SciPi

ICS5114 – Big Data Processing

Supplementary Material

June 2019

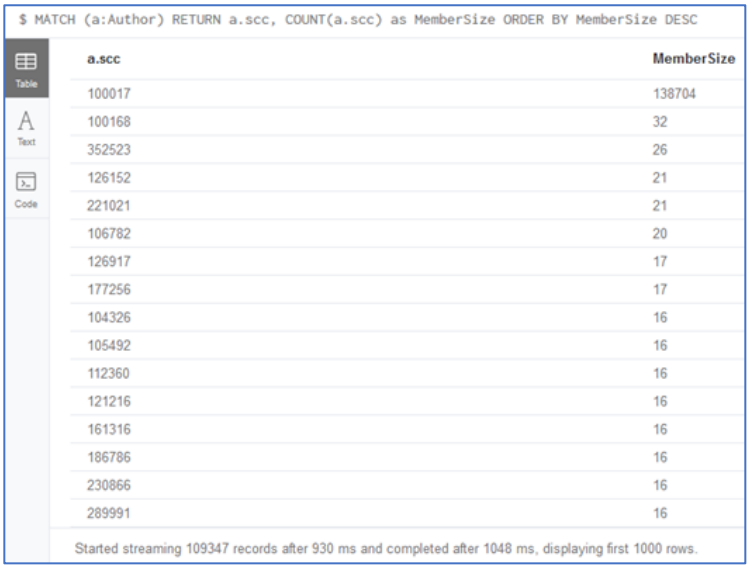
Joseph Azzopardi (474488M)

Andrew Cachia (211494M)

Supplementary Material – Neo4j Queries and Visualisation

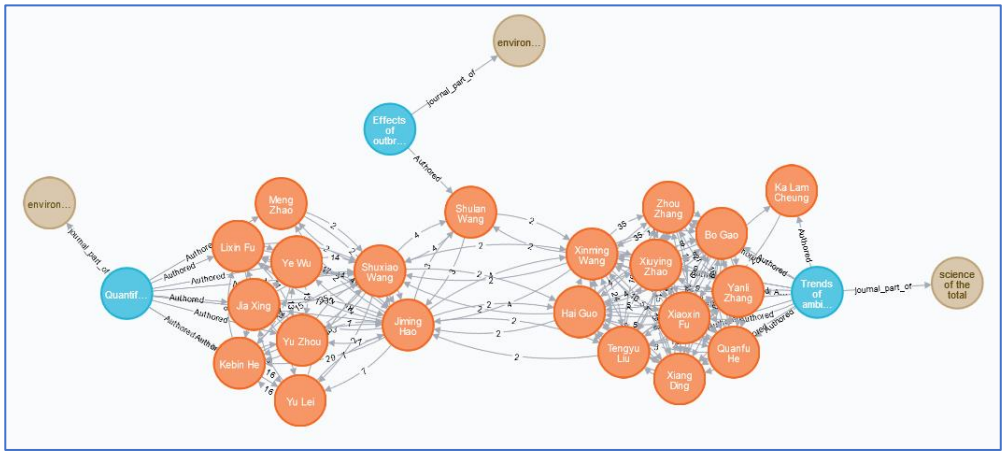
<table>
<tr><th>#</th><th>Query</th><th>Cyper Statement</th></tr>
<tr><td>1</td><td>Get all the groups discovered through Label Propagation and Strongly Connected Components, and order by member size.</td><td><pre>
Label Propagation
MATCH (a:Author)
RETURN a.label, COUNT(a.label) as MemberSize
ORDER BY MemberSize DESC

SCC
MATCH (a:Author)
RETURN a.scc, COUNT(a.scc) as MemberSize
ORDER BY MemberSize DESC
</pre></td></tr>
<tr><td></td><td colspan="2">

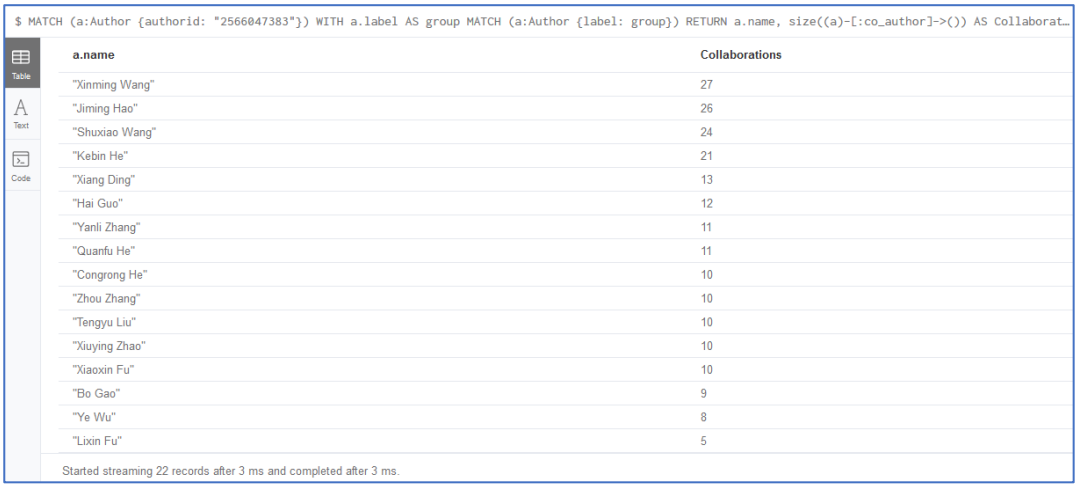


</td></tr>
<tr><td>2</td><td>List Authors within the same group, and the relationships to papers, and conference or journals.

Author: Bo Gao
authorid: 2566047383</td><td><pre>
MATCH (a:Author {authorid: "2566047383"})
WITH a.label AS group
MATCH (a:Author {label: group})<-[r]-
(p:Paper)-[rr]->(c)
WHERE c:Journal OR c:ConfInstance
RETURN a,r,p,rr,c
</pre></td></tr>
</table>

	label: 263328	
		

3	Get a cluster of authors starting from author “Bo Gao”. Get all members and compute their indegree
	
	Author: Bo Gao
	authorid: 2566047383
	label: 263328
		`MATCH (a:Author {authorid: "2566047383"})`
		`WITH a.label AS group`
		`MATCH (a:Author {label: group})`
		`RETURN a.name, size((a)-[:co_author]->()) AS Collaborations`
		`ORDER by Collaborations DESC`

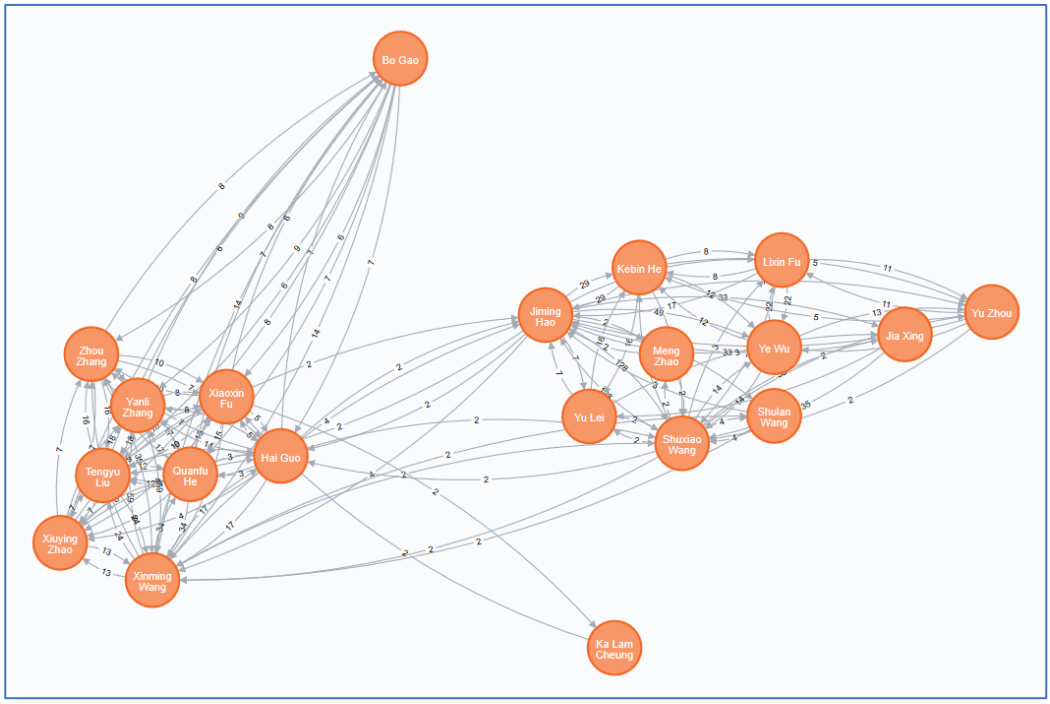
		

4	Get top 5 possible author collaborators within the same cluster for a *particular* author that he/she has not yet collaborated with.
	
	Author: Bo Gao
	authorid: 2566047383
	label: 263328
		`MATCH (b:Author {authorid: "2566047383"})`
		`WITH b.label AS group`
		`MATCH (a:Author {label: group})`
		`WHERE NOT((a)-[:co_author]-( {authorid: "2566047383"})) AND a.authorid <> "2566047383"`
		`RETURN a.name, size((a)-[:co_author]->()) AS Collaborations_in_Group`
		`ORDER by Collaborations_in_Group DESC Limit 5`

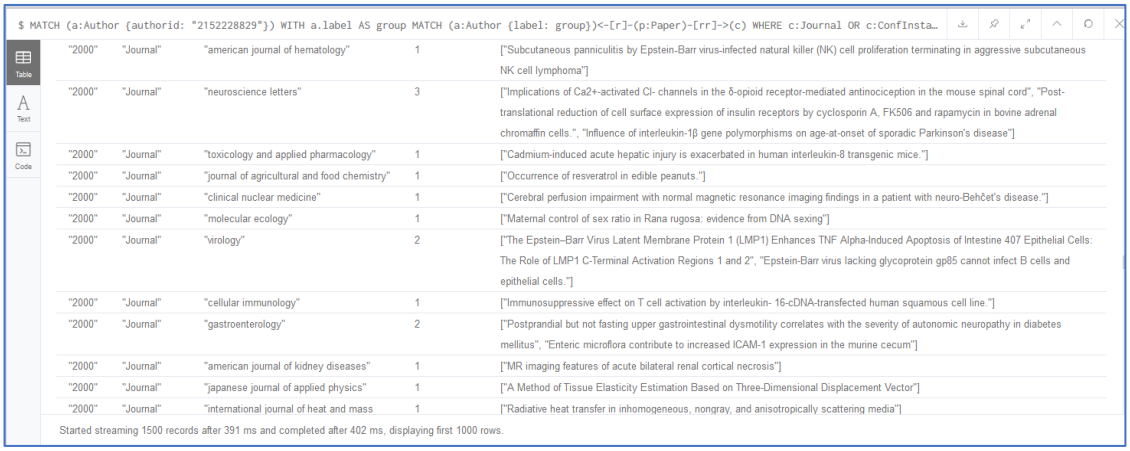
$ MATCH (b:Author {authorid: "2566047383"}) WITH b.label AS group MATCH (a:Author {label: group}) WHERE NOT((a)-[

Table | Text | Code

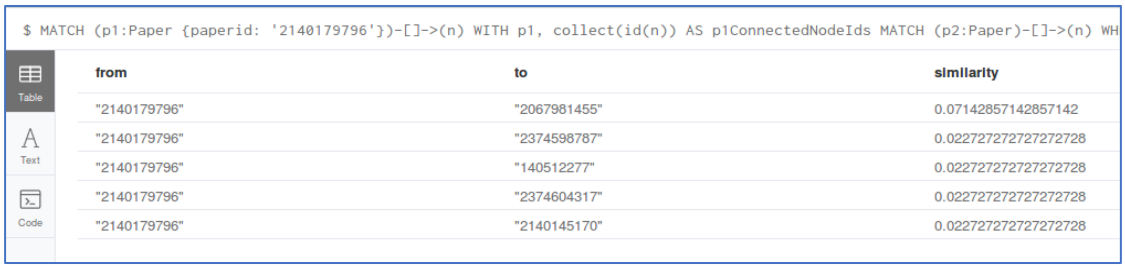
a.name	Collaborations_in_Group
"Jiming Hao"	26
"Shuxiao Wang"	24
"Kebin He"	21
"Congrong He"	10
"Ye Wu"	8



5	Dynamics from a publication Network.
	Given a collaborations pattern, group papers to the corresponding journal/conference, across the years.
		`MATCH (a:Author {authorid: "2152228829"})`
		`WITH a.label AS group`
		`MATCH (a:Author {label: group})<-[r]-`
		`(p:Paper)-[rr]->(c)`
		`WHERE c:Journal OR c:ConfInstance`
		`WITH p, p.doc_type as Type, p.year as Year,`
		`c.name as Event, collect(p) as temp`
		`RETURN Year, Type, Event, count(p.title) as`
		`Num_Papers, collect(p.title) as Papers`
		`ORDER BY Year ASC`



6	Get top 5 recommended papers given a particular paper id	`MATCH (p1:Paper {paperid: '2140179796'})-[]->(n)`
		`WITH p1, collect(id(n)) AS p1ConnectedNodeIds`
		`MATCH (p2:Paper)-[]->(n) WHERE p1 <> p2`
		`MATCH (p2)-[]->(n2)`
		`WITH p1, p1ConnectedNodeIds, p2,`
		`collect(id(n2)) AS p2ConnectedNodeIds`
		`RETURN p1.paperid AS from,`
		`p2.paperid AS to,`
		
		`algo.similarity.jaccard(p1ConnectedNodeIds, p2ConnectedNodeIds) AS similarity`
		`ORDER BY similarity DESC`
		`LIMIT 5`



7	Show relationship between two highly recommended papers	`MATCH (p1:Paper {paperid: '2140179796'})-[]->(a)`
		
		`MATCH (p2:Paper {paperid: '2067981455'})-[]->(b)`
		
		`RETURN p1,p2,a,b`

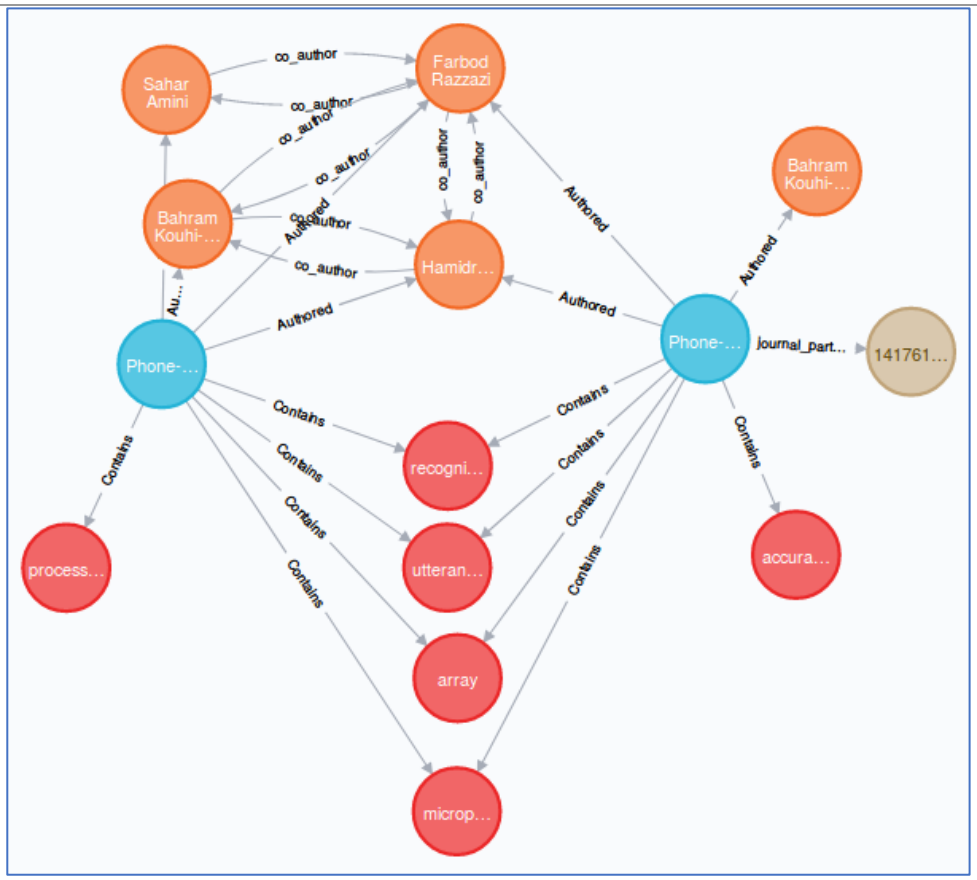


Table 1 – Graph Visualisation