Assignment 2: SQL - Connected Components and PageRank

- Aditi Balaji, ab231


```
CREATE PROCEDURE ConnectedComponent AS
BEGIN
   DECLARE @Current INT;
   DECLARE @CurrCompID INT = 1;

   DROP TABLE IF EXISTS CurrComp;
   CREATE TABLE CurrComp (NodeID INT PRIMARY KEY);

   DROP TABLE IF EXISTS NewPapers;
   CREATE TABLE NewPapers (ID INT);

   DROP TABLE IF EXISTS Components;
   CREATE TABLE Components(CompID INT, NodeID INT);

   DROP TABLE IF EXISTS Visited;
   CREATE TABLE Visited(NodeID INT PRIMARY KEY);

   -- Add unique papers into newpapers
   INSERT INTO NewPapers
   SELECT DISTINCT paperID AS ID FROM edges
   UNION
   SELECT DISTINCT citedPaperID AS ID FROM edges;

   -- Using cursor for bfs
   DECLARE node_cursor CURSOR FOR SELECT ID FROM NewPapers;
   OPEN node_cursor;

   FETCH NEXT FROM node_cursor INTO @Current;

   -- LOOP TO IDENTIFY COMPONENTS
   WHILE @@FETCH_STATUS = 0
   BEGIN
      IF NOT EXISTS (SELECT 1 FROM Visited WHERE NodeID = @Current)
      BEGIN
         TRUNCATE TABLE CurrComp;
         INSERT INTO CurrComp (NodeID) VALUES (@Current);
```

```
        WHILE 1 = 1
        BEGIN
            INSERT INTO CurrComp (NodeID)
            (
                SELECT e.paperID
                FROM edges e
                WHERE EXISTS (SELECT 1 FROM CurrComp cc WHERE cc.NodeID =
e.citedPaperID)
                    AND NOT EXISTS (SELECT 1 FROM CurrComp cc WHERE cc.NodeID =
e.paperID)
                UNION
                SELECT e.citedPaperID
                FROM edges e
                WHERE EXISTS (SELECT 1 FROM CurrComp cc WHERE cc.NodeID =
e.paperID)
                    AND NOT EXISTS (SELECT 1 FROM CurrComp cc WHERE cc.NodeID =
e.citedPaperID)
            );

            IF @@ROWCOUNT = 0 BREAK;
        END

        INSERT INTO Components (CompID, NodeID)
        SELECT @CurrCompID, NodeID FROM CurrComp;

        INSERT INTO Visited (NodeID)
        SELECT NodeID FROM CurrComp;

        SET @CurrCompID = @CurrCompID + 1;
    END

    FETCH NEXT FROM node_cursor INTO @Current;
END

CLOSE node_cursor;
DEALLOCATE node_cursor;

WITH complist AS (
    SELECT CompID, COUNT(*) AS numComp
    FROM Components
    GROUP BY CompID
    HAVING COUNT(*) > 4 AND COUNT(*) <= 10
)
SELECT c.CompID, d.NodeID AS PaperID, n.paperTitle
```

```
    FROM complist c
    JOIN Components d ON d.CompID = c.CompID
    JOIN nodes n ON n.paperID = d.NodeID
    ORDER BY c.CompID ASC;
END;

EXEC ConnectedComponent;
```

Final Output:

```
2,9509135,Classical and Quantum Mechanics of Non-Abelian Chern-Simons Particles
2,304155,Exact String-like Solutions of the Gauged Nonlinear O(3) Model
2,9805010,On the Gauged Non-compact Spin System
2,9507015,Topological and Nontopological Solitons in a Gauged O(3) Sigma Model
2,9703185,N=2 Supersymmetric Gauged O(3) Sigma Model
2,9506015,Statistical Mechanics of Non-Abelian Chern-Simons Particles
2,9303080,Non-Abelian Chern-Simons Quantum Mechanics
2,9707150,Bogomolnyi Solitons and Hermitian Symmetric Spaces
16,9502105,FIELD THEORETICAL AND QUANTUM MECHANICAL DESCRIPTIONS OF
COLLIDING AND
16,9703200,The Low Energy Limit of the Chern-Simons Theory Coupled to Fermions
16,9402020,Perturbative Bosonic End Anyon Spectra and Contact Interactions
16,7080,Relativistic scalar Aharonov-Bohm scattering
16,9703090,Perturbative Expansion in the Galilean Invariant Spin One-Half
16,9603185,The Aharonov-Bohm scattering : the role of the incident wave
16,9510085,Calculation of the Aharonov-Bohm wave function
16,9906170,Radiative Corrections to the Aharonov-Bohm Scattering
16,9411175,Aharonov-Bohm Scattering of a Localized Wave Packet: Analysis of the
16,9710025,On the Nonrelativistic Limit of the Scattering of Spin One-half
19,9212110,Three Dimensional Chern-Simons Theory as a Theory of Knots and Links III
19,9812105,Vassiliev Invariants in the Context of Chern-Simons Gauge Theory
19,9312215,Knot invariants from rational conformal field theories
19,9401095,Chirality of Knots 9_{42} and 10_{71} and Chern-Simons Theory
19,9607030,Vassiliev Invariants for Links from Chern-Simons Perturbation Theory
19,9807155,Combinatorial Formulae for Vassiliev Invariants from Chern-Simons Gauge
30,9706080,Moving Frames Hierarchy and BF Theory
30,9712255,Chiral solitons from dimensional reduction of Chern-Simons gauged
30,9709075,Chiral solitons from dimensional reduction of Chern-Simons gauged
30,9611185,A Nonrelativistic Chiral Soliton in One Dimension
30,9507110,Calogero-Sutherland model from excitations of Chern-Simons vortices
43,9904055,Finiteness following from underlying theory: a natural strategy
43,9906015,Two- and Three-particle States in a Nonrelativistic Four-fermion Model
43,9412050,Generalised Point Interactions for the Radial Schrodinger Equation via
43,5195,A differential equation approach for examining the subtraction schemes
43,3255,Dimensional Transmutation and Dimensional Regularization in Quantum
```

43,9511010,The regulated four parameter one dimensional point interaction
43,9706070,Non-perturbative regularization and renormalization: simple examples
48,8110,Understanding Skyrmions using Rational Maps
48,12215,Solitonic fullerene structures in light atomic nuclei
48,9904160,Spherically Symmetric Solutions of the SU(N) Skyrme Models
48,206160,Skyrmed Monopoles
48,210310,Homotopy of Rational Maps and the Quantization of Skyrmions
63,9611150,Dimensional Renormalization in phi^3 theory: ladders and rainbows
63,9805025,A dilogarithmic 3-dimensional Ising tetrahedron
63,9612010,Weight Systems from Feynman Diagrams
63,9712140,Non-zeta knots in the renormalization of the Wess-Zumino model?
63,9807125,How useful can knot and number theory be for loop calculations?
141,9511210,Modular Invariance and the Odderon
141,9508025,Quasiclassical QCD Pomeron
141,9802100,Solution of the Odderon Problem
141,9611025,Direct solution of the hard pomeron problem for arbitrary conformal
141,9805135,New Results on the Odderon in QCD

```
-- Pagerank
CREATE PROCEDURE PageRank AS
BEGIN
   DECLARE @Numnode INT;
   DECLARE @Damp FLOAT = 0.85;
   DECLARE @Conv FLOAT = 0.01;
   DECLARE @Iter INT = 0;
   DECLARE @Delta FLOAT = 0;

   SELECT @Numnode = COUNT(DISTINCT PaperID) FROM nodes;

   -- Create Page Rank Table
   CREATE TABLE #Pagerank (
   PaperID INT PRIMARY KEY,
   RankValue FLOAT
   );
   INSERT INTO #Pagerank (PaperID, RankValue)
   SELECT DISTINCT PaperID, 1.0 / @Numnode as RankValue
   FROM nodes;

   -- Remove Sink Nodes
   CREATE TABLE #edges_clean (
      PaperID INT,
      CitedID INT
   );
```

```
    INSERT INTO #edges_clean (PaperID, CitedID)
    SELECT e.PaperID, e.CitedPaperID
    FROM edges e
    UNION ALL
    SELECT sn.sink_paperid AS PAPERID, nd.paperid AS CITEDPAPERID
    FROM (
        SELECT DISTINCT e.citedPaperID AS sink_paperid
        FROM edges e
        WHERE e.citedPaperID NOT IN (
            SELECT DISTINCT e1.paperID
            FROM edges e1
            JOIN edges e2 ON e2.citedPaperID = e1.paperID
        )
    ) sn
    JOIN nodes nd ON sn.sink_paperid != no.paperID;

    -- Create #num_citations table
    CREATE TABLE #num_cite (
        PaperID INT PRIMARY KEY,
        CiteNum INT
    );

    INSERT INTO #num_cite (PaperID, CiteNum)
    SELECT paperID, COUNT(*) as Num_of_Citations
    FROM #edges_clean
    GROUP BY paperID;

    -- Declare Table Variables
DECLARE @iteration_Pagerank TABLE (
    PaperID INT PRIMARY KEY,
    PageRank FLOAT
);

-- Recursive CTE for iteration
WITH RecursivePagerank AS (
    SELECT
        p.PaperID,
        ((1 - @dampingFactor) / @n) + (@dampingFactor * SUM(pr.PageRank / c.Citations)) AS
NewPageRank
    FROM
        #Pagerank p
    JOIN
        #edges_updated e ON e.citedPaperID = p.PaperID
```

```
    JOIN
        #Pagerank pr ON pr.PaperID = e.PaperID
    JOIN
        #num_citations c ON c.PaperID = e.PaperID
    GROUP BY
        p.PaperID
),
PagerankDiff AS (
    SELECT
        SUM(ABS(pr.PageRank - r.NewPageRank)) AS Difference
    FROM
        #Pagerank pr
    JOIN
        RecursivePagerank r ON pr.PaperID = r.PaperID
)
INSERT INTO @iteration_Pagerank (PaperID, PageRank)
SELECT
    r.PaperID,
    r.NewPageRank
FROM
    RecursivePagerank r
CROSS JOIN
    PagerankDiff pd
WHERE
    pd.Difference > @convergenceThreshold;

-- Update #Pagerank in a single statement
UPDATE pr
SET
    pr.PageRank = ipr.PageRank
FROM
    #Pagerank pr
JOIN
    @iteration_Pagerank ipr ON pr.PaperID = ipr.PaperID;

-- Check for convergence
IF (SELECT TOP 1 Difference FROM PagerankDiff) <= @convergenceThreshold
BEGIN
    -- Convergence reached
    SET @iteration = @maxIterations;
END
ELSE
BEGIN
    -- Continue the iteration
```

```
    SET @iteration = @iteration + 1;
    -- Truncate instead of DELETE
    TRUNCATE TABLE @iteration_Pagerank;
END

-- Calculate sum once for better performance
DECLARE @Sum FLOAT;
SELECT @Sum = SUM(PageRank) FROM #Pagerank;

-- Select Top 10 results
SELECT TOP 10
    pr.PaperID,
    pr.PageRank / @Sum AS PageRank,
    n.paperTitle
FROM
    #Pagerank pr
JOIN
    nodes n ON n.paperID = pr.PaperID
ORDER BY
    pr.PageRank DESC;



end
```

Answer:
9504090,0.014724248584604112,Massless Black Holes and Conifolds in String Theory
9510135,0.014446305360815112,Bound States Of Strings And p-Branes
9711200,0.013647582829266357,The Large N Limit of Superconformal Field Theories and Supergravity
9802150,0.009697907266117285,Anti De Sitter Space And Holography
208020,0.008629895477463104,Open strings and their symmetry groups
9602065,0.007716301077518256,D--branes and Spinning Black Holes
9305185,0.007549767481589173,Duality Symmetries of 4D Heterotic Strings
9611050,0.007129378771709418,TASI Lectures on D-Branes
9501030,0.0058154548234567,Strong/Weak Coupling Duality from the Dual String
9602135,0.005416172003172842,Entropy and Temperature of Black 3-Branes