FR. CONCEICAO RODRIGUES COLLEGE OF ENGINEERING
Department of Computer Engineering

Course, Subject & Experiment Details

Assignment No:	2
Title:	An Organizational Comparative Analysis
Name of the Student:	Warren Fernandes
Roll No:	8940
Date of Performance:	24/03/2023
Date of Submission:	11/04/2023

Evaluation:

Sr. No.	Rubric	Grade
1	On time submission/completion (2)	
2	Preparedness (2)	
3	Skill (4)	
4	Output (2)	

Signature of the Teacher

Analyzing social media and Learning Through Content and Social Network Analysis: NASA & ISRO

To provide explanation and demonstration of my analytic strategy and framework, this section focuses on several analysis methods we rely upon and how they are used in combination to generate new insights about learning. For this case study, we use a sample of public tweets posted by NASA and ISRO

Text Analysis:

The first step in our case was to build concise summaries of the communal textual discourse present in the dataset by identifying frequently used words (mostly nouns). Figure 1 and 2 shows a word cloud visualization of the top 100 most frequently used words in the NASA and ISRO Twitter chat over the data collection period. The search keyword (#NASA and #ISRO) and other common words (also known as “stop-words”) such as “of”, “will” and “to” were automatically removed prior to building this visualization. The size of a word in the visualization is directly related to the number of times it appears in the dataset relative to the other words found in that same dataset. In Netlytic, this visualization allows users to click on any of the words in the cloud in order to explore the context(s) in which the word appears.

NASA:

By exploring the top 100 words, we can group words into four broad categories. The first category includes words relevant to the class but not necessarily unexpected, including “Telescope”, “Space”, “Mars”, “astronauts,” and “Hubble”. The most frequently mentioned word in this category (and in the whole dataset) is “NASA”. The second group of frequently used words includes Twitter hashtags: #artemis, #nasa. The third category includes a set of Twitter users frequently mentioned in the dataset such as @Astro_Jeremy, @BoeingSpace and @csa_asc. The fourth category of frequent words reveals what types of online content were found to be useful and shared within the class. For example, the presence of words like “time”, “today”, “testing” in the word cloud suggests that Twitter is in part being used to disseminate data. In addition to the four broad categories found in the dataset, we also observed the frequent use of the symbol “ko”, added manually or automatically to tweets when they are “retweeted” by others. The use of KOs may indicate the extent to which class participants paid attention to what others post; the prominence here suggests frequent attention to posts with retweeting content to their own followers fulfilling the “Feed Forward” action. It is important to note that there is no suggested “optimal” ratio of retweets or replies to original posts that one might want to see in successful class discussions on Twitter.

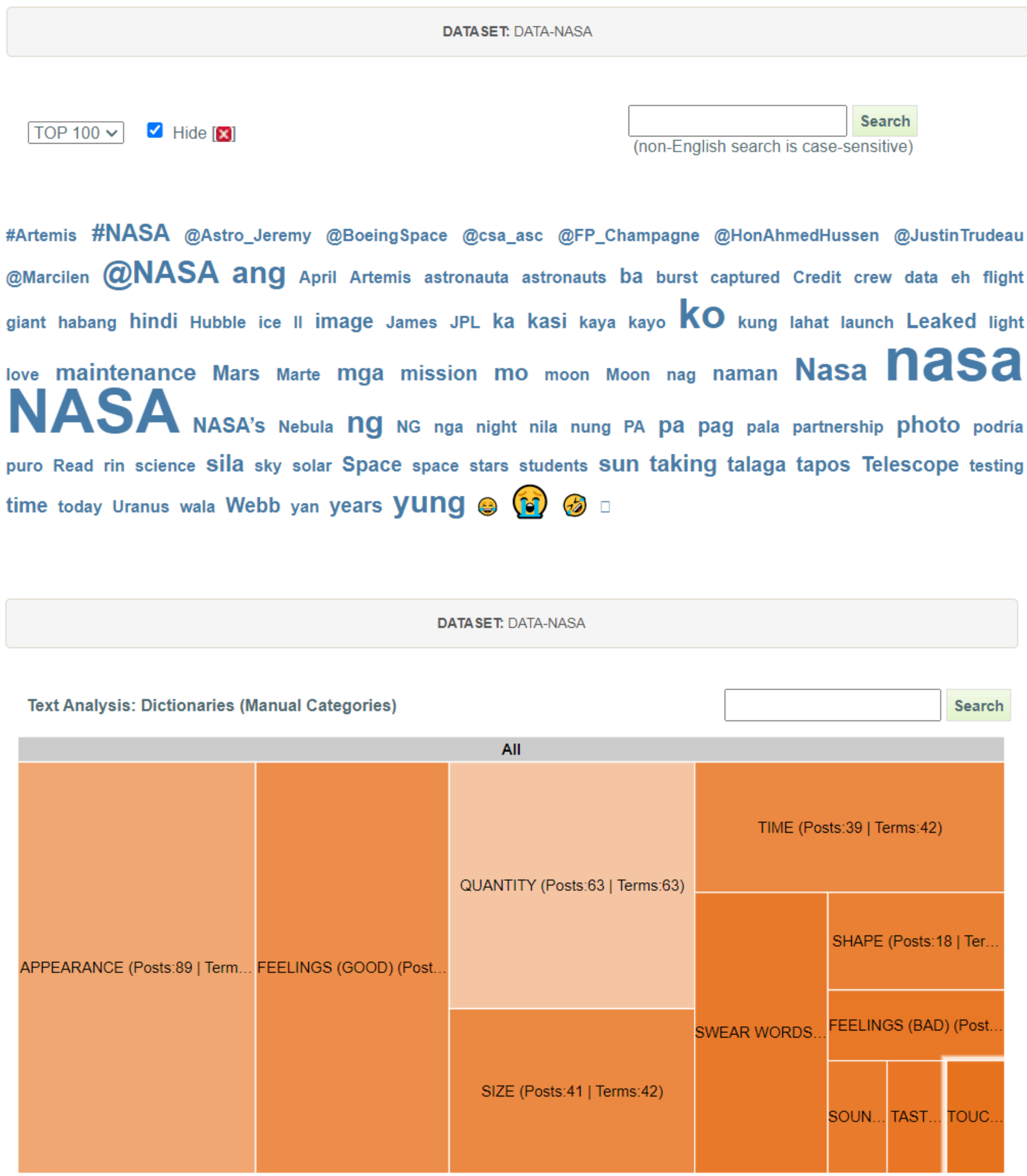
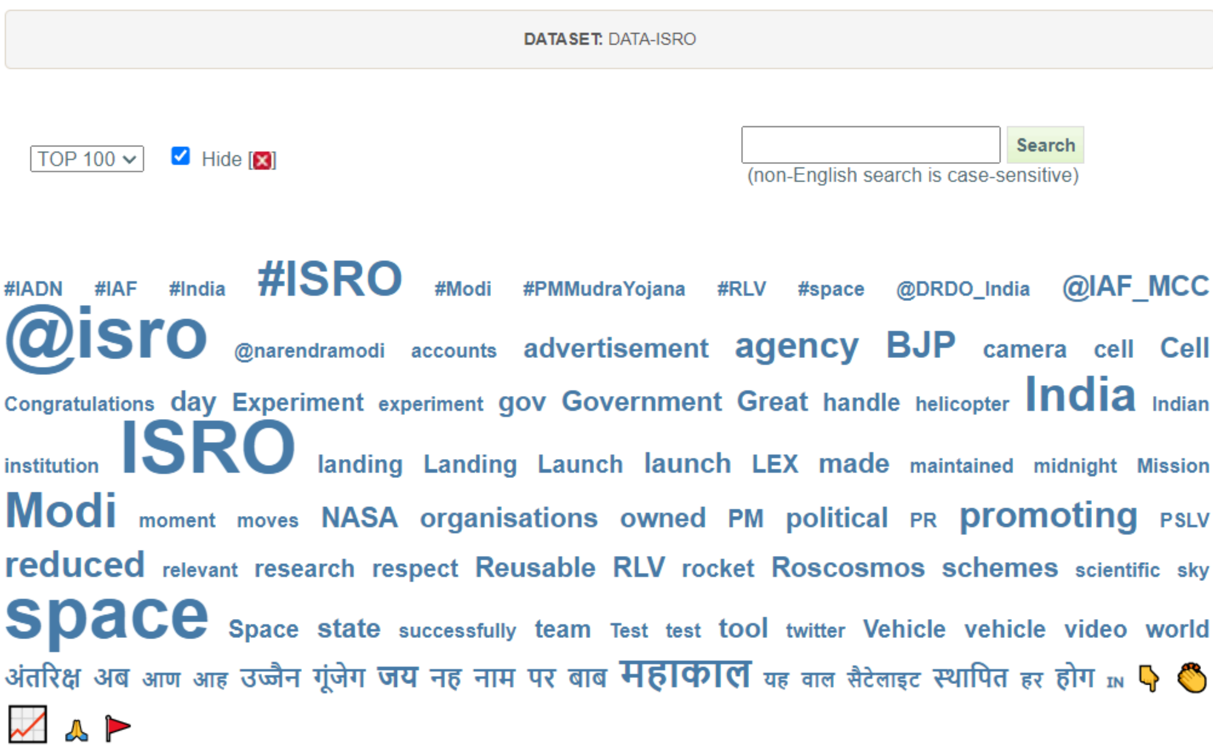


Figure 1

ISRO:

By exploring the top 100 words, we can group words into four broad categories. The first category includes words relevant to the class but not necessarily unexpected, including “Modi”, “Space”, “BJP”, “India,” and “promoting”. The most frequently mentioned word in this category (and in the whole dataset) is “ISRO”. The second group of frequently used words includes Twitter hashtags: #IADN, #IAF, #RLV, #PMMudraYojna. The third category includes a set of Twitter users frequently mentioned in the dataset such as @DRDO_India, @IAF_MCC and @narendramodi. The fourth category of frequent words reveals what types of online content were found to be useful and shared within the organization. For example, the presence of words like “vehicle”, “successfully”, “LEX” in the word cloud suggests that Twitter is in part being used to disseminate data.



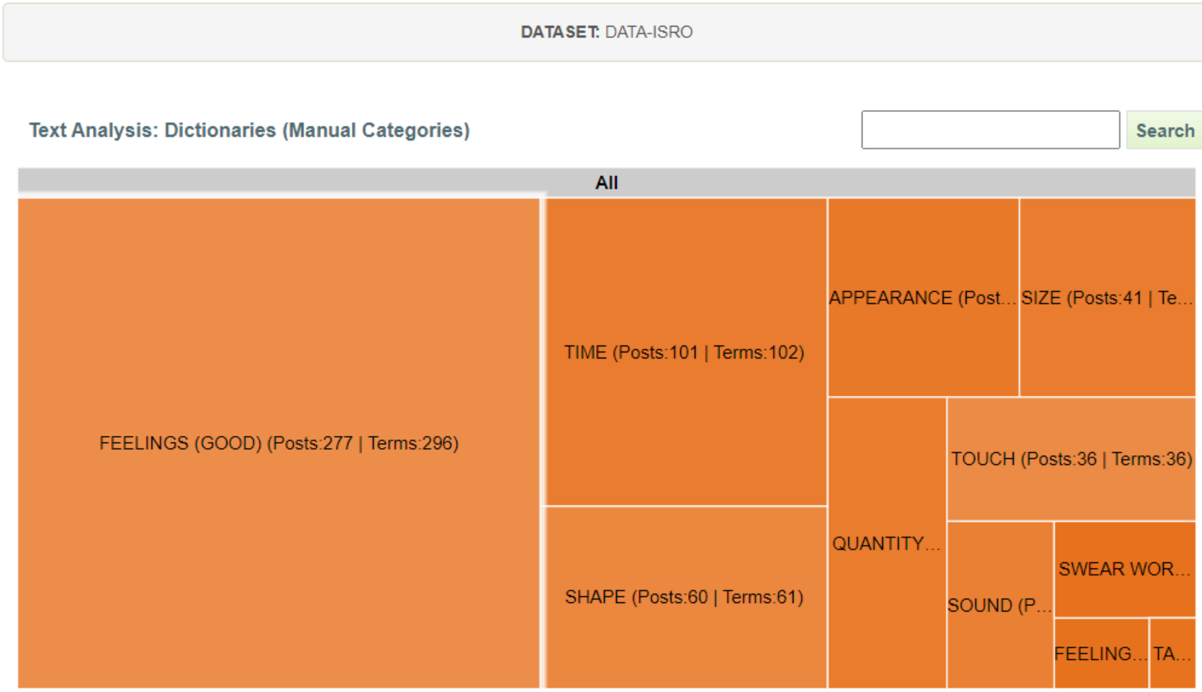


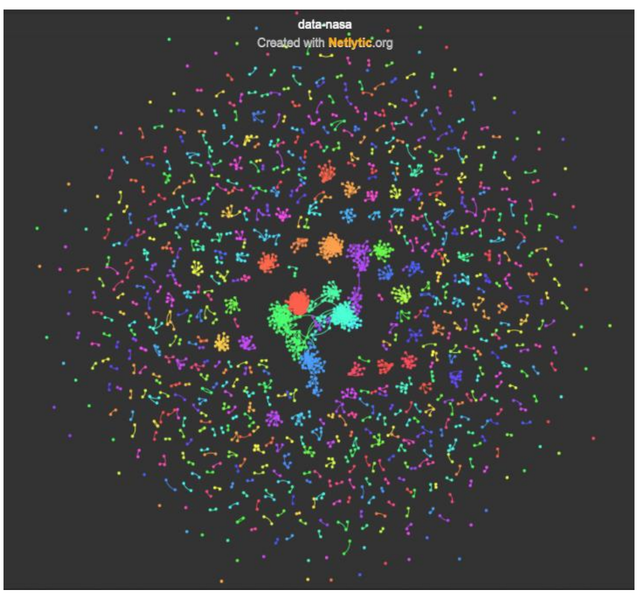
Figure 2

Network Analysis:

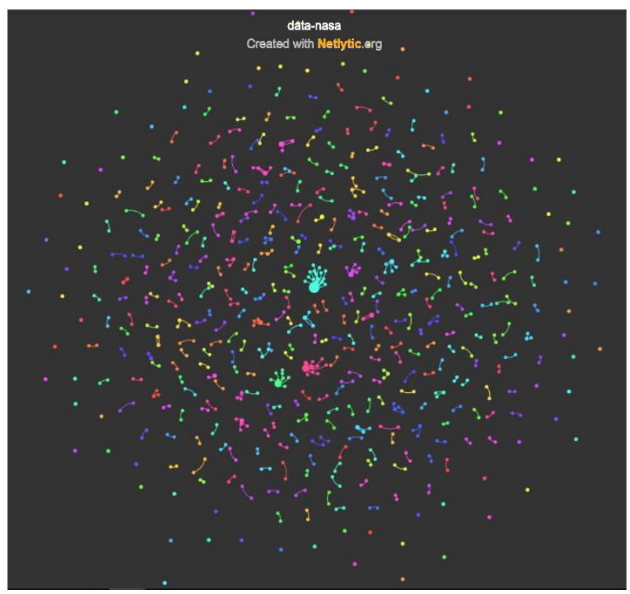
The next step is to explore the social connections underlying the online conversations being examined. Studying organizations from a network perspective allows us to see how knowledge is being co-constructed. In this step, we first discover how online participants are connected to one another (e.g., who is talking to whom), and then apply SNA to analyze the discovered networks. SNA allows us to judge whether the communication networks formed as part of the class are effectively supporting processes known to contribute to successful learning, such as information sharing, community building, and collaboration. To proceed with SNA, we built two types of communication networks: Name and Chain networks. The Name network shows connections between online participants based on direct interactions such as replies or indirect interactions such as mentions or retweets. In other words, two Twitter accounts will be connected in the Name network if one replies to, retweets, or mentions another in his/her message. By including indirect interactions such as mentions in addition to counting replies, we are able to capture instances when one person learns something from another as demonstrated by that person's retweets ("endorsement") or mentions ("acknowledgment"). The Chain network connects participants based on their posting behaviour and usually includes only direct interactions. In the case of Twitter, the Chain network is a subset of the Name network because it only connects people if one replied to another. Following the Twitter convention, this would be equivalent to starting a post with one's username, such as "@gruzd Thank you for sharing this link."

Figure 3 & 4 shows the Name and Chain networks built from the #NASA and #ISRO dataset. The node colours are assigned automatically (based on the "Fast Greedy" community detection algorithm); Each colour represents a group of nodes more likely to be connected to each other than with the rest of the network. In this manner, networks can be grouped into subsets, where each subset is densely connected internally relative to other nodes in the network. Such clustering can be useful in further research as communities correspond to clusters of nodes that may share common properties, interests, or have a similar role within a network. Based on the visual inspection of the networks, it is clear that the Chain network is less dense with fewer nodes. This is somewhat expected since it only represents direct replies between online participants. The Name network is denser and shows a number of overlapping groups of nodes (clusters) that highlight potentially interesting areas of the network to focus on in more detail. The clustering and network fragmentation aspects are discussed later in this section. Figure 4. Name network (on the left) and Chain network (on the right).

NASA:



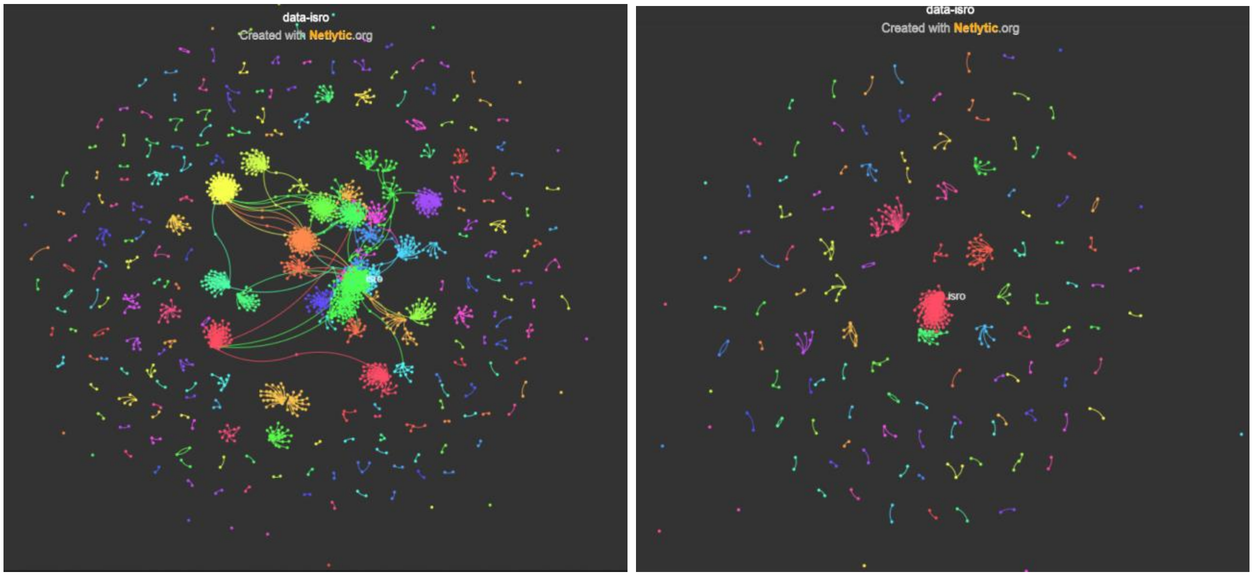
Name Network



Chain Network

Figure 3

ISRO:



Name Network Chain Network

Figure 4

Once the networks are discovered, we can use SNA to make sense of the emerging connections among online participants. With SNA, one can look at both micro-and macro-level measures to examine class interactions: micro-level measures provide insights at the individual node level; and macro-level measures capture the overall state of the network

Macro-level measures found to be useful when analyzing and comparing different social networks include density, reciprocity, centralization, and modularity Table 5 & 6 summarizes the values of these measures for both the Name and Chain networks.

NASA

Network Properties:

Diameter: 9
Density: 0.000314
Reciprocity: 0.023920
Centralization: 0.019680
Modularity: 0.974700

Name Network

Network Properties:

Diameter: 3
Density: 0.000720
Reciprocity: 0.064520
Centralization: 0.008243
Modularity: 0.995900

Chain Network

Table 5

ISRO

Network Properties:	Name Network	Chain Network
Diameter	12	9
Density	0.000478	0.001859
Reciprocity	0.019450	0.094960
Centralization	0.059670	0.105800
Modularity	0.904300	0.907600

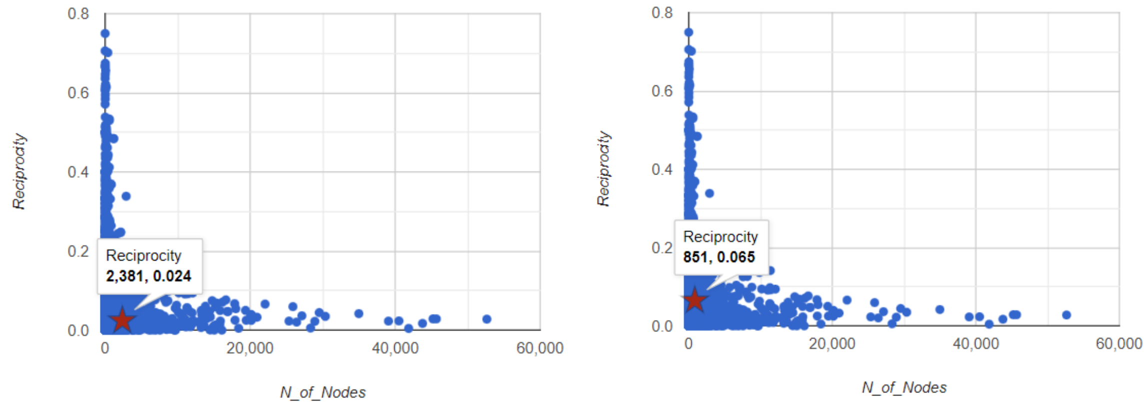
Table 6

Density indicates the overall connectivity in the network (the total number of connections divided by the total number of possible connections); it is equal to 1 when everyone is connected to everyone. In our case, the Chain network is almost three times denser than the Name network, but both networks have less than 1% of the total number of possible connections. Although it is generally useful to see how dense a particular network is, caution is needed when interpreting this measure because with an increasing number of nodes in the social network, the density value often drops because it is much harder to maintain many connections in larger networks. Diameter gives a general idea of how “wide” the network is; in other words, how many nodes information has to travel through between the two farthest nodes in the network. In mathematical terms, diameter is the longest of the shortest paths between any two nodes in the network. Smaller values for the diameter indicate a more highly connected network. The diameter measure is related to density; if density increases, we can expect diameter to reduce since there will be more paths for information to travel, thus potentially reducing the distance between online participants. In our case, the diameter is especially high and equal to 38 in the Name network. This means that it may take up to 38 connections for information to travel from one side of the network to the other. As a class facilitator, one may wish to keep the diameter low to ensure that information spreads efficiently in the network; however, when analyzing communication networks on social media, larger values of diameter may suggest that information originating inside the class also reaches people and communities far outside its core group of participants, which may be a positive sign. Like density, we need to exercise caution in interpreting the benefits of low diameter values, and, indeed, the two-mode nature of ties —strong for sharing, weak for new information. Reciprocity shows how many online participants are having two-way conversations. In a scenario when everyone replies to everyone, the reciprocity value will be 1. However, that almost never happens in social media conversations with hundreds or more online participants.

Finally, modularity provides an estimate of whether a network consists of one coherent group of participants engaged in the same conversation and paying attention to each other (modularity values closer to 0); or whether a network consists of different conversations and communities with a weak overlap (modularity values closer to 1). For more formal collaborative classes, the goal might be to achieve a network structure with a lower modularity value —i.e., everyone on the same topic attending to everyone else —potentially leading to a higher sense of community. At the same time, especially when designing a network to support informal learning, a network with a moderate number of overlapping communities (modularity values around 0.5) may be more desired as it would potentially expose participants to diverse sources of information, exercising the strength of weak ties while still maintaining the sense of community. In the case of #NASA, the Name network consists of both weak and strong ties as suggested by a moderate value of modularity (0.97). However, the modularity value of the Chain network is a bit higher and closer to 1 (0.99), suggesting that there are different groups of people having different conversations in the class. Higher values of modularity may be a sign of underlying homophilic tendencies of people to connect with other like-minded individuals. Based on the discussion above, it is clear that some measures such as centralization and modularity can be interpreted relatively easily; however, other measures, such as diameter or reciprocity, are more difficult to explain without a point of reference. To help with the interpretation, we can compare our values to the values of the same measures calculated for other Twitter networks of a similar size. We will use reciprocity as an example. The Name network's reciprocity level is 0.023, which means that about 2% of the total number of ties is reciprocal (or bi-directional). The Chain network's reciprocity is 0.064 (or about 6% of the total number of ties). It is expected that the Chain network will be more reciprocal since it includes only connections when one person replies to another.

For example, Figure 7 & 8 shows the scatter plot of the number of nodes versus the reciprocity values for about 100 communication networks built from various Twitter datasets. The plots reveal that in both cases, Name and Chain networks, the values for the NASA and ISRO class (marked with the red star), is somewhat higher than in the majority of other networks. This means that the NASA & ISRO class is reaching or exceeding the level of reciprocity that would normally be expected in Twitter data.

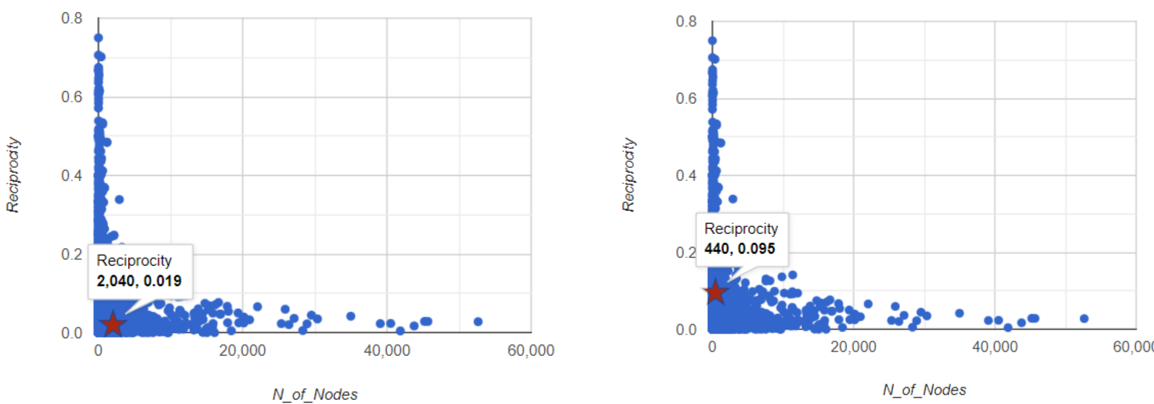
NASA



Name Network Chain Network

Figure 7

ISRO



Name Network Chain Network

Figure 8

To provide an explanation and demonstration of the analytic strategy and framework, this section focuses on several analysis methods we rely upon and how they are used in combination to generate new insights about learning. For this case study, we use a sample of comments on the most viewed videos posted by NASA and ISRO on their respective youtube channels

Text Analysis:

The first step in our case was to build concise summaries of the communal textual discourse present in the dataset by identifying frequently used words (mostly nouns). Figure 9 and 10 shows a word cloud visualization of the top 100 most frequently used words in the NASA and ISRO most viewed video's comments over the data collection period. The search keyword (#NASA and #ISRO) and other common words (also known as "stop-words") such as "of", "will" and "to" were automatically removed prior to building this visualization. The size of a word in the visualization is directly related to the number of times it appears in the dataset relative to the other words found in that same dataset. In Netlytic, this visualization allows users to click on any of the words in the cloud in order to explore the context(s) in which the word appears.

NASA:

By exploring the top 100 words, we can group words into two categories. The first category includes words relevant to the class but not necessarily unexpected, including "SpaceX", "Apollo", "Mars", "Moon," and "SLS". The most frequently mentioned word in this category (and in the whole dataset) is "NASA". The second group of frequently used words includes no words but rather emojis: , ♡. Since the comments were from the most viewed videos of NASA's youtube channel, the video is about the organization's successful attempt to send mankind to the moon. It is evident from the comments since it include words like, Artemis, Apollo, moon, years, mission, launch, lunar. Overall the majority of the words were positive, 252 positive words in 252 233 posts.

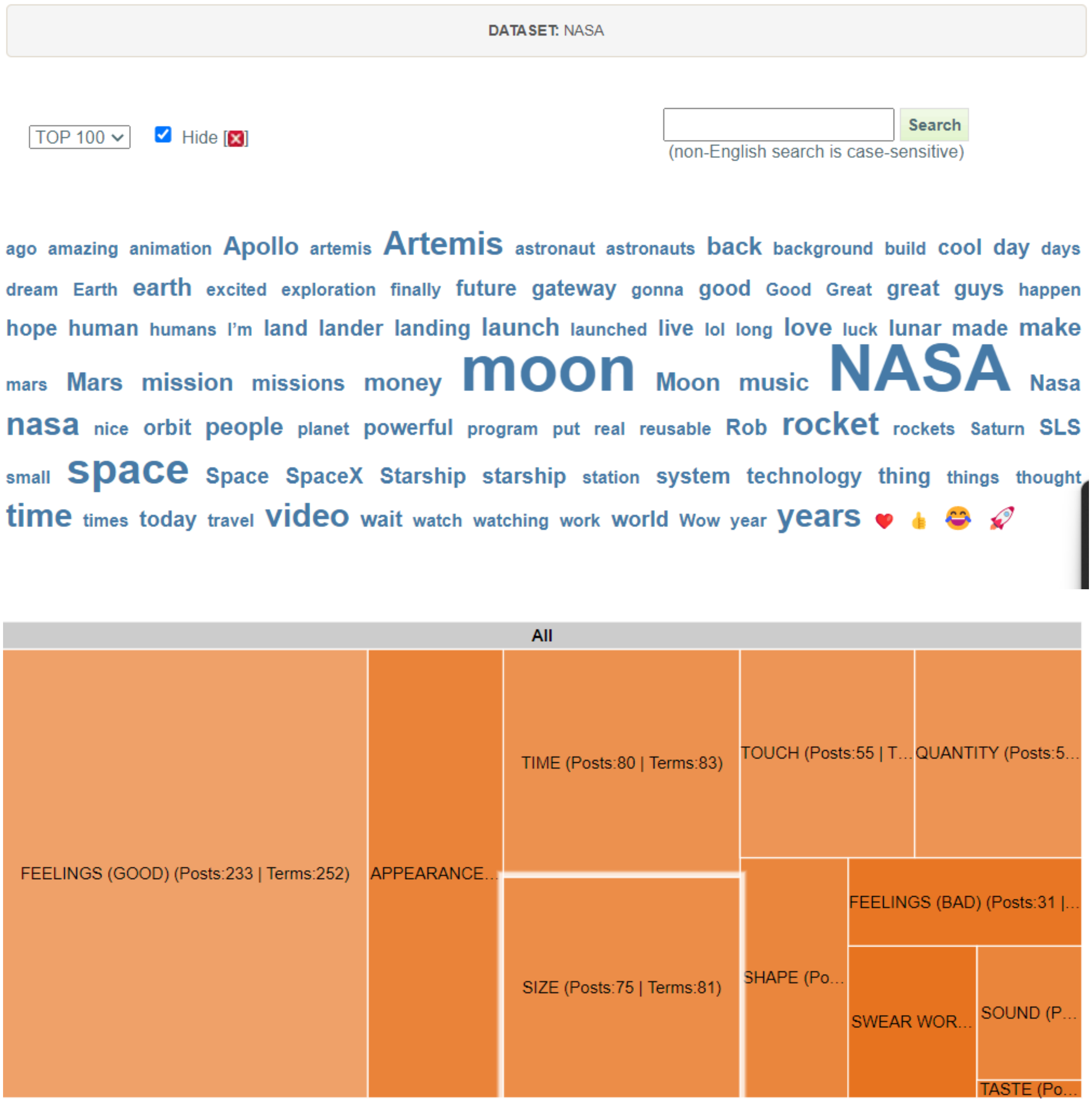
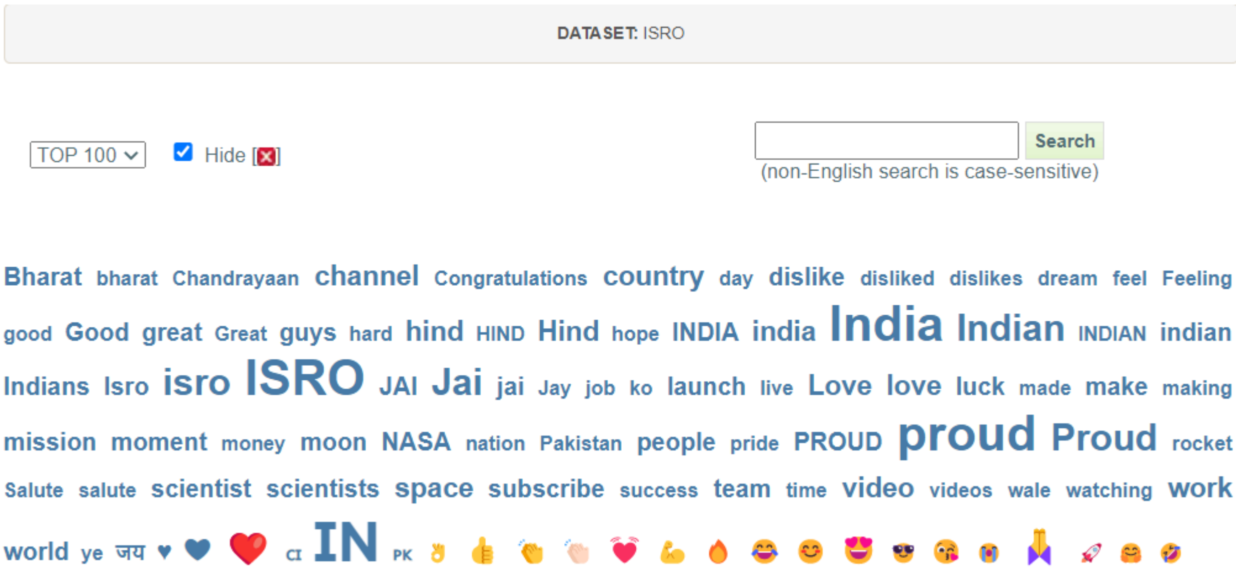


Figure 9

ISRO:

By exploring the top 100 words, we can group words into three categories. The first category includes words relevant to the class but not necessarily unexpected, including “ISRO”, “Indian”, “Proud”, “IN,” and “Indian”. The most frequently mentioned word in this category (and in the whole dataset) is “ISRO” and “India”. The second group of frequently used words includes words typed in Hindi and Hinglish: “Jai”, “Wale”, “ye”, “जय” . The third category includes emojis , ♡, 👍. Overall the majority of the words were positive, 824 positive words in 766 posts.



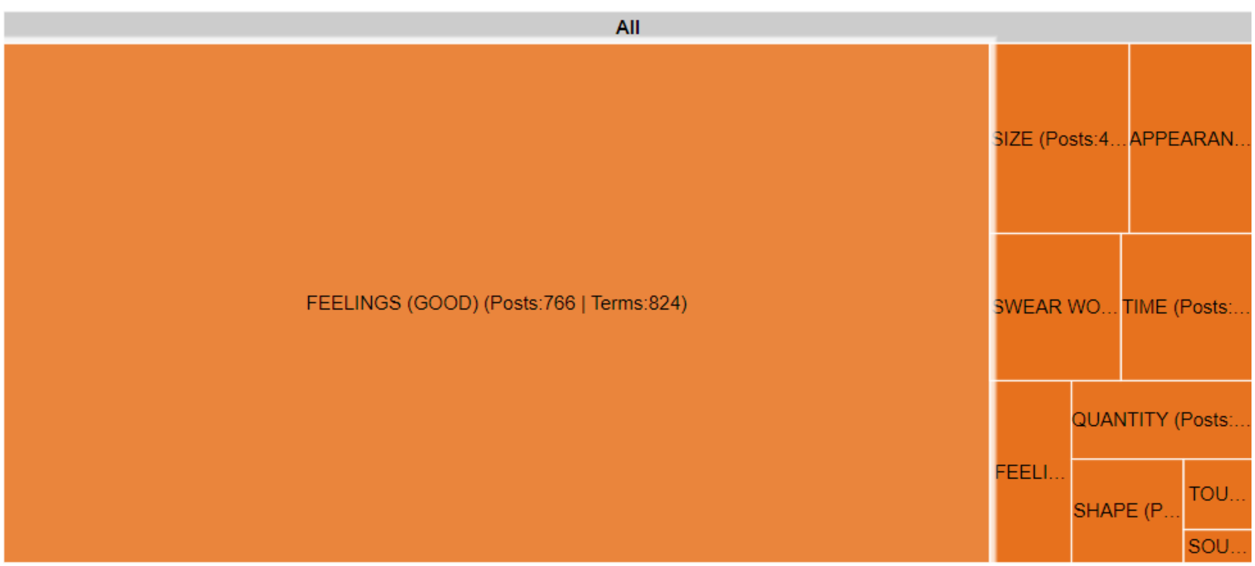


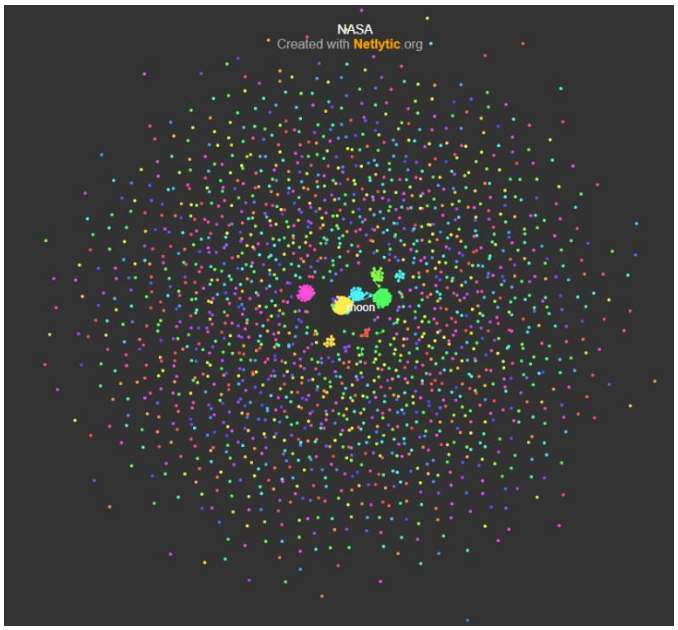
Figure 10

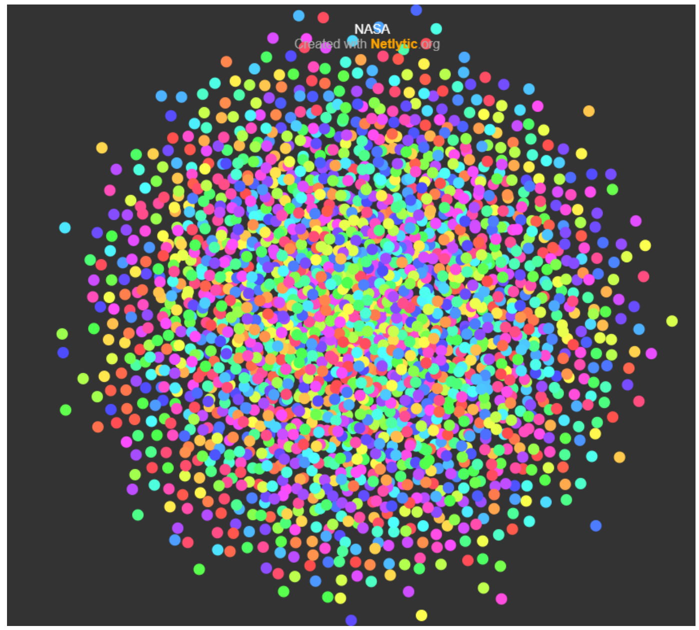
Network Analysis:

The next step is to explore the social connections underlying the online conversations being examined. Studying organizations from a network perspective allows us to see how knowledge is being co-constructed. In this step, we first discover how online participants are connected to one another (e.g., who is talking to whom), and then apply SNA to analyze the discovered networks. SNA allows us to judge whether the communication networks formed as part of the class are effectively supporting processes known to contribute to successful learning, such as information sharing, community building, and collaboration. To proceed with SNA, we built two types of communication networks: Name and Chain networks. The Name network shows connections between online participants based on direct interactions such as replies or indirect interactions such as mentions or replies. In other words, two Youtube or Google accounts will be connected in the Name network if one replies to or mentions another in his/her message. By including indirect interactions such as mentions in addition to counting replies, we are able to capture instances when one person learns something from another as demonstrated by that person's retweets ("endorsement") or mentions ("acknowledgment"). The Chain network connects participants based on their posting behaviour and usually includes only direct interactions. In the case of Youtube, the Chain network is a subset of the Name network because it only connects people if one replied to another. Following the Youtube comment, this would be equivalent to starting a post with one's username, such as "@gruzd Thank you for sharing this video."

Figure 3 & 4 shows the Name and Chain networks built from the #NASA and #ISRO dataset. The node colours are assigned automatically (based on the "Fast Greedy" community detection algorithm); Each colour represents a group of nodes more likely to be connected to each other than with the rest of the network. In this manner, networks can be grouped into subsets, where each subset is densely connected internally relative to other nodes in the network. Such clustering can be useful in further research as communities correspond to clusters of nodes that may share common properties, interests, or have a similar role within a network. Based on the visual inspection of the networks, it is clear that the Chain network is less dense with fewer nodes. This is somewhat expected since it only represents direct replies between online participants. The Name network is denser and shows a number of overlapping groups of nodes (clusters) that highlight potentially interesting areas of the network to focus on in more detail. The clustering and network fragmentation aspects are discussed later in this section. Figure 4. Name network (on the left) and Chain network (on the right).

NASA:

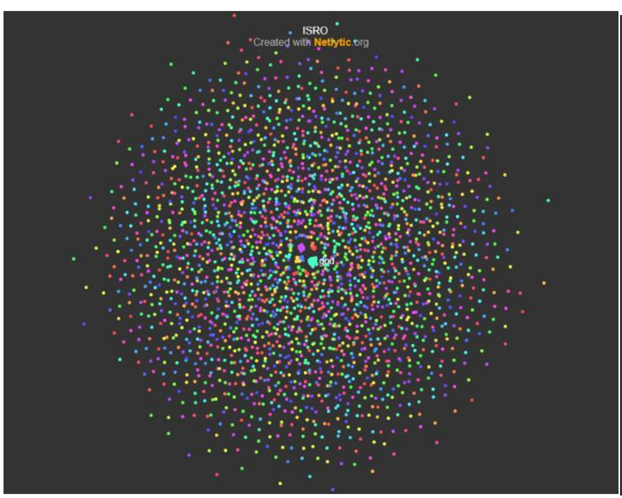


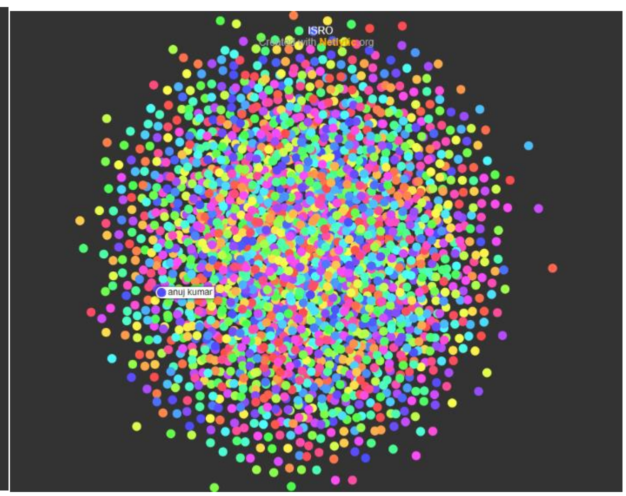


Name Network Chain Network

Figure 11

ISRO**:**





Name Network Chain Network

Figure 12

Once the networks are discovered, we can use SNA to make sense of the emerging connections among online participants. With SNA, one can look at both micro-and macro-level measures to examine class interactions: micro- level measures provide insights at the individual node level; and macro- level measures capture the overall state of the network

Macro-level measures found to be useful when analyzing and comparing different social networks include density, reciprocity, centralization, and modularity Table 13 & 14 summarizes the values of these measures for both the Name and Chain networks.

NASA

Network Properties:	Network Properties:
Diameter: 2	Diameter: 0
Density: 0.000551	Density: 0.000443
Reciprocity: 0.000000	Reciprocity: NaN
Centralization: 0.068690	Centralization: 0.000000
Modularity: 0.941100	Modularity: 0.999600
Name Network	Chain Network

Table 13

ISRO

Network Properties:	Network Properties:
Diameter: 1	Diameter: 0
Density: 0.000424	Density: 0.000418
Reciprocity: 0.000000	Reciprocity: NaN
Centralization: 0.003131	Centralization: 0.000000
Modularity: 0.999400	Modularity: 0.999600
Name Network	Chain Network

Table 14

Density indicates the overall connectivity in the network (the total number of connections divided by the total number of possible connections); it is equal to 1 when everyone is connected to everyone. In our case, the Chain network is almost as dense as the Name network, but both networks have less than 1% of the total number of possible connections. Although it is generally useful to see how dense a particular network is, caution is needed when interpreting this measure because with an increasing number of nodes in the social network, the density value often drops because it is much harder to maintain many connections in larger networks. Diameter gives a general idea of how “wide” the network is; in other words, how many nodes information has to travel through between the two farthest nodes in the network. In mathematical terms, diameter is the longest of the shortest paths between any two nodes in the network. Smaller values for the diameter indicate a more highly connected network. The diameter measure is related to density; if density increases, we can expect diameter to reduce since there will be more paths for information to travel, thus potentially reducing the distance between online participants. In our case, the diameter is especially equal to 1 in the Name network. This means that it may take up to 1 connection for information to travel from one side of the network to the other. As a class facilitator, one may wish to keep the diameter low to ensure that information spreads efficiently in the network; however, when analyzing communication networks on social media, larger values of diameter may suggest that information originating inside the class also reaches people and communities far outside its core group of participants, which may be a positive sign. Like density, we need to exercise caution in interpreting the benefits of low diameter values, and, indeed, the two-mode nature of ties — strong for sharing, weak for new information. Reciprocity shows how many online participants are having two-way conversations. In a scenario when everyone replies to everyone, the reciprocity value will be 1. However,

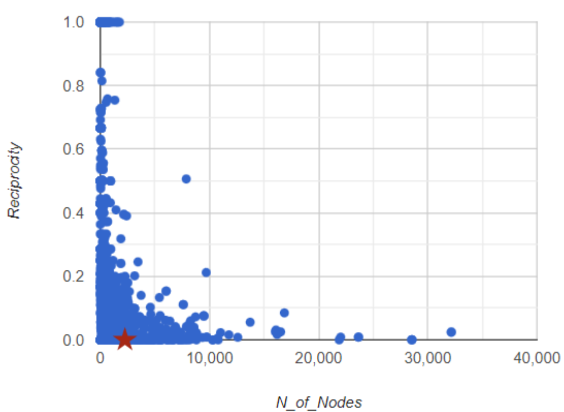
that almost never happens in social media conversations with hundreds or more online participants.

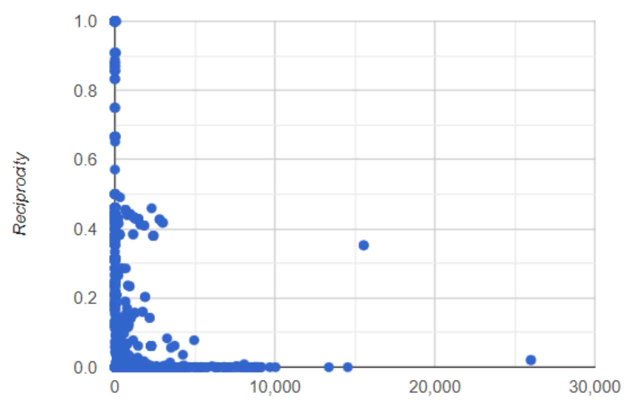
Finally, modularity provides an estimate of whether a network consists of one coherent group of participants engaged in the same conversation and paying attention to each other (modularity values closer to 0); or whether a network consists of different conversations and communities with a weak overlap (modularity values closer to 1). For more formal collaborative classes, the goal might be to achieve a network structure with a lower modularity value —i.e., everyone on the same topic attending to everyone else
—potentially leading to a higher sense of community. At the same time, especially when designing a network to support informal learning, a network with a moderate number of overlapping communities (modularity values around 0.5) may be more desired as it would potentially expose participants to diverse sources of information, exercising the strength of weak ties while still maintaining the sense of community. In the case of #NASA, the Name network consists of both weak and strong ties as suggested by a moderate value of modularity (0.97). However, the modularity value of the Chain

network is a bit higher and closer to 1 (0.99), suggesting that there are different groups of people having different conversations in the class. Higher values of modularity may be a sign of underlying homophilic tendencies of people to connect with other like-minded individuals. Based on the discussion above, it is clear that some measures such as centralization and modularity can be interpreted relatively easily; however, other measures, such as diameter or reciprocity, are more difficult to explain without a point of reference. To help with the interpretation, we can compare our values to the values of the same measures calculated for other Youtube networks of a similar size. We will use reciprocity as an example. The Name network's reciprocity level is 0.023, which means that about 2% of the total number of ties is reciprocal (or bi-directional). The Chain network's reciprocity is 0.064 (or about 6% of the total number of ties). It is expected that the Chain network will be more reciprocal since it includes only connections when one person replies to another.

For example, Figure 15 & 16 shows the scatter plot of the number of nodes versus the reciprocity values for about 100 communication networks built from various Youtube comments datasets. The plots reveal that in both cases, Name and Chain networks, the values for the NASA and ISRO class (marked with the red star), is somewhat higher than in the majority of other networks. This means that the NASA & ISRO class is reaching or exceeding the level of reciprocity that would normally be expected in Youtube comments data.

NASA



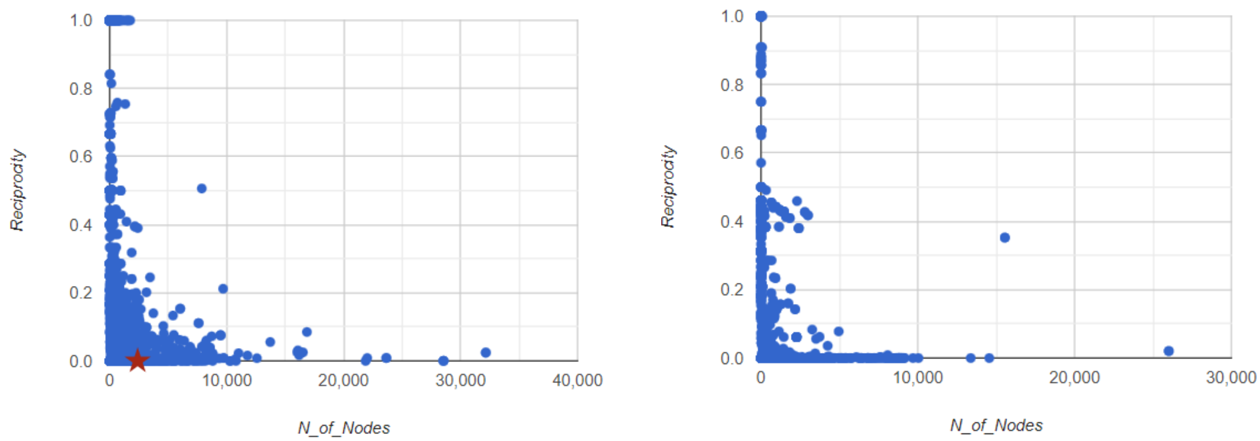


Name Network

Chain Network

Figure 15

ISRO



Name Network Chain Network

Figure 16