

# Detecting Leadership in Peer-moderated Online Collaborative Learning: Text Mining and Social Network Analysis for Learning Analytics

Kui Xie, Gennaro Di Tosto, Young Suk Cho, Lin Lu

The Research Laboratory for Digital Learning, College of Education and Human Ecology

## Introduction

**Structured tasks** and **peer-moderated online discussions** are pedagogical models that have shown unique benefits in **discussion-based learning activities**. When students are appointed with leadership roles to stimulate discussion, they seem to be able to positively affect the dynamics in their groups by engaging participants, raising questions and advancing problem solving.

Social processes within groups of students are often complex. In the case of **technology-based learning environments**, they are also often latent and not explicitly observed, due to the fact that students are physically isolated and only interact through computer-mediated communications.

The automatic processing of the contents of learning management systems, however, can aid the development of targeted interventions in technology based environment. And, with the added understanding of leadership dynamics and their effects on learning outcomes, it can harness the role of peers in enhancing the learning experience and environment.

## Aims

In this study, we are interested in examining the **emergence of leadership** in a collaborative, online learning setting.

We define **leadership** in terms of the individual contributions made to the online discussion, as well as the effects discussions have on group structure. To this aim we developed a computational technique capable of identifying patterns of interactions associated with messages that display elements of leadership, regardless of whether the author of a message is an appointed moderator or a regular student.

The following **research questions** guided the design of this study:

1. To what extent can data mining extract linguistic features to detect leadership at the message level?
2. To what extent can social network analysis model leadership at the person level?
3. To what extent does our computational approach provide evidence of emergent leadership?

## Dataset

The dataset was built mining the contents of asynchronous online discussions among students of one university course offered entirely online. Student-facilitated discussions over the topics assigned by the instructor were the major learning task. For each discussion session, up to two students were assigned to design discussion questions and led the class discussions.

This study included 57 course participants (gender: 11 M; 46 F — age: range 19-53; median 29). Four sections generated 4083 posts in total. Text, hyperlinks, and emotions were the main content of each post. Information such as posting time, de-identified author ID, and the post ID it responded to were aligned with each post, and are all used to prepare for the following analysis.

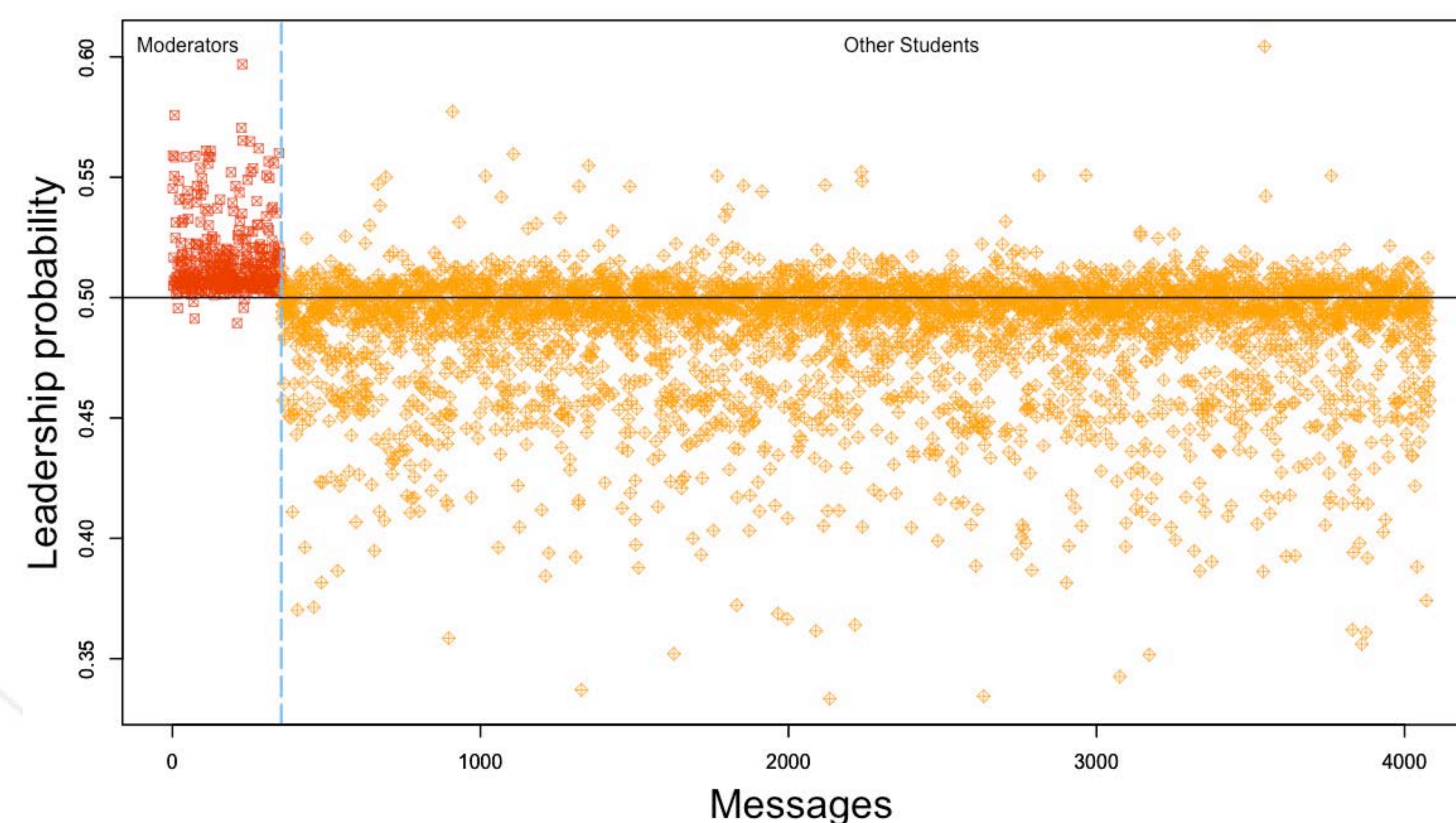
## Methods

The methodology designed is organized in two steps: (1) Processing of online communication and message categorization; (2) Student ranking and leadership detection.

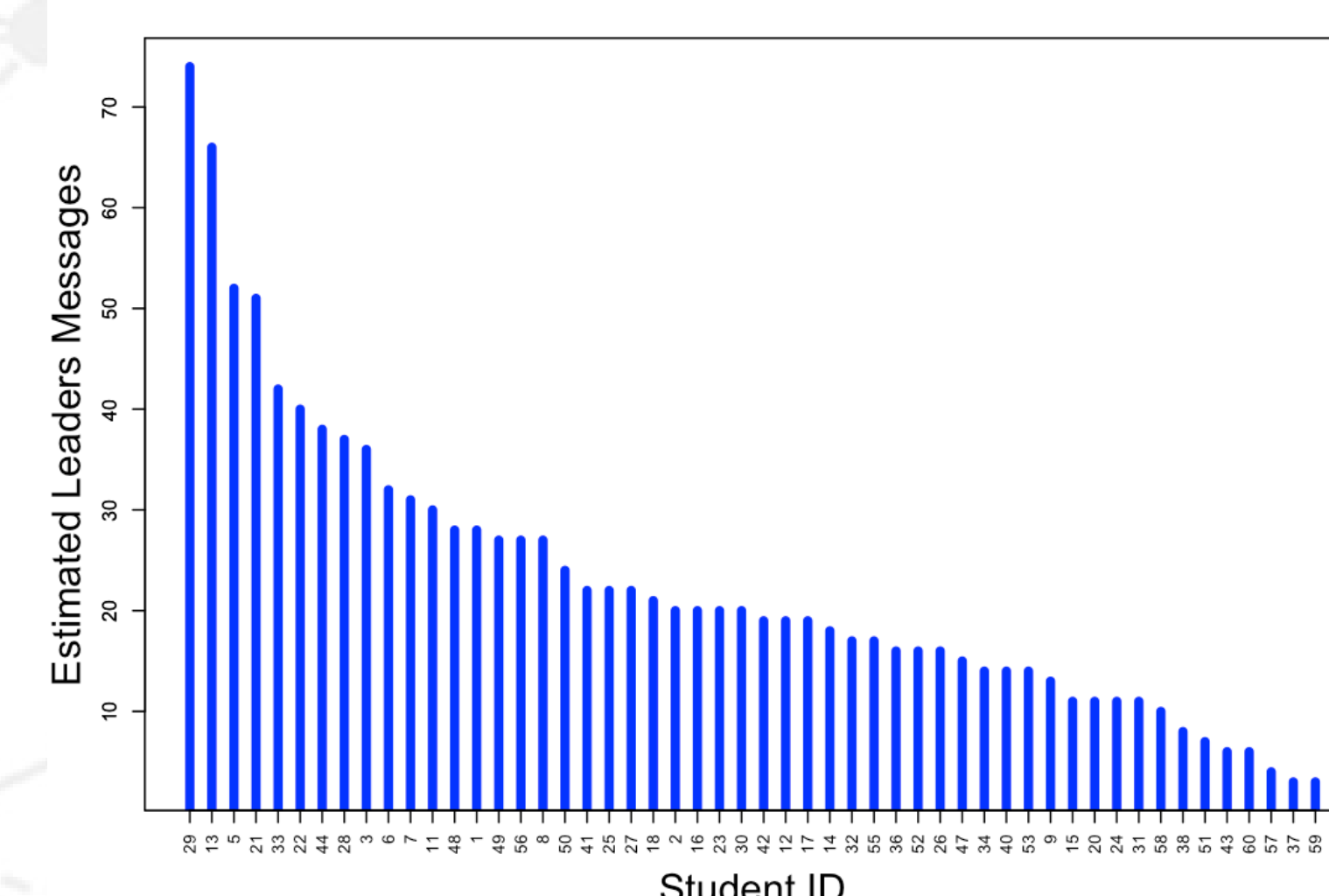
The first step makes use of **natural language processing** to analyze the content of students' online interactions and, consequently, machine learning techniques to identify those messages that play a moderating role in the discussions. We label these messages as **Leader Messages**.

The second step adopts **social network analysis** to identify structural relationships between the senders and recipients of Leader Messages and to calculate their network centrality in order to rank their contribution to moderating discussions. We label this person-level measure based on network centrality **Leadership Index**. The Leadership Index represents our outcome in the analysis of peer-moderated online discussion and leadership detection.

## Results (Message-level analysis)



Individual posts were numbered and converted to feature vector tables, then appropriately marked the author was a student formally asked to lead the weekly discussion (Appointed Leaders Postings). Subsequently, a probabilistic model, implemented using an AdaBoost algorithm, was used to classify messages in a binary category: Leaders vs. Others.

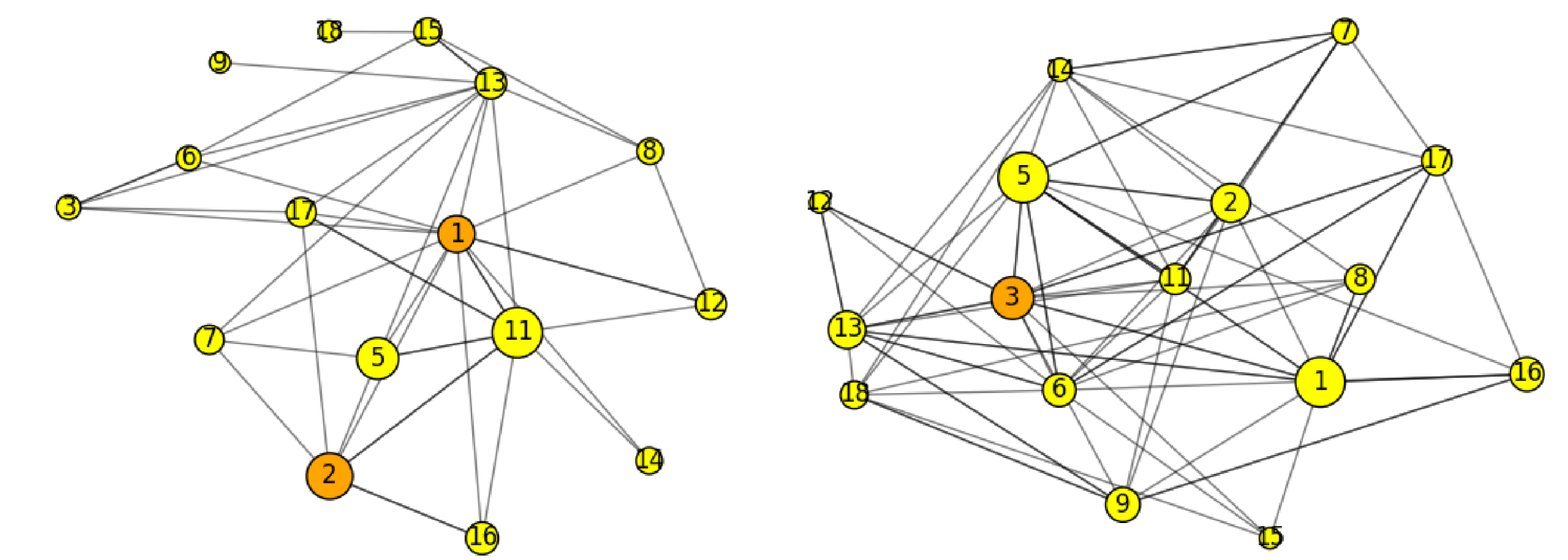


All the messages scoring above the threshold line are considered **Leaders Messages** (accuracy 88%). Messages posted by students who were not formally asked to lead discussion (above, right) that scored above the threshold level are used to identify candidates for **Emergent Leadership**.

## Results (Person-level analysis)

All the postings identified as Leader Messages were carried on in the analysis with the goal to further study students' moderation activity. Senders and recipients information were used to build a graph representing the weekly leadership patterns for each chapter's discussion across the four recorded sessions of the course.

In the directed graph built using Leader Messages, each node is a student and each edge represents a reply received in a discussion thread. Using the PageRank algorithm we calculated the centrality of the nodes and used it as a measure of the **Leadership Index** for that student.



ID	L	Centrality
11		0.1659
2	*	0.1375
5	*	0.1113
1	*	0.0851
16		0.0628
13		0.0581
12		0.0551
17		0.0512
7		0.0492
15		0.0431
14		0.0391
8		0.0377
6		0.0319
3		0.0308
18		0.0217
9		0.0195

ID	L	Centrality
5		0.1311
1	*	0.1281
3	*	0.0942
2		0.0796
13		0.0741
9		0.0636
16		0.0616
6		0.0588
11		0.0510
8		0.0486
17		0.0481
18		0.0423
7		0.0365
14		0.0312
15		0.0262
12		0.0250

The above selection of graphs represents a sample of the weekly discussions. The size of the nodes are proportional to their centrality. The corresponding tables on the left show the ranking of the students. An asterisk marks the appointed leader for the week (orange in the graph). As the examples show, it is not uncommon for a regular student to emerge as a more important leader in the network, obtaining higher centrality scores, with some students consistently receiving high Leadership Index (e.g. ID 5).

## Results (Person-level analysis)

Based on these findings, the following questions stand out as significant for guiding further research:

1. What are the characteristics that appointed moderators and emergent leaders have in common in their patterns of interactions?
2. Can we design a system of early warning signs on the basis of the leadership detection mechanism?
3. In what way peer-moderation can better utilize the activity of emergent leaders in the development of cooperative tasks?