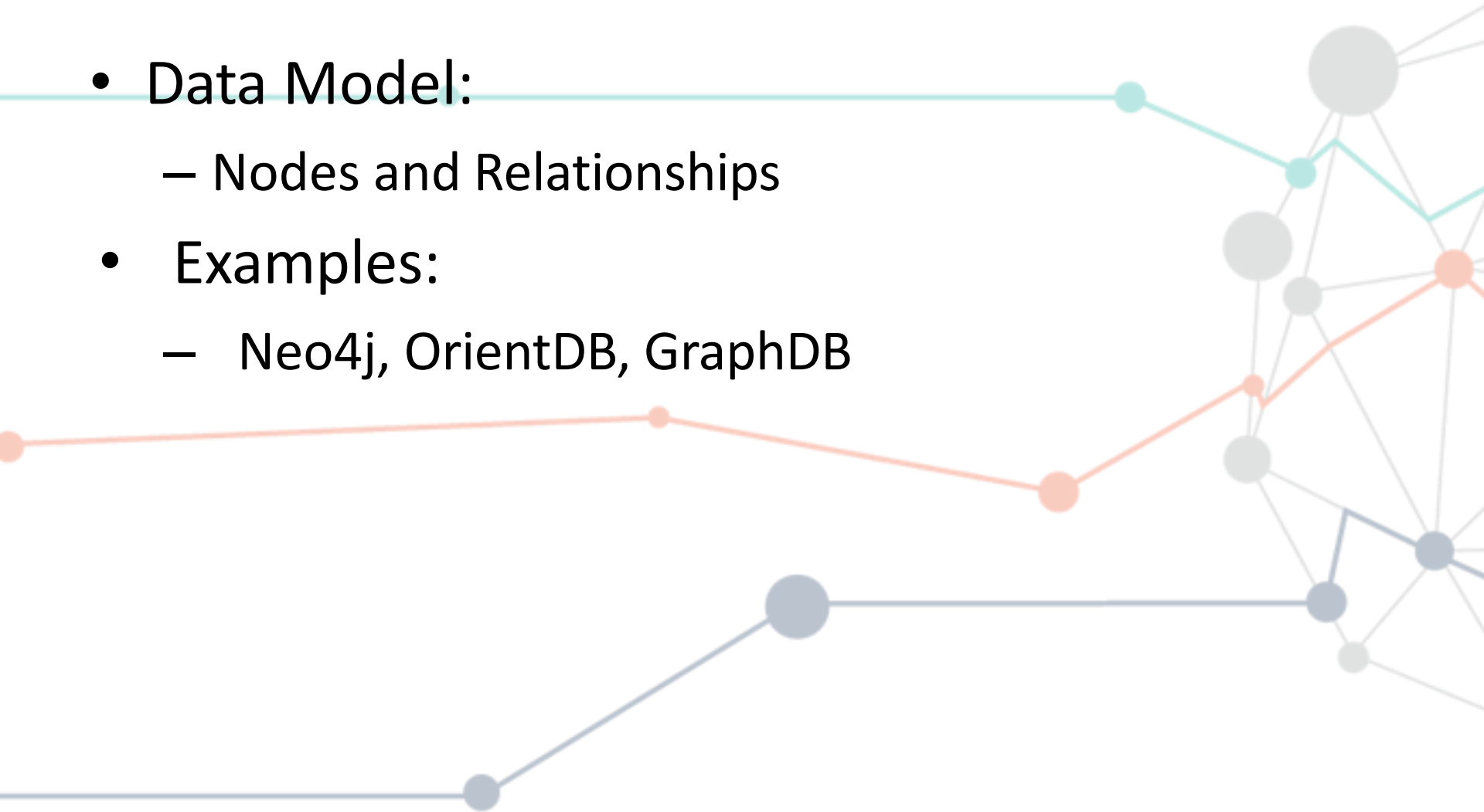
# Document Databases: Pros and Cons

- Pros:
  - Simple, powerful data model
  - Scalable
- Cons
  - Poor for interconnected data
  - Query model limited to keys and indexes
  - Map reduce for larger queries



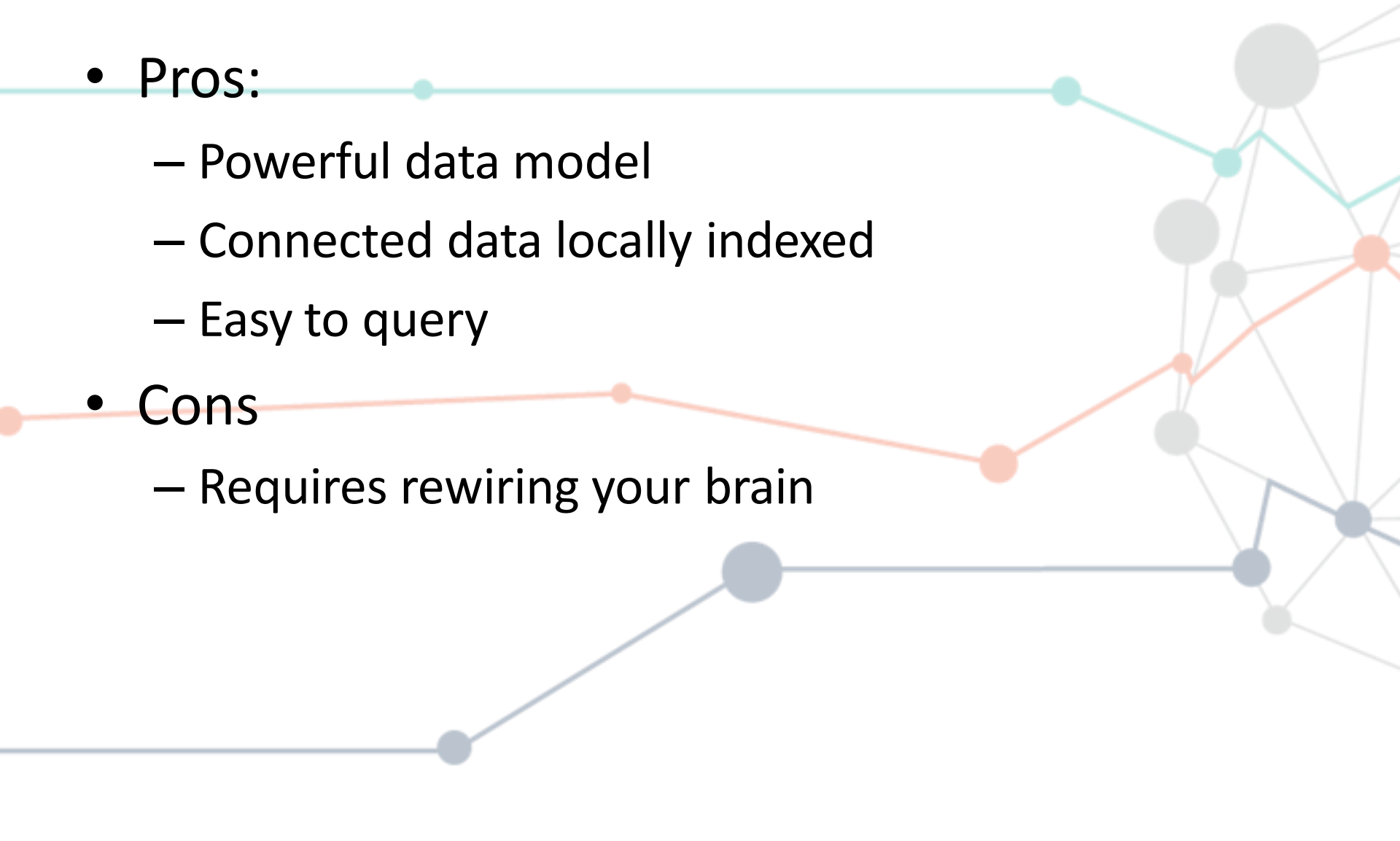
# Graph Databases

- Data Model:
  - Nodes and Relationships
- Examples:
  - Neo4j, OrientDB, GraphDB

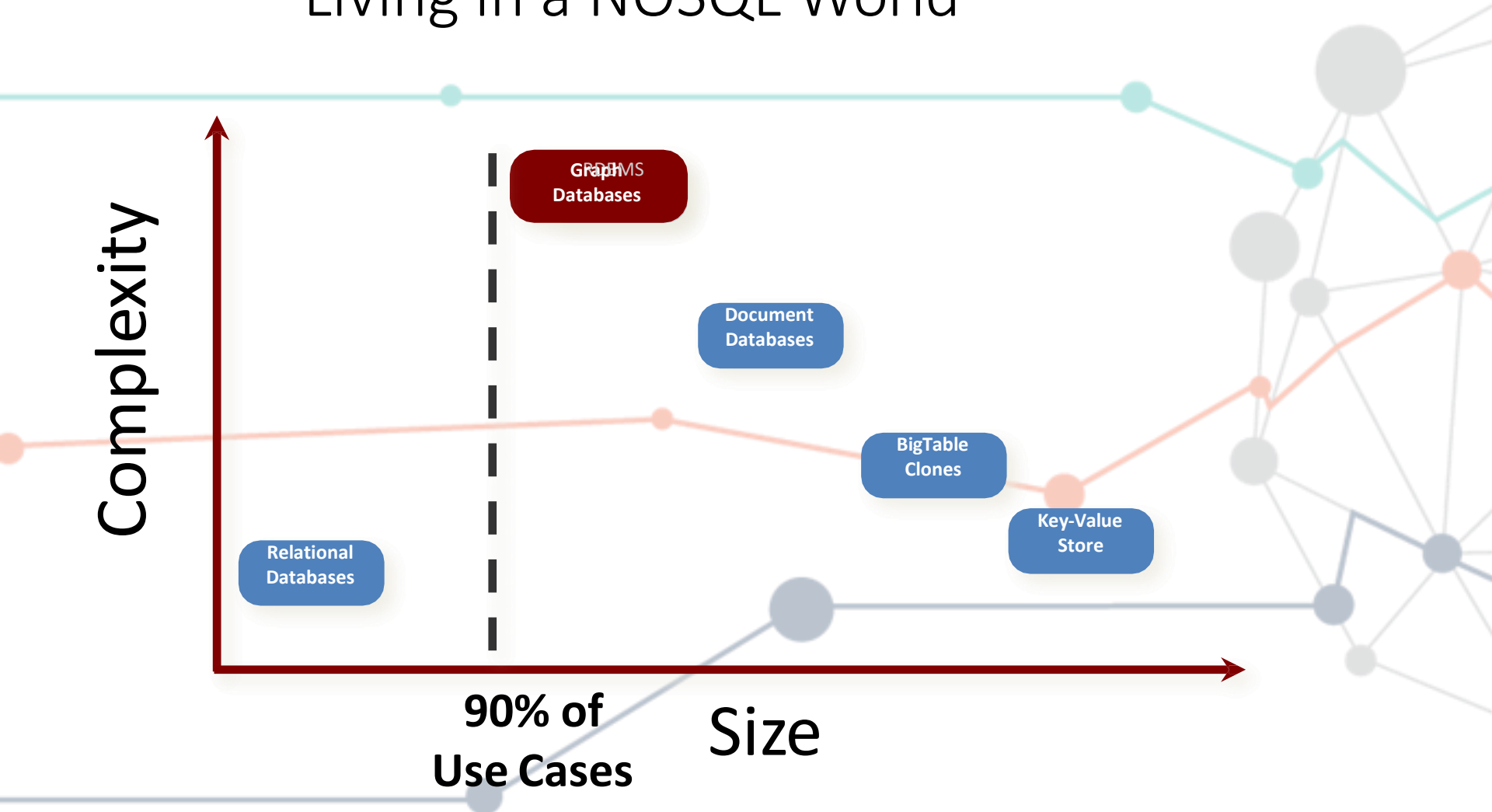


# Graph Databases: Pros and Cons

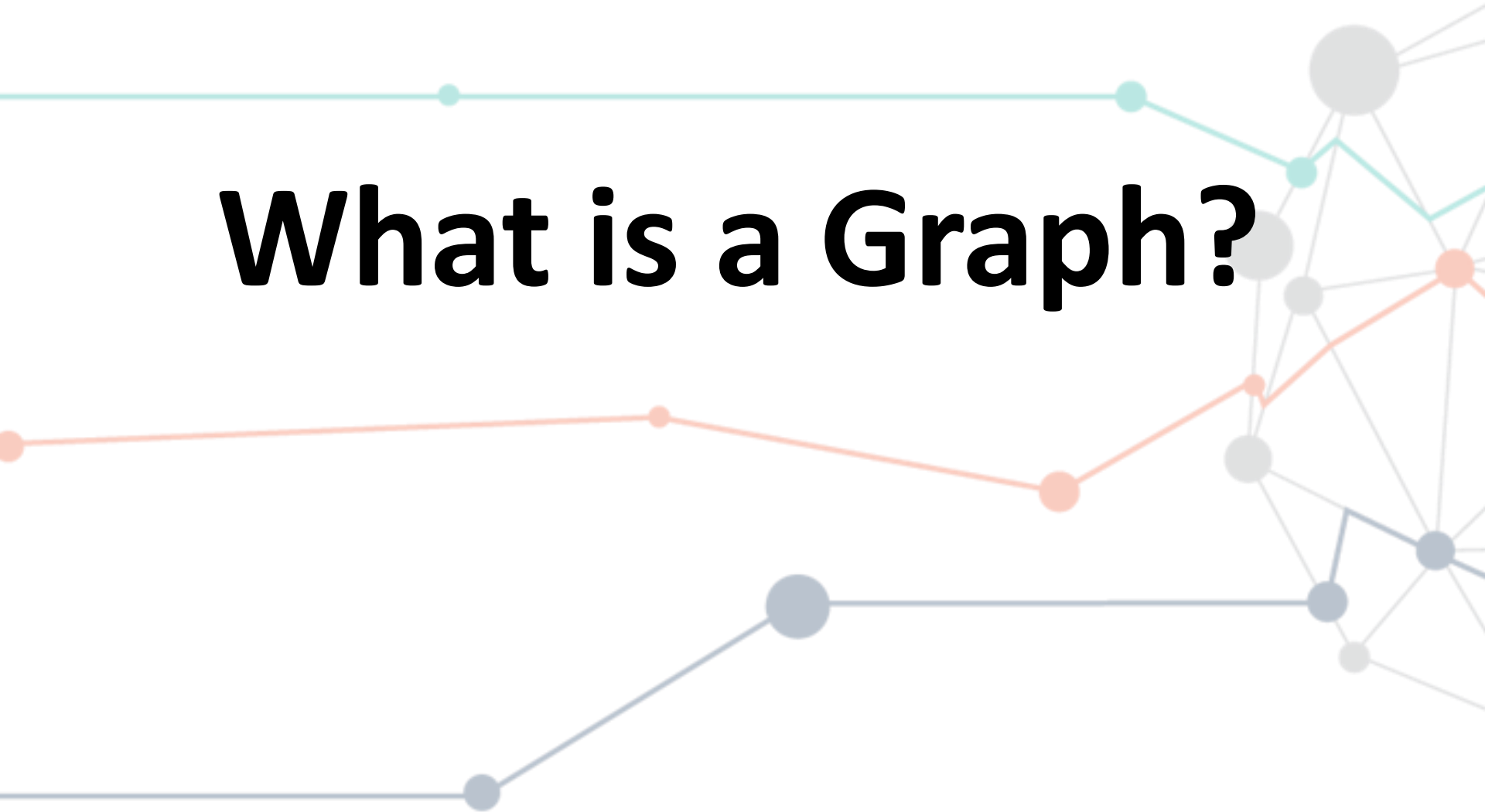
- Pros:
  - Powerful data model
  - Connected data locally indexed
  - Easy to query
- Cons
  - Requires rewiring your brain



# Living in a NOSQL World



# What is a Graph?



# What is a Graph?

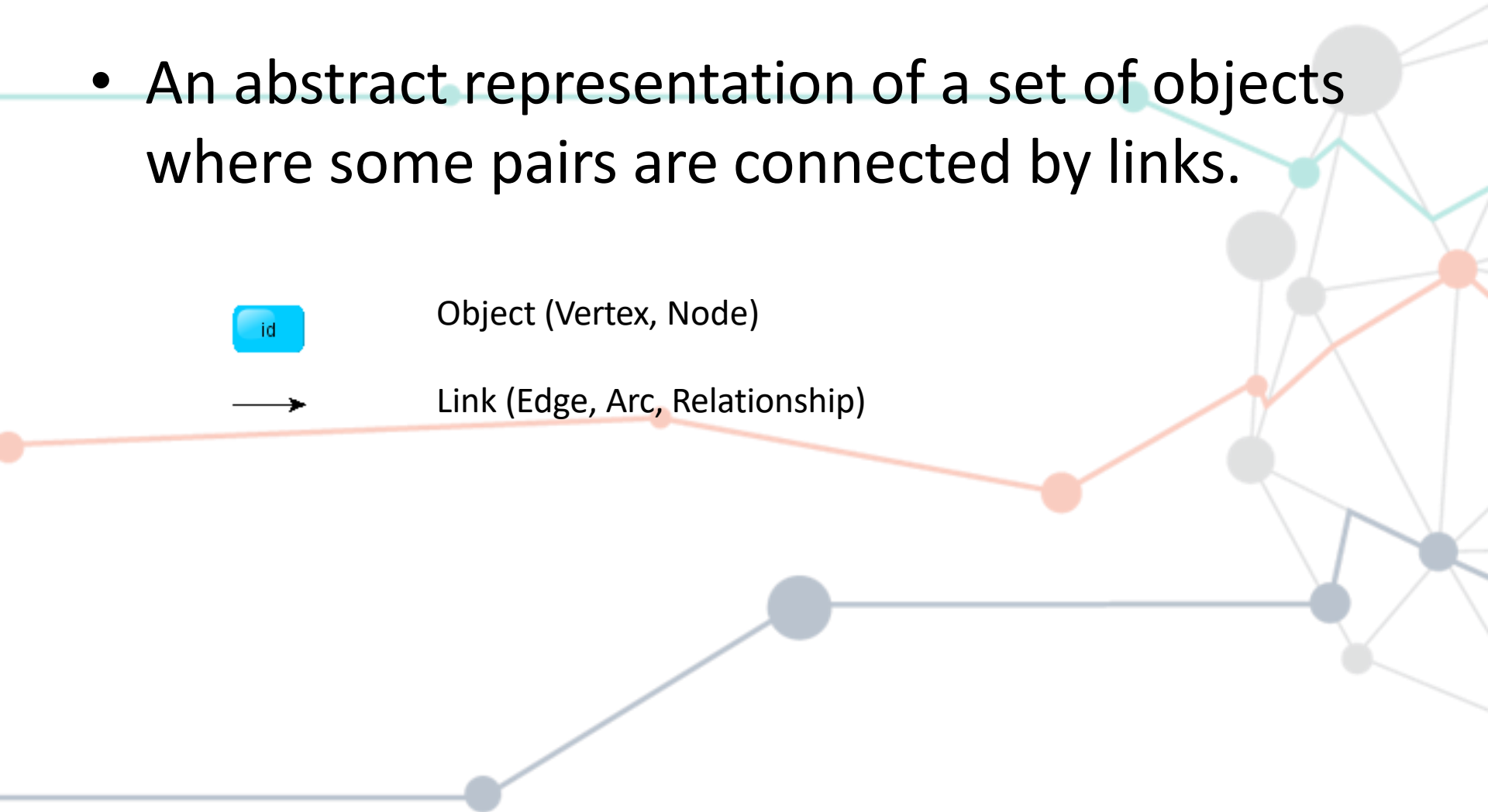
- An abstract representation of a set of objects where some pairs are connected by links.



Object (Vertex, Node)



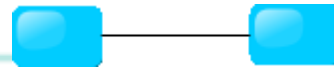
Link (Edge, Arc, Relationship)





# Different Kinds of Graphs

- Undirected Graph



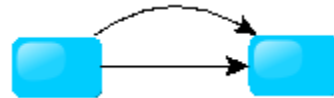
- Directed Graph



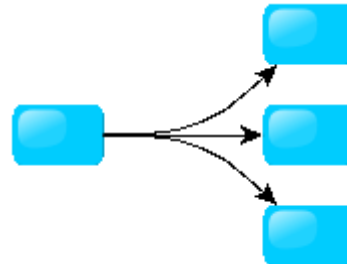
- Pseudo Graph



- Multi Graph

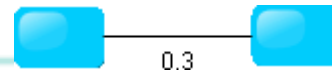


- Hyper Graph

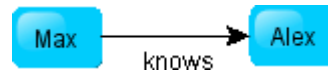


# More Kinds of Graphs

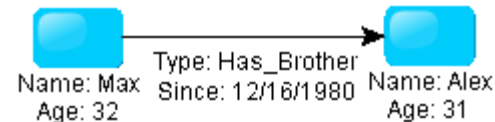
- Weighted Graph



- Labeled Graph

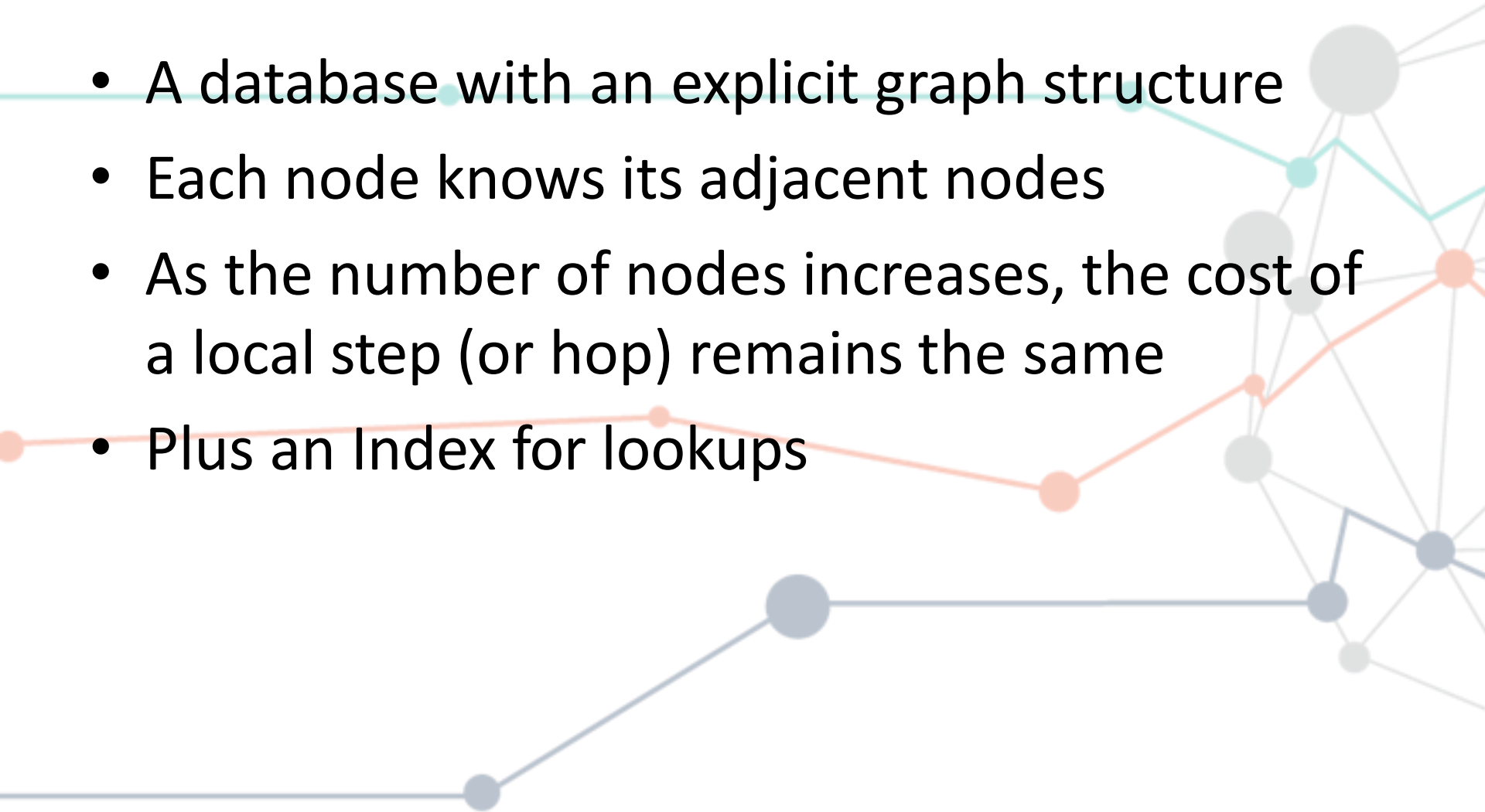


- Property Graph



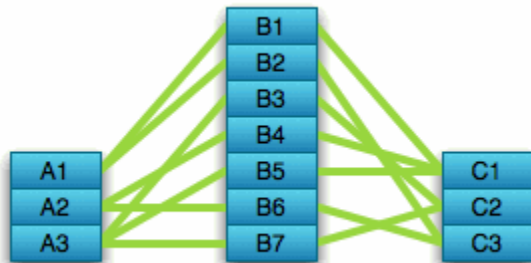
# What is a Graph Database?

- A database with an explicit graph structure
- Each node knows its adjacent nodes
- As the number of nodes increases, the cost of a local step (or hop) remains the same
- Plus an Index for lookups

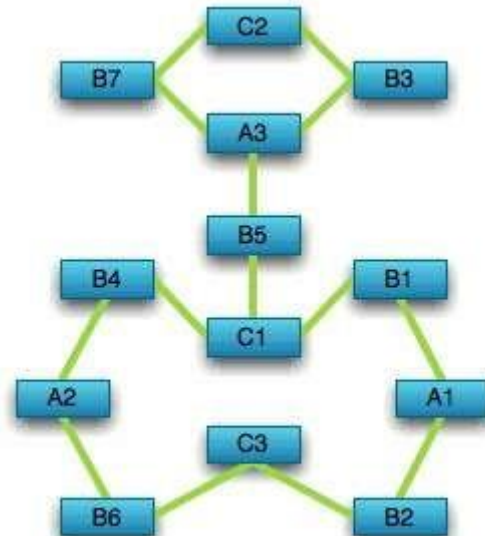


# Compared to Relational Databases

Optimized for aggregation

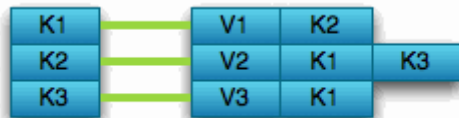


Optimized for connections

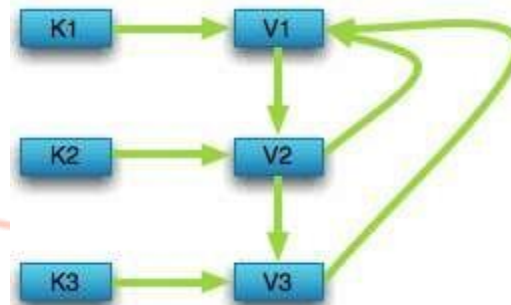


# Compared to Key Value Stores

Optimized for simple look-ups



Optimized for traversing connected data

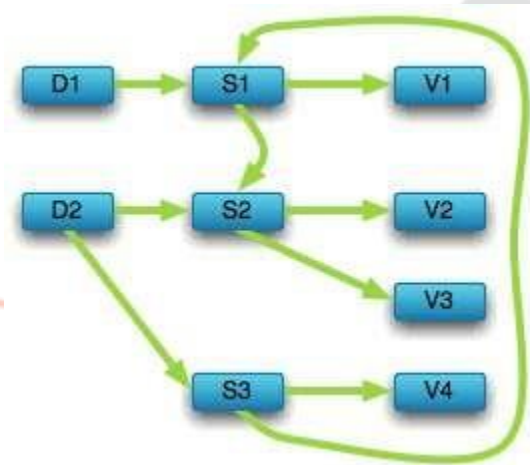


# Compared to Key Value Stores

Optimized for “trees” of data



Optimized for seeing the forest and the trees, and the branches, and the trunks

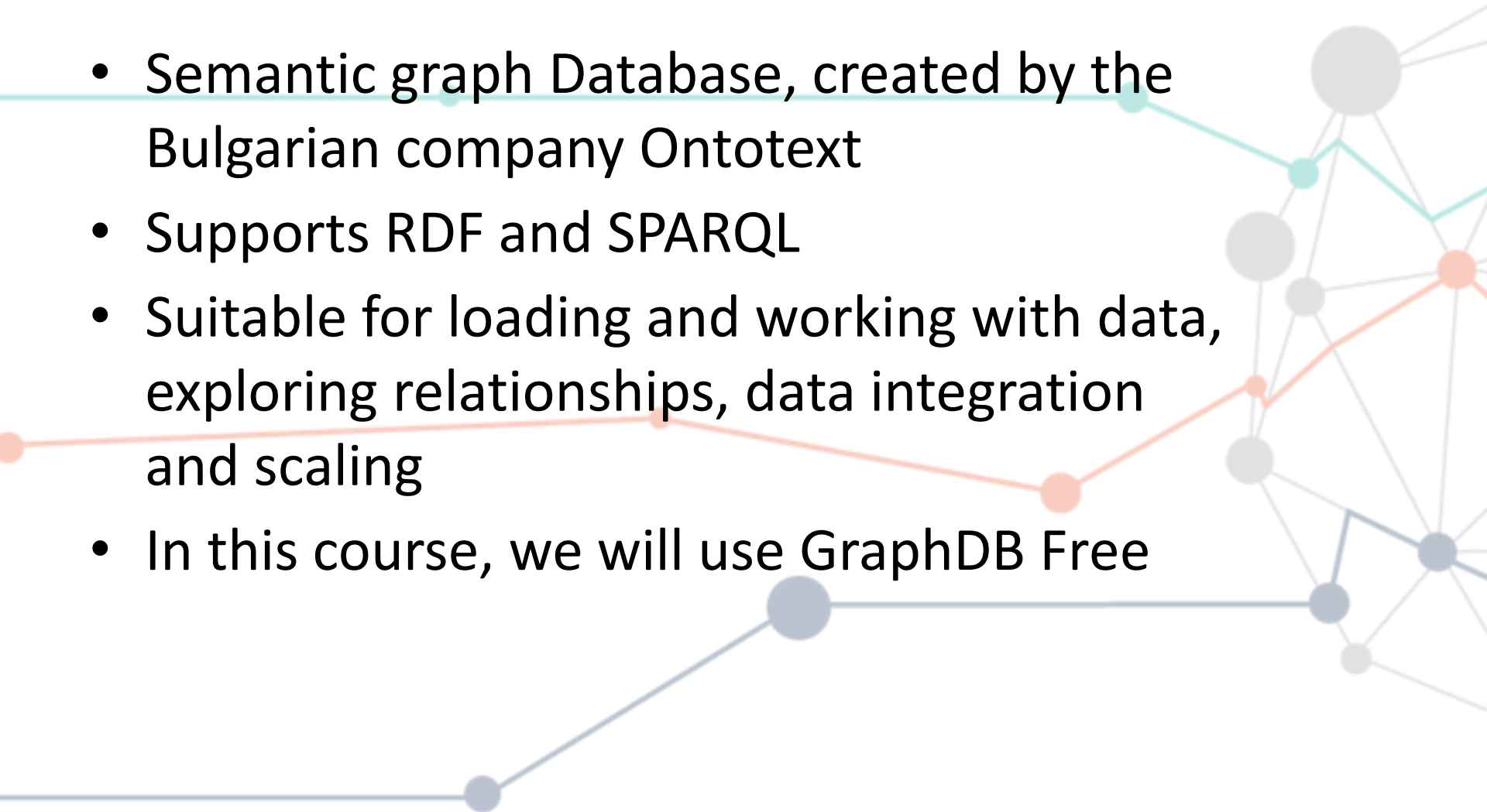


# What is GraphDB?

The background features a complex network of lines and nodes. A prominent teal line runs horizontally across the top, with two small teal nodes. Below it, an orange line curves across the middle, with three orange nodes. At the bottom, a blue line runs horizontally, with a small blue node. To the right, a dense cluster of grey nodes is interconnected by thin grey lines, with some nodes also connected to the teal, orange, and blue lines.

# What is GraphDB?

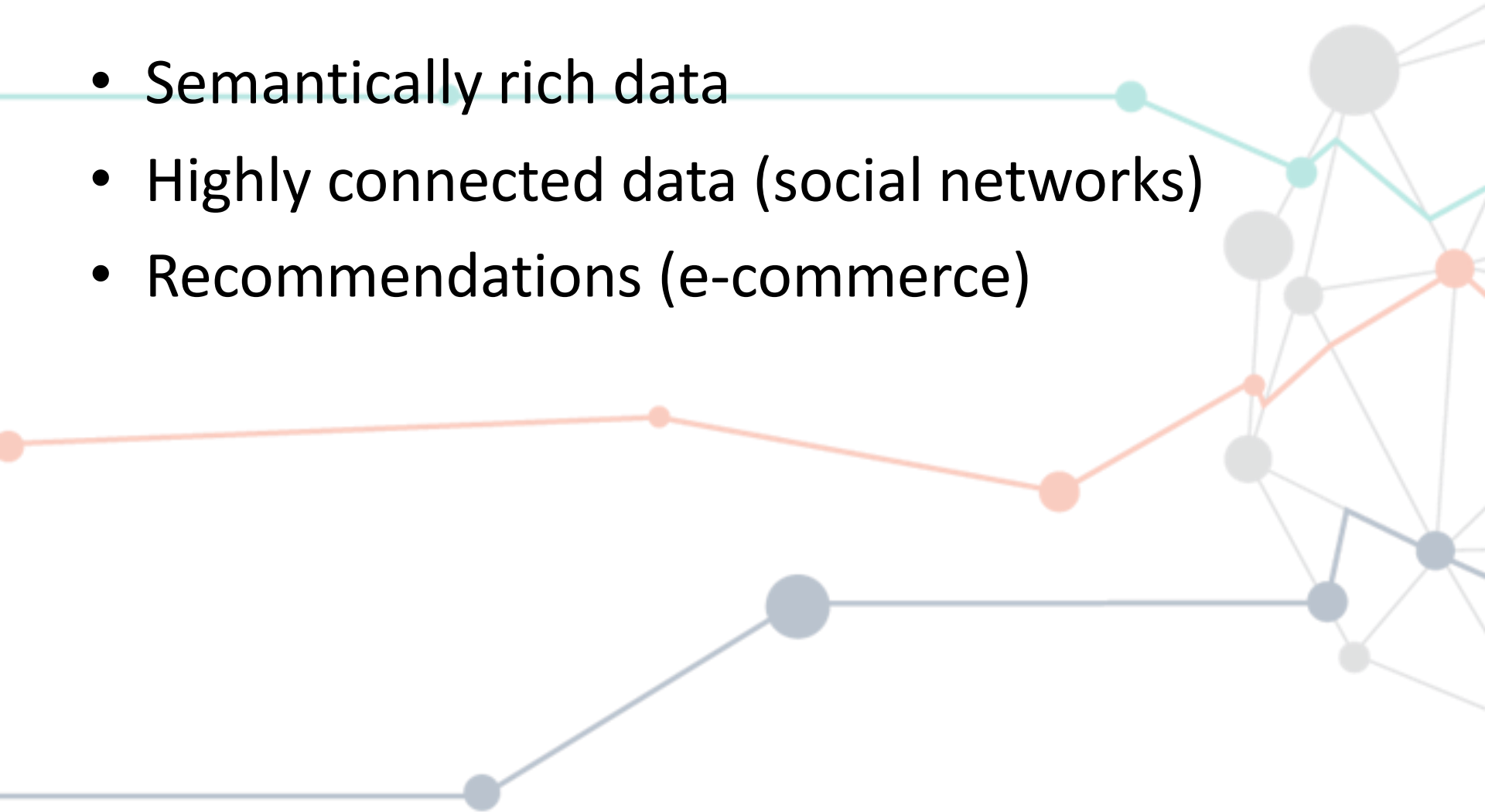
- Semantic graph Database, created by the Bulgarian company Ontotext
- Supports RDF and SPARQL
- Suitable for loading and working with data, exploring relationships, data integration and scaling
- In this course, we will use GraphDB Free





Good for

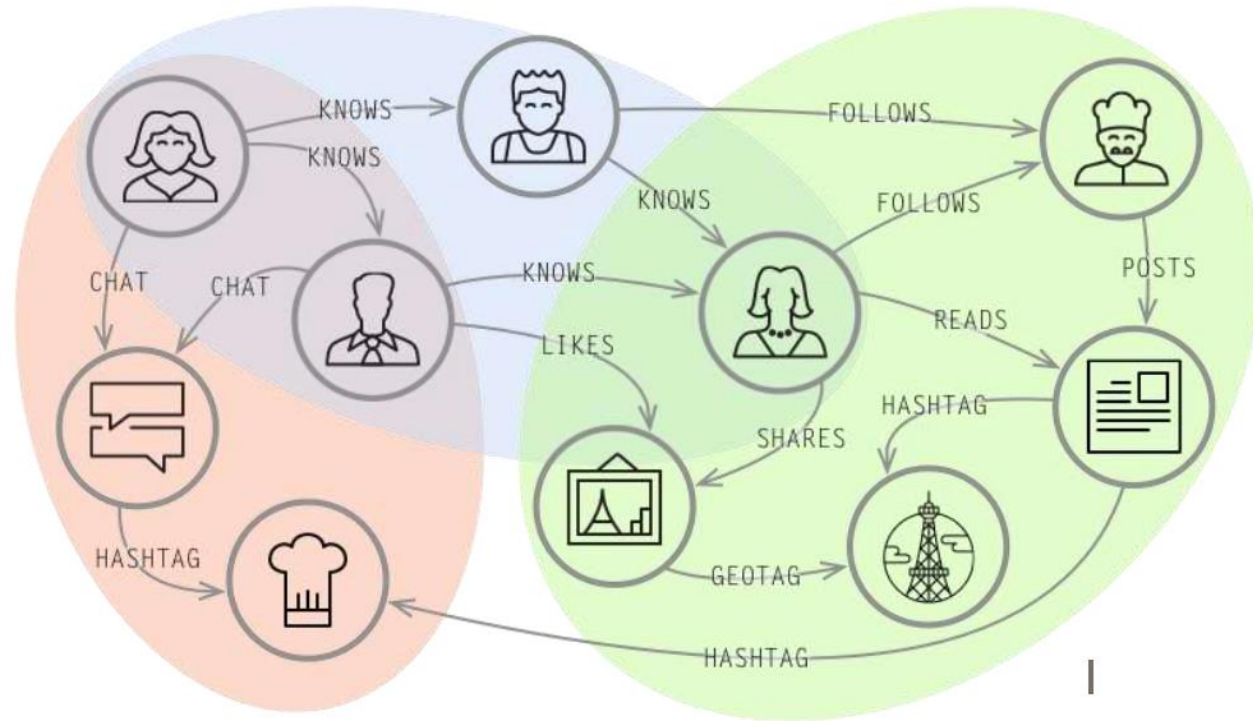
- Semantically rich data
- Highly connected data (social networks)
- Recommendations (e-commerce)



# Social Networks

Used for:

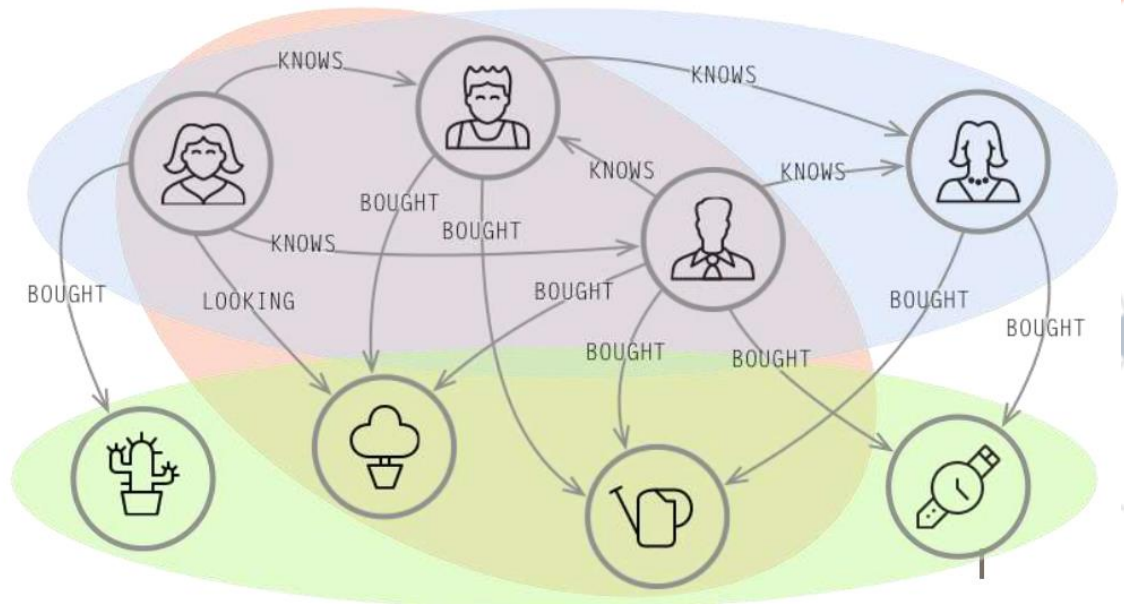
- Recommendation of friends
- Analysis of influence
- Implicit user clusters



# Recommendations

Used for:

- Recommendation of products, services and content



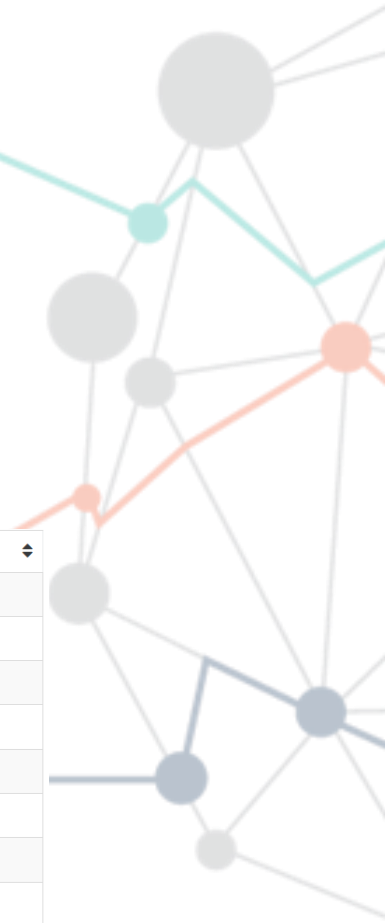
- <http://sparql.cwrc.ca/ontologies/genre.rdf>
- Find all the literature classes and the hierarchy between them

- <http://sparql.cwrc.ca/ontologies/genre.rdf>
- Find all the literature classes and the hierarchy between them

# Example: Genre ontology

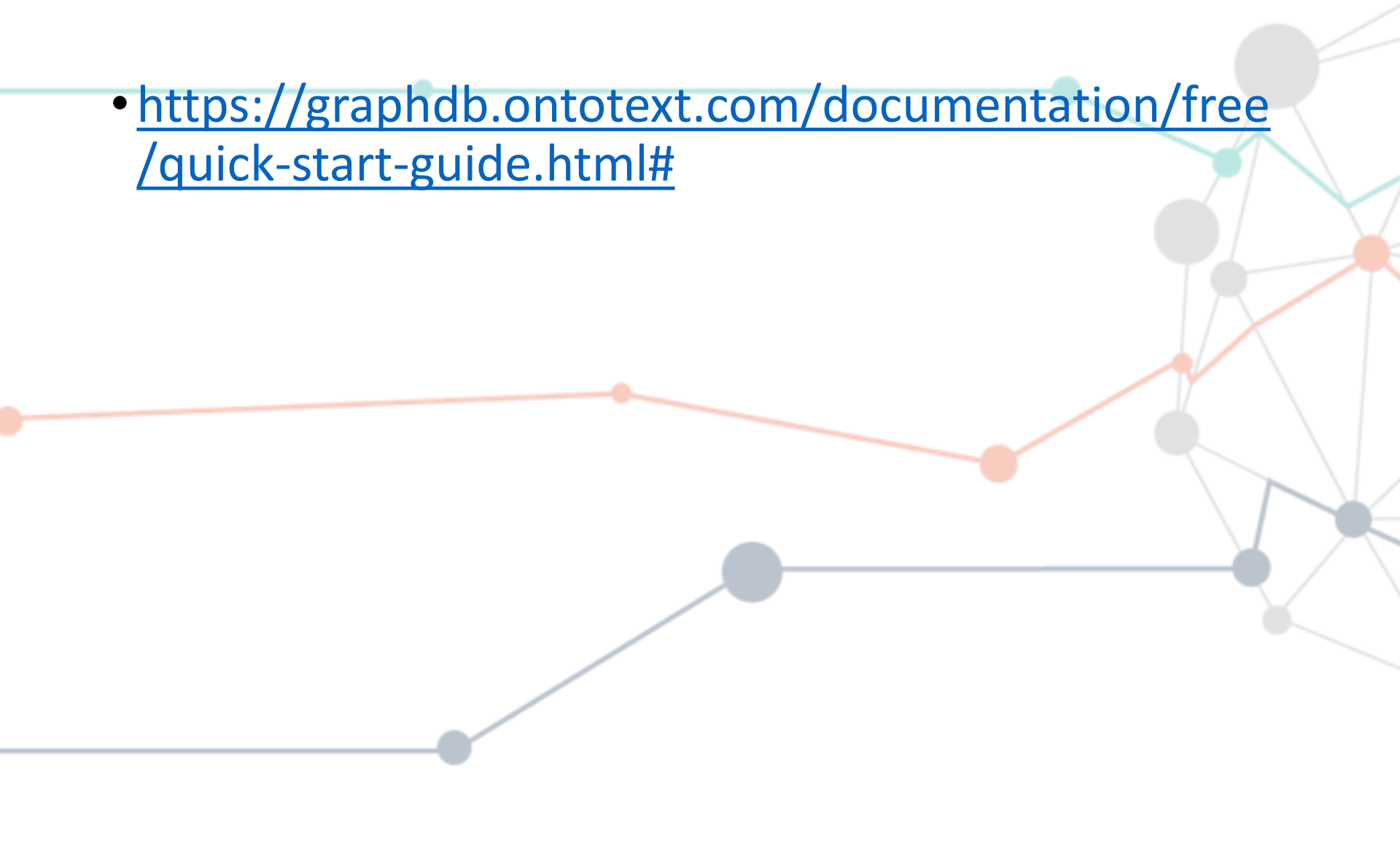
	genre	label
1	genre:realist	"realist"@en
2	genre:fairytale	"fairytale"@en
3	genre:guidebook	"guidebook"@en
4	genre:treatise	"treatise"@en
5	genre:novella	"novella"@en
6	genre:polemic	"polemic"@en
7	genre:magicRealist	"magic realist"@en
8	genre:panegyric	"panegyric"@en
9	genre:slaveNarrative	"slave narrative"@en
10	genre:dramaticMonologue	"dramatic monologue"@en

	Genre_classes
1	genre:Genre
2	genre:FictionalGenre
3	genre:LiteraryGenre
4	genre:NarrativeGenre
5	genre:ThematicGenre
6	genre:ScholarlyGenre
7	genre:InformationalGenre
8	genre:NovelisticGenre
9	genre:DialogueOrDebateGenre
10	genre:PrintMedium



# Graph DB installation guide

- <https://graphdb.ontotext.com/documentation/free/quick-start-guide.html#>



# Questions?



The background features a network of thin grey lines connecting various circular nodes. Some nodes are larger than others. Three distinct paths are highlighted: a teal line at the top, an orange line in the middle, and a blue line at the bottom. The teal line starts on the left, goes right, and then angles down towards a cluster of nodes on the right. The orange line starts on the left, goes right, dips down, and then angles up towards the same cluster. The blue line starts on the left, goes right, dips down, and then angles up towards the cluster. The text 'Thank you!' is centered in the upper half of the image.

Thank you!