グラフ・データベースで 実践 Graph Neural Networks！

Oracle Developer Days 2022 - Spring

山中 遼太
Product Manager Asia-Pacific
Spatial and Graph, Product Development
May 20, 2022

Copyright © 2022, Oracle and/or its affiliates

Agenda

グラフ DBMS を使う理由
- 違い
- 動機

映画のリコメンデーション
- 表からグラフへの変換
- グラフパターンマッチング
- リコメンデーション（Personalized PageRank, SALSA）

記事のトピックと引用
- 辿るクエリ
- コミュニティ検出 （Label Propagation）
- ノードのベクトル表現（DeepWalk, Graph Neural Networks）

Copyright © 2022, Oracle and/or its affiliates

グラフ DBMS を使う理由

クエリ言語
- アスキーアートを用いたパターンの記述
- パスの検索

データモデル
- ノードとエッジという二種類の構造を用いたモデルの記述
- スキーマレス？

実装
- グラフを辿る処理に対する最適化
- グラフアルゴリズムの実行に対する最適化

Copyright © 2022, Oracle and/or its affiliates

グラフ DBMS を使う理由

モチベーションの例１

- 表に落とすことが困難な複雑なデータがある
- グラフを使ったモデリングをしたい
- リレーショナルのモデリングはスキップ
- それでもデータベースを作ることができる、可視化もできる

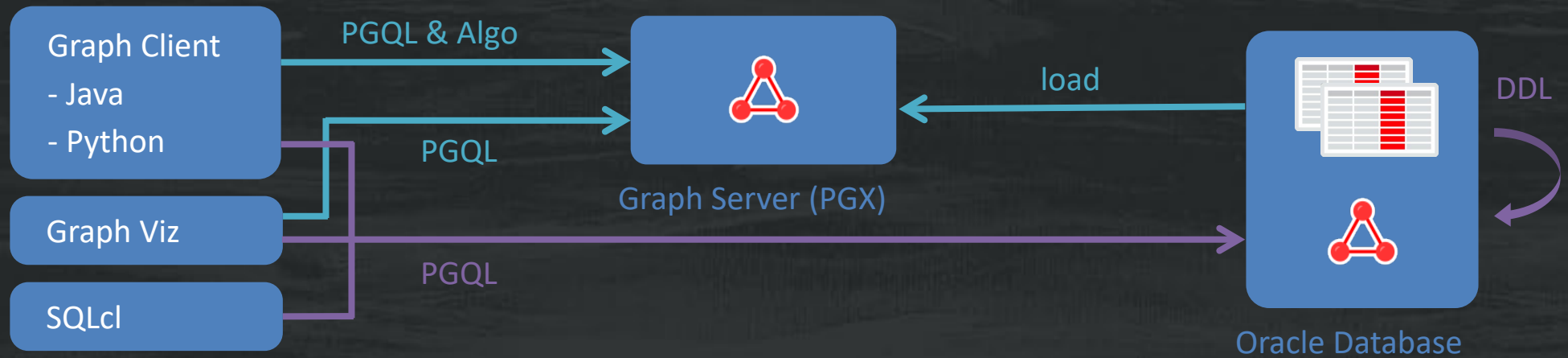
モチベーションの例２

- データはすでに表に入っている
- よく考えるとグラフを使ったモデルに展開することもできる
- そうすることで「**今までできなかった分析**」ができる　　<-- 今回はここに注目！

パスの検索、リンク予測、コミュニティ検出、関係性を加味した機械学習、など

Copyright © 2022, Oracle and/or its affiliates

Oracle Graph アーキテクチャ



Copyright © 2022, Oracle and/or its affiliates

映画のリコメンデーション

Copyright © 2022, Oracle and/or its affiliates

Dataset

Moviestream データセット

- 5,119 顧客
- 3,800 映画タイトル
- 906,117 レンタル履歴（顧客が映画をレンタルした記録）

こちらのワークショップからダウンロード

- https://apexapps.oracle.com/pls/apex/dbpm/r/livelabs/workshop-attendee-2?p210_workshop_id=889&p210_type=3

他の顧客のレンタルの傾向から、顧客がまだ観ていない映画を推薦できるか

Copyright © 2022, Oracle and/or its affiliates

Dataset

SELECT * FROM customer WHERE ROWNUM <= 5;

CUST_ID	LAST_NAME	FIRST_NAME	EMAIL
1018668	Mccall	Norris	norris.mccall@oraclemail.com
1018680	Cole	Rene	rene.cole@oraclemail.com
1018685	Ortega	Phoebe	phoebe.ortega@oraclemail.com
1018694	Harrington	Beatriz	beatriz.harrington@oraclemail.com
1018702	Rosales	Kelvin	kelvin.rosales@oraclemail.com

Copyright © 2022, Oracle and/or its affiliates

Dataset

```
SELECT * FROM movie WHERE ROWNUM <= 5;

 MOVIE_ID                TITLE    YEAR
___________ ____________________________ _______
      1070 El Mariachi                    1992
      1071 El Norte                       1983
      1072 El juego de Arcibel            2003
      1073 El pasajero clandestino        1995
      1074 El perro                       2004
```


Copyright © 2022, Oracle and/or its affiliates

Dataset

SELECT * FROM rental WHERE ROWNUM <= 5;

RENTAL_ID	CUST_ID	MOVIE_ID	RENTAL_DATE
1	1000450	3023	19-07-02
2	1000450	3023	20-10-04
3	1000450	3024	20-12-14
4	1000450	3036	19-07-20
5	1000450	3036	20-04-14

Copyright © 2022, Oracle and/or its affiliates

Create Graph

```
CREATE PROPERTY GRAPH graph1
 VERTEX TABLES (
  customer
    KEY (cust_id) LABEL customer
    PROPERTIES (cust_id, first_name, last_name, email)
 , movie
    KEY (movie_id) LABEL movie
    PROPERTIES (movie_id, title, year)
 )
 EDGE TABLES (
  rental
    KEY (sales_id)
    SOURCE KEY(cust_id) REFERENCES customer
    DESTINATION KEY(movie_id) REFERENCES movie
    LABEL rented PROPERTIES (day_id)
 )
```


Copyright © 2022, Oracle and/or its affiliates

Query Graph

-- ノード

```
SELECT LABEL(n), COUNT(n)
FROM MATCH (n) ON graph1
GROUP BY LABEL(n)
```

-- エッジ

```
SELECT LABEL(e), COUNT(e)
FROM MATCH ()-[e]-() ON graph1
GROUP BY LABEL(e)
```


Copyright © 2022, Oracle and/or its affiliates

Query Graph

```
-- 顧客と映画

SELECT DISTINCT c.first_name, m.title
FROM MATCH (c)->(m) ON graph1
LIMIT 5

-- 顧客と映画の関係（レンタルした日付）

SELECT DISTINCT c.first_name, r.rental_date, m.title
FROM MATCH (c)-[r]->(m) ON graph1
LIMIT 5
```


Copyright © 2022, Oracle and/or its affiliates

Query Graph

```
-- Ricky Rogers が借りた映画

SELECT DISTINCT c.first_name, r.rental_date, m.title
FROM MATCH (c)-[r]->(m) ON graph1
WHERE c.first_name = 'Ricky'
AND c.last_name = 'Rogers'
LIMIT 5
```


Copyright © 2022, Oracle and/or its affiliates

Query Graph

```
-- Ricky Rogers と同じ映画を借りた人

SELECT c1.first_name || ' ' || c1.last_name AS c1_name
    , c2.first_name || ' ' || c2.last_name AS c2_name
    , COUNT(DISTINCT m) AS cnt
FROM MATCH (c1)->(m)<-(c2) ON graph1
WHERE c1.first_name = 'Ricky'
  AND c1.last_name = 'Rogers'
  AND ALL_DIFFERENT(c1, c2)
GROUP BY c1, c2
ORDER BY cnt DESC
LIMIT 5
```


Copyright © 2022, Oracle and/or its affiliates

Query Graph

```
-- Ricky Rogers と同じ映画を借りた人が借りた他の映画

SELECT c1.first_name || ' ' || c1.last_name AS c1_name
    , m2.title
    , COUNT(DISTINCT c2) AS cnt
FROM MATCH (c1)->(m1)<-(c2)->(m2) ON graph1
WHERE c1.first_name = 'Ricky'
  AND c1.last_name = 'Rogers'
  AND ALL_DIFFERENT(c1, c2)
  AND ALL_DIFFERENT(m1, m2)
GROUP BY c1, m2
ORDER BY cnt DESC
LIMIT 10
```


Copyright © 2022, Oracle and/or its affiliates

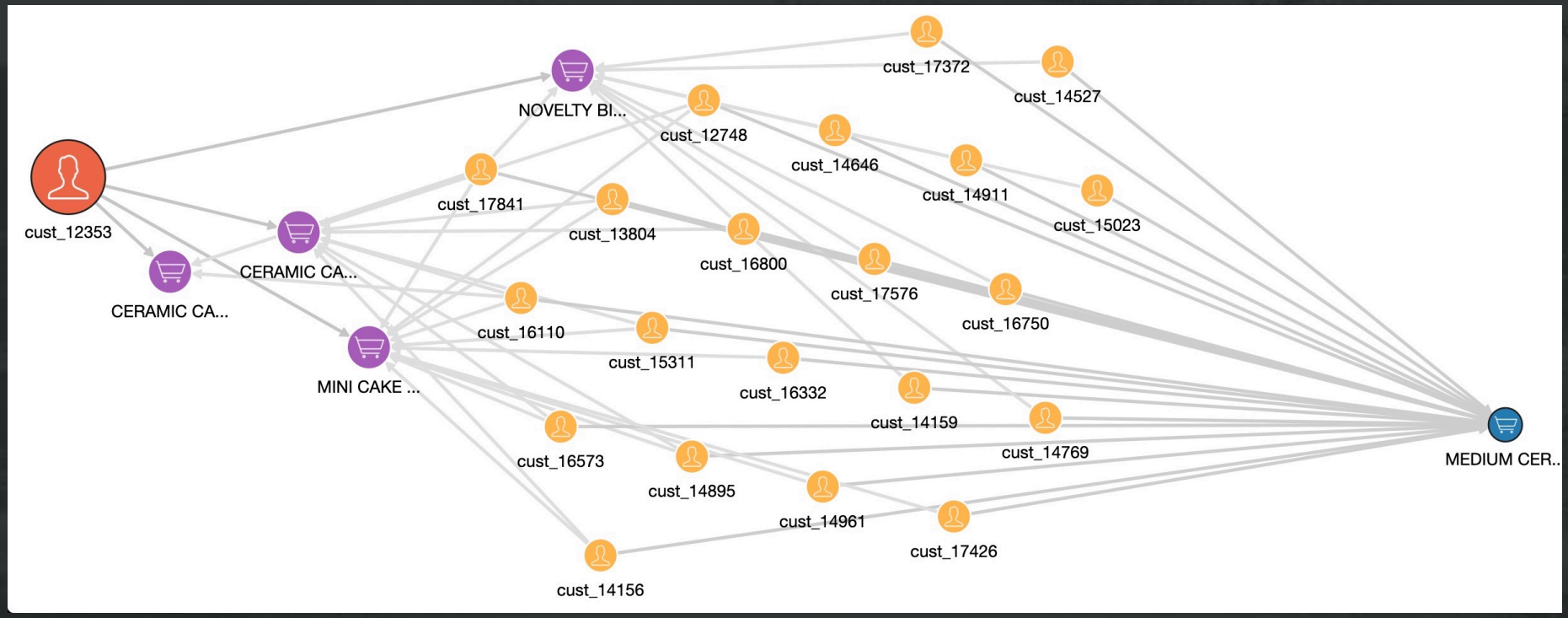
Query Graph

C1_NAME	TITLE	CNT
Ricky Rogers	Avengers: Endgame	5048
Ricky Rogers	Star Wars Episode IX: The Rise of Skywalker	4371
Ricky Rogers	Captain Marvel	4258
Ricky Rogers	Spider-Man: Far from Home	4251
Ricky Rogers	Aladdin	3722
Ricky Rogers	Aquaman	3475
Ricky Rogers	Avengers: Infinity War	3347
Ricky Rogers	The Lion King	3286
Ricky Rogers	Black Panther	3110
Ricky Rogers	Bohemian Rhapsody	2963

10 rows selected.

Copyright © 2022, Oracle and/or its affiliates

Personalized PageRank (PPR)



Copyright © 2022, Oracle and/or its affiliates

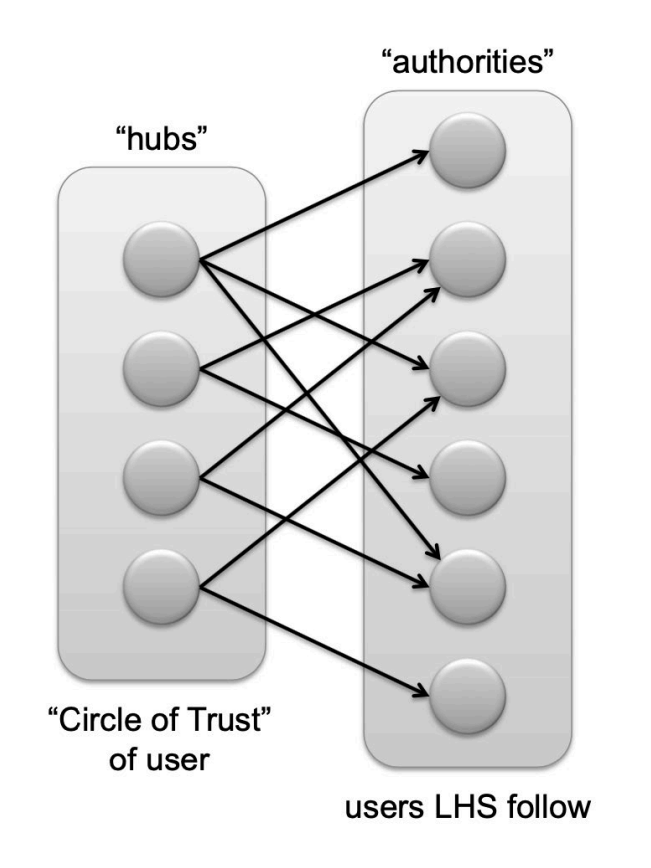
Personalized PageRank

```
analyst.personalized_pagerank(graph2, cust, rank="ppr")
```

```
SELECT m.movie_id, m.title, m.ppr
  FROM MATCH (m) ON graph2
 WHERE LABEL(m) = 'MOVIE'
   AND NOT EXISTS (
     SELECT *
       FROM MATCH (c)-[:rented]->(m) ON graph2
      WHERE c.first_name = 'Ricky'
        AND c.last_name = 'Rogers'
   )
 ORDER BY m.ppr DESC
 LIMIT 10
```


Copyright © 2022, Oracle and/or its affiliates

SALSA



SALSA = Stochastic Approach for Link-Structure Analysis

- https://graal.ens-lyon.fr/~lmarchal/art-docs-0506/SALSA.pdf
- https://stanford.edu/~rezab/papers/wtf_overview.pdf

こちらのワークショップ（Lab 7）に実行例を掲載

- https://apexapps.oracle.com/pls/apex/dbpm/r/livelabs/workshop-attendee-2?p210_workshop_id=889&p210_type=3

Copyright © 2022, Oracle and/or its affiliates

記事のトピックと引用

Copyright © 2022, Oracle and/or its affiliates

Dataset

Cora データセット

- 2,708 記事（7 つのトピックに分類）
- 5,429 引用関係
- ある単語が含まれているかどうか（0 か 1）

https://linqs.soe.ucsc.edu/data

引用関係や含まれている単語から、その記事のトピックを予測できるか

Copyright © 2022, Oracle and/or its affiliates

Create Graph

```
CREATE PROPERTY GRAPH cora
 VERTEX TABLES (
  cora_content
    KEY (id)
    LABEL content
    PROPERTIES ARE ALL COLUMNS
 )
 EDGE TABLES (
  cora_cites
    KEY (id)
    SOURCE KEY(src) REFERENCES cora_content
    DESTINATION KEY(dst) REFERENCES cora_content
    LABEL cites
 )
```


Copyright © 2022, Oracle and/or its affiliates

Query Graph

-- ノード

```
SELECT LABEL(n), COUNT(n)
FROM MATCH (n) ON cora
GROUP BY LABEL(n)
```

-- エッジ

```
SELECT LABEL(e), COUNT(e)
FROM MATCH ()-[e]-() ON cora
GROUP BY LABEL(e)
```


Copyright © 2022, Oracle and/or its affiliates

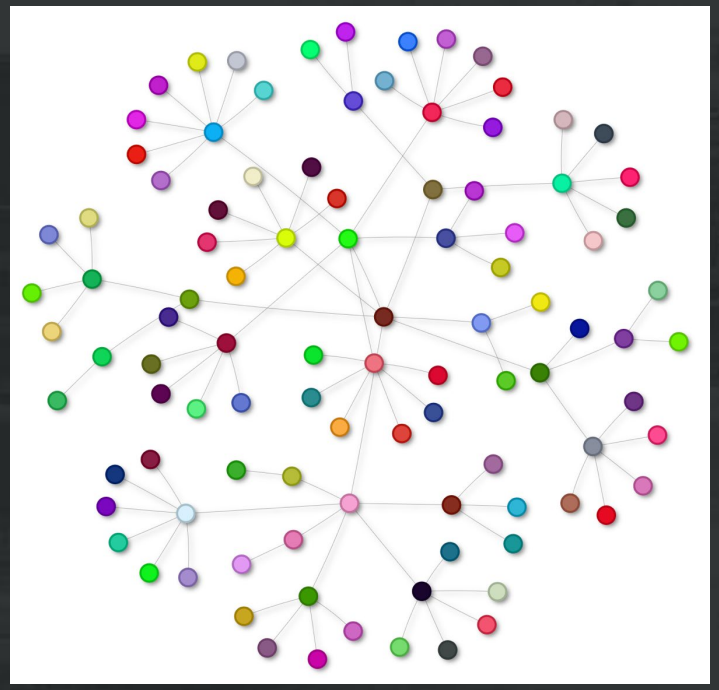
Query Graph

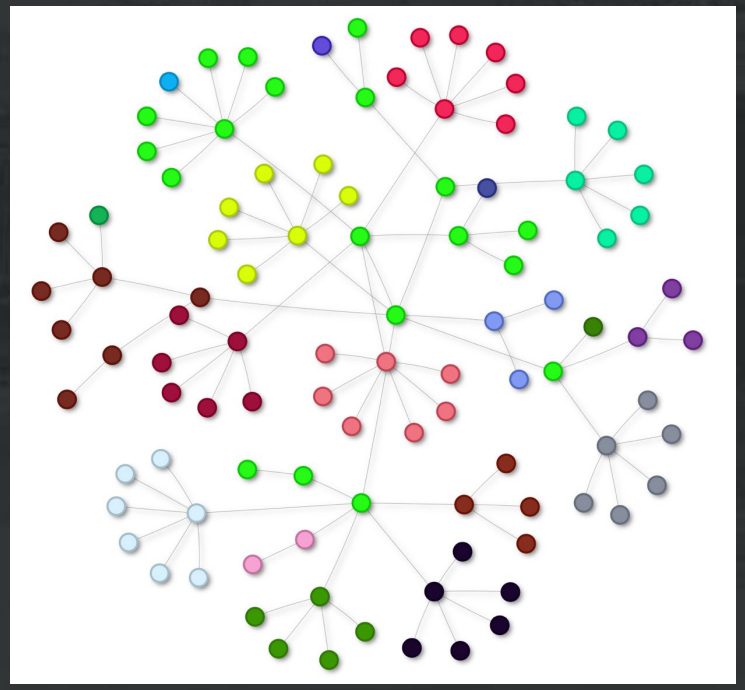
```
-- 同じトピックの記事を引用

SELECT p1.topic, COUNT(ID(p1))
FROM MATCH (p1)->(p2) ON cora
WHERE p1.topic = p2.topic
GROUP BY p1.topic
LIMIT 5

-- 異なるトピックの記事を引用

SELECT p1.topic, COUNT(ID(p1))
FROM MATCH (p1)->(p2) ON cora
WHERE p1.topic != p2.topic
GROUP BY p1.topic
LIMIT 5
```

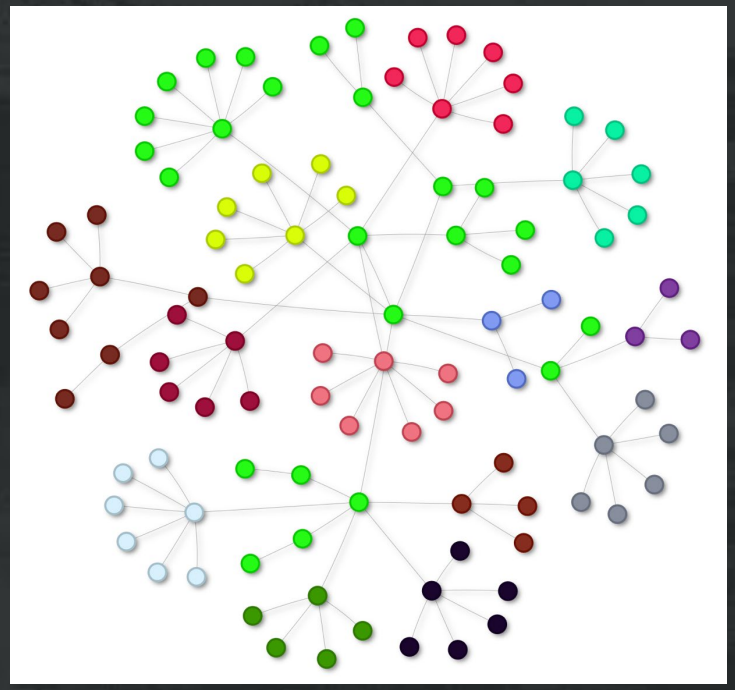

Copyright © 2022, Oracle and/or its affiliates

Label Propagation

エッジによる繋がりを用いたコミュニティ検出







https://graph-algorithm-showcase.info/label-propagation.html

Copyright © 2022, Oracle and/or its affiliates

Label Propagation

-- アルゴリズム実行

```
analyst.communities_label_propagation(graph1, label='community_id')
```

-- 結果の確認

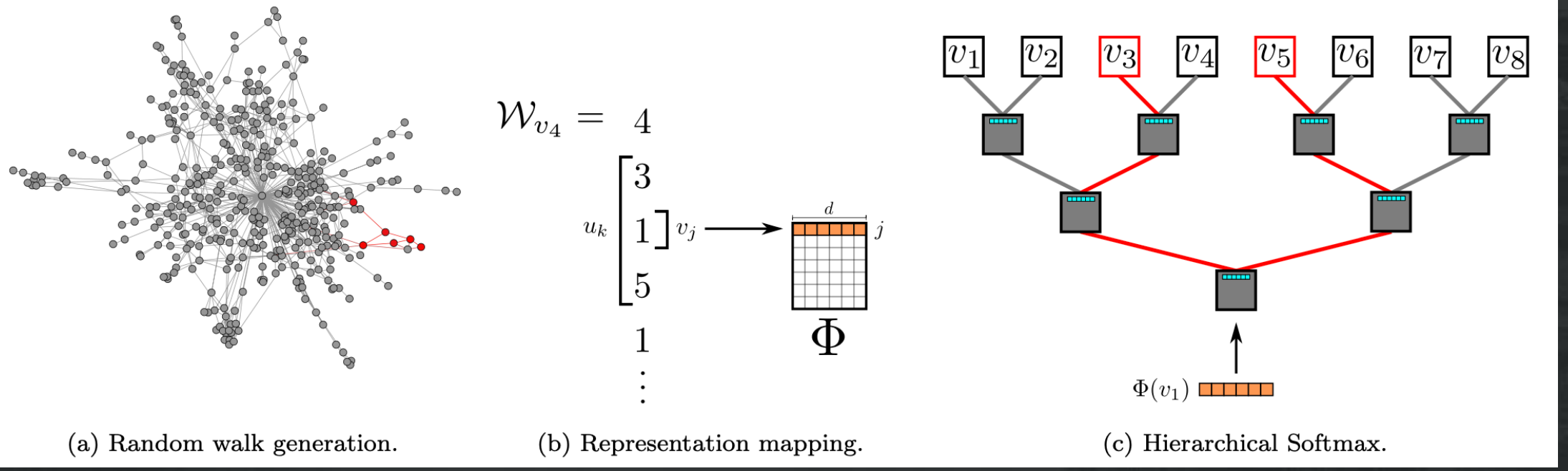
```
SELECT ID(v) AS id, v.topic, v.community_id
FROM MATCH (v)
LIMIT 5
```

```
| id                  | topic           | community_id1 |
+-----------------------------------------------------+
| CORA_CONTENT(1133028) | Theory          | 0           |
| CORA_CONTENT(578780)  | Genetic_Algorithms | 1        |
| CORA_CONTENT(1110768) | Neural_Networks    | 2        |
```


Copyright © 2022, Oracle and/or its affiliates

DeepWalk

各ノードのベクトル表現（embedding）を求める



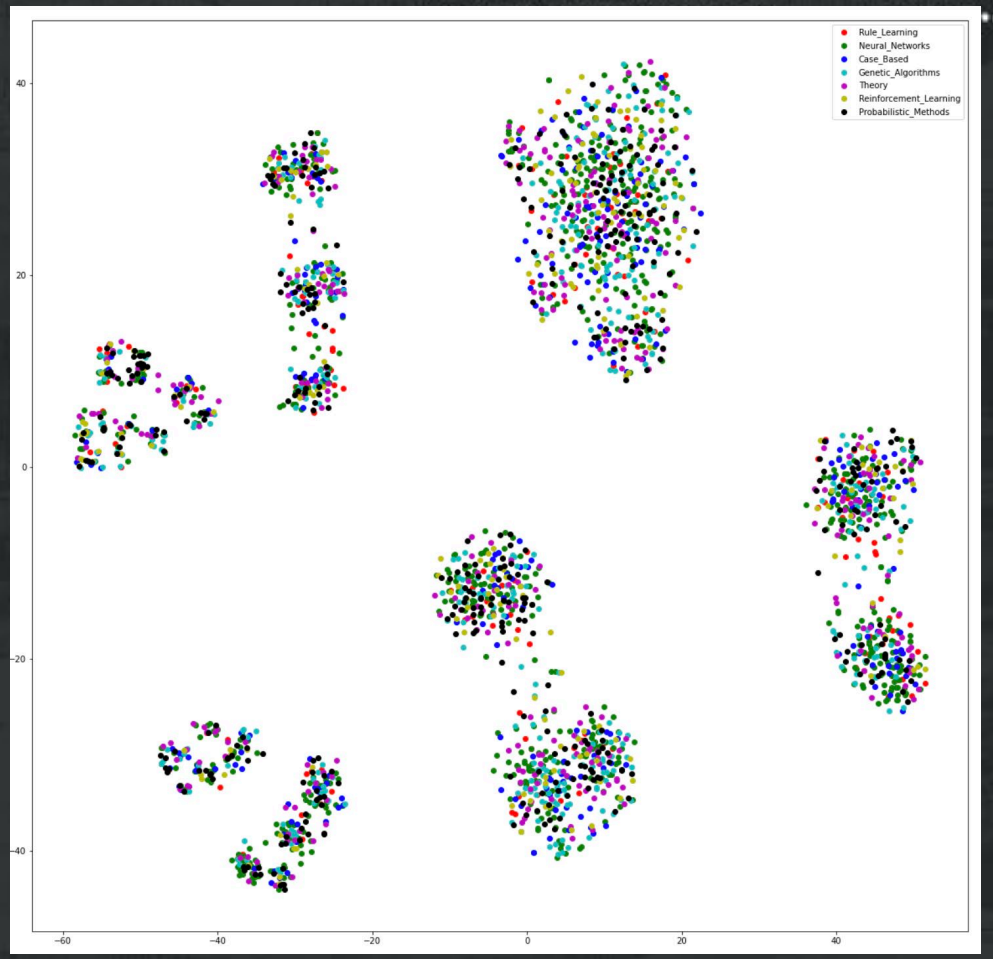
(a) Random walk generation. (b) Representation mapping. (c) Hierarchical Softmax.

https://arxiv.org/pdf/1403.6652.pdf

Copyright © 2022, Oracle and/or its affiliates

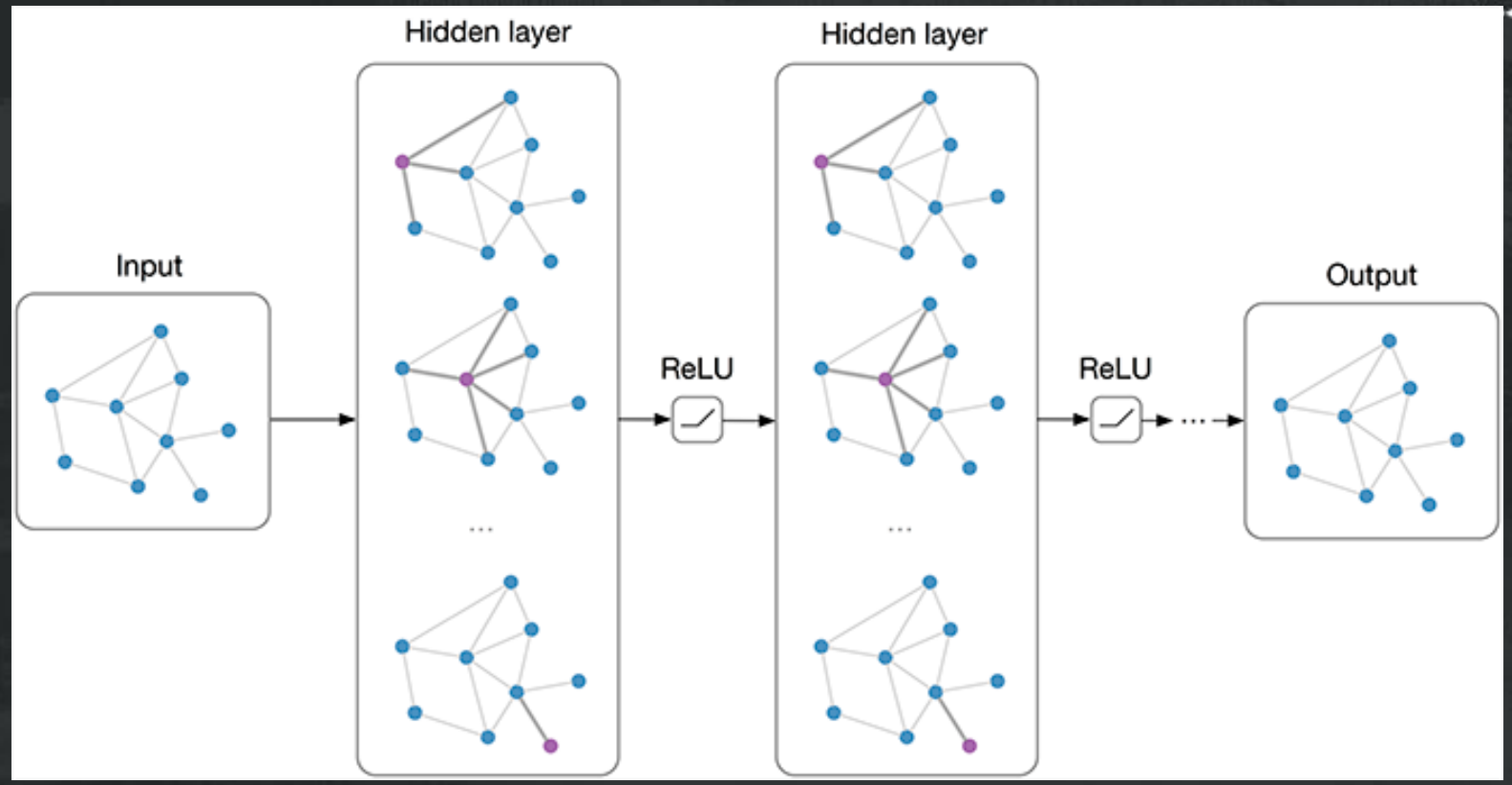
DeepWalk

```
model = analyst.deepwalk_builder(
    window_size = 3,
    walks_per_vertex = 6,
    walk_length = 4,
    num_epochs = 10
    )

model.fit(graph1)

vertex_vectors =
    model.trained_vectors
```



Copyright © 2022, Oracle and/or its affiliates

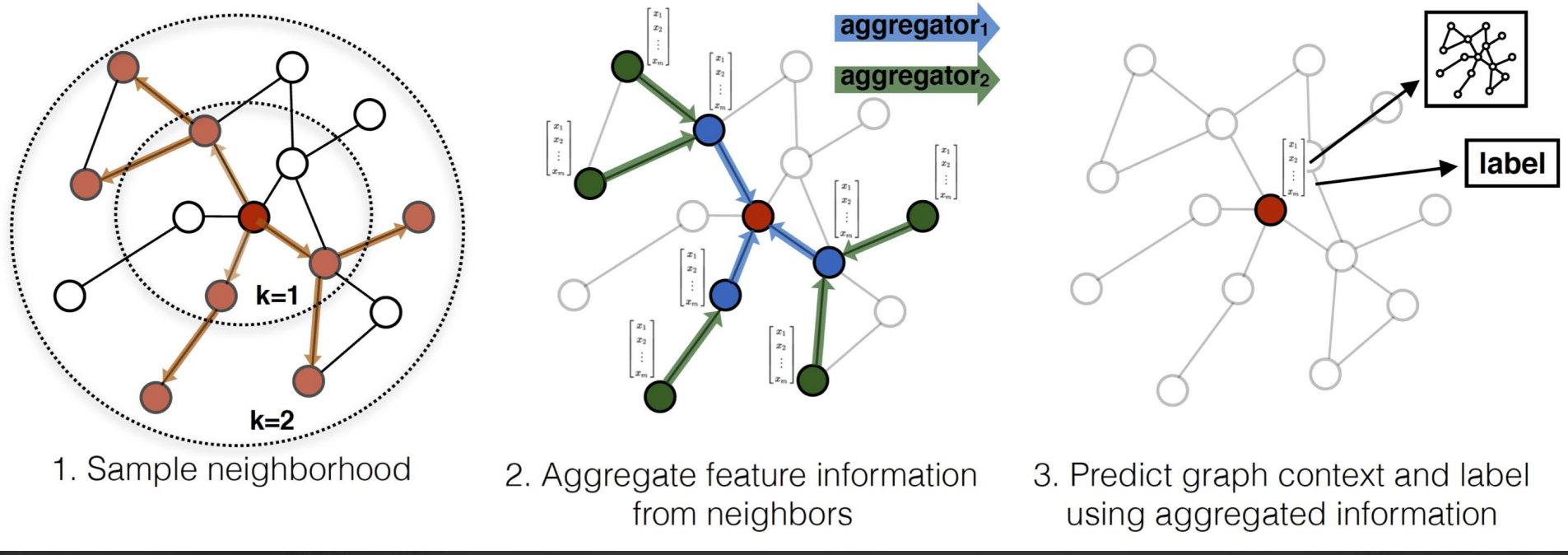
Graph Neural Networks



http://tkipf.github.io/graph-convolutional-networks/

Copyright © 2022, Oracle and/or its affiliates

Graph Neural Networks



1. Sample neighborhood

2. Aggregate feature information from neighbors

3. Predict graph context and label using aggregated information

https://cs.stanford.edu/people/jure/pubs/graphsage-nips17.pdf

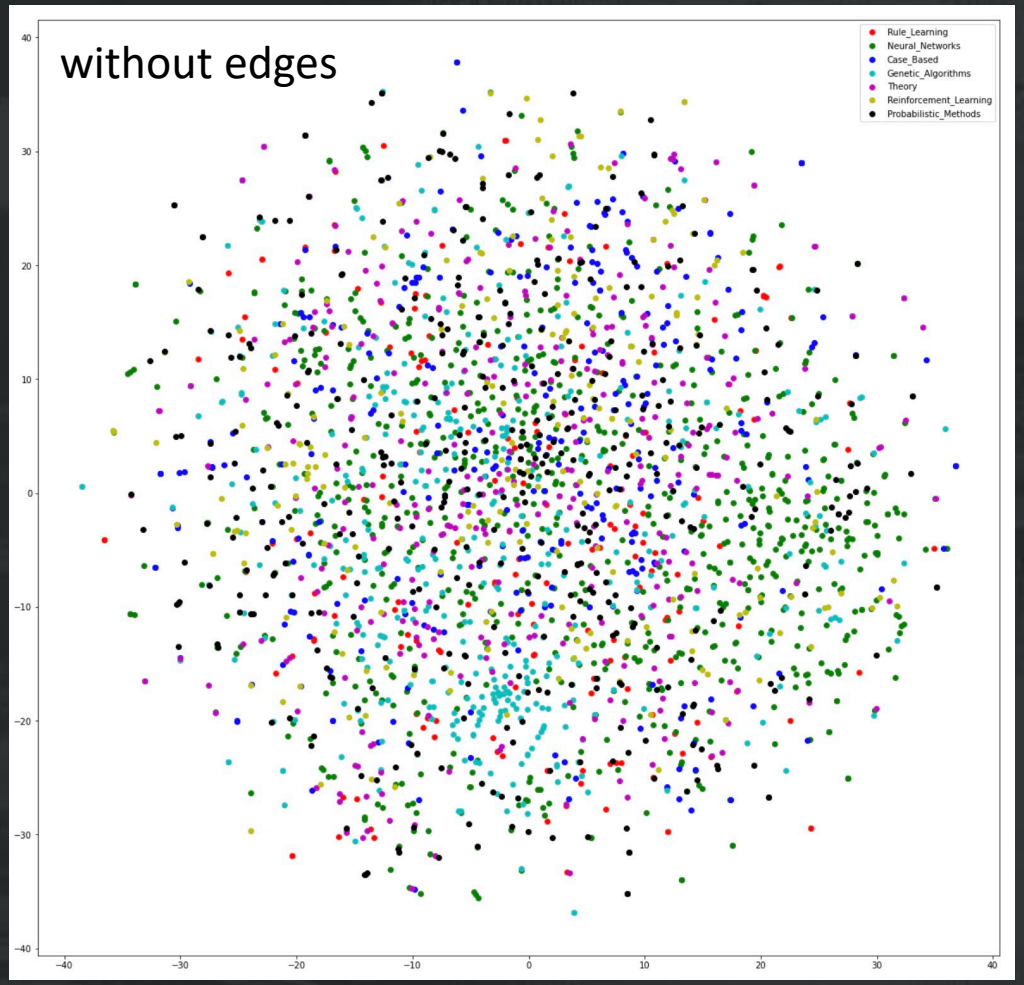
Copyright © 2022, Oracle and/or its affiliates

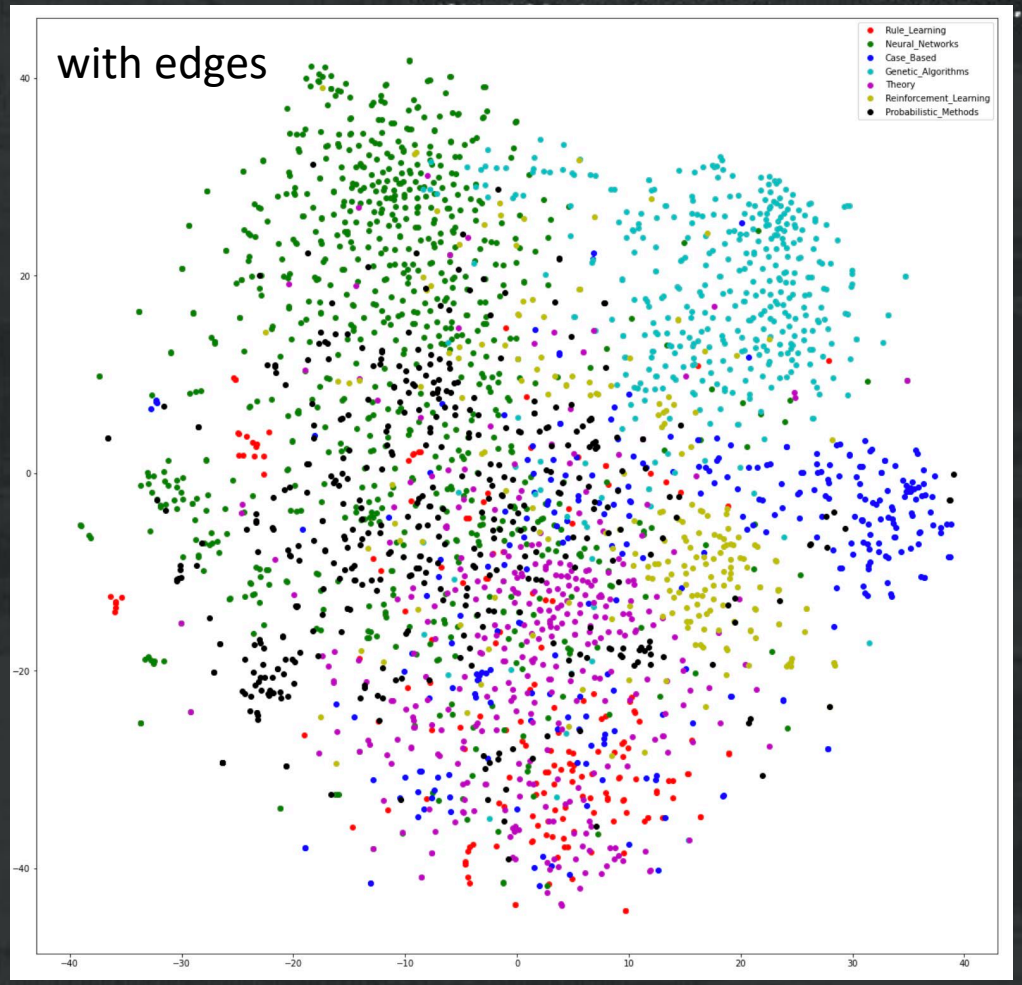
Unsupervised GraphWise

```
property_names = []

for i in range(900):
   property_names.append("F" + str(i + 1))

model = analyst.unsupervised_graphwise_builder(
   vertex_input_property_names = property_names)

model.fit(graph1)

vertex_vectors = model.infer_embeddings(graph1, graph1.get_vertices())
```


Copyright © 2022, Oracle and/or its affiliates

Unsupervised GraphWise





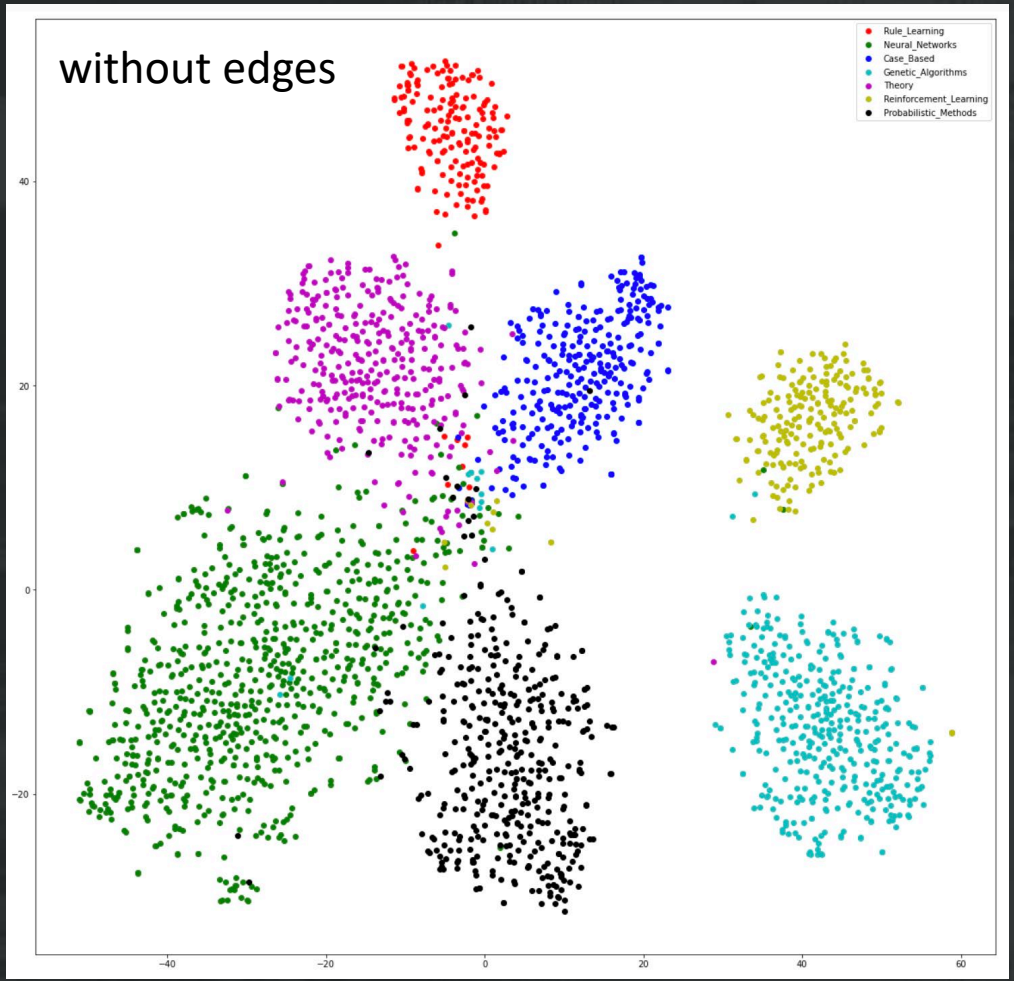
Copyright © 2022, Oracle and/or its affiliates

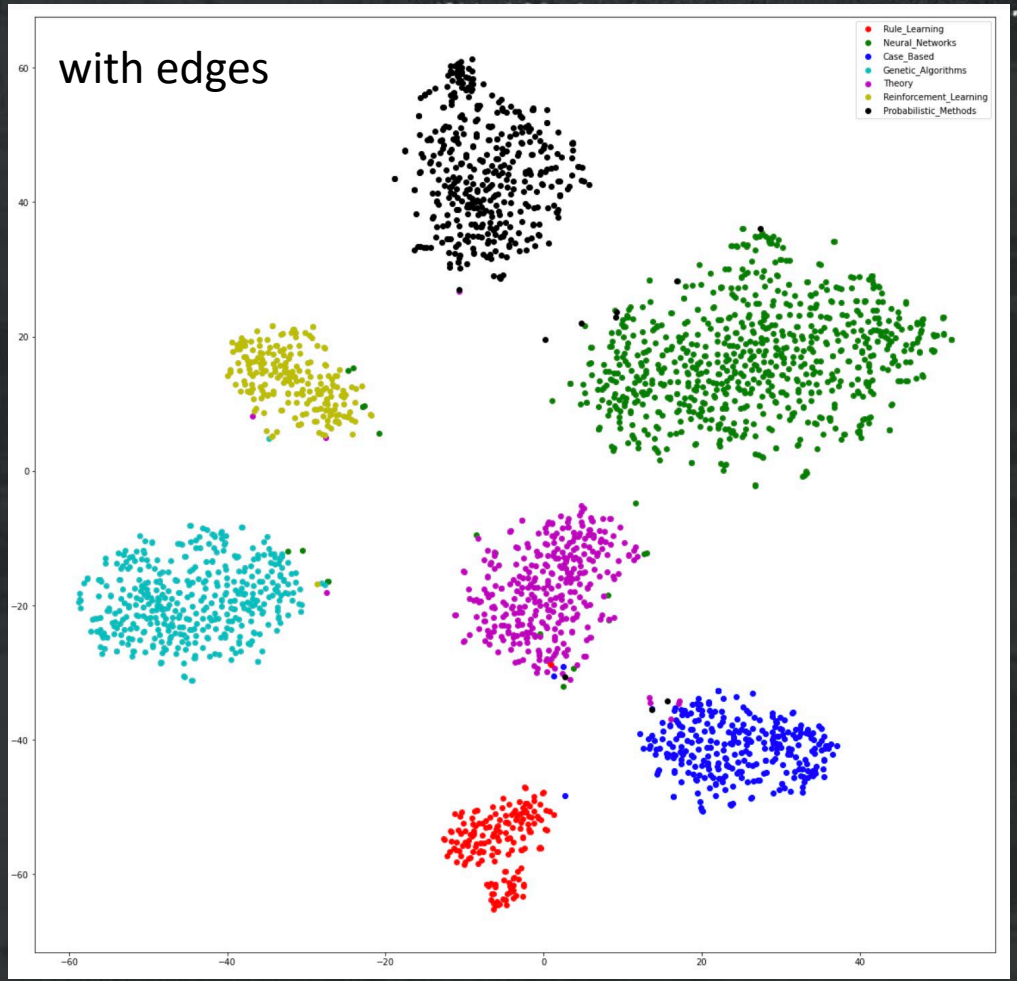
Supervised GraphWise

```
property_names = []

for i in range(900):
   property_names.append("F" + str(i + 1))

params = dict(vertex_target_property_name = "TOPIC",
              vertex_input_property_names = property_names)

model = analyst.supervised_graphwise_builder(**params)

model.fit(graph1)

vertex_vectors = model.infer_embeddings(graph1, graph1.get_vertices())
```


Copyright © 2022, Oracle and/or its affiliates

Supervised GraphWise





Copyright © 2022, Oracle and/or its affiliates

まとめと参考資料

- まとめ
  - 既存の表データからグラフを作成することができる
  - 特に「関係性」を含むデータではグラフを使った分析手法が有用である
- ブログ
  - Qiita（日本語）https://qiita.com/tags/oraclegraph
  - Medium（英語）https://medium.com/tag/oracle-graph
- Slack
  - OracleDevRel (= invitation link, please visit #oracle-db-graph)

Copyright © 2022, Oracle and/or its affiliates

Thank you

Copyright © 2022, Oracle and/or its affiliates