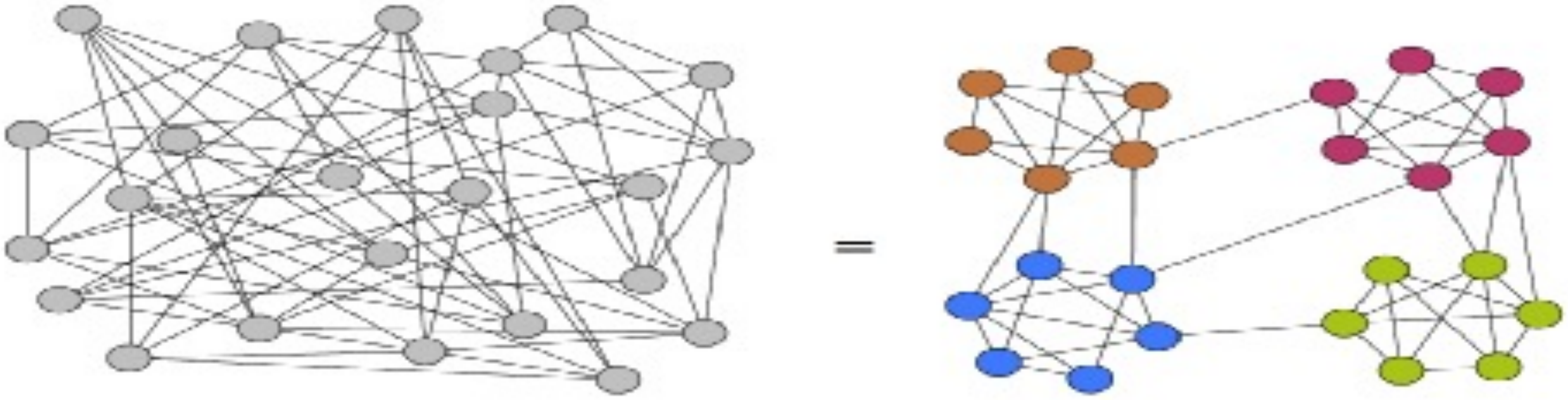


Enhancing Group Formation in Online Discussions with the Louvain Algorithm for Group Detection

Minkyung Lee and Priya Sharma
Penn State University

PennState

Group Formation

Group formation has been challenging in online settings as instructors must capture different needs and characteristics of the large number of students to create balanced groups.

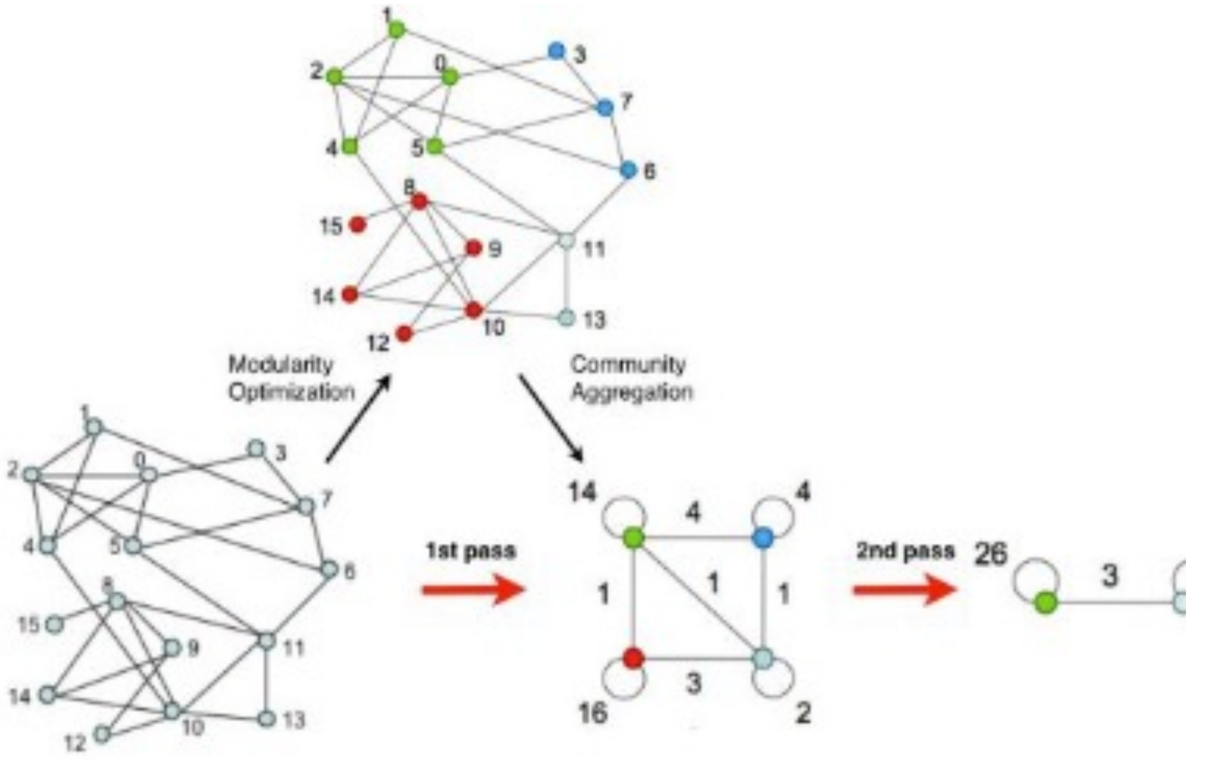
Group Formation

- Different methods exist to form groups:
  - random selection
  - self-selection
  - teacher selection
  - automatic selection (Felder & Brent, 2001).

- Algorithmic group formation
  - Uses algorithms to assign students to groups based on criteria such as skills, interests, or learning needs.
  - Methods can be effective and economically beneficial, as it can generate optimal groupings quickly with a given set of criteria (Muller et al., 2022).

Group Formation

- Group formation by frequency of interaction: Louvain algorithm (Blondel et al., 2008)
- The most popular modularity optimization method (Menczer et al., 2018)
- The Louvain algorithm iteratively optimizes modularity by reassigning nodes to communities and aggregating them into super-nodes, resulting in a hierarchical community structure.



(Blondel et al., 2008)

Community detection to examine learning

- Jan & Vlachopoulos (2018) - They suggest that community-based learning and structural similarity between networks and communities make SNA a natural choice for deeper understanding of group interaction. The study substantiated their method as an effective framework for structural identification of a Community of Inquiry (CoI) and Community of practice (CoP).

- Yassine (2020) used the label propagation algorithm and Louvain algorithm for community detection and found those different community detection algorithms can be implemented on learning networks and detect communities.

Purpose of the study

To explore the potential of the Louvain algorithm for group formation in online learning environments, concentrating on identifying groups based on their connectivity patterns.

Our research questions are as follows:

1. How can the Louvain algorithm be applied to identify groups within online learning environments based on their connectivity patterns?
2. What are the global network characteristics of courses and the local network characteristics of the groups identified by the algorithm?

Data collection

The raw data of an online undergraduate business course was anonymized and used for analysis.

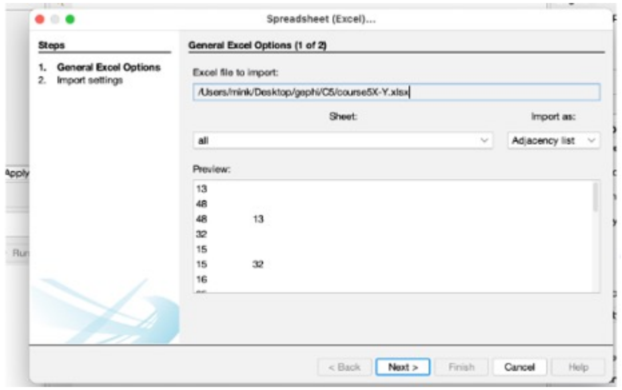
	# of students	# of posts	# of topics
Course4	53	808	9
Course5	50	764	9
Course6	51	839	9

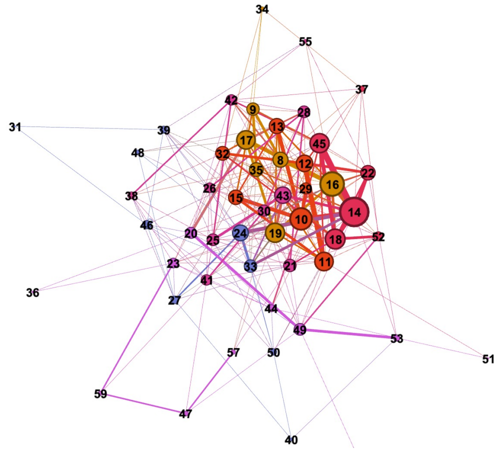
Students were given different reflective questions for nine weeks and required to write their opinions as well as replies to others.

Data preparation

course n	course canv	term nam	student ca	discussion entry id	parent discussion	topic id	topic	group	topic message	discussion message	message	workflow	reply d	topic assignment id	score	points
MBADM	1933434	2188 - 20	6881588	10500000024068311		10500000012157160	M5 Discussion: New Thinking		<p class="page-title">Th	<p>Volunteer Line Judge - Girl's Volleyball</p>\n<p>To challenge n	2402	active	1	10500000010228279	25	25
MBADM	1933434	2188 - 20	6636858	10500000024076724		10500000012157160	M5 Discussion: New Thinking		<p class="page-title">Th	<p>The service that I decided to utilize to incorporate new thinking	2391	active	1	10500000010228279	25	25
MBADM	1933434	2188 - 20	6636858	10500000024076771	105000000240683	10500000012157160	M5 Discussion: New Thinking		<p class="page-title">Th	<p>Hi Neville,</p>\n<p>This is a great analysis of your new way of	976	active	2	10500000010228279	25	25
MBADM	1933434	2188 - 20	6901221	10500000024081680		10500000012157160	M5 Discussion: New Thinking		<p class="page-title">Th	<p>M5 Discussion-New Thinking</p>\n<p>I have an idea for New	2818	active	1	10500000010228279	25	25
MBADM	1933434	2188 - 20	6868764	10500000024083666		10500000012157160	M5 Discussion: New Thinking		<p class="page-title">Th	<p>My object is a wallet…because I need a new one.</p>\n<p>Sin	2866	active	1	10500000010228279	25	25
MBADM	1933434	2188 - 20	6910663	10500000024084150		10500000012157160	M5 Discussion: New Thinking		<p class="page-title">Th	<p>An activity I have identified that would benefit from new thinking	2973	active	1	10500000010228279	25	25
MBADM	1933434	2188 - 20	6896477	10500000024085451		10500000012157160	M5 Discussion: New Thinking		<p class="page-title">Th	<p>While running today, I asked myself, "how much do i need to ru	2588	active	1	10500000010228279	25	25

udent n	Respondi	scussion message	topic	topic mes	created a
58	53	†Hi <<name>>,I totally agree with you on the fact that keeping these different shots together was typical of Griffith, eeping the narrative and melodramatic perspective of good and evil in the movie. I personally think that the crowd of eople were black since this white family was abundantly fearful, but they could have also just feared the intensity of the tuation and the stampede could have been white people or the KKK. Would love to hear your opinion!Best,<<name>> †¬†¬†	Lesson 1 Discussion	The followin	51
105	53	i <<name>>!I really enjoyed reading your response and must say that I agree with a lot of your ideas! This clip was efinitely typical of Griffith and his style. I believe that the people in the valley below are antagonists... I think I saw a urning house which I assumed was the house of the protagonist woman and her children. I think you were right that the oman and children were afraid of them and I believe that may be why!Thanks for your response,	Lesson 1 Discussion	The followin	70
87	53	ello <<name>>,I think it's a great point that you mentioned that the panoramic shot ending in a long shot shows the normity of the group of men fleeing. I did not think about it that way but I totally agree with you, the long shot puts a ocus on how large the group of people is below the family. I thought the long shot contrasting the shot on the family hows just how open and vulnerable the family is.¬†	Lesson 1 Discussion	The followin	91
83	53	agree 100% what you say and it is also somewhat terrifying to see. I enjoyed reading your response because you detailed verything so well and interpreted the messages from the clip as they should be!¬†	Lesson 3 Discussion	Watch the cl	164
56	53	<name>>,I agree with you that the plot of the film would have not worked if Dexter and Tracy were poor. Like you entioned the political aspects of the film are there and are unavoidable. I thought Tracy was going to chose <<name>> the end because of his growth and how they were together. Dexter and Tracy ending up together made me think they nly chose each other because they are both upper class.	Lesson 5 Discussion	Watch The P	264
58	53	i <<name>>,Thanks for your response and that is an interesting observation you made about <<name>>. I would agree hat that could be a reason they added a character like him, I think the balance of both sides of the spectrum of class eeps the story more relatable.Best,	Lesson 5 Discussion	Watch The P	269
		hat's an interesting perspective about <<name>> representing the lower class' envy of the elite.¬† I consider his haracter there to point out the kind of admiration that Tracy is searching for at the beginning, someone who will admire er and be the "high priest" who worships her as the goddess she feels she is.¬† He maintains that persona throughout, ving the viewers the stark contrast between him and Dexter after her education, ensuring that the her love interest hould be Dext.¬† Now that you mention it, it makes sense that <<name>> represents the lower class, even though he has ade it out into the elite class.¬† This is even evident by Dexter stepping in to protect him from himself when <<name>>			



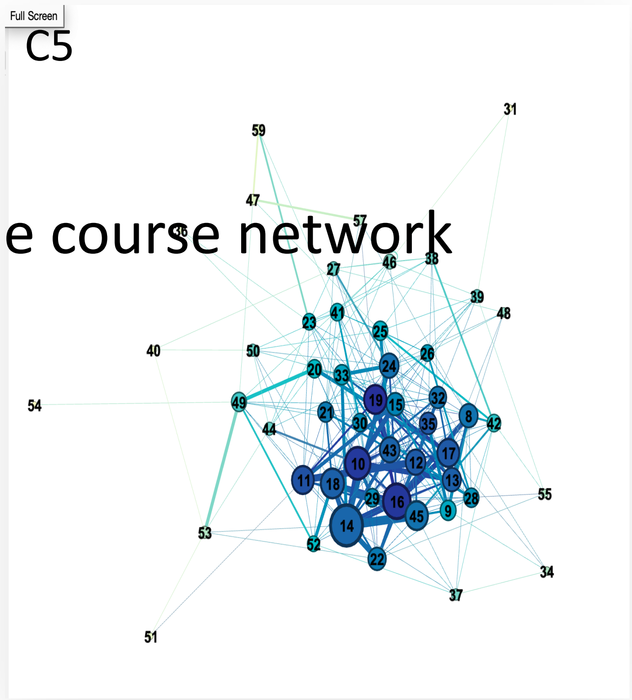


Data preparation

- We extracted the source and target data from each course and created adjacency lists to generate sociograms.
- This data was imported into Gephi, a tool for data visualization and analysis, where we ran various network statistics such as weighted degree, diameter, and density, including modularity function to identify communities within the network.
- We applied the same procedure to all courses for consistency and comparability.

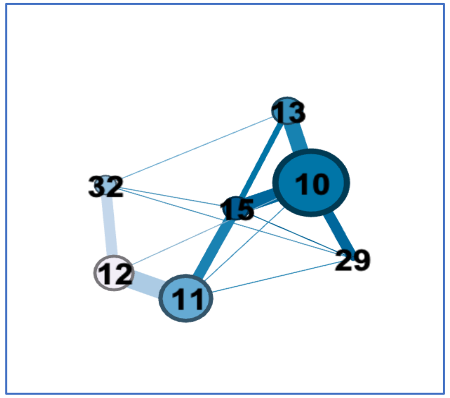
*We used the default setting of modularity resolution and chose the undirected graphs.

Results



a) Explored the entire network at a global level to understand the course network structures with specific measures

b) Investigated one of the courses (C5) to understand the interaction within groups generated by the Louvain algorithm



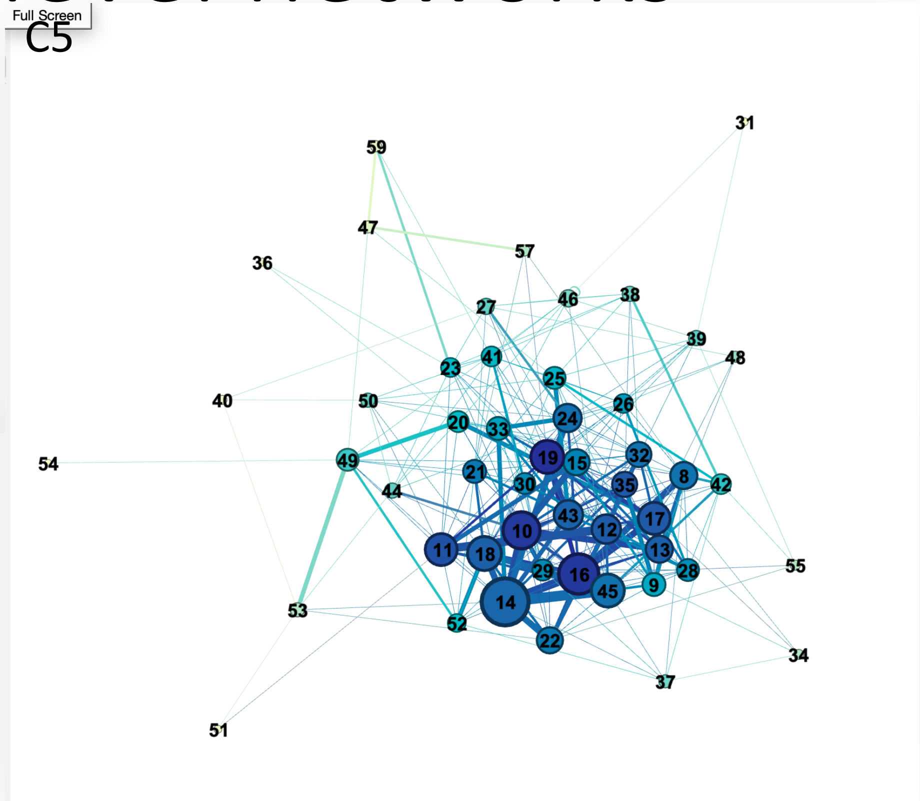
Results – Global level

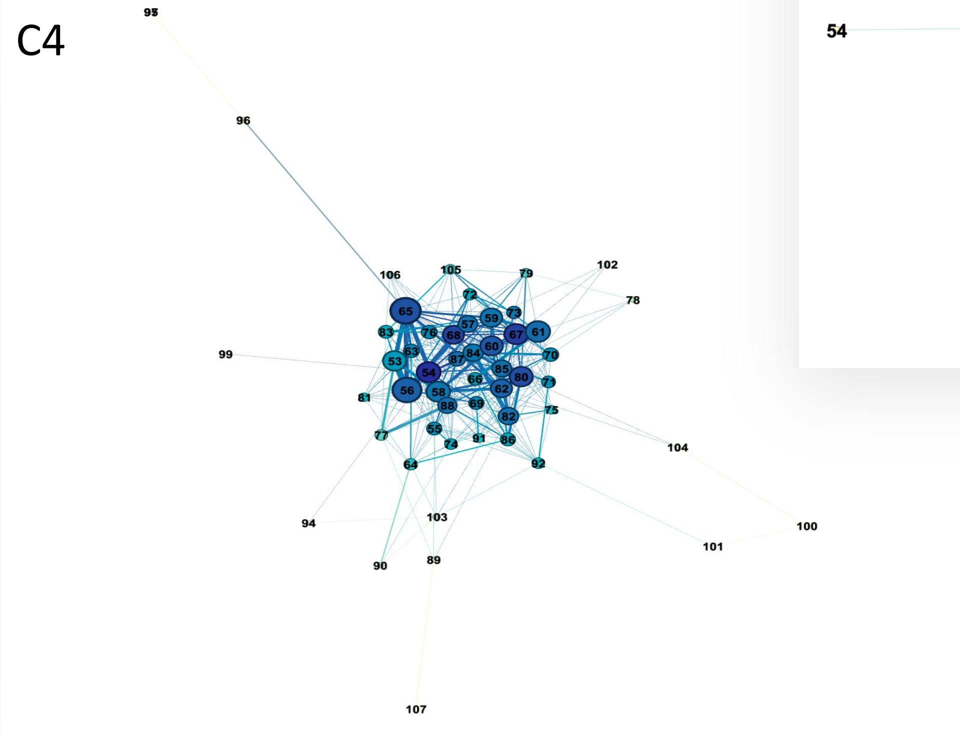
- Global metrics describe characteristics, topology, and structures (Morrison et al., 2022).

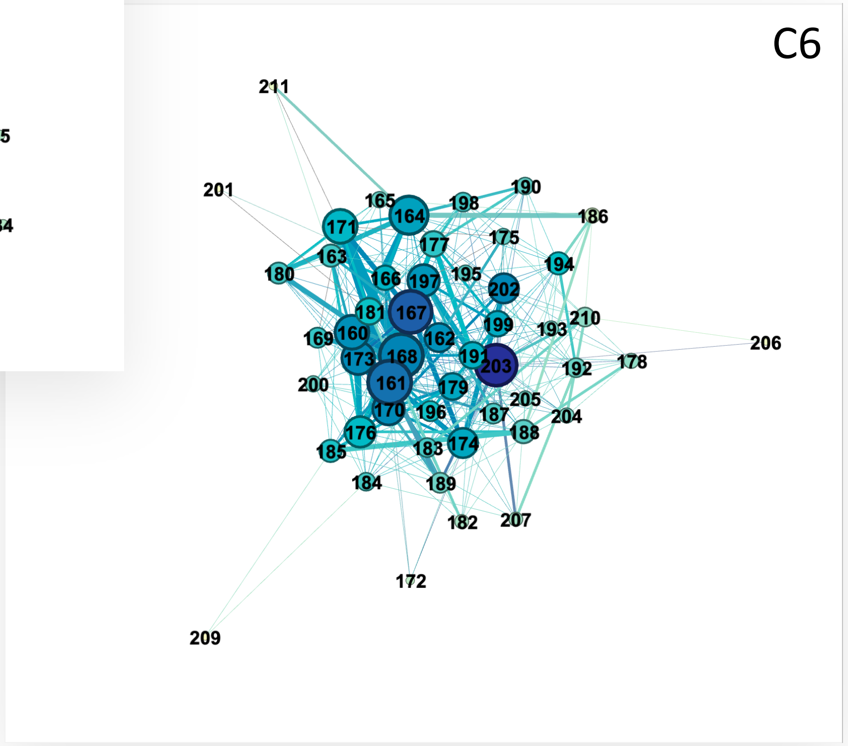
	# of nodes	# of edges	Graph density	weighted degree	path length	Diameter	Clustering coefficient
C4	53	314	0.228	16.679	2.111	5	0.345
C5	50	261	0.213	14.96	1.996	4	0.282
C6	51	340	0.267	17.294	1.807	4	0.372

- Course 6(C6) has the most active and interconnected student network, with the highest levels of interaction, communication intensity, and cohesion. Course 4(C4) shows moderate levels of interaction, while Course 5 (C5) appears to have the least active network.

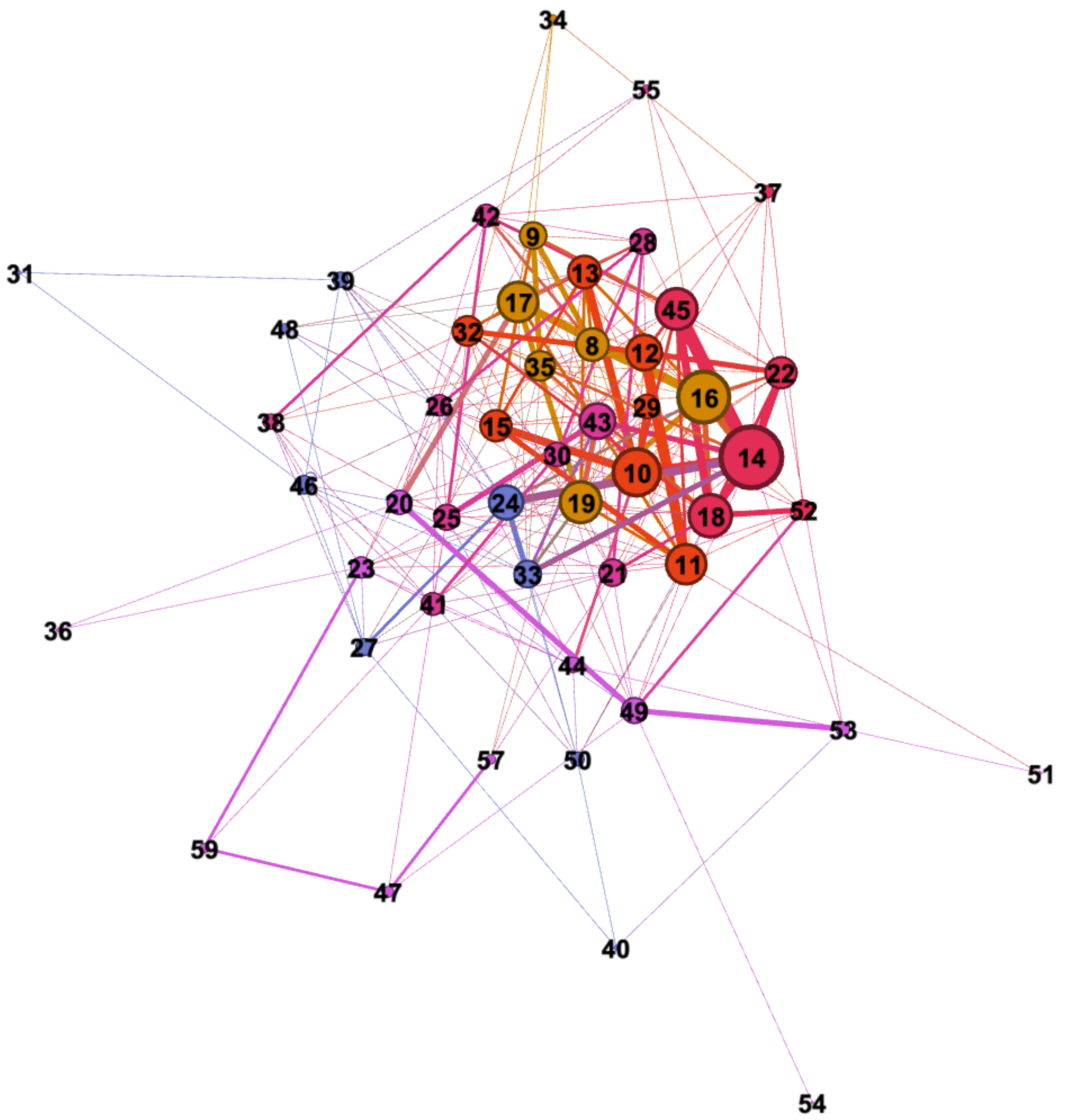
Results – Global level networks







Node Size: Weighted degree
Color: Relative connectivity as a hub



Group 0 –orange
Group 1- purple
group 2 – yellow
Group 3- pink
Group 4- blue
Group 5- light violet

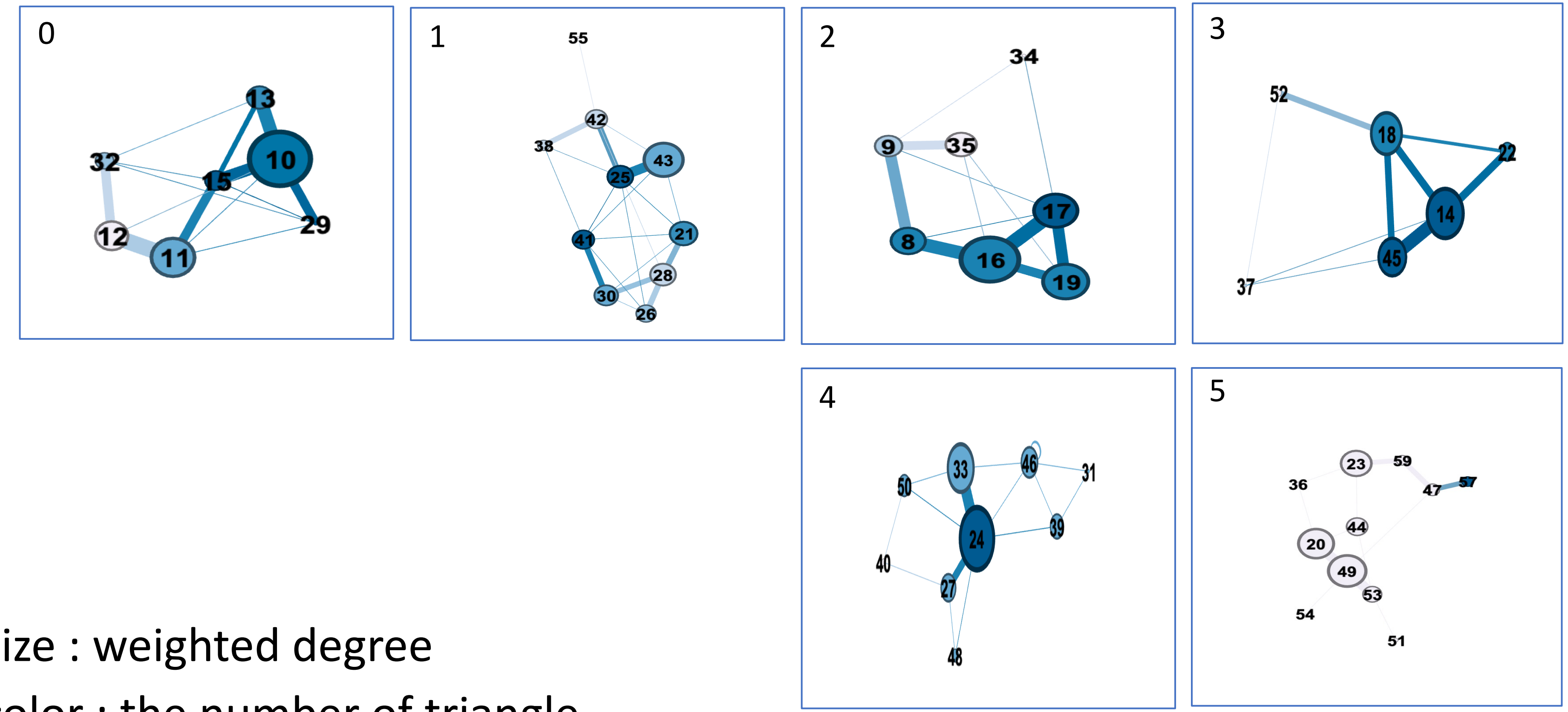
Results – Local level

- Local metrics are measures describing the attributes of a network at the node or edge level (Morrison et al., 2022).

	Degree	Weighted Degree	Clustering coefficient	Triangles
group0	**4.286**	**9.143**	**0.619**	**4.714**
group1	4.200	5.800	0.463	3.700
group2	3.333	9.000	0.528	2.500
group3	3.714	7.714	0.595	3.000
group4	3.333	4.000	0.385	1.000
group5	2.364	3.636	0.015	0.091
C5	3.538	6.549	0.434	2.501

- The nodes in group 0 are more connected in groups, more influential(weighted degree) and more densely connected(clustering coefficient, triangle) to their neighbors than the nodes in the other groups.
- We may imply Group 0 could be more central or influential within the C5 network compared to the other groups.

Results – Local level networks



- Node size : weighted degree
- Node color : the number of triangle

Implication

Global level- the course network structures

Local level- the group network structures by Louvain algorithm

We think this research..

a. expand the ways to form groups using algorithm for online learning as well as classroom learning
b. give insight on group interaction network to anticipate the group cohesiveness

Limitations

- Implementation requires technical expertise and understanding.
- Online learning environments are dynamic, causing constant group adjustment.
- Louvain algorithm is more often used with large networks.

Future research

- Conduct content analysis to integrate the results from the network data
- Conduct controlled experiments to compare group dynamics and learning outcomes of Louvain algorithm with traditional methods to evaluate its effectiveness
- Develop user-friendly tools to make Louvain algorithm accessible for instructors in various learning settings

Thank you and we welcome any questions and feedback!

Minkyung Lee mzl263@psu.edu
Priya Sharma pus3@psu.edu

Correlations						
		Degree	Weighted Degree	Clustering Coefficient	Triangles	Eigenvector Centrality
Degree	Pearson Correlation	1	.964**	.752	.965**	.935*
	Sig. (2-tailed)		.008	.142	.008	.020
	N	5	5	5	5	5
Weighted Degree	Pearson Correlation	.964**	1	.733	.993**	.873
	Sig. (2-tailed)	.008		.159	<.001	.053
	N	5	5	5	5	5
Clustering Coefficient	Pearson Correlation	.752	.733	1	.750	.896*
	Sig. (2-tailed)	.142	.159		.144	.040
	N	5	5	5	5	5
Triangles	Pearson Correlation	.965**	.993**	.750	1	.867
	Sig. (2-tailed)	.008	<.001	.144		.057
	N	5	5	5	5	5
Eigenvector Centrality	Pearson Correlation	.935*	.873	.896*	.867	1
	Sig. (2-tailed)	.020	.053	.040	.057	
	N	5	5	5	5	5

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).