USING TECHNOLOGY TO ENGAGE UNDERGRADUATE STUDENTS IN NETWORK SCIENCE RESEARCH

Garrett Dancik & Megan Heenehan
Eastern Connecticut State University Computer Science & Mathematical Sciences Departments
83 Windham Street
Willimantic, CT 06226
dancikg@easternct.edu & heenehanm@easternct.edu

1. Introduction

In this paper, we will describe how we used technology to allow undergraduate students to complete a network science research project over the course of a five-day Summer Research Institute (SRI) at Eastern Connecticut State University (ECSU) in Willimantic, Connecticut. In Section 2, we will describe the motivation and inspiration for the SRI and goals for the institute. Section 3 will describe the schedule for the SRI (3.1) and the background in network science (3.2), programming (3.3), and Gephi (3.4) provided to the students. Section 3.5 describes the research project students complete and a discussion of research questions. We also describe introducing presentation skills and employability skills to students in Sections 3.6 and 3.7, respectively. Samples of student work are provided in Section 4 and pre- and post- SRI student survey results are provided in Section 5. In Section 6, we discuss challenges of and future goals for the SRI. In Section 7, we end with some final thoughts and additional resources.

2. Research Institute Background

At ECSU there is a push to create more undergraduate research opportunities. Recently there has been a focus on undergraduate research opportunities for first- and second-year students in order to increase retention, create strong ties between students and faculty early on in students' academic careers, and to peak interest in a field of study. To create these experiences ECSU asked for applications to run week-long Summer Research Institutes (SRIs) during 2018 and 2019. In each of these institutes ten students, one peer mentor, and faculty mentors would spend five days completing undergraduate research projects. The question for the authors became, what research projects in math and computer science would be appropriate for students so early on in their academic careers?

Inspired by Beveridge and Shan's 2016 article "Network of Thrones," [1] we decided to create an SRI in which students would analyze a movie using network science. Network science is an interdisciplinary and growing field that utilizes graph theory and data mining/data science (subfields of computer science and mathematics/statistics) in order to characterize, visualize, and analyze complex networks. A network can be defined as a collection of interconnected objects. Networks are central to many complex systems across a variety of fields. Examples of networks include biological networks that are composed of interacting genes or proteins and computer networks that are composed of interconnected computers which share data with one another. Additional types of networks include social networks, financial networks, and communication networks. Applied to television and film, network science has been used to evaluate films and the relationships between them. In general, network science can be used to identify important network components, to characterize a network's structure, and to measure network robustness (the ability of a network to withstand a failure in one of its components). Because complex networks are found across a wide range of disciplines, network science has a broad impact and appeals to many students.

The goal of the SRI was to enable a student to perform a network analysis of character interactions in a movie, in order to evaluate a hypothesis about the movie's social structure. Students were taught foundational concepts in network science, including basic programming (scripting) for data collection, statistical concepts for summarizing and analyzing data, and graph theory for measuring important network characteristics. Common tools for network visualization and analysis were described, and students completed hands-on training involving all aspects of the network analysis process (data collection, network construction, visualization, and analysis). Students used their newly acquired skills to programmatically extract character interactions from a movie script, create a character interaction network, and identify relevant network characteristics such as "centrality" and other network science measures appropriate to their research hypotheses.

3. Running the SRI

3.1 Schedule

During the summers of 2018 and 2019 ten rising sophomores and juniors participated in a five-day Network Science SRI which met from 9 a.m. – 5 p.m. with one hour for lunch. Each day had a topic and activity in the morning and another in the afternoon. **Table 1** gives a brief outline of the schedule and objectives for the week. Each summer we also had a peer mentor who had experience with both programming and network science.

3.2 Network Science Background Our Application & Network Measures

Before we discuss the technology used and students' work, we give a brief background on networks and network science. Informally, a network is a set of points and a relationship between pairs of points. The points are called nodes or vertices and the relationship between points is represented by a link or edge. Networks are pervasive in society. For example, the World Wide Web is a network in which nodes are webpages and links are hyperlinks, collaboration on research articles forms a network in which nodes are authors and links represent co-authorship, protein interactions can be represented as networks with proteins as nodes and links representing binding interactions. We can represent a movie as a network in which nodes are characters and

Table 1. Outline of Summer Research Institute Schedule. Morning and afternoon activities are listed including the objectives for each day.

<table>
<tr><th></th><th>Morning</th><th>Afternoon</th></tr>
<tr><td rowspan="2">Day 1</td><td>Introduction to Network Science
Watch movie and manually collect data</td><td>Overview of measures
Data entry and analysis of movie data
Introduction to Gephi</td></tr>
<tr><td>Objectives
• Gain an understanding of Network Science and the common measures used to analyze networks.
• Collect Data</td><td>Objectives
• Gain an understanding of Network Science and the common measures used to analyze networks.
• Meaningfully display data as a network.</td></tr>
<tr><td rowspan="2">Day 2</td><td>Details on measures
Students choose a movie to analyze
Make hypotheses about the movie</td><td>Introduction to Programming concepts and writing scripts in R</td></tr>
<tr><td>Objectives
• Gain an understanding of Network Science and the common measures used to analyze networks.
• Make hypotheses.</td><td>Objectives
• Develop basic programming skills.</td></tr>
<tr><td rowspan="2">Day 3</td><td>Write scripts for data retrieval
Analysis of data – word counts, word cloud, sentiment analyses</td><td>Visualization using Gephi
Continued work on data retrieval and analysis – centrality measures</td></tr>
<tr><td>Objectives
• Develop basic programming skills in R.
• Collect and analyze data.</td><td>Objectives
• Develop basic programming skills in R.
• Collect and analyze data.
• Meaningfully display data as a network.
• Evaluate hypotheses.</td></tr>
<tr><td rowspan="2">Day 4</td><td colspan="2">Continued analysis and presentation preparation</td></tr>
<tr><td colspan="2">Objectives
• Collect and analyze data.
• Evaluate hypotheses.
• Meaningfully display data as a network.
• Effectively communicate results of research.</td></tr>
<tr><td rowspan="2">Day 5</td><td>Presentation Preparation</td><td>Student presentations</td></tr>
<tr><td>Objectives
• Meaningfully display data as a network.
• Effectively communicate results of research.</td><td>Objectives
• Effectively communicate results of research.</td></tr>
</table>

links are interactions between characters. Students must determine what is meant by an "interaction" of two characters. For simplicity sake in the remainder of this paper we say two characters interact if they appear in the same scene. We can then analyze the network to determine which characters are important to the network structure and thus important in the movie. To do this we use network centrality measures.

Centrality measures help to determine which node(s) are important based on the network. As students are introduced to the different measures they perform calculations on small networks, discuss what it means for a node to have high or low centrality for that measure, discuss what it means for a character in a movie to have high or low centrality, and create networks highlighting the measure in Gephi using data we provide from the movie *Princess Bride*. Next we briefly describe some of the measures introduced to the students.

The simplest measure introduced to the students is degree centrality. A node's degree centrality is the number of links connected to the node. In a movie network the degree centrality of a character indicates the number of different characters in which the character represented by the node appears in a scene. In **Figure 1** we see a network created using scene data from the movie *Princess Bride*. In this network large dark green nodes have high degree centrality while small dark pink nodes have low degree centrality.

A related network measure is weighted degree centrality. Weighted degree centrality is the sum of the weights of the links incident to a node. In our movie interaction network the weight of a link is the number of times two characters appear in a scene together. Therefore, the weighted degree centrality of node is the total number of scenes in which the character appears. In **Figure 1** thicker edges between nodes indicate characters were in more scenes together.

The next few centrality measures involve more complex mathematics to which the students in the SRI are introduced, but which is beyond the scope of this paper. We instead provide brief descriptions of the remaining centrality measures to which students are introduced. Eigenvector centrality gives nodes a boost for being connected to other nodes with high weighted degree. In [1] they describe eigenvector centrality as "weighted degree with a feedback loop."

- In PageRank centrality every node starts with an inherent importance for just appearing in the network and then gets a boost for being connected to other important nodes. A nodes importance is divided evenly among its neighbors.
- Closeness centrality measures the average distance to all other nodes in the network.
- Betweenness centrality measures how many shortest paths connecting other nodes in the network go through the given node.

For additional information about these measures and questions to ask of students see [2-3].

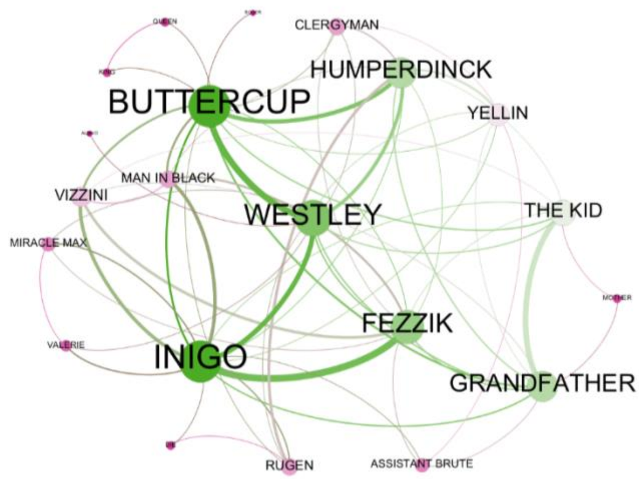


Figure 1. Interaction network for the *Princess Bride* Nodes are colored and sized by degree centrality. Large dark green nodes have high degree centrality, while small dark pink nodes have low degree centrality.

There are several other centrality measures and network measures that we do not introduce to all students. However, if their research question cannot be answered using the measures described we encourage and help them to find and understand other network measures. The other tool we introduce to all students is community detection.

In a network, a community is a group of nodes that is more likely to be connected (by links) to each other than to nodes outside the community. Students are briefly introduced to two methods for detecting communities (1) the Girvan-Newman Algorithm and (2) modularity. The mathematics of the Girvan-Newman algorithm is described during the SRI. However, the methods for modularity are beyond the scope of the students in the SRI so the basics are described and then Gephi is used to perform the calculations. For an activity on community detection see [3]. In **Figure 2** the *Princess Bride* network's nodes are colored by their community. The figure on the left in **Figure 2** highlights modularity-based communities and the figure on the right in **Figure 2** highlights Girvan-Newman based communities. Both were calculated using Gephi which we will discuss later in this paper.

3.3 R programming for data collection & analysis

Students are introduced to the *R* programming language for data collection and analysis. Specifically, students are taught programming for purposes of *natural language processing*, so that they can extract characters from a movie script in order to generate a character interaction edge list; and to extract dialogue to carry out textual analysis including word counts and sentiment analyses (whether text is positive, neutral, or negative).

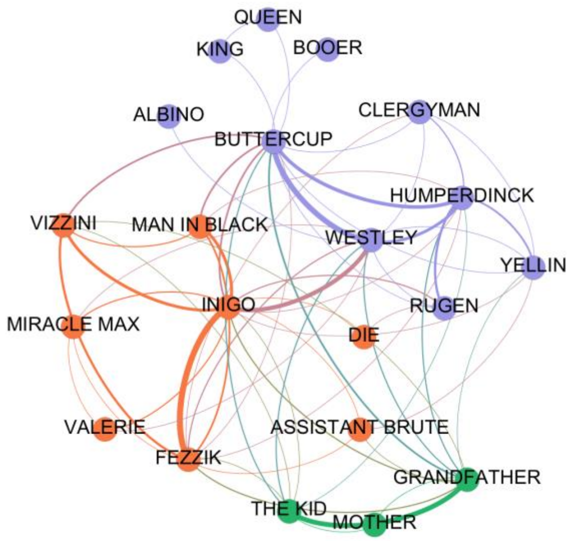


Figure 2. The *Princess Bride* interaction network colored by community. The network on the left is colored by modularity-based communities. The figure on the right is colored by Girvan-Newman based communities. The communities were calculated using Gephi.

To develop *R* code, students are taught to use *R* Studio (https://rstudio.com/), an integrated development environment. Example *R* scripts are used to demonstrate the following concepts:

- Data structures (vectors, lists, matrices, and data frames)
- Expressions
- Plotting using *ggplot* (scatter plots, bar graphs, and boxplots)
- Branching (*if...else* statements)
- Functions, loops, and functional programming (*apply, lapply, sapply*)
- String manipulation (*stringr* package) and regular expressions
- Text analysis including word frequencies, word counts (using the *wordcloud2* package), and sentiment analysis (using the *sentimentr* package)

One of the key concepts covered is that of regular expressions. A *regular expression* is a sequence of characters that define a search pattern, and for our purposes is used to identify characters, scene changes, and dialogue in movie scripts. We cover fundamental regular expression concepts including literals, character classes, anchors, and qualifiers. For example, the string “a” is a literal which matches every occurrence of the letter “a”; the character class “[[:upper:]]” will match any uppercase letter, the anchor “$” matches the end of a line; and the qualifier “+” matches 1 or more occurrences of the character or character class preceding it. More details about regular expressions can be found in reference [2].

A specific example of a regular expression is given in **Figure 3**, where the regular expression "\s{10,}[[:upper:]]+$" is used to match at least 10 white space characters (spaces, newlines, or tabs; "\s{10,}"), followed by 1 or more upper case letters ("[[:upper:]]+"), followed by the end of the line ("$"). In other words, the regular expression identifies lines ending with one or more uppercase words, and no other text. Such a regular expression can be used to identify characters with speaking roles from the script of *The Princess Bride*. This is illustrated in **Figure 3 (left)**, which also demonstrates how the website RegEx 101 (https://regex101.com) can be used for testing regular expressions. We recommend that students use RegEx 101 to construct and test regular expressions. The website also provides an "explanation" for the regular expression which facilitates learning (**Figure 3, right**).

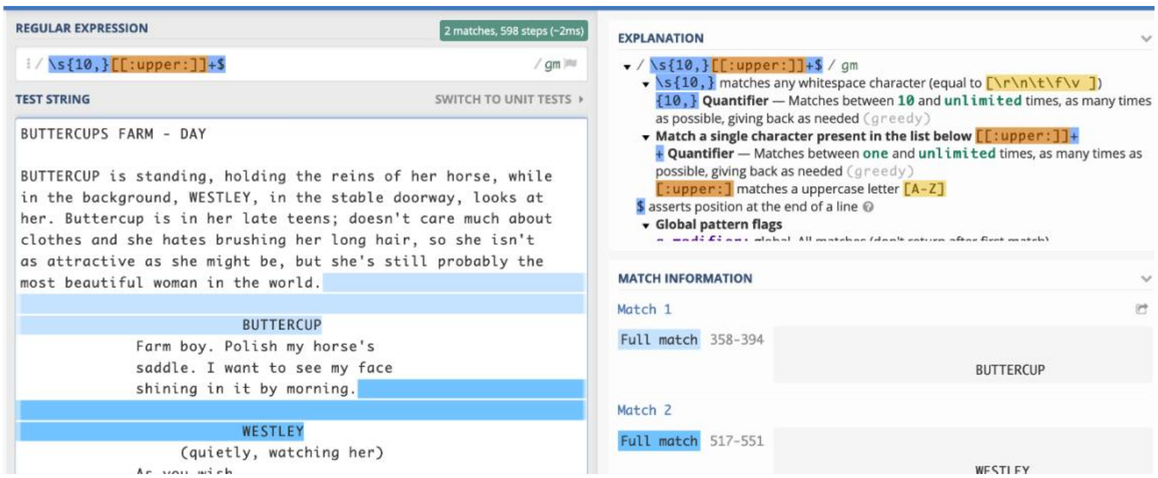


Figure 3. Screen shot of regular expression testing using RegEx 101. (https://regex101.com). This regular expression identifies characters (in shades of blue) with speaking roles from the movie script for *The Princess Bride* (downloaded from the Internet Movie Script Database, https://www.imsdb.com/). The top right of the web page includes an explanation for the regular expression.

Once the appropriate regular expressions are determined, students use regular expressions to extract characters and create edge lists, such as the one in **Table 2**.

Table 2. Example character interaction edge list extracted using regular expressions. *Source* and *Target* characters appear in the same scene together. Because our network will be undirected, the choice of Source and Target is arbitrary.

Source	Target
Buttercup	Westley
Fezzik	Vizzini
Fezzik	Man in Black
Vizzini	Man in Black
King	Queen
Rugen	Humperdinck

3.4 Network Analysis & Visualization Using Gephi

Gephi (https://gephi.org/) is an open source software for analyzing and visualizing networks. Data for a network can be imported into Gephi as either an edge list or an adjacency matrix. As we saw in Section 3.3, students write a script to extract character interactions and form an edge list. Once the edge list is imported into Gephi the software creates the network and produces a random layout as seen in **Figure 4**. As can be seen in **Figure 4**, the center panel in the Gephi window is the "Graph" panel. In this panel labels for nodes and/or links can be turned on and off and resized using the buttons on the bottom. The buttons on the side of the "Graph" panel can be used to select nodes, manually drag nodes to new locations, color and size individual nodes, add new nodes and links, highlight the shortest path between nodes or neighborhood of a node.

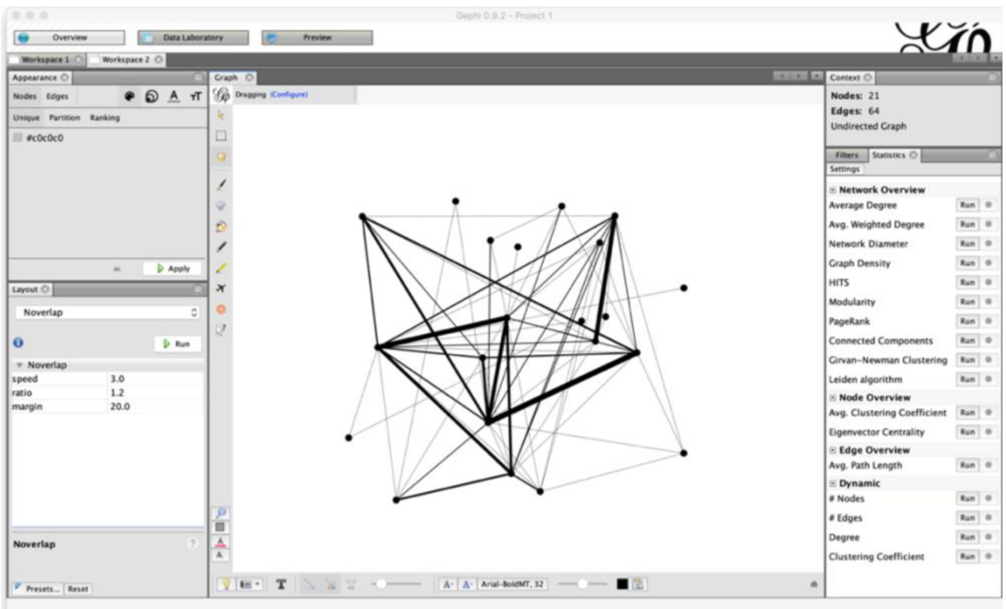


Figure 4. Screenshot of Gephi window. The network shown was generated after importing an edge list from the *Princess Bride*.

On the right of the Gephi window in **Figure 4** we see the "Statistics" panel. From this panel we can run reports for the various centrality measures discussed in Section 3.2. Betweenness centrality and closeness centrality as well as several other measures are found by running "Avg. Path Length." When we run the program for a measure a report is produced giving the distribution of the measure. For example, when we run "Average Degree" a chart is produced showing the degree distribution for the network, that is the number of nodes of each degree. The report also displays the average degree of the network. **Figure 5**, shows the degree distribution for the *Princess Bride* interaction network. When a measure is run a column is added to the "Data Laboratory" in Gephi showing the value of the measure for each node and/or link. See **Figure 6** for the "Data Laboratory" for the *Princess Bride* interaction network.

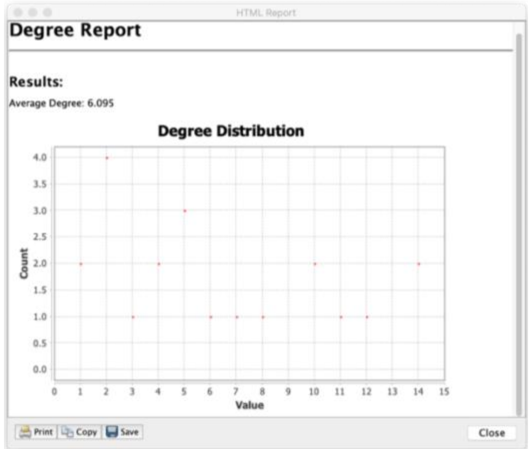
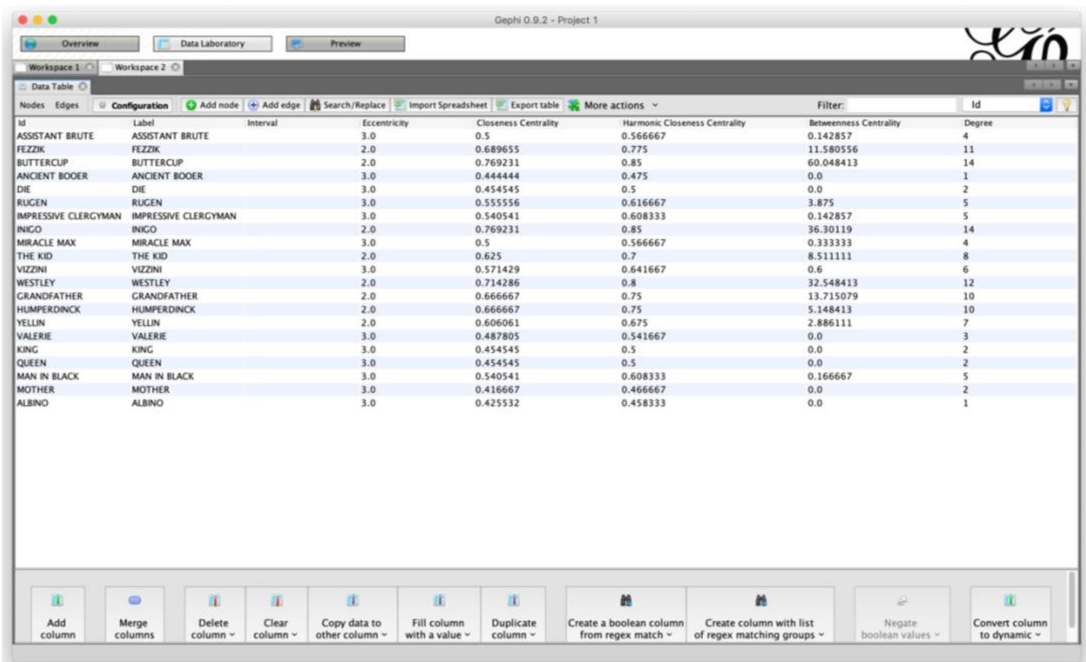


Figure 5. Degree distribution for the *Princess Bride* interaction network. This report is generated automatically by Gephi when "Average Degree" is run from the statistics panel.



Id	Label	Interval	Eccentricity	Closeness Centrality	Harmonic Closeness Centrality	Betweenness Centrality	Degree
ASSISTANT BRUTE	ASSISTANT BRUTE		3.0	0.5	0.566667	0.142857	4
FEZZIK	FEZZIK		2.0	0.689655	0.775	11.580556	11
BUTTERCUP	BUTTERCUP		2.0	0.769231	0.85	60.048413	14
ANCIENT BOOER	ANCIENT BOOER		3.0	0.444444	0.475	0.0	1
DIE	DIE		3.0	0.454545	0.5	0.0	2
RUGEN	RUGEN		3.0	0.555556	0.616667	3.875	5
IMPRESSIVE CLERGYMAN	IMPRESSIVE CLERGYMAN		3.0	0.540541	0.608333	0.142857	5
INIGO	INIGO		2.0	0.769231	0.85	36.30119	14
MIRACLE MAX	MIRACLE MAX		3.0	0.5	0.566667	0.333333	4
THE KID	THE KID		2.0	0.625	0.7	8.511111	8
VIZZINI	VIZZINI		3.0	0.571429	0.641667	0.6	6
WESTLEY	WESTLEY		2.0	0.714286	0.8	32.548413	12
GRANDFATHER	GRANDFATHER		2.0	0.666667	0.75	13.715079	10
HUMPERDINCK	HUMPERDINCK		2.0	0.666667	0.75	5.148413	10
YELLIN	YELLIN		2.0	0.606061	0.675	2.886111	7
VALERIE	VALERIE		3.0	0.487805	0.541667	0.0	3
KING	KING		3.0	0.454545	0.5	0.0	2
QUEEN	QUEEN		3.0	0.454545	0.5	0.0	2
MAN IN BLACK	MAN IN BLACK		3.0	0.540541	0.608333	0.166667	5
MOTHER	MOTHER		3.0	0.416667	0.466667	0.0	2
ALBINO	ALBINO		3.0	0.425532	0.458333	0.0	1

Figure 6. Data Laboratory for the *Princess Bride* interaction network.

In addition to centrality measures, community detection algorithms can be run from the "Statistics" panel. Modularity is built into Gephi, but the "Girvan-Newman Clustering" must be added as a plugin by using the "Tools" menu. When a community detection algorithm is run a column is added to the "Data Laboratory." Each node is assigned a number, nodes that are in the same community are assigned the same number.

Once measures have been run for the network, the "Appearance" panel can be used to size and color nodes based on one or more measures and/or nodes can be partitioned based on a measure or their community.

The “Layout” panel can be used to change the layout of the nodes. Different layouts are not explicitly covered in the SRI, but students are told to research the pros and cons of the layouts and choose appropriately. **Figure 7** shows the Gephi window with the *Princess Bride* interaction network where nodes have been sized and colored by betweenness centrality, node labels have been turned on and the layout has been switched to “OpenOrd” and then expanded using the layout “Expansion.”

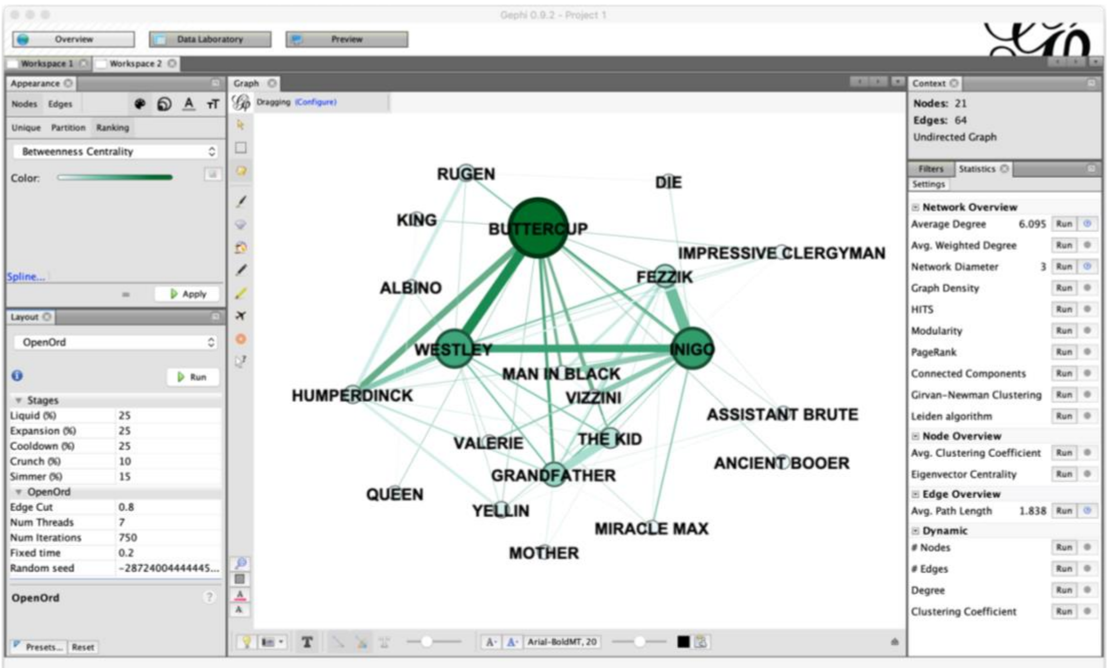


Figure 7. The *Princess Bride* interaction network with nodes sized and colored by betweenness centrality. Node labels have been turned on and the layout was determined using “OpenOrd” followed by “Expansion.”

Gephi has many features that are beyond the scope of this paper. The website https://gephi.org/users/ has many resources for learning how to use Gephi and its many features.

3.5 Student Research Projects & Research Questions

The goal of the SRI is for students to carry out a research project involving a network analysis of character interactions in a movie, in order to evaluate a hypothesis about the movie’s social structure. In order to complete their research project, students must (1) choose a movie, (2) generate research questions, (3) decide what data they need to answer the questions, (4) decide how to use network science to analyze the data, (5) analyze the data, (6) create network representations, graphs, and tables as part of their analysis, and (7) give a 10-12 minute presentation of their project on the last day of the SRI.

On the third day of the SRI, students give a brief presentation that introduces the movie to be analyzed and their proposed research questions. Students typically want to identify the most “important” character in the movie, which leads to a discussion of what makes a character important, the network science measures most appropriate for measuring importance (based on how students define “importance” in their movie), and the data that would be needed in order to create the appropriate network. In addition, students have been interested in identifying communities of “good” and “bad” characters, or of characters from different physical communities; as well as determining if important characters or communities change over time. Some ambitious groups have also been interested in comparing characters across different movies, such as multiple movies in a series.

3.6 Presenting Research

As part of our SRI, we include a module covering basic research presentation skills. We emphasize the fact that students began the research project with one or more hypotheses or objectives, and designed data collection and analysis to test their hypotheses or meet their objectives. **Table 3**, which is similar to a handout provided to students, summarizes the structure of the presentation and our expectations. Examples of student research projects are shown in **Section 4**.

Table 3. Overview of research presentation.

Section	Description
Introduction (1-2 slides)	• Clear and concise description of the movie(s) • Objective / hypotheses/research questions are stated
Methods (1 – 2 slides)	• Methodology used to carry out the project • Includes software, languages, and packages used • Description of network science measures used (can be combined with results)
Results (2 – 4 slides)	• An unbiased presentation of the results • Methods and results are clearly connected to research questions
Summary / Conclusion (1 – 2 slides)	• Major findings are summarized • Limitations and possible extensions or future work are discussed

In addition to presenting their research at the end of the SRI, we encourage students to present their research to our math and computer science clubs, at regional conferences, at the National Conference on Undergraduate Research, and at an annual campus-wide conference at ECSU. We also encourage students to continue to carry out research with us through Independent Studies.

3.7 Employability skills

At ECSU we are beginning to explicitly connect educational experiences with employability skills so that students can articulate how their classroom experiences have prepared them for the workplace. Anecdotally, we have found that while students develop important skills in class that are desired by employers, they often are unable to connect their experiences with the terms and concepts that employers are looking for. We emphasize that during the SRI, students completed the following:

- Designed, planned, and implemented a project from beginning to end
- Engaged in critical thinking at each stage of the project
- Worked collaboratively, which involved dividing tasks, communicating ideas between team members, and presenting as a group
- Learned time management while working intensively to complete a research project in a short amount of time
- Developed skills, including R programming, network analysis, critical thinking, teamwork, and presentation skills

4. Student Work

Here we present two select examples of student work, in order to demonstrate the types of research questions asked, the analyses carried out, and the quality of the work that students are able to produce over the 5-day long SRI.

One group analyzed the movie *Mean Girls* with the objective of characterizing how communities change throughout pivotal points in the movie, identifying the most negative and positive characters, and identifying the character with the most connections to other characters. Some of their results are presented in **Figure 8**.

Another group analyzed the Matrix Trilogy, looked at the sentiment of the main characters over the course of each movie, and generated character interaction networks for each movie in order to identify important characters (**Figure 9**). The students in this group had a strong programming background, which allowed them to apply their analyses across multiple movies. It should be noted that this analysis was later presented at the National Conference on Undergraduate Research [4].

Character interaction networks, with node size corresponding to weighted degree, node color corresponding to betweenness centrality (the darker the color, the higher the betweenness). Edge color and thickness reflect the number of interactions between characters.

In addition, students have analyzed *The Chronicles of Narnia: The Lion, The Witch, and The Wardrobe*, to determine if the communities captured different sets of characters depending on whether they were inside or outside of Narnia; *Batman & Robin* and *The Dark Knight* in order to compare the corresponding villains ("Mr. Freeze" and "The Joker"); *Avengers: Infinity War* to identify characters who were "information brokers" and to compare the villains and the heroes; and *Mulan* to identify important characters, communities, and to compare the sentiment of the main characters.

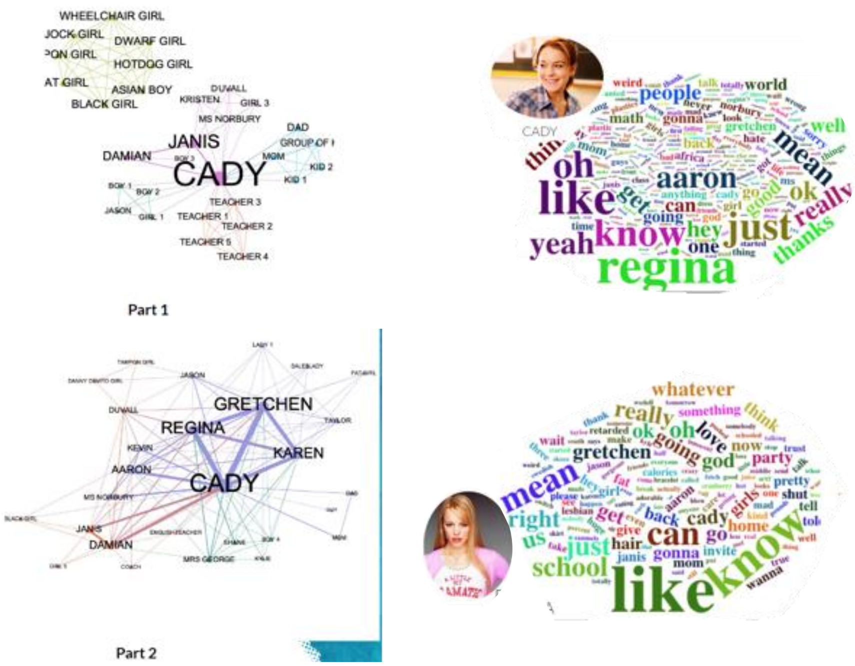


Figure 8. Analysis of *Mean Girls*. The movie was divided into four parts and networks were constructed for each (networks for the first two parts are shown). Word clouds from the dialogue of Cady and Regina are also shown.

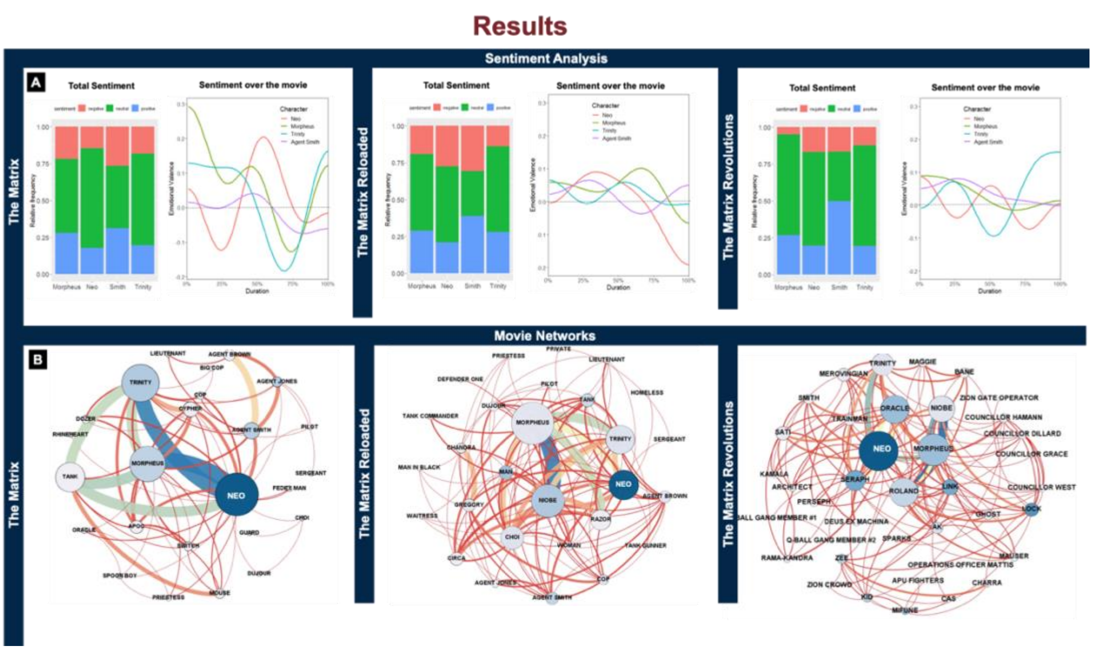


Figure 9. Analysis of The Matrix Trilogy. A) Sentiment analysis of main characters Neo, Trinity, Morpheus, and Agent Smith in The Matrix, The Matrix Reloaded, and The Matrix Revolutions. **B)** Character interaction networks, with node size corresponding to weighted degree, node color corresponding to betweenness centrality (the darker the color, the higher the betweenness). Edge color and thickness reflect the number of interactions between characters.

5. Survey results

At the beginning and end of each SRI we survey our students to collect demographic information and to obtain self-reported feedback on student research-related abilities. In this section we report the results from our 2019 SRI. The demographic breakdown of students is given in **Table 4**, which shows that we had a good balance of females and males, and that the majority of students were entering their junior year (recall that we intentionally target students who will be entering their junior year).

Students are asked to rate themselves on their quantitative skills, critical thinking, and other abilities defined in **Table 5,** using the scale in **Table 6**.

Table 4. Demographic information for SRI 2019

Gender	N
Female	4
Male	6
Class status	N
Freshman	0
Sophomore	1
Junior	8
Senior	1
Honors	1

Table 5. Abilities self-reported by students at the beginning and end of each SRI.

Quantitative Skills – the application of mathematics and logic to analyze and interpret information
Critical Thinking – the ability to think in a self-directed, disciplined, and objective manner
Computer Programming – the ability to write computer code to solve problems
Development of Research Questions – the ability to identify interesting, useful, and clear questions that contribute knowledge or value to a discipline
Oral Communication and Presentation – the ability to clearly articulate ideas and findings to small or large groups

Table 6. Numeric ratings for student self-reported abilities.

1	**2**	**3**	**4**	**5**	**6**
Very Poor	Poor	Fair	Good	Very Good	Excellent

The results of our survey for SRI 2019 are provided in **Figure 10**. In general, self-reported student abilities increased from around Good (4) to Very Good (5). The pre/post survey differences for Quantitative Skills, Critical Thinking, and Computer Programming are statistically significant at $P < 0.05$ (two-sample t-test). All students also rated their overall SRI experience as “excellent”.

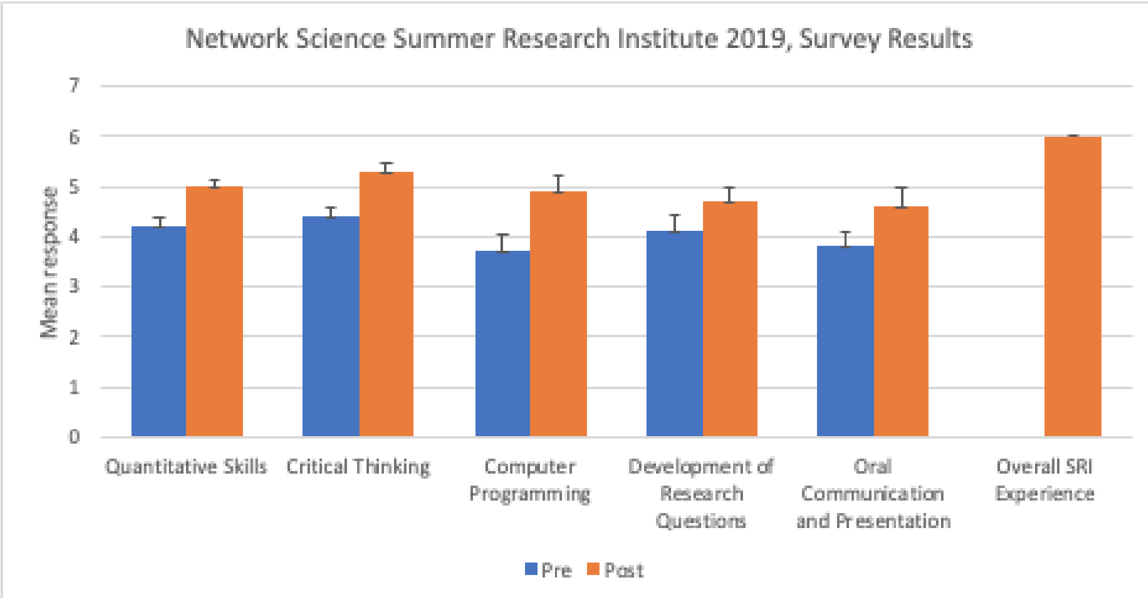


Figure 10. Summary of self-reported abilities for SRI 2019.

We also ask the students about their favorite part of the SRI and for suggestions on how we can improve. From student responses, common themes emerge. Below we list our questions as well as representative responses.

-*What was your favorite part of the SRI?*

- “Learning to code in *R* was a fun and frustrating experience. In the end we learned methods that are extremely powerful in extracting data...”
- “Seeing graphs in Gephi”
- “Meeting and connecting with other Network Science students...”

If you could change something, what would it be?

- "Extend the length of the SRI..."
- "I wouldn't change a single thing..."

6. Challenges & Future goals

Based on student feedback and our own observations, the biggest challenge has been carrying out the SRI in five days. During this time, students must learn the necessary background, and then design and carry out a research project. Network science background and *R* programming are covered in about 2 days, which is an intense experience. Students also have limited time (about 1 day) to work on their presentations once their analyses are complete. In addition, students often struggle to come up with their own research questions and ideas for analyses, which is partly because students must come up with a project idea in a short amount of time.

In the future we hope to secure external funding so that we can run a longer SRI. This would allow students to have more time to practice and complete exercises covering background material prior to their research project. It would also allow students to explore and pursue a wider variety of research projects that apply network science on other datasets, such as on biological networks or Twitter data.

7. Conclusions

In this paper we have described our experiences with using technology to engage undergraduate students in network science research during a 5-day long SRI held at ECSU during the summers of 2018 and 2019. During the SRI, students learn important concepts in network science and gain hands-on experience in network analysis, including data collection, network construction, visualization, and analysis. Students apply these skills in order to perform a network analysis of character interactions in a movie, in order to evaluate a hypothesis about the movie's social structure.

The technology used – *R* programming and Gephi – allows students to collect and analyze data in order to answer novel research questions while reinforcing underlying network science theory. In short, students gain experience in applied network science. We have found that the technology we use is accessible to mathematics and computer science students who will be entering their junior year. Both *R* and Gephi are free tools and therefore are easily used in other classes or in projects outside of school. *R* is also a core statistical programming language that is used in data science, which is a growing field.

Handouts, example code, and other resources used in our SRI can be found at https://gdancik.github.io/NS-SRI. In addition, we have published two articles which contain additional background and student activities for modeling movies using network science [2-3].

References

1. Beveridge, A., & Shan, J. (2016, April). Network of Thrones. *Math Horizons*, 18-22.

2. Dancik, G.M., Davis, M., Heenehan, M (2019). "Modeling Movies Using Network Science, Part I", in Froelich, G. (ed.) *CONSORTIUM*: *The Newsletter of the Consortium for Mathematics and Its Applications*, Number 117, COMAP, Inc (MA), pp. 8-36.

3. Dancik, G.M., Davis, M., Heenehan, M (2020). "Modeling Movies Using Network Science, Part II", in Froelich, G. (ed.) *CONSORTIUM*: *The Newsletter of the Consortium for Mathematics and Its Applications*, Number 118, COMAP, Inc (MA), pp. 11-30.

4. Ramanenko, N., Johnson, C., Marchand, S., Heenehan, M., Dancik, G.M., 2019. Network Analysis of the Matrix Trilogy. National Conference on Undergraduate Research. Kennesaw, GA, USA (poster).