Applied Social Network Analysis in Education
## Contents

### List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Tables</td>
<td>5</td>
</tr>
<tr>
<td>List of Figures</td>
<td>7</td>
</tr>
<tr>
<td>Preface</td>
<td>7</td>
</tr>
<tr>
<td>Why not Canvas for this course?</td>
<td>7</td>
</tr>
<tr>
<td>Two tracks of participation</td>
<td>8</td>
</tr>
<tr>
<td>Ways to contribute to this site</td>
<td>8</td>
</tr>
<tr>
<td>Syllabus</td>
<td>9</td>
</tr>
<tr>
<td>0.1 Introduction</td>
<td>9</td>
</tr>
<tr>
<td>0.1.1 What is Social Network Analysis?</td>
<td>9</td>
</tr>
<tr>
<td>0.1.2 Why Social Network Analysis?</td>
<td>9</td>
</tr>
<tr>
<td>0.1.3 Syllabus</td>
<td>10</td>
</tr>
<tr>
<td>0.1.4 Digital learning environments</td>
<td>10</td>
</tr>
<tr>
<td>0.1.5 Week 1 Activities</td>
<td>12</td>
</tr>
<tr>
<td>0.2 Basic Concepts</td>
<td>13</td>
</tr>
<tr>
<td>0.2.1 Understanding SNA in Educational Research</td>
<td>14</td>
</tr>
<tr>
<td>0.2.2 Basic Concepts</td>
<td>15</td>
</tr>
<tr>
<td>0.2.3 Week 2 Activities</td>
<td>15</td>
</tr>
<tr>
<td>0.3 Applications and Examples I</td>
<td>16</td>
</tr>
<tr>
<td>0.3.1 Applications and Examples of SNA</td>
<td>18</td>
</tr>
<tr>
<td>0.3.2 Week 3 Activities</td>
<td>18</td>
</tr>
<tr>
<td>0.4 Collect and Manage Network Data</td>
<td>20</td>
</tr>
<tr>
<td>0.4.1 Understanding SNA Data</td>
<td>20</td>
</tr>
<tr>
<td>0.4.2 Ethics in SNA research</td>
<td>22</td>
</tr>
<tr>
<td>0.4.3 Linear algebra basics</td>
<td>22</td>
</tr>
<tr>
<td>0.4.4 Managing SNA data</td>
<td>23</td>
</tr>
<tr>
<td>0.4.5 Import data into software</td>
<td>26</td>
</tr>
<tr>
<td>0.4.6 Week 4 Activities</td>
<td>27</td>
</tr>
<tr>
<td>0.5 Sociograms and Density</td>
<td>28</td>
</tr>
<tr>
<td>0.5.1 Sociograms and network visualizations</td>
<td>29</td>
</tr>
<tr>
<td>0.5.2 Density and Other Structural Measures for Complete Networks</td>
<td>33</td>
</tr>
<tr>
<td>0.5.3 Assignments and Activities</td>
<td>36</td>
</tr>
<tr>
<td>0.6 Centrality and Centralization</td>
<td>37</td>
</tr>
<tr>
<td>0.6.1 Centrality and Centralization: An Overview</td>
<td>39</td>
</tr>
<tr>
<td>0.6.2 Betweenness Centrality</td>
<td>39</td>
</tr>
<tr>
<td>0.6.3 Eigenvector Centrality</td>
<td>39</td>
</tr>
<tr>
<td>0.6.4 Week 6 Activities</td>
<td>40</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>--------------------------</td>
<td>------</td>
</tr>
<tr>
<td>0.7 Components and Cliques</td>
<td>42</td>
</tr>
<tr>
<td>0.7.1 Introduction to Components and Cliques</td>
<td>42</td>
</tr>
<tr>
<td>0.7.2 From Data to Analyses: A Demo</td>
<td>42</td>
</tr>
<tr>
<td>0.7.3 Week 7 Activities</td>
<td>42</td>
</tr>
</tbody>
</table>
List of Tables

0.1 Table 1. A table of nodes. ........................................... 23
0.2 Table 2. A table of weighted ties .................................. 23
0.3 Table 3. A table of raw data of ties. ................................. 24
0.4 Table 4. Sports teams in the school. ................................. 24
0.5 Table 5. Student affiliation with sports teams. ..................... 25
0.6 Table 6. Sports teams in the school (version 2). ................. 25
0.7 Table 5.1: Different types of networks. ............................ 34
List of Figures

1 Gephi Graph Overview. .......................................................... 41
2 Gephi Data Laboratory. .......................................................... 41

Preface

This site is the course portal of CI 8371 - Applied Social Network Analysis in Education, taught by Prof. Bodong Chen at the University of Minnesota in Spring ’19.

Content on this site is actively built and refined throughout the semester.

Key links for the course

- Textbook Carolan (2014). Note: UMN library access required.
- Slack team: We use Slack for course “management”, aka. announcements, communications, chats, small group work, etc. You can also get connected with course alumni via public Slack channels.
- Hypothes.is: We interact on top of web materials (including the textbook) during this course. Hypothesis is a cutting-edge, open-source web annotation tool, which is used more and more broadly by educators in the past two years.
- Optional – Twitter hashtag #SNAEd: You can tweet about this course using #SNAEd.

This site or book is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

Last update: 2019-03-04

Why not Canvas for this course?

To learn social network analysis (SNA) – and many other stuff – I believe the best way is to ‘live’ with it. That is, the best way to learn SNA is to seize every opportunity of applying SNA in one’s own learning journey. This philosophy leads to two major reasons I’m piecing together a learning environment by myself.¹

First, as an instructor I wish to foster social interactions as much as possible. Traditional learning management systems, like Moodle and Canvas, do not prioritize this goal.

Second, to practice SNA with our own ‘learning traces’, system log data\(^2\) need to be somehow exposed to users. Traditional systems are less transparent in this area.

If you haven’t used these tools mentioned above, no worries. I will provide detailed guidelines to get you started (see Section 0.1.4). Many of us would be venturing out of our comfort zones. (SO DO I – because I cannot stop trying new things :-)). I hope we will all learn a great deal about SNA and beyond.

Have fun!

Two tracks of participation

This course website is openly available online, reflecting my commitment to ‘open scholarship’. So there are two tracks of participation designed for this class.

1. **UMN class**: For UMN graduate students enrolled in the class, your participation in this class is not that different from taking another online class. You are expected to meet course requirements outlined in the course syllabus. You will have private spaces (on Slack and Hypothes.is) for within-class communication.

2. **Open participation**: For ‘Open Participants’ who are following the class, you can access the same materials I curate on this website. You may need to buy the required textbook if you do not have a library access. And ways for you to participate include 1) making public Hypothes.is annotations using tag **SNAEd**, and 2) tweeting out your ideas or links to your blog posts using hashtag **#SNAEd**. Open Participants accomplishing major course requirements will be offered Digital Badges.

If you have any feedback on the current design, please get in touch. I’d love to hear your thoughts.

Ways to contribute to this site

This is the 2nd time I am offering this course and the 3rd time teaching in such an ‘open’ manner. By making this course site open, I hope it could be useful beyond the formal class. So I’d appreciate your help to continually improve the site.

1. The best way to contribute to this site is to annotate it using Hypothes.is. For example, to point out a possible spelling error, you can highlight the error and include two tags – **SNAEd** and **issues** – in your annotation.
2. The second best way is to report a New Issue on this website’s Github issues page.
3. If you’re Github-savvy, please consider forking the Github repository, making edits, and sending me a pull request.

4. If none of those options work for you, simply drop me an email at chenbd@umn.edu.

Syllabus

The most up-to-date syllabus is here. Please post Hypothes.is annotations on the syllabus if you have any questions.

0.1 Introduction

This week we will:

1. Get to know each other
2. Walk through the Syllabus
3. Get to know course logistics

0.1.1 What is Social Network Analysis?

0.1.1.1 A light intro video made by colleagues from Duke.

Credit: ModU: Powerful Concepts in Social Science

0.1.1.2 A deeper dive: The hidden influence of social networks

0.1.2 Why Social Network Analysis?

0.1.2.1 Perspectives from colleagues

We come together to learn (more) about social network analysis (SNA) – for various reasons. In preparation for the class, I invited three colleagues who’re all doing interesting work with SNA to share a bit about their experiences (note some videos are only visible for UMN viewers):

Alejandro Andrade, PhD from Indiana University, now Data Scientist at ACT

Michael Brown, PhD from the University of Michigan, now Assistant Professor at Iowa State

Fan Ouyang, PhD from the University of Minnesota, now Assistant Professor at Zhejiang University
0.1.2.2 Networks are everywhere! Getting inspired?

Post your initial ideas of using social network analysis below, by clicking on the + button.

0.1.3 Syllabus

The most current syllabus will always be available online.

In Week 1, please read the Syllabus in full. Ask questions, if any, on the #q_and_a channel on Slack or annotate the syllabus using Hypothes.is.

0.1.4 Digital learning environments

This is a **fully online** class. An online course does not need to be about watching videos and answering quizzes. To foster rich learning experiences, we will be using several technological tools to foster social, collaborative learning. Depending on your familiarity with these tools, there could be a learning curve. So please spend time this week to familiarize yourself with them.

Below is a list of them and we could tweak the list when moving forward.

**Bookdown course site.** This website you’re reading is built by myself using **Bookdown**. **Bookdown** is designed for writing books. While this site may evolve into an open textbook, I’m using it as our course portal.

Course resources, usually organized by weeks, will be published here. Each chapter maps onto a week. Weekly course announcements will be pushed to the class through Slack (more below). (If you are an ‘Open Participant’, just check this website every Tuesday!)

**Slack** (UMN Students Only). You will receive a link that invites you to join our Slack community. Course announcements are made on Slack. All sorts of class discussion will take place there as well. First time using Slack? Watch the video below and/or read this brief introduction to Slack.

To get started with Slack, get familiar with a few **Slack terms** you need to know:³

- **Messages** or **Updates** — the basic message or status of Slack. Similar to Facebook updates, tweets, chats, etc.
- **Channels** — like separate rooms within the domain. They can be **public** or **private**.
- **Direct Messages** — like DMs or private messages anywhere else.
- **Posts** — longer than a status update. Similar to a blog post or a Word document. Once it’s shared to a **Channel**, folks can **comment** on it or even **edit** it if the owner allows.
- **Snippets** — Chunks of syntax-highlighted code for when that’s the thing you need to share. We might use it.

³See [http://kateto.net/network-visualization](http://kateto.net/network-visualization)
0.1 Introduction

- **@replies** — much like on Twitter, used when you want to mention someone. You can @everyone or @channel.
- **emoji** — just like what you see everywhere.
- **Integrations** — Slack can ‘talk’ with many other systems. For example, you can post a doc from your Google Drive when submitting an assignment; you can type /giphy to share random gifs. I will explain more later.

To learn more, check out the Slack tutorials.

**Hypothes.is.** If you’re a formal participant, you should have received a link that invites you to join a private Hypothes.is group. Hypothes.is is a web annotation tool that enables us to discuss readings in a contextual manner. When annotating, please make sure our group name (**SNAEd**) is properly selected. **Pro Tip:** Please try to use a same (or at least a similar) username for Slack and Hypo so that we can match you in two spaces.

If you are an Open Participant, please annotate publicly and make sure to tag your annotations with **SNAEd**. Otherwise, it will not be picked up by our class feed.

To learn more, check out Hypothesis Quick Start Guide for Students, and a more complete Resource Guide for Students.

**Zoom.** We will host virtual meetings on Zoom from Week 7. Please make sure your computer has Zoom installed (see the video below). Please test your Zoom setup in advance to make sure audio and video configurations work properly. You will receive a calendar invitation with details about joining our Zoom meetings.

**Twitter** (optional). If you tweet, like me, please tag your tweets with **#SNAEd** as well.
To demonstrate how Twitter could be useful, I bet you can learn more from TechPizza than from this course (set Topic to be, for example, ‘Text, Network Analysis’ -> ‘network:gephi’).

**Google Drive.** Google Drive will be used for major assignment submissions. I will provide detailed feedback directly in your Google Docs, so it is important you use Google Drive. UMN offers free Drive access and unlimited storage. If you need help, check this OIT page: [https://it.umn.edu/technology/google-drive](https://it.umn.edu/technology/google-drive)

To submit an assignment through Slack, head to the #assignment Slack channel and click on the ‘+’ button to pull out a menu with options of adding Drive files. You may need to authenticate Google Drive on Slack. See help page.

### 0.1.5 Week 1 Activities

Below are a list of activities for Week 1. They are all **due by Jan 23, 23:59PM**, but I strongly encourage you to get started as early as possible!

#### 0.1.5.1 Get to know each other

**UMN Participants.** Please record a short Flipgrid video to introduce yourself to the class. This activity is important for this primarily online class. You can either follow [this link](#) or directly record below.

After recording your self-intro video, please check each other’s videos and respond. Don’t feel shy about pointing out your shared interests or one classmate’s nice wall decorations :-). (Pro Tip: Test your sound to make sure it works.)

**Open Participants.** Please leave a Hypothes.is annotation to **SAY HI**!

#### 0.1.5.2 Readings & Annotations

Read and annotate the following texts

- Carolan, ch. 1

When annotating, please try to do at least one entry for each of the following:

- Annotate an **SNA term** you find interesting/useful, and provide a definition. Use tags term in your annotation
- Annotate an **example application** of SNA in readings, and explain how you find it useful or interesting. Use tags application in your annotation
- **Reply to annotations** made by other participants

Have a wonderful week!
0.2 Basic Concepts

You’ve done an absolutely wonderful job on Flipgrid and Hypothesis! Each of you brings very unique interests and expertise to the community. I am excited about learning with and from you in the world of SNA!

Our interaction network on Hypothes.is

Your contributions on Hypothes.is have been very thoughtful in so many ways. You introduced domain knowledge (e.g., about families) to contextualize our shared texts. You asked important questions. You took time to search for terms and share what you found via annotations. Keep up with your good work!

Based on our social interactions on Hypothes.is, I constructed a simply network visualization below. In the visualization below, each node represents one of us, and each directed edge represents the existence of at least one reply. So far 7 colleagues in our communities have connected during the first week. In Week 2, I believe 1) more colleagues will join the conversation, 2) there will be more back-and-forth dialogues (or increased reciprocity in SNA terms), and 3) we will have a denser network. Will we be able to conquer these goals?

In terms of specific learning goals, in this week, we will:
1. Explore social network perspectives
2. Become familiar with basic SNA concepts
3. Start to draft an initial SNA project idea

0.2.1 Understanding SNA in Educational Research

0.2.1.1 Three levels of considerations

In the previous video, I described three levels of considerations that educational researchers often need to be aware of. You could find more information in text such as (Niglas 2010) or some other research methodology courses. Please note that this framework was constructed to help us grapple with the complex terrain of research, and it is highly debatable.

Below, I try to re-iterate the key message of the video – but from the bottom up:

- **m**, or methods/techniques: The “small m” in SNA constitutes methods or techniques we apply in SNA research. Imagine we are using SNA to investigate friendship of a network of high-schoolers (think about “Gossip Girls” if you’ve watched that TV series). A technique in this SNA research could be a questionnaire used to collect friendship data among students; it could be the force-directed layout we use to visualize this network; it could be the measure of betweenness centrality we use to characterize high-schoolers; it could also be a network modeling algorithm we apply to model the flow of gossips. In a nutshell, these techniques are more about what we concretely do in an SNA research.

- **M**: When SNA is referred to as a “big M” Methodology, it is treated as a systematic approach of investigating a phenomenon. Beyond simply applying these techniques, a methodology is also concerned with why a technique gets used. In other words, understanding SNA as a methodology means learning to make informed decisions in any stage of an SNA project. For example, why using a questionnaire instead of observations or interviews? In which cases should one use a circle layout instead of a force-directed layout? Why a specific SNA measure is appropriate for addressing a research question? In a nutshell, the big M is concerned with how knowledge could be best gained by following many SNA methodologists and researchers have created so far.

- **P**, worldviews, philosophical schools of thoughts, paradigms: In SNA, some scholars go further to argue SNA offers a unique way of “seeing the world.” In Carolan (2014) chapter 2 you will read about the relational perspective that represent a particular worldview that emphasizes relations instead of attributes. You will also read about relational realism that is referred to as an ontology grounding SNA. To a great extent, SNA offers a new research paradigm. As put by Barry Wellman, a guru in SNA from the University of Toronto, “It is a comprehensive paradigmatic way of taking social structure seriously by studying directly how patterns of ties allocate resources in a social system” (see, Carolan, 2014, p. 33).

In this course, we are mostly concerned with the “big M” level. We will not dive too deep into the P level, and we will not settle with specific techniques. Together, we will learn how to apply various techniques to systematically produce knowledge about a phenomenon.
0.2.2 Basic Concepts

Last week, you’ve identified a wide range of terms from course readings. Thanks to many of you who contributed your definitions of important terms. See the Hypothes.is search results below:

```shell
# PhantomJS not found. You can install it with webshot::install_phantomjs(). If it is installed, please make sure the phantomjs executable can be found via the PATH variable.
```

This week, we will be immersed in a number of key SNA concepts. In the following video, I “glide over” some basic terms that are often used in SNA. You do not need to memorize them all. As a matter of fact, I see those terms existing as a network (see the image below); as you “unlock” one term, you are also activating others. So spend time on some terms, and you’re implicitly learning about others.

(Credit: Small World of Words project)

0.2.3 Week 2 Activities

Due by Feb 4, 11:59PM

0.2.3.1 Read & Annotate

Read:
• Carolan, ch. 2, starting from “The Integration of Theory and Method” (pp. 32–42)
• Carolan, ch. 3

Pro Tip:
• Please use the links I provided below (i.e., with .ezp1.lib.umn.edu). Please also annotate the webpage directly instead of PDFs. Otherwise, our annotations/convos may get scattered among different pages.

Annotate using Hypothes.is. Even though I do not require you to make a specific number of annotations, I encourage you to accomplish the following “ABC” this week:

1. **Ask a question**
2. **Brag about your understanding** about an SNA term, a domain-specific theory, a cool tool, etc.
3. When you reply to one peer, **Connect another peer’s ideas** in your annotation

### 0.2.3.2 Start to put together an intial project idea

From your Flipgrid videos, I see most of you already having fascinating project ideas. Even though we are only in Week 2, I encourage you to take your ideas to the next level when you engage with our readings and annotations.

Specifically, you can consider the following: (1) what is the central **problem** of your research and why it is worth investigating, (2) what **phenomenon** you are studying, in which **context**, and what key **question(s)** you have about the phenomenon, and (3) why SNA is potentially fruitful based on your current understanding.

We will share out our initial project ideas by the end of Week 4.

Have a wonderful week!

---

### 0.3 Applications and Examples I

Thank you all for doing the **ABC**—Ask a question, Brag about your understanding, and Connect with a peer’s idea—when annotating our readings! Below is a visualization of our interactions on Hypothesis in the past two weeks. It is slightly more complex than the one from last week, as I used a node-level measure named **indegree** to scale the size of nodes.
In Week 2, I challenged our community with the following goals:

1) more colleagues will join the conversation, 2) there will be more back-and-forth dialogues (or increased reciprocity in SNA terms), and 3) we will have a denser network.

Based on what I see in this new network, I can claim with confidence we have more colleagues participating in our conversation on Hypothesis (from 7 to 10) and higher reciprocity (from 0.18 to 0.3). But because of the increase of participants, our network did not get denser (from 0.26 to 0.22). So I would like to congratulate on our progress and encourage you to interact even more with different colleagues in our community!

Cannot wait to make such a graph on your own? We will start exploring SNA data and tools very soon. But I see our current exploration of theories, contexts, and research questions even more critical than manipulating tools because tools are the easy part!

In this week, we will:

- Examine some concrete examples of SNA research
- ‘Present’ an example of your choice to the class
- ‘Take’ ideas from each other to inform your project ideas

*Pro Tip:* You can adjust the speed of Youtube videos by clicking on the gear icon at the bottom of the video player. *Note:* Please refer to the Week 3 Activities page for deadlines.
0.3.1 Applications and Examples of SNA

After the first two weeks of exploration, I am quite sure we’ve come to a share recognition on the potential of SNA in various domains. In this page, I want to share two applications of SNA – outside of education – that I found inspiring.

The first example is Suzanne Simard’s TED talk on “How trees talk to each other” that I stumbled upon when listening to NPR. (Canadian accent alert!) The reason I picked this example is Simard’s work demonstrates a paradigm shift (discussed by Barry Wellman in last week’s readings) in the study of trees and forests by highlighting the importance of ties among trees.

The second example is Amar Dhand’s work on “stroke patients’ health behaviors” that is based on a specific type of social network named ego-networks. The reason I picked this example is Dhand did a great job presenting the motivation of their work and went on to investigate a solution powered by SNA. This presentation could be a great model for you to follow for this week’s class activity (see next page for details).

Did you find these examples inspiring? Feel free to share on the Slack ideas channel, or make Hypo annotations on this page using hashtags SNAEd & ideas.

0.3.2 Week 3 Activities

- **Read & Share**, due by **Friday Feb 8, 11:59PM**
- **Take / comment**, due by **Monday Feb 11, 11:59PM**

0.3.2.1 Readings

1) **Skim** Carolan, ch. 10, on “Social Capital”, and annotate concepts you recognize / don’t know

2) **Closely examine** an empirical study of your choice. You could either (a) **claim** one article from below – by annotating it using Hypo, or (b) **search** for one article from your research area. I would recommend option #b so that this reading activity is adding to your project.

When examining the article, please try to discern those three levels of considerations I presented in Week 2 (i.e., Philosophical, Methodological, and methods).

0.3 Applications and Examples I


0.3.2.2 Share & Take

1) **Share** the article you read in a concise manner to the class. The video on “stroke patients” in the previous page could be a great example to follow. In your sharing, you should cover:

- research problem and questions
- collection of SNA data
- specific use of SNA (e.g., specific SNA measures, visualizations)
- key findings
- one “praise” and/or one “push”

You can choose between these **two formats**:

- a **screencast** (5-min max): A video presentation of you talking through your example. Post a link to your video to the Slack **assignments** channel. Potential video capturing tools include Jing, QuickTime, and Screencast-O-Matic.
- a **Slack post** (500-word max): Post a text-based post to the Slack **assignments** channel as we did in Week 2.

2) **Take** ideas from each other’s presentations to inform your project. **Comment** on
each other’s posts to acknowledge ideas you take from them, seek clarifications, etc. Continue to craft your project ideas.

Have a wonderful week!

0.4 Collect and Manage Network Data

In this week, we will:

- Get to know some key decisions SNA researchers need to make on data collection
- Understand several data collection techniques
- Understand processes involved in storage and transformation of SNA data
- Practice importing data into SNA software
- Become aware of ethical issues involved in SNA research

0.4.1 Understanding SNA Data

Blankly speaking, data collection in SNA research is concerned with two types of data: (1) **relational data** that describe *ties*, and (2) **attribute data** that describe *nodes*. For example, if I want to study friendship in a high school class, depending on my research questions I may choose to collect attribute data of each student (such as gender, race, GPA), and relational data of every possible pair of students (such as whether Student A texts Student B, or how many times A texts B). Quite simple, right?

However, real-world SNA projects in education demand a number of critical decisions to be made by the researcher. Just to list a few examples:

- how to gain access to the research “field” (e.g., a student fraternity, an intimate parent group)
- from whom are data collected (and who are excluded from data collection)
- which instruments are used for data collection
- how are data structured and stored
- how to transform data to different “shapes” to address specific research questions

In this week’s reading – Carolan (2014), Ch 4 – you will read detailed suggestions from the author. Below I briefly comment on a few key points, before engaging you in detailed techniques in later sections.

0.4.1.1 Sources of SNA data

SNA data may be obtained in a variety of ways—from historical archives, questionnaires, ethnographic studies, system logs of online platforms, Medicaid claims, etc. For example, from public records researchers could analyze *co-sponsorship of legislation* in the U.S. Senate;
0.4 Collect and Manage Network Data

from a classic Chinese novel, *Dream of the Red Chamber*, researchers could use SNA to estimate relationships between characters based on their co-occurrence (Zhao and Weko 2016). In these two cases, we can see that in SNA research some relational data are natural or readily available (e.g., co-sponsoring of a legislature), while some other need to be derived (e.g., relationship between novel characters based on the co-occurrence of names in a same sentence).

No matter how relational data get gathered or derived, I want to emphasize the importance of making sound justification on the collection/creation of relational data. For example, I reviewed a manuscript that analyzed “co-location networks” based on students’ simultaneous access to wifi hotspots on a university campus. As a critical reviewer, I would pay special attention to any strong claims made on “social” connections among students, because accessing a same wifi hotspot does not imply any social interaction. However, if the study looks at pairs of students simultaneously accessing 10 hotspots on the campus every day, it would become a totally different story as such intense co-location could be an indicator of (potential) social ties. Therefore, in SNA studies we need to constantly reflect on the contextual definition(s) of ties and the operationalization of the definitions in data collection.

0.4.1.2 Totality and sampling

In some cases, we are able to collect a whole social network. Imagine the year-long NASA simulation of a Mars mission in an isolated dome, researchers would have a better chance of studying the social network of all six scientists in its totality. In other cases, when it becomes impossible to study a whole network (e.g., terrorist networks), researchers will need to apply specific techniques of sampling. As you will read this week, sampling in SNA research is different from sampling we commonly discuss in an introductory research methods course. This is because SNA research is concerned with both the nodes and the ties. Simply put, a representative sample of nodes does not naturally guarantee a meaningful sample of ties. SNA researchers need to be especially aware of the impact of sampling on relational data. For example, in a study we may systematically sample every 5th student from a school based on student IDs (sampling applied on nodes), and we may also ask each student to name up to 3 friends in this school (sampling applied on ties). Ego-network also implies an interesting sampling mechanism, as it cares about a focal ego and the nodes to whom the ego is directly connected to plus the ties. This could be intuitively understood by sampling based on distance from the ego.

How we sample nodes and/or ties depends on how we specify the boundaries of networks. The definition of network boundaries is highly critical for any SNA research and requires systematic considerations on research questions, theoretical perspectives, availability of data, etc. To quote Scott (2012):

\[\ldots\] the determination of network boundaries is not simply a matter of identifying the apparently natural or obvious boundaries of the situation under investigation. Although ‘natural’ boundaries may, indeed, exist, the determination of boundaries in a re-

search project is the outcome of a theoretically informed decision about what is significant in the situation under investigation... Researchers are involved in a process of conceptual elaboration and model building, not a simple process of collecting pre-formed data (pp. 44-45)

This important recognition speaks back to my Week 1 video on the importance of learning to make decisions in such a research methodology class. Defining the boundaries of networks is certainly the most central decision for an SNA project. When we start to inspect these decisions, it may looks like we’re opening “a can of worms”, as many decisions may look artificial, or slippery at best. In this class, I hope we see decision points as “a bag of diamonds”—each worth staring at from different angles.⁵

0.4.2 Ethics in SNA research

Ethical considerations are important for both research and practice. We won’t spent much time on this topic this week, but I want to encourage you to attend to ethical concerns in your specific research contexts. Does your research involve venerable populations (e.g., young children)? Does your research involve sensitive data (e.g., health data)? Is your research field online or offline? Because SNA is applied in all kinds of educational research, it is almost impossible to come up with a unified ethics guideline. I strongly encourage you to skim this chapter from an Open Textbook produced by the UMN Libraries. What are the possible ethical concerns in your project? I encourage you to start taking notes on how you could address them.

I also want to mention that research ethics is never a static topic at all. In many emerging research spaces (e.g., the Internet), research ethics remains highly debatable (Rivers and Lewis 2014; Kraut et al. 2004). Feel free to share your thoughts via Hypothesis by adding the ethics hashtag, or in the Slack home channel.

If you’re a UMN student and wish to conduct an actual SNA project as part of the class project (which is not required), please check this UMN IRB website for further information. I’m willing to sponsor your project and we can meet individually to talk it through.

0.4.3 Linear algebra basics

(You can skip this section if you’re already familiar with linear algebra.)

Matrix is all you need to know at this point. This class will focus on mathematical intimations and let the computer do heavy-lifting computations for us.

If you need a quick refresher, check out this video on matrix:

If you want to dive deeper in linear algebra, start from this intro video and move onto its learning sequence. Again, you don’t need to watch them all and we will introduce mathematical concepts later when necessary.

⁵See https://en.wikipedia.org/wiki/Force-directed_graph_drawing
0.4 Collect and Manage Network Data

0.4.4 Managing SNA data

On this page, I introduce recommended ways of structuring and handling SNA data. Here I especially consider principles of tidy data (Wickham 2014), which may question things you encounter in textbooks. The principles of tidy data are very simple (p. 4), and I will explain each principle below with examples.

1. Each variable forms a column.
2. Each observation forms a row.
3. Each type of observational unit forms a table.

0.4.4.1 Basic representations

Note that SNA data typically include attribute data about nodes and relational data about edges. So the most straightforward way to represent a network is to have two separate tables.

For example, consider a student group with four students (nodes). Table 1 contains attribute data of each student. This table is tidy because each variable form a column, each observation (i.e., student) form a row, and it contains a single observational unit student.

TABLE 0.1: Table 1. A table of nodes.

<table>
<thead>
<tr>
<th>name</th>
<th>gender</th>
<th>age</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>F</td>
<td>15</td>
</tr>
<tr>
<td>B</td>
<td>M</td>
<td>14</td>
</tr>
<tr>
<td>C</td>
<td>M</td>
<td>13</td>
</tr>
<tr>
<td>D</td>
<td>F</td>
<td>14</td>
</tr>
</tbody>
</table>

Table 2 describes book lending activities (ties) among students. For example, Row 1 means Student A lent one book to B, while Row 4 shows B lent 2 books to C. It is also a tidy dataset.

TABLE 0.2: Table 2. A table of weighted ties

<table>
<thead>
<tr>
<th>source</th>
<th>target</th>
<th>weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>C</td>
<td>0</td>
</tr>
<tr>
<td>A</td>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>2</td>
</tr>
<tr>
<td>B</td>
<td>D</td>
<td>1</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>0</td>
</tr>
</tbody>
</table>
Note that *weight* is not always required for networks. In a study that only cares about the existence of a tie, Column 3 will contain only 0 and 1. Or, rows with having a value of 0 in Column 3 will be simply removed from this table.

Table 3 could be the original record from which Table 2 is constructed. In Table 3, each row represents a book lending action, with its date recorded in Column 3. Here, you get a sense how researchers may need to **transform** data from its original observations (Table 3) to a specific format (Table 2), even though most SNA software can handle both formats.

**TABLE 0.3: Table 3. A table of raw data of ties.**

<table>
<thead>
<tr>
<th>source</th>
<th>target</th>
<th>date</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>B</td>
<td>2017-02-03</td>
</tr>
<tr>
<td>A</td>
<td>C</td>
<td>2017-02-04</td>
</tr>
<tr>
<td>A</td>
<td>D</td>
<td>2017-02-05</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>2017-02-06</td>
</tr>
<tr>
<td>B</td>
<td>D</td>
<td>2017-02-07</td>
</tr>
<tr>
<td>C</td>
<td>D</td>
<td>2017-02-08</td>
</tr>
<tr>
<td>B</td>
<td>C</td>
<td>2017-02-09</td>
</tr>
</tbody>
</table>

Additionally, in situations you do not care about **node attributes** (in Table 1), you can simply only use relational data – only about edges – to construct a network. In this case, you will only have a table of relational data, which already contain the most basic information (identifiers) of nodes. Take Table 2 for example, all unique node identifers in columns *source* and *target* will be extracted to create a list of nodes, with no further information about their attributes.

### 0.4.4.2 Two-mode data

Imagine the research project is actually more complicated: We are also interested in the relationship between book-lending behaviors and student affiliations with sports teams. In this case, you may have two additional tables below.

**TABLE 0.4: Table 4. Sports teams in the school.**

<table>
<thead>
<tr>
<th>sports_teams</th>
<th>pratice_day</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseball</td>
<td>Tue</td>
</tr>
<tr>
<td>basketball</td>
<td>Mon</td>
</tr>
<tr>
<td>volleyball</td>
<td>Fri</td>
</tr>
</tbody>
</table>
TABLE 0.5: Table 5. Student affiliation with sports teams.

<table>
<thead>
<tr>
<th>student</th>
<th>team</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>basketball</td>
</tr>
<tr>
<td>A</td>
<td>volleyball</td>
</tr>
<tr>
<td>B</td>
<td>baseball</td>
</tr>
<tr>
<td>C</td>
<td>basketball</td>
</tr>
<tr>
<td>D</td>
<td>baseball</td>
</tr>
<tr>
<td>D</td>
<td>volleyball</td>
</tr>
</tbody>
</table>

Like what I just mentioned, you could ignore Table 4 if Table 5 already contains all information about sports teams. But if there is a football team not covered by Table 5, you will need to include Table 6 as well.

TABLE 0.6: Table 6. Sports teams in the school (version 2).

<table>
<thead>
<tr>
<th>sports_teams</th>
<th>pratice_day</th>
</tr>
</thead>
<tbody>
<tr>
<td>baseball</td>
<td>Tue</td>
</tr>
<tr>
<td>basketball</td>
<td>Mon</td>
</tr>
<tr>
<td>volleyball</td>
<td>Fri</td>
</tr>
<tr>
<td>football</td>
<td>Wed</td>
</tr>
</tbody>
</table>

Using Table 5, you could construct a **two-mode network** – also called as an affiliation network – with students and sports teams as two types of actors in the network. In contrast, Table 2 only has one mode – students.

Finally, if your research project is concerned with friendship in general – which covers both book lending and sports affiliation – you could even merge two types of relational data together (with solid justification). For example, from Table 5 we can tell A and C are both in the basketball team. We can then adjust the weight between A and B in Table 2 accordingly. This is another type of **transformation** you may need to do in SNA research. Knowing basic data transformation techniques – either in spreadsheet software or in R – would be helpful for work in this class.

To summarize, this page provides a basic overview of how SNA data could be structured. You may encounter different ways of representing SNA data, such as a relationship matrix with rows and columns representing the same set of actors (see the Harry Potter support networks for example). Such representations could all be derived from a *tidy* dataset discussed above. In data collection, we will strive for keeping as much raw information as possible (such as timestamp), to enable analyses that only come to your mind afterwards.
0.4.5 Import data into software

After all these explorations in the past weeks, it’s time to play with some SNA data and generate some graphs!!

**Open Dataset:** Les Miserables, containing co-appearance weighted network of characters in the novel *Les Miserables*. Note: The dataset is the 5th on the page.

**Two Tracks.** To accommodate your learning preferences, I provide two tracks of learning based on what SNA software you wish to learn.

- **R** is a statistical language that involves a little bit of coding. It may have a steep learning curve if you’ve never coded before, but I believe everyone of us can learn it and the benefits of knowing R is huge. So I recommend R in this course.
- **Gephi** is a wonderful SNA software package coming with a graphical user interface. If you’re more comfortable with tools like MS Word or SPSS, and if you’ll never want to try R, it would be a good option for you. There are many similar software applications out there, many of which you’ve already encountered in readings. You can use them, but they’re not officially supported in the class.

If you have questions, please post in the *home* channel.

0.4.5.1 Track R

If you’ve not used R before, please take Module 1 of this free course on Introduction to R. If you have questions, post them in the *home* channel or DM Bodong.

Below are R codes to get you started. I’ve invited you all to an RStudio Cloud space where you can open a project named *Week 4*, which has the R code for you to get started. See the screenshot below.

![RStudio Cloud screenshot](https://rstudio.cloud/spaces/8406/projects)

You can go far beyond the codes. And make sure you share what you make to our Slack *assignment* channel.
# Load igraph
library(igraph)

# Load data
# check the head (first 6 rows) of the dataset
head(lesmis)

# Create a graph using the graph_from_data_frame function
g <- graph_from_data_frame(lesmis)

# Plot the graph
plot(g)
# make the graph a little prettier
plot(g, edge.arrow.size=.2, vertex.label=NA, vertex.size=8)

Below is a great video made by James Cook (a professor from University of Maine) that would be helpful as well.

0.4.5.2 Track Gephi

To begin, download the .gml file directly from the dataset page. You can directly open this .gml file using Gephi.

Follow this really nice tutorial on network visualization and analysis with Gephi. You only need to get through the first 10 minutes. Given the nature of the dataset, you may want to filter edges to leave out those edges with a weight of 0.

Like folks from Track R, please share what you make to our Slack assignment channel.

0.4.6 Week 4 Activities

0.4.6.1 Read & Annotate

- Carolan, ch. 4
- Book chapter on Ethical Considerations With Social Network Analysis

Even though I do not require you to make a specific number of annotations, please continue to do the “ABC” of social annotation in this community:

1. Ask a question
2. Brag about your understanding about an SNA term, a domain-specific theory, a cool tool, etc.
3. When you reply to one peer, Connect another peer’s ideas in your annotation
0.4.6.2 Project idea share-out

After spending the past few weeks on SNA basics and examples, we will share out our initial project ideas to the community. Specifically, you can post the following: (1) what is the central problem of your research and why it is worth investigating, (2) what phenomenon you are studying, in which context, and what key question(s) you have about the phenomenon, and (3) why SNA is potentially fruitful based on your current understanding.

- To share, please create a Slack post under the assignment channel, due by Friday Feb 15, 11:59PM.
- Please spend time reading our colleagues’ posts and leave your constructive comments by Monday Feb 18, 11:59PM.

0.4.6.3 Play & Share

- Import a practice dataset into the SNA software of your choice
- Get familiar with the software
- Share the network visualization you make in the assignment channel, due by Monday Feb 18, 11:59PM

Have fun, everyone!!

0.5 Sociograms and Density

In this week, we will:

- Understand how network visualization works
- Become familiar with a number of network-level measures
- Compute network-level measures using concrete data

An overview of Week 5 activities:

- Read and build knowledge
  - Carolan, ch. 5
  - Annotate measures with mathematical/computational solutions
  - Identify network-level measures not covered by the book and share to the class
- Complete the data collection hands-on assignment
- Compute network-level measures and share to the class
0.5 Sociograms and Density

0.5.1 Sociograms and network visualizations

A sociogram is a graphic representation of a social network. Its goal is to transform mathematical representations of a network – e.g., an adjacency matrix – into a network drawing.

The creation a sociogram depends on network layout algorithms. Network layouts are simply algorithms that return coordinates for each node in a network. Before the invention of such algorithms, researchers had to use graph drawing heuristics to produce network graphs by hand, which could involve multiple iterations of drawing. Nowadays we ask the computer to do the job for us :).

Decision Point. How you make a sociogram depends on what information, such as network structures and patterns, you hope to visually convey. Even though sociograms we’ve seen in readings are predominantly force directed, it does not mean force-directed layouts are always the best choices.

The following knowledge and understanding are important for making a sociogram:

- Know common network visualization goals (e.g., revealing segregation, showing hierarchy)
- Know network visualization types (e.g., heat maps, network maps, hive plots), and corresponding layout algorithms
- Know visualization controls (e.g., size, color, shape, width, position)

Putting knowledge into action, when producing a sociogram you need to:

1. Identify a network visualization goal (or goals)
2. Choose a network visualization type
3. Apply proper visualization controls
4. If needed, repeat 1-3.

Below, I present a few network visualization examples that do not draw on force-directed layouts.

---

7 See https://en.wikipedia.org/wiki/Force-directed_graph_drawing
8 See http://kateto.net/network-visualization
0.5.1.1 Example 1: Global Flight Network

1. Goal: Visualize patterns of the global flight network, for example, to see intensive flight activities in certain areas.
2. Type: The coordinates are determined by the origin and destination of each flight. In other words, no additional algorithmic decisions are made beyond the use of a world map.
3. Controls: Line color represents duration of a flight. Brightness of each line adds up to each other to represent the intensity of flight activities in certain areas.

Note: How will this visualization look differently if the map is not eurocentric? Does it matter? What ‘world-views’ might have been built into such a visualization?

0.5.1.2 Example 2: Networks of interaction in mobile animal groups

(See the article)

Here the concern is network dynamics, and the placement of each node (representing each individual animal) is based on its (projected) physical location in the real world.
0.5 Sociograms and Density

Note: In this movie, which information naturally existed and which information was constructed by researchers?

0.5.1.3 Example 3: Migration Network

This visualization attempts to visualize migration flows into Minnesota across time. This visualization could be seen as a series of ego-centric networks of Minnesota in terms of migration into the state.

Note: Like Example 2, which information naturally existed and which information was constructed by researchers? Which decisions did the ‘data journalist’ make in this graph?
0.5.1.4 Example 4: Light Up the Curriculum

This is a visualization made to demonstrate semantic similarity among three types of documents—curriculum standards, students’ online posts, and their verbal discussions. This visualization technique is called *hive plots*. Each axis represents a type of document. This visualization was chosen because researchers only cared about cross-type linkages and didn’t care about the links within each document type.

0.5.1.5 Example 5: The Food Network Backbone

If you are a foodie, you may appreciate this visualization of flavors backed up by some hardcore science. Do you find the use of graphical devices by the authors illuminating in this example?
0.5.1.6 Play more with network visualizations

This part is optional. I’m providing more materials for those who hope to explore network visualization a bit more. Also, note that there are some “advanced” concepts in these materials that we will discuss in later weeks.

**Track R**: I encourage you to follow [this tutorial](#) I’ve already mentioned above. You can download data and R scripts used in this tutorial from the author’s Github repository.

**Track Gephi**: I encourage you to read the first section of [this tutorial](#). Also, check out the following video that demonstrates Scientific Literature Analysis using Gephi. You are also encouraged to check out the first 3 sections of [this Gephi tutorial](#).

As always, please report back to our Slack home channel.

0.5.2 Density and Other Structural Measures for Complete Networks

Carolan (2014), chapter 5 does a great job introducing a list of commonly used network-level measures.
Below, I make several extra points that are not covered:

First, the chapter is based on a directed, non-weighted network. We need to recognize that there are several types of networks based on the kinds of ties they have (see Table 5.1⁹). Take the measure of density for example, the number of possible links are different for a directed network and an undirected network with the same size: a directed network has more possible links because each pair of nodes have two possible links. There are different algorithms (built in SNA tools) that would account for network types when calculating these measures.

**TABLE 0.7: Table 5.1: Different types of networks.**

<table>
<thead>
<tr>
<th></th>
<th>Undirected</th>
<th>Directed</th>
</tr>
</thead>
<tbody>
<tr>
<td>Unweighted</td>
<td><img src="image" alt="Unweighted Diagram" /></td>
<td></td>
</tr>
<tr>
<td>Directed</td>
<td><img src="image" alt="Directed Diagram" /></td>
<td></td>
</tr>
</tbody>
</table>

⁹See the R code I wrote to produce these toy networks.
Second, the researcher sometimes need to make important decisions when computing network measures, even if a same algorithm is applied. For example, for a weighted network, it is a common practice to set a threshold of weight (based on some justified reasons) to filter links weaker than the threshold (see Example 4 from the previous page). Then network density could be computed the same way as an unweighted network. In this case, the calculated measure of density is highly sensible to the researcher’s decisions, which are influenced by her/his assumptions of the network. (Decision Point)

“It is important that a researcher does not use a measure simply because it is available in a standard program. A researcher must always be perfectly clear about the assumptions involved in any particular procedure, and must report these along with the density measures calculated” (Scott, 2013, p. 73)

As we’re diving into specific SNA techniques – such as density, average degree – I want to remind you of the “3 levels of considerations” we explored in Week 2. Decision points we’ve covered this week, such as which layout to use and what threshold to apply when computing density of a weighted network, are often informed by theoretical frameworks we adopt in a
study or contextualized knowledge we have about a research problem. It is important for us to make our theoretical assumptions explicit in an SNA study.

0.5.3 Assignments and Activities

Due by Monday, 2/25, 11:59PM

0.5.3.1 Read and build knowledge

Read Carolan (2014) Chapter 5. Make Hypothesis annotations as we normally do using proper hashtags (e.g., question, idea). Don’t forget to include our SNAEd course hashtag.

Build knowledge as a group: When reading through the chapter, identify at least one network-level measure, conduct research on how it could be computed using R or Gephi, and share your solution using Hypothesis. Include at least three hashtags in your Hypothesis annotation: SNAEd, compute and R/Gephi. - If you found someone else has already covered your ‘favorite’ measure, try to build on the existing solution by, for example, introducing a parameter/mechanism to deal with weighted networks.

Finally, if you find a network-level measures not captured in the book, please annotate HERE on this page. Briefly introduce the measure and provide a computation technique.

0.5.3.2 Assignment: Data collection hands-on

Last week, we used the Les Miserables sample dataset. This week, you are encouraged to identify a dataset – public or your own – and organize it according to the “tidy data” principles discussed in Week 4.

- The Network Data Repository is a good place to look at, or you can build a network based on public data (e.g., from a novel, Twitter). Take a look at NodeXL if you’re interested in public Twitter data.
- You’re not expected to accomplish an SNA study by the end of this semester, but you are encouraged to work on an authentic study of your own when possible.
  - Please note that if you plan to conduct a study that involves human subjects, an IRB approval is often required. The IRB office’s webpage is a great starting point for you to decide whether you need an IRB approval.
- If the public dataset you find is not tidy, you will need to transform it to a tidy dataset.

When submitting this assignment, depending on whether your dataset contains sensitive data or not, you can choose to share it publicly within the class on the assignment Slack channel, or only share computed results (more below) with the class on Slack.

0.5.3.3 Compute network-level measures and share

No matter which dataset you use, please try to compute network-level measures and share back to the class on Slack.
You are encouraged to share scripts (for Track R) or mini-videos (for Track Gephi and R) even though they are not required. You are also encouraged to ‘triangulate’ measures with visualizations.

Enjoy a great week!

### 0.6 Centrality and Centralization

In Week 5, we explored a range of network-level measures, including **size**, **density**, **diameter**, **average path length**, **centralization**, **reciprocity**, **transitivity**, **clustering coefficient**, and so on. There are other interesting measures not covered by the textbook but worth your attention:

- **Cohesion**: The degree to which actors are connected directly to each other by cohesive bonds.
- **Structural cohesion**: The minimum number of members who, if removed from a group, would disconnect the group.

You have done a phenomenal job digging into different concepts and measures and sharing your findings back to the group via annotations. A shout-out to **HopeinJ** for asking an important question:

@bech0079, I don’t quite understand the difference btw the network.density between the network.density and edge.density, both gives the ratio between extant edges/potential edges.https://www.rdocumentation.org/packages/network/versions/1.13.0/topics/network.density
Do you have any idea on that?

I want to take this opportunity to mention that there is a “universe” of R packages, the dependencies among which can be visualized as a giant graph (see this page). Among these packages, there are many packages developed for SNA. The **network** package mentioned by **HopeinJ** is another R package that can be used for SNA. In **igraph**, please use [edge_density](https://www.rdocumentation.org/packages/igraph/versions/1.2.4/topics/edge_density) to compute network density.

---

Please make sure to check our Hypothes.is annotations to learn about contributions from other colleagues. If there are burning questions, please continue the discussion on Slack.

This week, we will:

- Engage with node-level measures in social networks
- Understand how network-level and node-level measures are connected
- Learn to calculate node-level measures using SNA applications
0.6 Centrality and Centralization

0.6.1 Centrality and Centralization: An Overview

When we consider the importance of a node in a social network, how central it is usually an important consideration. In Week 4, we were able to use sociograms to identify central nodes in a network. How can we identify those central nodes mathematically in case they are not easily visually identifiable?

Centrality is a key measure in SNA developed to achieve this goal. SNA researchers have developed many ways to quantify centrality in a network. Below, I curate a list of quality resources for you to explore different centrality measures. I selected these resources because of varied ways they present centralities – equations vs. intuitions, real-world examples vs. toy networks, step-by-step demonstrations vs. one-step computation.

First, review the following PDF presentation presenting four main centrality measures: Degree Centrality, Betweenness Centrality, Closeness Centrality, and Eigenvector Centrality. Note that the author introduces node-level centrality and network-level centralization together in this presentation. These two concepts are often mistakenly treated as the same by researchers. Now you see their differences and connections. Also, if you could follow the math equations, that would be great; if not, please focus on the intuitions.

Second, watch the following video from Lada Adamic covering Degree Centrality, Betweenness Centrality, and Closeness Centrality with concrete examples. She also made an attempt to distinguish centralities from centralization. She also noted that when considering centrality, it is very important to be clear about the scope, or the boundary like we discussed in Week 0.4. A little tweak will make a difference.

0.6.2 Betweenness Centrality

Diving deeper, betweenness centrality is probably the most popular centrality measure I’ve personally seen in educational research. The notion of residing in-between has strong implications for many scenarios, either for facilitating/blocking access to resources or encouraging creativity. So I think betweenness centrality is worth further exploration.

Below is a great video made by James Cook to explain betweenness centrality. Enjoy!

0.6.3 Eigenvector Centrality

Eigenvector centrality is another centrality measure that is well aligned with the social capital theory. Watch the following video by Lada to find out more.
0.6.4 Week 6 Activities

0.6.4.1 Readings

None. The textbook does not provide a specific introduction to centraly measures for a whole network.

But you’re encouraged to look back into example papers you selected in Week 3 to see whether you’ve developed any fresh understanding after diving deeper into these measures. Please share your findings with the class on Slack.

0.6.4.2 Compute Node-level Measures

Like last week, you can continue to use a dataset you have, either your own dataset or a public one.

0.6.4.2.1 Track R

Search “centrality” on the igraph doc page. For example, you will find that \texttt{betweenness(g)} is the function for computing \textit{betweenness centrality}. Apply these centrality functions on your dataset from Week 5. Tweak the parameters, such as \texttt{directed}, to see how the results might change. Share your findings on Slack.

0.6.4.2.2 Track Gephi

Some of you have already played with Gephi’s Statistics panel in Week 4. For this week,

- Check page 12 of this Gephi tutorial to see how centralities could be computed using Gephi. See Figure 1.
- Toggle to Data Laboratory to see results of Gephi computations in the Data Table tab. For example, if you click on Network Diameter in the Statistics panel, a number of centrality measures (including betweenness centrality) will be saved in the Data Table. You can click on the header of each column to sort the Data Table. See Figure 2.
- Share your findings on Slack.

Like earlier weeks, if you have any questions or ideas, share in corresponding channels on Slack. Enjoy a great week!
FIGURE 1 Gephi Graph Overview.

FIGURE 2 Gephi Data Laboratory.
0.7 Components and Cliques

Our goals for Week 7 include:

- Understanding components and cliques
- Becoming familiar with several popular approaches to identifying groups
- Using SNA software to identify components and detect cliques

Important note: Our first Zoom meeting is taking place on Monday, March 11, 5:30-7pm. Please check your email for the invitation.

0.7.1 Introduction to Components and Cliques

Please watch the following video before reading Chapter 6 of Carolan (2014). This video provides intuitions that could be useful before diving into mathematical representations.

0.7.2 From Data to Analyses: A Demo

In the following video, I demonstrate a process of going from a secondary dataset towards some simple analyses covered in the past few weeks. This demo is especially useful if you are in Track R. But some principles, such as tidy data and reproducible analysis, should also apply for Track Gephi.

R code used in this demo can be found here. Please feel free to reuse and extend.

0.7.3 Week 7 Activities

0.7.3.1 Read, Annotate, and Share

- **Read** Carolan (2014), ch. 6
- **Annotate** as we normally do using proper hashtags (e.g., question, idea) and doing our ABCs (i.e., “Ask a question”, “Brag about your understanding”, and “Connect another peer’s ideas”).
- **Share** tips and tricks related to component/clique analysis on Slack:
  - For Track R, post Code Snippets from your analysis on the R channel. You can post anything that others may find useful. For example, you can show off your code that detects strong components in a directed network.
  - For Track Gephi, post description, snapshots, or posts on the Gephi channel.
- (Optional) **Find** and **Share** an educational study that deals with network components or cliques. Point out specific techniques applied in that study.
- Don’t forget to respond to each other’s contributions!
0.7 Components and Cliques

0.7.3.2 Project Checkpoint 2

In this week, there is a project checkpoint assignment designed to keep you “on track”. The description of this checkpoint reads as following:

Students will share their refined project ideas, with a concrete data collection plan fleshed out.

At this point, we have diverged tremendously on the final projects! I understand that you are all at different stages of your projects. So I will not require formal submission of this assignment and instead encourage you to make an appointment with me to discuss your project if needed.

Have a great week, and I look forward to seeing you in Zoom on March 11!


