And one last check to make sure that my clicker is working here. Oh, we're good! Fabulous.

Thank you so much for joining us today. My name is Mary Ton, I'm the Digital Humanities Librarian.

And today I'm joined by Cadence Cordell, who will be managing the chat, answering questions as they arise.

Today's topic is network analysis. If you are really excited about any of the resources that we talk about today or want to go back and check some of the projects that we're mentioning, be sure to go to today's slides. The link is go.illinois.edu/networks. The "networks" is all lowercase. It's go.illinois.edu/networks.

Our tasks are first to describe what are the key components of a network. What are the ingredients that we're baking with?

The second thing is to discuss the structure of linked information. How do we need to record and capture information as if we're interested in using this mode of analysis?

And then finally, identify commonly used tools.

But before we get into that, I'd like for you to take a moment to think about what kinds of materials, evidence or sources do you tend to work with?What questions do you hope these materials will help you answer? So take a moment to think about your sources.

What kinds of things are you excited about? What kinds of questions or research interests do you have? If you feel comfortable putting those in the chat, please do. We'll give folks a moment. (Long moment of silence, apologies online viewers)

I'd say for me, I'm a 19th century historian and literary scholar by trade. I'm interested specifically in missionaries who were traveling in the global south. Who were they interacting with? Who was supporting their missionary work? What kinds of local expertise did they need in order to do what they were doing?

These networks of people, places, institutions are truly global. I've been thinking about how do I represent both the extent and the global nature of these networks. And also celebrating local expertise, indigenous communities, and the kinds of local networks of support that these communities formed around.

In thinking about - I always find that when you're working in digital humanities, it can be helpful to look at examples of projects and the kinds of questions or the kinds of things that previous projects have done.

Viral Texts is a project from a professor in the iSchool here, Ryan Cordell. What we're seeing is how information traveled through 19th century newspapers.

You're seeing a network that connects the story that was published in New York with its reappearance in California, and how that grows and changes over time.

One of my other favorite network analysis projects is The Six Degrees of Francis Bacon. And yes, this is a play on the game of Six Degrees of Kevin Bacon, where you try to figure out the connection between two people, a person and Kevin Bacon, through various people.

But this is actually about early modern correspondence networks and points of connection among scientists, explorers, royalty.

Finally, we have one of my favorite research tools, Connected Papers, which is looking at networks of citation.

You can take a paper, put it into this tool and it will show you not just the papers that it has cited, but also the papers that those are citing, and can give you a very quick view of conversations as they've changed over time.

This tool has been absolutely indispensable for literature reviews for me, and I recommend it often.

All of these are using networks as a way to visualize the connection between people, places, documents, and thinking about how these connections change over time.

The first ingredient that we think about when we're thinking about networks are nodes. These are the things in your network. Most often, the nodes in your network are going to be people. But they can also be documents, academic articles, newspaper articles, et cetera.

What are the things in my network? And then we got to think about scope. Who or what is in my network? And who or what is not? Having a sense of limitations not only makes your project doable, but also gives you a sense of what the limitations are in terms of the claims that you can make once you have done this visualization.

Here you see a Web of Thrones, which is a visualization of all of the characters in Game of Thrones and their connections. Apologies for the spoilers for those of you who haven't caught up.

But this is just one example of how networks can help you track the relationship between people.

In this case, the nodes in this diagram are people. It's an example of a network that has only one type of node and that node is people.

There are networks that can include multiple different types of things. We can have a network that connects people and institutions, or people and documents.

These networks might be bimodal if they have two different kinds of nodes, or multimodal if they have three or more.

The next thing that we need to think about are edges. How are these things in our network connected? We might also ask how strong are these connections? These are the weight of the edges.

This might be calculated by number. Two people interact often. They might also be two people tend to cite each other often. So we can see that weight increase as the number of documents piles up and weighs it down.

We can also think about, as the Francis Bacon Project does, if these connections are documented or inferred.

Here's where I have my historian hat on. What I really like about the six degrees of Francis Bacon Project is that they have different visual vocabulary for connections that are documented versus connections that are implied. In this case, we can see that Walter has some implied connections to Francis Bacon because they were credentialed at the same inn or stayed at the same inn.

All right, once we have our nodes, once we have our edges, and once we bring them together, we can ask questions about the network itself. We tend to look at three different things.

The first is density. Is this a highly interconnected network? Are all the nodes, do they have multiple points of connections to each other or is this a sparse network? Are there only a just few key people in this network that are holding it together?

A highly dense network means that there are a lot of different connections between people, places, things, and if you remove one of those nodes, the network will still remain connected.

But if it's a sparse network, that might lead you to ask questions about who is central to this network, who's important, and who is so important that if I remove them, the whole network will fall apart?

These are the kinds of questions that we ask when we're thinking about centrality. Who is most important, who is central in this network? But we might also ask, who are the outliers? Why might they be outliers?

I'll give you an example of an instance where an outlier didn't necessarily mean that someone was insignificant.

There was a project at Indiana University that was looking at networks of spies in the Spanish American War. Project was carried out by Arlene Diaz and Kalani Craig. When they were looking at spies, there was one person who was off to the side that they hadn't seen referenced in any letters or any other correspondence.

It turns out that one person was the spy master who was running these networks of spies in the Spanish-American War.

When we think about centrality, it can often indicate importance or someone that's highly connected to other people or at least has highly documented connections. But we also want to think about the outliers. Who has only tangential connections and why might that be important?

The final thing that we're thinking about is directionality. Sometimes we're not so much interested in - we're more interested in connection. But there are certain types of research questions that might want to take into account directionality.

For example, if I'm interested in how missionary societies were funded, I might say I'm interested in how money traveled from one node in central London, a donor, to a missionary outpost in Rarotonga.

In that case, I want to indicate the directionality of that gift so that I can see how much money is flowing out, and in the case of missionary societies, how much money is also coming back to London and thinking about funding missionary societies elsewhere around the globe.

Let's think about a fun example from fan fiction.

So a research question, an example of a research question that network analysis might help us answer: Who gets shipped the most in Star Trek fan fiction? I've been addicted to Star Trek: Lower Decks recently. Super fun. I have Cadence to thank for the idea for this question.

When we're thinking about ships, the first kind of question that we would ask in this research project is where would we look for this information?

Does anybody have any ideas where we might look for information about fanfiction? If you're in person, you see those black boxes on your table. You can hit the button in the center to speak. Or, our friends online can put things in the chat. Where would you look for information? (Long silence)

[Cadence] Archive of Our Own is correct. Are there any other places?

[Mary] I think Archive of One's Own would be my go to place, the first place that I would check. I might look for things in the metadata that mention specific characters and that would mention Star Trek specifically.

So in this network, who would our nodes be? Who or what would be the things in our network? And I see another one, Fanfiction.net. Yea. (Another long silence)

Who or what would be the nodes in this network? (More silence)

Thank you, Dani, for being so responsive. Star Trek characters? Yeah, we would list Star Trek characters. Absolutely. Because we're interested in the people who are getting put together in fanfiction, like imaginary relationships. I should've defined what shipping is for those of you who don't know this delightful form of fanfiction.

We definitely would want to be looking for specific characters. But we might also be interested to see if we had research questions about, are people writing fanfiction across different eras of Star Trek?

We might want to include information about the series. We might want to include information about the different ships that they're traveling on.

A network with only one mode would have just the characters in it, but a multimodal network would have information about their ships, the series, potentially.

To support this research question, how are they connected? What we might be thinking about are the number of fanfictions written about two specific characters. The more fanfiction that has been written about a specific pairing, the more weight that that edge will have, the more central in our network those two characters will become.

Finally, we might ask, what other kinds of connections do we want to document? Or information about our nodes? If we're interested in reading Star Trek fanfiction through a LGBTQ+ lens, we might want to include information about the gender identity of the characters.

We might also - if we're interested if certain types of pairings have changed over time, we might be interested in including information about when those specific pieces of fanfiction were published to see if there are characters who have a moment, who are more popular at certain points of time in the franchise.

These are the kinds of things that we would consider when we're designing a network analysis project.

Now, structuring your data. When we work with network analysis, we need to think about how we're representing information about those nodes and those edges.

This is where the structure of triplets comes in. Normally, we want to have a node - usually, this is where we would represent information as a spreadsheet.

We have a node, Brad Boimler, who is in Lower Decks, and we have a node of the Cerritos, the ship that he serves on.

We can also add more information about how those two nodes are connected. We can add information about the type of edge that this is. We can say this person is a crew member of this particular ship.

If we were interested in, say, are crew members of the same ship most likely to get shipped together, I realize that's a lot of ships in one sentence. We might include this kind of information so that we can filter the network based on crew type.

Linked open data takes this concept one step further, in that it gives information a specific structure, this triplet structure. And does so in a way that makes information easier to share. It has a subject, a predicate and an object.

In this case, both the subject and the object have what's called a Uniform Resource identifier. This is a number that makes sure that any instance of Brad Boimler or a real person remains the same and that all the information about that particular subject or object can be linked together.

There's also a standardized vocabulary for the predicate, so "isPartOf". These standardized vocabularies allow people to see different types of relationships. This might be published, author, or gave money to, or is a member of a particular organization.

These standardized vocabularies are called ontologies and there are many different kinds of ontologies. They give people a shared vocabulary to describe relationships. So that when we're working together, we can use the same vocabulary to describe the same patterns and trends and people and their connections.

Why is linked open data important, especially when we're thinking about networks of information and knowledge graphs? Well, linked open data connects information about people, places, institutions, documents.

It's one way to do this. It's not necessarily the same as network analysis, but it's related and adjacent and is structuring and thinking about information in much the same way.

The second thing about linked open data that makes it special is that it uses standardized vocabulary so that you can re-use and expand previous datasets.

Now you might need something that's more specific for your research project. And it's okay if you're not using a standardized vocabulary, but thinking about the relationship between people, places, and things in a standardized way invites collaboration in a way that's not possible with a customized project.

Finally, it creates opportunities for ongoing collaboration through open access resources. Two really outstanding examples of thinking about information, networks of people, places, institutions, and making that information available.

The first of these examples is a project called Enslaved, that is trying to recover peoples of the historical slave trade. Thinking about and creating links and foregrounding stories in things that are very dehumanizing as the ships, log books and bills of sale. Using linked open data to humanize and foreground the stories of African Americans and enslaved peoples from this period as a recovery project.

There's also graphic possibilities, another project out of Michigan State that uses linked open data to represent information about comic books, their authors and their publishers.

If you go to this website, you can see how they've used linked open data to create visualizations of publishing networks, authors, their illustrators, and all the creators who contribute to comic books. And as a shameless plug for upcoming events, you can contribute to this project and think about these networks, incredibly vibrant networks of publishers and comic books.

We're having a Wikidata Edit-A-Thon, which is a great introduction to linked open data, to network analysis and thinking about information in a structured way. It will be happening on October 10, from 9:30-4:30, and it's drop in. You're not required to attend from 9:30-4:30, but we welcome you whenever you're free, as you explore thinking about information in a structured way.

With that, we are going to stop the recording to open for questions, ideas, concerns. So give me just one moment.