SciPi

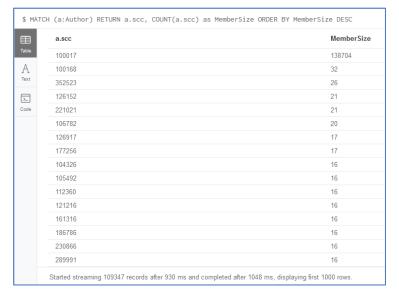
ICS5114 - Big Data Processing

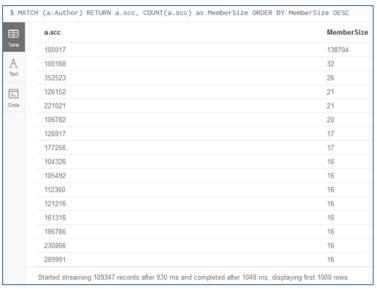
Supplementary Material

June 2019 Joseph Azzopardi (474488M) Andrew Cachia (211494M)

Supplementary Material - Neo4j Queries and Visualisation

#	Query	Cyper Statement
1	Get all the groups discovered through Label Propagation and Strongly Connected Components, and order by member size.	# 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

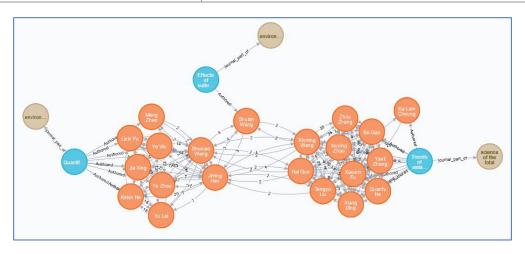




List Authors within the same group, and the relationships to papers, and conference or journals.

Author: Bo Gao authorid: 2566047383 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

label: 263328



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

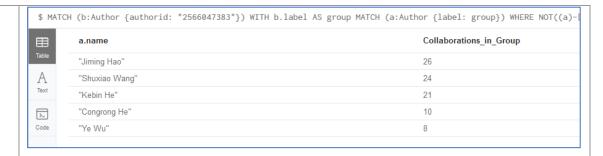
a.name	Collaborations
"Xinming Wang"	27
"Jiming Hao"	26
"Shuxiao Wang"	24
"Kebin He"	21
"Xiang Ding"	13
"Hai Guo"	12
"Yanli Zhang"	11
"Quanfu He"	11
"Congrong He"	10
"Zhou Zhang"	10
"Tengyu Liu"	10
"Xiuying Zhao"	10
"Xiaoxin Fu"	10
"Bo Gao"	9
"Ye Wu"	8
"Lixin Fu"	5

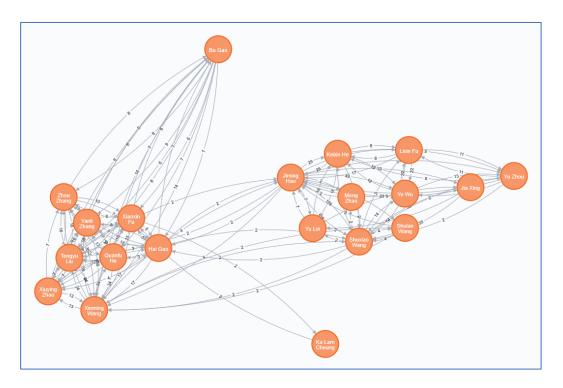
Get top 5 possible author collaborators within the same cluster for a *particular* author that he/she has not yet collaborated with.

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





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

\$ MA	"2000"	or {authorid "Journal"	"american journal of hematology"	ip MATCH (a	:Author {labe1: group})<-[r]-(p:Paper)-[rr]->(c) WHERE c:Journal OR c:Cou ["Subcutaneous panniculitis by Epstein-Barr virus-infected natural killer (NK) cell prolifera	
Table	"2000"	"Journal"	"neuroscience letters"	3	NK cell lymphoma"] ["Implications of Ca2+-activated CI- channels in the δ-opioid receptor-mediated antinocice	untion in the mouse opinal cord" "Post
Α	2000	Journal	Hedroscience retters	3	translational reduction of cell surface expression of insulin receptors by cyclosporin A, FI	
Text					chromaffin cells.", "Influence of interleukin-1β gene polymorphisms on age-at-onset of spo	
N. Code	"2000"	"Journal"	"toxicology and applied pharmacology"	1	["Cadmium-induced acute hepatic injury is exacerbated in human interleukin-8 transgenion	mice."]
Code	"2000"	"Journal"	"journal of agricultural and food chemistry"	1	["Occurrence of resveratrol in edible peanuts."]	
	"2000"	"Journal"	"clinical nuclear medicine"	1	["Cerebral perfusion impairment with normal magnetic resonance imaging findings in a pa	tient with neuro-Behĉet's disease."]
	"2000"	"Journal"	"molecular ecology"	1	["Maternal control of sex ratio in Rana rugosa: evidence from DNA sexing"]	
	"2000"	"Journal"	"virology"	2	["The Epstein-Barr Virus Latent Membrane Protein 1 (LMP1) Enhances TNF Alpha-Induc The Role of LMP1 C-Terminal Activation Regions 1 and 2", "Epstein-Barr virus lacking gly epithelial cells."]	
	"2000"	"Journal"	"cellular immunology"	1	["Immunosuppressive effect on T cell activation by interleukin- 16-cDNA-transfected huma	in squamous cell line."]
	"2000"	"Journal"	"gastroenterology"	2	["Postprandial but not fasting upper gastrointestinal dysmotility correlates with the severit	y of autonomic neuropathy in diabetes
					mellitus", "Enteric microflora contribute to increased ICAM-1 expression in the murine ce	cum"]
	"2000"	"Journal"	"american journal of kidney diseases"	1	["MR imaging features of acute bilateral renal cortical necrosis"]	
	"2000"	"Journal"	"japanese journal of applied physics"	1	["A Method of Tissue Elasticity Estimation Based on Three-Dimensional Displacement Vo	ector"]
	"2000"	"Journal"	"international journal of heat and mass	1	["Radiative heat transfer in inhomogeneous, nongray, and anisotropically scattering media	1"]
					WITH p1, collect(id(n)) AS p1Co MATCH (p2:Paper)-[]->(n) WHERE MATCH (p2)-[]->(n2)	
					WITH p1, p1ConnectedNodeIds, p. collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds,
\$	MATCH (p1:Paper	<pre>fpaperid: '2140179796'})</pre>		collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity
\$		p1:Paper	{paperid: '2140179796'})		collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity
	a fr				collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity MATCH (p2:Paper)-[]->(n)
Tab	f fi	rom	341		collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity MATCH (p2:Paper)-[]->(n) similarity
Œ	a fr	rom 2140179796 2140179796	5"		collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity MATCH (p2:Paper)-[]->(n) similarity 0.07142857142857142 0.022727272727272728
Tab	file "	rom 2140179796 2140179796 2140179796	5" 5"		collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity MATCH (p2:Paper)-[]->(n) similarity 0.07142857142857142 0.022727272727272728 0.022727272727272728
Tab	file "	rom 2140179796 2140179796	5" 5"		collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity MATCH (p2:Paper)-[]->(n) similarity 0.07142857142857142 0.022727272727272728
Tab	file "	rom 2140179796 2140179796 2140179796	5 ³⁷ 5 ³⁷ 5 ³⁷		collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity MATCH (p2:Paper)-[]->(n) similarity 0.07142857142857142 0.022727272727272728 0.022727272727272728
Tabl A Tex	file "	2140179796 2140179796 2140179796 2140179796	5 ³⁷ 5 ³⁷ 5 ³⁷		collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity MATCH (p2:Paper)-[]->(n) similarity 0.07142857142857142 0.0227272727272728 0.022727272727272728 0.022727272727272728
Table A Text	ft	2140179796 2140179796 2140179796 2140179796 2140179796	5" 5" 5" 5"	7)<-[3->	collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity MATCH (p2:Paper)-[]->(n) similarity 0.07142857142857142 0.0227272727272728 0.0227272727272728 0.0227272727272728
Table A Text	find the state of	2140179796 2140179796 2140179796 2140179796 2140179796	sy s	[]->(r	collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity MATCH (p2:Paper)-[]->(n) similarity 0.07142857142857142 0.0227272727272728 0.0227272727272728 0.0227272727272728
Table A Text	find the state of	2140179796 2140179796 2140179796 2140179796 2140179796	5" 5" 5" 5"	[]->(r	collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity MATCH (p2:Paper)-[]->(n) similarity 0.07142857142857142 0.0227272727272728 0.0227272727272728 0.0227272727272728
Table A Text	find the state of	2140179796 2140179796 2140179796 2140179796 2140179796	sy s	[]->(r	collect(id(n2)) AS p2Connected RETURN p1.paperid AS from,	NodeIds ectedNodeIds, ity MATCH (p2:Paper)-[]->(n) similarity 0.0217272727272728 0.0227272727272728 0.0227272727272728 0.0227272727272728



Table 1 – Graph Visualisation