These transcriptions may contain errors, especially in spelling of names. These are unfortunate, and we regret that we do not have the resources to fix these errors. Still we believe these transcripts will be valuable to many users.

Mapping social media spaces

Marc Smith

>> All right, my dear friend, my buddy, Marc Smith, PhD Sociology, University of California UCLA, and a co-author with the book on NodeXL, but I just had great fun with Marc. We've been working together for five years in the NodeXL project. Every Wednesday for an hour, we meet on the phone with about six or seven other people and argue vehemently about what's the next feature to be added to NodeXL. And we--so when we get together, we continue to argue vehemently but it's just wonderful--

>> Respectfully.

>> --a dear buddy, I really love it. And you should know, it was his birthday this Sunday, so we took him out to dinner too, so we're having fun too. [Applause]

>> I'm 47 [laughter]. All right, welcome back. It's good to see you all here. How many people, just a quick show of hands, were not here on Monday? Wow! Who was here on Monday? Okay. Why don't you guys go outside? [Laughter] Okay, 'cause I'm going to do some of the same stuff. It will be a little different but it's going to be somewhat similar on Monday afternoon for those who could come in early. Alan and I were here with Bernie Hogan and we did a much longer hands-on demo and workshopy kind of thing, a tutorial on this thing, NodeXL. And we're going to move a little bit more quickly, you know, in a higher rate, but the main theme here is that we have designed a tool for the non-algorithmic, so another show of hands, please. How many coders are in the room? Okay. You guys can leave--[laughter]. All right, well, my argument is that while the coders may be a little roll their own, they often and seem to, they--

>> I already suggest that we smoke it. [Laughter]

>> If you can figure out how. [Laughter] Just tell me how it goes. That what--

>> You've heard of smoking code. [Laughter]

>> That what we're trying to do is develop a tool for the non-programmers, and maybe for the programmers who have to talk to non-programmers and I would argue that the distribution of programmer to non-programmer on Earth runs about 99.9 percent non-coder to about 0.01 percent coder. And so coders tend to come together homo-folly, remember, homo-folly. [Laughter] And so they often don't realize how non-coding the rest of us are. And I have often been told, well, why don't you just learn to code? And we've had some discussions about that. And one of the answers is that I'm really bad at it, and I don't know about you but you may not be all that good at it either. I had a son who is a musician. He has been spending seven years learning how to play the saxophone. I've watched him. And so I know that you blow in one part of this thing, you run your fingers up and down on the buttons. That doesn't make me a saxophone player. And talking to a lot of Python developers and knowing that you write words on the line and hit return doesn't make me a software developer either. But I'm a social scientist and I'm a geeky social scientist. I'm really interested in what is happening online. I call this stuff computer-mediated collective action. I am interested in collective action. Collective action is the core concern of the social sciences. I would argue anyway that sociology certainly is mostly concerned with what happened when two or more people get together and do something that would either be very difficult or almost impossible for people to do on their

own. And so it takes two to tango and it takes three or four people to move a couch. It takes groups of people to accomplish things. And when we look at the internet, what we see is an enormous domain of collective action. We just heard all about Wikipedia and what is Wikipedia made out of. It's made out of people. It's made out of not just one person. It's not that the person, some, you know, Mr. Wiki wrote the Wikipedia. It's about, well, I think the answer is it's about 100,000 active contributors and about a million and a half less active more casual contributors. And so the internet seems to be about harnessing the swarm, bringing collective action into focus and getting over all the obstacles to collective action and there are many obstacles to collective action. In fact, the study of collective action often talks about the tragedy of collective action, that there is a failure to accomplish successful collective action. And the studies of failure usually come down to one thing. Why do some groups succeed and why do some groups fail and the answer is information about other people's choices. And so the internet has made sharing information about other people's behavior so much easier that we are now in a situation where it is possible to coordinate your behavior with others, other people who aren't in your time zone, who may or may not actually speak your language. It allows for the aggregation of a critical mass of a labor. And when that happens, Wikipedias are born, operating systems are created, browsers are written, message ports are populated with useful questions and answers, lots of really nifty resources. We talked about Flickr earlier today. One of my favorite things, perhaps you will help me by collectively adding your name to your photo when you go to my photo screen on Flickr. So how do these things work? And how do we study them? How do we study them in a, actually empirical, quantifiable, objective, robust way? For that matter, even if those weren't our goals, how do we cover these topics with any sense of scale? I don't know about you but I started in this business by reading messages. And there is a point when you realize you will never read enough messages. It's not possible, or maybe you have mastered one domain. I wonder is--Nancy, are you here? Nancy read her messages about soap operas. So many so that she really understood the nature of that community. But I ask you, what if she had to then move to the Battlestar Galactica community? [Laughter] Have to earn those stripes again. What if you then move to some other kind of fan, you know, media topic? What happens when you move to the people who are talking about pharmaceuticals? You need a tool. Now, the challenge I think is that the tools that are out there were not designed for you. They're really designed for software developers. And I'm hoping that I'm going to--anytime now, advance, maybe not. There we go, okay. So the tools that are out there are really for software developers. Even the graphical user interface versions of these tools are really for very, very advanced users. There is a tool and they're fine tools. There's Easy Inet. There is Gephi, G-E-P-H-I, it's actually developed by a group of francophones, so it's Gephi, but I say Gephi. And there's Pliek, there's Gephi, there's Easy Inet, there are several. There is in fact a Wikipedia page managed, well, founded by Jana Diesner who will be here on Friday to speak to you, and she's got 40 or 50 tools listed. So what do we need another one for? Well, Gephi has a motto, and its motto is that it is the Photoshop of graphs. We also have a motto. And our motto is that we are MS paintographs. [Laughter] And our goal is that we're the tool that you're actually going to use, not the one that you're thinking about learning. And so we may or may not have achieved that goal but it is our intent to be the 80 percent solution for graph analysis. Does that mean that we will have the feature that we need to do everything you're going to do in very sophisticated graph analysis? There's a good chance that it isn't the case. We're trying to be the 80 percent case tool. However, we do think that if you run off the edge of the NodeXL real estate, when you then take the next step into the next tool, you will have a much more motivated reason for being there, rather than trying to climb a very steep learning curve. So, I am trying to do the social media thing myself. I'm working as an independent consultant and also helping to lead the Social Media Research Foundation. We are a 501 c3 not for profit California Corporation. Our goal is to convert money into software. I know a lot of other people would like to turn software into money but we have a different goal. And we will also then make that software free and open. And so that's free as in speech, free as in beer, it's free. The group of people that we've gathered, I have to say is a motivation in and of itself. There's some remarkably talented people who for one reason or another have been corrupted and been forced to engage in this project. And they come from--importantly, not just a wide variety of institutions and time zones but they also come from a wide variety of disciplines. It's not just

computer scientists. I love computer scientists, I really do, but it's not just computer scientists. So by bringing social sciences and user interface designers and computer scientists together, what we've done is created a tool that meets the needs of users rather than developers of that tool. So, we have a goal. Money converts to software and we want to become not unlike Firefox, free and open tool to browse a really important data structure. But we're different from Firefox because Firefox is not a web browser and NodeXL is. How could that be? Well, Firefox is not a web browser. Firefox is a page browser. You'll never see webs in Firefox. You see pages. You see a node. You see the vertex. It renders the content of a vertex, but it never shows you the associations between multiple pages. It can't. It's a page browser. NodeXL is a web browser. Okay, I now give up, because we're not going to win that battle. Obviously, yes, Firefox is a web browser and no, we are not one. Maybe we are a net browser. I'll take that. So what we want to do is be like Firefox but for some other data structure, not HTML, and this other data structure is known as GraphML. And if you Google up GraphML, you'll find at graphml.graphdrawing.org that there is this really nice, simple, clean, open, basic XML schema that is now widely being used for a variety of tools to interoperate, and what is it? It's the file format for storing a collection of connections, a graph, a bunch of things and the relationships amongst them. And our argument is that while the world clearly understands the value of HTML, it will soon come to recognize the value of GraphML, that we are all on the path to network consciousness, an awareness that none of us are really isolates. We're not individuals. As a social scientist, I'm here to say, we're not individuals. We're actually deeply interdependent. If we were individuals, we'd all speak our own private language. And that's actually considered to be a pathology, so don't do that. So, we want to make it so that in the same way the Firefox stepped up the game from, you may remember, links, maybe you don't, about the text version of a web browser before Mosaic there was links. I remember, you know, the sign of my great capacity for technology prognostication. I remember seeing links and thinking--I saw a URL for the very first time and I said, "That'll never work. [Laughter] Forget about it. It's not going to happen." And although it happened and so our goal is now to build this tool that lets you as non-coders and as coders to communicate to non-coders. How do I get a graph? Where do we get it from? Where do I put it in once I get it? What do I do with it once I have it? How do I analyze it? Where do I--what those network analysis even entail? And then what do I do to visualize it? And having visualized it, how do I find an insight into it and how do I publish that insight? And that's the tool chain, the data chain, the flow of data and operations on that data that I believe anyone who is engaging in network analysis has to perform. And there's a challenge. And the challenge is that pretty much any second year computer science graduate student can do all those operations in Python or Ruby or something. And if I say to the computer--I sense that they were over on this side, so if I say to the computer scientist, you know, can you use Python to open a socket to communicate to an API to do a query, to grab out a bunch of rows of data, maybe a few hundred million rows of data and then put it into a scalable, let's say, MongoDB and maybe you're going to build a Dub Cluster. If you're losing me, that's my point [laughter] but these guys know. And then you're going to take queries out of that database, run them through other tools like igraph networkX or SNA, then take that data, populate the database with the results and then take those results and then pass them to some other process to visualize them. You don't have to go to school to learn all of that. [Laughter] But wait a second, you are going to school to learn something else and so there's an opportunity crossing. You really can't do both. Some people can and I--it's a fine distinction. If you can, you should. But if you don't feel like that's the path that you can walk down, I am here to say I don't think I can go there. I wrote my dissertation work in Pearl and then I met real developers, and they looked at my code and went. And you know, maybe it was a good experience but I think maybe the best experience is to learn enough, to know how to talk to a talented software developer and then find one in partner. Or we have a tool for you. So we are trying to build this tool and we're trying to fulfill the foundations motto, open tools, open data, open scholarship. And we think that these are three goals that are really important for us as a culture as we wholesale, move human societies into cyberspace. If you were an ethnographer and you wanted to understand the civic engagement, you might go to the public square and watch people talk to each other and it was a physical space and you could occupy it with a notepad and a pen, you could do good work. But the civic square is now in cyberspace. So, no matter how many people actually were in the Tahrir Square, and there are many

Tahrir Squares, so even if we added up all the people who ever were in all of the Tahrir Squares, more people talked about those people in Tahrir Square in the cyber-Tahrirs of the world. What is the magnitude for it? Millions of us talked about Tahrir, hundreds of thousands of people were in those Tahrirs. And so, how do we now as social scientists, as people interested in what's happening to human cultures in an age of technology, what do we do? How do we get our data and do some meaningful analysis of it? And so open tools, open data, open scholarship, our goal then is first, to open tools and we have built a tool. And NodeXL is that tool, it's the network overview, discovery and exploration add-in for XL, NodeXL. Someone once told us that acronyms cannot be trademark and we can be sued if we did that. We believed them, so, we are NodeXL. We were something else before and we got sued, so we're not that anymore. [Laughter] We're really sorry. So NodeXL is this tool, you get it, you plug it into XL, and yes there are caveats. It doesn't run on a Mac. It's caveat, I have a Mac, but it runs on my Mac, and that's because I'm running Windows on the Mac, but, you know, you can use a virtual machine. There aren't problems, and we'd love to run on a Mac and to the software developers in the room. [Laughter] We should talk. And of course, we're really thinking that we're just going to skip the Mac and we're going straight to the web, and so we should talk, because there's a lot of overhead in getting to the web and leaving our good foundation XL behind. So, we built this tool, we believe it is a friendly end user tool. It allows you to do many of the very sophisticated kinds of analysis that you've seen this morning where there were networks and metrics and all sorts of stuff going on and maybe you felt a little left out. Maybe you looked at not that side of the room, that side of the room, maybe you felt like, "Wow, that's really cool but I can't play in that space." And so we want to say to you, you know, there was a time where the brochure, the newsletter, the short novel, the report, it also was something that was probably not within your school skill set. It was also something that a professional was necessary. You had to go to an illustrator, you had to go to a printer and printers didn't do just, you know, here is the car key, go to the Xerox, print what you like. They actually did what you do now in front of Illustrator Word or any of the other tools that you're going to use to put together some text. And so, what happened to illustrators? They moved to higher ground. They do the fancier stuff now. So can't software follow the same path for networks? I would argue it can, that we are the desktop publishing of networks, that we are the tool that pretty much anybody can use, and that gives the 80 percent solution to a very large population. There are certainly things that you can't do, but maybe they aren't the things that you need to do. So, we've also been working on this concept of open data. Many people, as we--they get into the space of social media, research discover, getting the raw material of that research is not easy. Twitter has things called rate limits and API call budgets. Facebook has permission requirements. All of them require some kind of authentication and many of them will limit the amount of information they will give you. Worse, services like Twitter are happy to give large amounts of information to anybody but you, so long as they have a big fat checkbook. And so if you have 60,000 dollars, you could probably get your data. If you have 300,000 dollars you could get your data. But if you have zero dollars, you might have a problem. Some members of the NodeXL community have been grandfathered into this data access tier where they can get more data and so we have built a service, we call it the NodeXL Graph Gallery, call it Flickr for graphs and email. And in that gallery, there are now a thousand graphs. And as the user community for NodeXL grows, it is our hope and aspiration that there will be a million graphs in the NodeXL Graph Gallery. And that this will allow you to pretty much ask almost any question, any topic, and you might find that there's a graph in there already. And if there isn't, I do encourage you to request the graph. Go to our message boards, ask if someone could get that graph and upload it for you. And so we would then have addressed two out of the three problems, open tools, don't have a tool, now you have the tool, no data, now you have data. And so our next goal is open scholarship and of course, we're doing events like this and others because we think it's really important to teach people not only the--just nuts and bolts of using any tool, but by learning those features of that tool, really learn the larger set of network concepts to help you, as I'll say later, think link. Think in the form of a network. And when you do, you'll walk out into the world and you'll never see it the same way again. You will always be thinking, "What are the patterns of connections amongst these people? Who knows who? How do they know each other? Who's the bridge? Who's the cutpoint? Who's the hub?" And we're going to look for these patterns, and I think once you start thinking in

a network way, you will have a very hard time avoiding, seeing those patterns. So in addition to the training, we're also publishing and we've been publishing, I think, in two separate tracks. One track is a CSS, UI, UX kind of track where we are attempting to build tools and features that improve the way that people handle networks. Not only the tools to get the data and calculate metrics about the data, we actually have outsourced that to the very talented Dr. Jure Leskovec whose SNAP library is our calculation engine. In some ways, you could argue that NodeXL is nothing more than a GUI over SNAP, which is otherwise a C plus-plus library. But we think we're a little bit more than just a GUI over it because we've also been working on innovations in the way that we represent the visualization of the graph. So we have some goals in the project. One of those goals is to address the number of lines of code USR user will have to master in order to use the tool, and that goal is zero. We've reached that goal. We have another goal and that is the number of human seconds necessary to process a graph before it crosses a threshold that we call good enough, and that goal is zero. Zero seconds of human involvement. Once you'd kicked off a graph, we believe that you can get a graph that is good enough in zero human seconds of investment. Now, does that mean that your work is, therefore, done? We have a joke that there's going to be a new button in NodeXL called write paper. There's another button after that called publish paper and a button right next to that called get tenure. [Laughter] But no. We believe that once we've given you these zero human seconds of involvement work product, then that's where your work begins. This is where the storytelling begins. This is where the insight begins because the machine can count things, but it doesn't know how to tell a story. If you watched the Hollywood movie written recently, you know that that's true. They don't know how to tell stories. I believe that technically, all of the movies actually must have been written by machines. So an example of this, so track one actually tried to build better software to make it easier to do this. Track two, use the software, and I think track two is really the social science track. And that is, apply it, we built a microscope, what can you see? We have a telescope, what can you see? Now, once you say the word "telescope," you invite comparisons to Galileo. We are not Galileo, but we are definitely seeing things that have not been seen before. Whether or not these are in fact the moons of Jupiter, I don't know, but we are seeing patterns in the data that were very difficult to get to unless you were a developer. And even then, very few developers have used their toolkit to build data at the scale that we have so that we have now thousands. In fact, I know half about a quarter of a million slices of graph from a variety of social media data sources. And so we're learning things by looking in at microscope or into that telescope, and things like the way that a lot of political discussion is pulverized. Now, it's probably not surprising to you, at least for the Americans. Maybe the Canadians who watch American TV, they know, were a little polarized in America, just a little bit. And, you know, I think we intuit this but can we measure this? And so the image you see here is a discussion in Twitter amongst people who all said the word GOP, Grand Old Party, the Republicans, and what do you see? I see two clusters with very sparse connection between them. We see polarization. We see a way of actually quantifying polarization as essentially graphs that had multiple clusters that had very little connection to each other versus sub-communities which are essentially clustered graphs that have strong connections to each other but are still sub-communities. And so, our goal is to both contribute to--let's build a better tool, and then let's apply that tool and do better social science. Let's understand social media. And by that, I mean, how many different kinds of social media are there? And I don't mean Twitter, Facebook, Flickr, but within Twitter, how many kinds of hashtags are there? How many kinds of users are there? And can we taxonomize them? Can we come up with categories? Now, when everybody is using the tool, we do believe that we will accomplish a goal not unlike this project, this is the Allen Very Large Telescope Array, Paul Allen, co-founder of Microsoft, has spent his money in interesting ways. One of which is to build the world's largest array of radio telescopes. It's in the New Mexico desert and it--rather than trying to build a really, really, really big dish, which structural engineers have said basically, "We're at the limit and I can build a dish any bigger than the one they have in Puerto Rico," they're not getting any bigger. And the answer to that is "Fine. I will take 1,400 small ones" and they arranged them in this big cruciform out in the desert and that thing has the resolution power of essentially a multi mile-wide dish. Our inspiration then is to encourage people to also set up NodeXL and NodeXL data collectors and collect data in all of the regions of the world that they are doing research on the topics that motivate you, and then

to share that data. And in doing so, create a triangulated map of social media to actually let us see all the different kinds of social media there are. So, all right, social media has already, in some ways, peaked, it was on the cover of Time magazine. I take some issue with this issue, if you will. Partially, the issue for me is that English is ambiguous in this pronoun and that it could be in the individual, right, it was individual and plural. Now, I was born in New York and raised in Philadelphia, we speak a dialect. And you go to the--a shout out to the Drexel people "Go Dragons!" And in Philadelphia, when you say you, you mean one and when you mean to say more than one, you mean that would be "youse" [laughter]. Now, if Time had actually put on the cover "yous," then I'd had no problem with it because what it suggests to me is that when you say you, what you mean is the individual, that the internet was really about people and it is, but not individual people. It's about youse or if you're Southern that would be, I guess, "y'all." It's incorrect, I was corrected, I was in the South and I was told that "y'all" is of course singular. It's all y'all. All y'all is plural. [Laughter] So all y'all is what I mean about the phenomena, it's not about individual. It is about prominent people within a community but it's not communities, it's about collective action. It's about more than one person. It is true that often there is a very small group of the maximal contributors and we look at that, you know, you'll see millions of editors in Wikipedia but you really see on the low order of tens of thousands who are the core of Wikipedia. That's true. But even those very few editors are not alone and they form a community. And so, it is about the collective that we are really most interested. So social media is inherently, well, social. It's about the exchange of objects between people, turned out that this network thing was not about the efficient use of super computing resources, it was about exchanging photos and music, and text messages, and URLs. And essentially, every time you do one of those things, you form a connection. And in aggregate, those are collections of connections and collections of connections are networks. And, of course, the other big issue for us and opportunity is that humans have been social prior to social networking services. They were actually social before electricity. So humans have been doing this social thing for a very, very long time. But throughout all of human history, being social has been an ephemeral phenomena. But with few exceptions, most of the time when you talk to somebody and walk away, the conversation has evaporated and if no one saw it happen, it might as well had not happened except for the memories of the people who participated and those memories are flawed and you can't quite recall who you talked to yesterday. You know, one of the greats in social science studies is to simply ask people what they had for breakfast yesterday and most people can't tell you. Maybe breakfast is more stable but lunch, they can't remember. We don't remember these things. And so, on the net, the great opportunity is that silicon, even in its most raw form, is really good at storing the foot prints. And the difference between silicon and, well, unprocessed and processed silicon is in eight hours, the tide is going to come in and wash this all away. With digital footprints, there is no tide. Nothing is going to erase the digital footprint at this point. I thought it was interesting that it was now newsworthy that Facebook has now committed to the ability of users to actually delete a photograph. That's news. You can delete something. And they promised they'll get it done in 45 days or less. [Laughter] And I'm not too sure that they can even guarantee it. When you say delete, what you meant was don't show it to me anymore, make it less findable, but you know there's a back up tape. You know that there's another drive array somewhere with these bits. And so bits may be more immortal than humans. I think there's a kind of law like regularity. The only bits you cannot find or the ones you really need right now, the only bits you cannot get rid of are the ones that are most embarrassing to you. So patterns are left behind and these patterns are the patterns of ties and there are lots of kinds of ties. And these ties are all of the internet verbs. These are the things you do when you're clicking, when you're dragging and dropping, when you're hitting that carriage return, what have you done? You liked, you linked, you replied, you reviewed, I can't read, let's say the list, like, link, reply, rate, review, favorite, friend, follow, forward, edit, tag, comment, check in, I think send is missing. Maybe we should add send, a direct line. There are that many others but every one of these is an edge. It's a link. It's a tie. It's a connection. And in aggregate, they form networks. And so our goal here is to encourage you to think link. To think about, look, there all these entities in the world and they have relationships with each other and not all of those relationships are the same. They come in flavors. And in fact, not all of the relationships even are of this type which is unidirectional that, you know, it goes from A to B. We talked a little bit yesterday about how it

is the case that some relationships are bidirectional and others are unidirectional. We call them directed or undirected. So directed edges are often not reciprocated. So when you do something with someone, that doesn't mean that they're going to do it with you. I followed you on Twitter, it doesn't mean that you follow me. Other relationships are reciprocated. If I'm you're friend on Facebook, you must be my friend. It can't be any other way. So some relationships are balanced, others are not. I'd like to note that, you know, some relationships, you give the ride to somebody to the airport but they tend not to be the person to give you a ride to the airport. You lend them money, they tend not to be the person who lends you money. But some relationships are inherently bidirectional and we talked yesterday about marriage, right? If you find yourself married to somebody who's not married to you. [Laughter] It's probably good to find counseling. But other relationships are that way too. If I live near you, you cannot live far from me, right? Unless there was one-way streets [inaudible] but it's really not possible. And so relationships come in flavors, there's unidirectional, bidirectional, they come in types, there's friend, follow, favorite, all of those kinds of things. And this is a good thing for us because we now live in a world of--and this of course is a very abbreviated, was there're all of these different kinds of social media. And theoretically, the challenge is, how do we make sense of all the diversity that's out there on the internet? And there is good news. Despite the fact that blogs are not Wikis and Wikis are not Facebook, and Facebook are not Flickr, and Flickr is not YouTube, although it is sort of like YouTube only the pictures don't move, but for the most part, these different services. They all have one thing in common and that is that they all encode networks. Your email has a reply graph, your Facebook has a friendship graph and a comment graph and a like graph. The universal data structure that underpins the entire social web is a network. And if you are in a piece of software that calls itself social software or social media, and there is no way for people to connect the people, it's not social media. So I've heard many marketers talk about "Oh, you get our app and it's social media 'cause you can look at our brochure" and that's not social media. That's digital media but it's not social media. Social media involves people connecting to people. All be it, it is possible that they connect indirectly. So I added a Wiki page, you added a Wiki page. It doesn't mean that we added each other but we had a shared intermediate connection. And so when you think about network theory, and I hope you do, as you learn network theory, it is a challenge and that there are many new concepts and these concepts tend to be interrelated, that the meaning of one is actually related to the meaning of another and so until you've learned them all, you haven't really learned any. But I will offer you the shortcut and that is that network theory has a lot to do with real estate. You may recall the three most important things in real estate. And that's good financing, the foundation, and available schools, no. That would be location, location, location. And network theory doesn't quite have location, location, location 'cause location suggests that there is a north, there is a south, and there isn't in networks. But there is in networks, is position, position, position. You want the prior image?

>> No.

>> Okay. So, what we're trying to do then is make it possible for me to achieve a goal that I started to have in my mind in 1991. I bought this off the news stand, I'm sure you all did too. And you know, I buy Scientific American for the articles but I like the pictures too. And what I really found compelling was just the cover. I think this is Alvin [inaudible] from Australian National University. It's 1991 and this is a network visualization at that time. I was looking at BITNET. I was looking at BBSs. I was looking at message boards, and I looked at this image and I said, "I would really like peppermint sticks" but no, I want a picture like this. That's what I thought. I thought I would really like to see my data in this way. And, you know, any day now, we're going to get there. But it has been what? That's--oh my goodness. [Laughter] Is it really? It's 20 something years since then. Yeah, okay, so it's 20 something years I'm working on this. Proof that obsessive-compulsive behavior is really a good career developer. Okay, so our goal is to get to this, to understand these phenomena and in part because now, you know, doing research the old way is now too hard. This is the work of William H. Whyte whose book "City: Rediscovering the Center" is a great book, also related to his shorter work, The Social Life of Small Urban Spaces. I strongly recommend reading these books. They're

wonderful. And in this image, what we're seeing is data that was collected about where do people stop to talk to each other? Where does an edge form? Only it's a street corner in New York City in Manhattan. It's where Saks Fifth Avenue is. And it's worth noting that on these city streets, 10, 000 New Yorkers will walk down the streets. Now if I told you that 10, 000 New Yorkers are behind you, and that you need to now stop and talk to somebody that you've just encountered on the street, what would you do? You might think, "Oh, I will get out of their way." The reality is, you will stand stuck still right where you meet that person. You will not move and you will, instead of causing a blockage kind of human cholesterol, you will actually be like a rock in the stream and they will just flow around you. You would imagine that if, you know, there are 10, 000 New Yorkers behind you. So actually, looking at what people do matters 'cause they do things that you wouldn't expect, but the data collection mechanism here, well, was somewhat organic. It was graduate students, and graduate students are wonderful people but they have short duty cycles and they have this unfortunate habit of graduating, if all goes well. And so, what we can now do is actually collect information about this off of your cellphone. And in fact, companies like Sense Networks are doing exactly this. We are building a worldwide model of who is where and who is interacting with who at the petabyte scale and people like Sandy Pamplin and somebody is working with Sandy here, there you are, you know, these are guys are building these kinds of models and they're not using graduate students at the data collection interface. They're using them at the data analysis level. And so we are at the cusp of a real change in the nature of social science research. It used to be that a data set of 30 was large. It used to be that an observation window of a few hours was long. It used to be that, you know, studying a region was a lot. And now we are going to be studying the mobile connected user-base of earth, which I believe, is now about to crust 2 billion units. And so we're going to have a real time model of where we all are. After all, who here, and I think I know how to ask this question, who here does not have a phone on them? Where did you leave it sir?

>> I don't have a phone.

>> This is our outlier. [Laughter & Inaudible Remark] But, you know, a few weeks ago, my wife left the house said, you know, "Have a nice day," went off to work, 8 minutes later she's back, why?

>> Left her phone at home.

>> She forgot her phone. How many other things would you actually turn around and come back to the house for? Your wallet, your keys, no, you're in the car, you couldn't have forgotten the keys. So, you know, it's California, we're all driving. So you know, this device has become essentially--and there was an article recently, "Don't Call It a Phone" was the article. There's a ranked ordered list of which applications are used most often. Do you know where a phone is on that list? Fifth. And if you've ever tried to actually have a conversation on an iPhone, you know why. [Laughter] It's not a phone, it's a web browser, yes, phone, no. And so these things--the proposal was that's not a phone, that's a tracker. We should call them trackers. There is that. This thing knows where I am, it knows who I'm with. It knows when I landed. It knows where I take off from. It knows, you know, pretty much knows my entire pattern of life. In fact, I believe it was your group who--was it? You'll clarify this. Research team found that you don't need to have somebody's social security number to have a grid for them, a unique identifier for them. All you really need to know is the GPS coordinates of their phone at 2 a.m. and 2 p.m. 'cause there are rarely two people on earth who have those same coordinates. Maybe if you work with your spouse, but that's relatively rare and then we only have one false positive. So it's hard to be in that bed with more that two people. I don't know. [Laughter] Maybe you have like, I don't know. I mean, now all of them are going to have phones, maybe if there are artists. I don't know but--[laughter] so we're now in a phase of social science research where we've got the data or at least the government has the data. At least the phone company has the data. At least these guys have the data. Ask them for graphs. The issue is, how will we manage all this data. And it's true that these are--this is data about our connections and our connections come in flavors and we know that for example that there are

people who have strong ties and Barry Wellman has really done an enormous amount of work identifying just how many strong ties do you have and did the internet change this number and I believe the answer is no. The number of strong ties in our networks is not growing. It's about same. And Barry very poetically says that the number of strong ties you have is about the same number of people as you have chairs that could be placed around a dinner table on a holiday night and you put the leaf in. That's how many people you know really well. The people that you will send a birthday card to, call on a weekly basis, you know, when they're ill, you go and visit them or you call them, these are your strong ties. And the internet has been criticized because it is now seen as the land of weak ties. And yet, you know what the punch line is, it's true that not all 750 people that I have on my Facebook list are really my friends. And I appreciate even, you know, two days ago was my birthday and I was really pleased, I got all these happy birthday greetings, but I don't really know all those people. And it was nice and I said thank you and I think you should, like and comment and say thank you to every single birthday greeting. It's just a nice thing to do but all of those ties we're not really strong ties. In fact, the issue of these ties are very ambiguous. In fact, technologists, and we're going to that side of the room now, technologists have taken the most analog thing on earth, human relationships. And they've turned it into a binary. Oh yeah, you know. [Laughter] Are you my friend? Yes or no. But it could be lots of things. I like you, I kind of like you, I really like you, I feel socially obligated to link you, I'm into you, I know you, you know, it could be that I, you know, we met at a conference, it seemed like the thing to do, you know. There are lots of meanings and nuances. And so, are you my friend? Yes, no doesn't quite cut it. So what we do know though is that while there is this preponderance of weak ties, there is some strength in all those weak ties. And so we'll invoke Mark Granovetter, seminal article, "The Strength of Weak Ties" and he asked the question, where do you get a job? Hypothesis, you get a job from the people who know you best, your strong ties. Answer: No, incorrect, you get job leads from people who don't know you all that well. Why would they keep you job leads? Why would they be the source of information? They don't know you that well. Well, because they don't know you that well, they know a lot of people and things that you don't know. They have news. They are from a distant part of the network. This is, by the way, why we go to conferences. This is what conferences are for, to sustain weak ties and why do you go? Because that's where your next job is coming from. It's where your next collaboration that's going to get funded by the NSF because you need that diversity. It's why we go to conferences. It's why we go to events to meet new people because--or reestablish connections with people we've met before who are acquaintances 'cause they are the source of a great deal of value. Here, Nash [phonetic] is pointing to an article from, I believe it's 1933, it's part of the--really the first public awareness of network analysis. I'm sorry?

>> Great. It showed likes and dislikes at the end of the result.

>> That's right.

>> Will I expect that?

>> Yeah, well, [laughter] they call them favorites but yeah, yes. [Laughter] So Jacob Marino comes from Romania in the '30s and he goes to NYU. He's in Manhattan. And before he has a psychotic break with reality, he founds a whole new discipline of social science which is essentially the social science application of topology to human relationships, and he uses the name psychological geography. And he actually starts mapping out. Something that I just saw on the web and maybe Bernie knows what it is, it's the movie network database not the--

>> Movie gallery.

>> Movie gallery, so this is which characters and which movie actually will share the scene together. And so you have a tie when you're on the screen together. We have an example in the book, a similar example of

which characters interact with which characters in Les Miserables, the same kind of idea. Psychological geography then gets re-titled Sociometry. Sociometry then pretty much dies largely because Marino begins to tell people that he is in personal communication with the Almighty. And this was really bad for having graduate students 'cause as you may know, most graduates think that their adviser is the Almighty [laughter]. And that therefore, it is--it's a category violation. How could there be another? And so, this kind of dies out for a while but these original graphs, these hand drawn graphs, really are the origin point of social network analysis. This was the map of American Football Team and who liked or disliked other people on the team. This was an early map of people in the Western Electric Wiring room, a manufacturing plant outside of Chicago where in the old days, humans made phone companies switch devices, physical, mechanical relay devices. And these were who worked with who, who helped who in that environment. And so, these very prude drawings are really the origin of network visualization and we're now on a path where we want to make network visualization, the visualization of social media data, a lot easier. And I would argue that we are still in 1959. It was only, what? 40, 52 years ago that we took our first ever photograph of the planet we live on. So, that was 1960. Could you imagine if you are a meteorologist in 1959? You have never seen you planet. You have never seen a gestalt image of the entire phenomena. That is what internet research is today. And unfortunately, next year is 1960, next year is probably like 1923, we are long way away from seeing a holistic view of the whole internet, of the whole social web. And so, what I'm going to suggest, our partial fragmentary steps towards that goal. So here, for example, is a map, this was made by NodeXL and it's a map of relationships in Twitter. And this is one of--I hope to show you several of these maps because it's not enough to see one, you really have to see multiples in order to see in what way are they the same or different, what are the categories or the varieties of them. And everybody's got WiFi now, right? Yesterday we didn't have WiFi. If you go to the website at NodeXLGraphGallery, NodeXLGraphGallery, all one word, dot org, you will see what Edward Tufte referred to as small multiples. You'll see a whole bunch of liquor-like images of graphs. And it's my hope that simply, well, well gosh, let's just go look at it, right? [Laughter] Why should I talk about what data I will show you or what page?

[Pause]

And here is that page, right. So if you go to NodeXLGraphGallery.org, you'll see pictures of you. How did you get in there? Well, you tweeted, didn't you? So these are all maps of various social media networks, still that can be social to you, there's networks. They just tend to be social media networks. It's what's interesting a lot now. And I'll know that you as a NodeXL user, you can contribute to this in the same way that pretty much anybody can upload a picture to Flickr, you can upload a NodeXL graph to the graph gallery. And it's important to us to get lots of these graphs in order to illustrate, am I running late?

>> You're on 38 minutes.

>> Oh yeah, thank you. Thank you very much. The goal here is to really say, "Look, that one is kind of like that one but this one is more like that one." And overtime, I believe that we will have a taxonomy of social media. We will actually be able to say, "How many different kinds of hashtags are there?" And not just hashtags, really, query terms, anything that halls out a collection of social media content, how is it structured. And we're going to see that there are different patterns in these maps. For example, here is a--as of this afternoon, that would be us. This is SSW12, this is 52 people who's recent tweets contained the SSW12, created at 1358 UTC so subtract 5 from that so that was what, what is that? That's like 10 o'clock this morning, something like that? And what this caption will tell you is what am I looking at. What are the edges? Green edges are follows and blue edges are mentions or replies. And the edge thickness is an indication of how many messages were exchanged between those two people. And you'll see that we have been clustered and that some of us are in different clusters, and the size of the image, PJ doesn't like the sizing but we're going to have a good methodological dispute later. I make the size to indicate follower account because I believe that the location in the graph is conveying between this. But I think a good

argument can be made that you'd actually want them to be sized by between this not by follower account. If you wanted to focus on only the things that are present in the graph rather than external forces coming into the graph, actually, I kind of like having external data coming into the graph because this allows me to see that, for example, not so much in this graph but in some graphs, you'll find a very large image indicating somebody with many followers. And yet they are very peripheral in the graph. And I like to refer to these people as visiting dignitaries. If you are so lucky, it has to have Oprah to read about your topic, that's great, and no doubt at all it will attract an enormous amount of attention. But that doesn't mean that it will represent the center of your graph. She probably isn't. And so, people with a lot of followers are not necessarily the people with the influence within a bounded community, within some space of discussion. And so we can then use the metric tools here. Let's see if I can make that scroll. There we go. And we provide a whole bunch of metadata about the graph. We're going to tell you things like, here's the number of vertices, there's 52 people on this graph. How many connections are there? What's the density of the graph? Graph density is very high, it's 0.19. That's really high. We're also going to tell you who the top people are. That would be the Webshop account, Ben, Jenna, Jenna, there you are. You are among the most central people in our graph. And then that would be me, then PJ, Blessy [inaudible], there you are. In the top ten, Mediamum [phonetic], hello, there you are, and Hussein, hello sir, Jen Preece. That's Jenny. You made to top ten. And Jenny Corn [phonetic], there you are. Yeah, so guys have more than the others, not tweeted a lot. Not had more followers than the other, but connected to each other more. So these are our connectors. There are hubs. And so we can say something about the nature of our internal community here and I've been making these maps before, during, I'll even do them after, and of course we have theoretical reasons, imagine that we will see densification of the graph over time, in part because edges don't die. And so any edge that's added will always be there. Unless Twitter stops giving us data about last week in which case, some edges will disappear after a while. So we will also tell you who got replied to a lot, but not just in the entire graph, but in each of those clusters. So, you might not be the king or the mayor of the whole hashtag but you may be the mayor or the most central person in one of the subgroups. We also--we'll analyze the tweets. So these are the top tweets, our schedule, the video, the graph gallery, some tweets, an image, more of the Webshop links, so these are all of the links that are being passed around in those tweets and they're ranked by frequency I've mentioned. And then we break them out by groups so that you could see the different subgroups are actually tweeting on different topics, pointing to different resources. We also then tell you about the top hashtags. So these are the top hashtags in the entire graph, SSW12, not surprising. And then what do we have? Tu1 that was the Tuesday 1 session. It looks to me like the--well, it depends on the time of the graph, so I'll have to make another graph, but maybe the Tuesday 6 is not getting tweeted about, tweet quickly. Help yourself. [Laughter] HCIL, [inaudible] survival, ASA, NodeXL, all of these are the topics of discussion right now, the top topics of your discussion but not exclusively. In each of the subgroups, you may be talking about slightly different things. And in fact, we have applied these kinds of techniques to the analysis of politically-charged to polarized topics. And we've created some interesting findings as a result. I shouldn't steal the thunder of our final speaker on Friday, Professor Itai Himelboim from the University of Georgia Department of Communications and Journalism. But working with Itai and using NodeXL, we've been studying highly polarized topics in Twitter. Topics like contraception or GOP, or Medicare, there are lots of them and we get the two big groups, kind of thing, and then we ask a question. What is the rate of overlap between the use of hashtags, words, and URLs? We're both talking about the same thing, do we point out into the world in the same way? So what percentage of overlap do you expect there to be in some of these graphs? That's a--this is a question of how many URLs that are mentioned in one group or one cluster are also going to be mentioned in another cluster? Well, in a highly polarized graph, the answer is none, none. So that means that no one in, let's say the liberal cluster pointed at FOX News and nobody in the conservative cluster pointed at npr.org. None. And this for me is actually a fairly concerning finding, that we are talking about the same things and we don't even talk about it in the same way, and that we don't have any common referent for that topic, is a concern. And so, being able to analyze what is being talked about is a very useful addition to our analysis. And then what happens if it's not a hashtag, it's not a URL, and it's not in that name, then it's words and then we were giving you word pairs.

What are the words that have gone and some are social, makes sense, right? It does make sense that those things would go together. But if you are dipping into a domain that you are not familiar with, getting this kind of a report can be very handy, and I'll show you what that might look like in a live version. So I made the SSW map of--right before I started speaking. And so, this is that map and we are now in NodeXL, and so how did I get this data? Look, it's really--it's getting bigger. And TU2, 3, 4, and 5, all got mentioned but not 6 but I wasn't talking at the time so, maybe you'll tweet about TU6. So what I did was to go to the NodeXL menu and you'll see that we have NXL of ribbon and I like ribbons. And in the ribbon, we have a whole bunch of sections, the Data section, the Graph section, the Visual Property section, Analysis option, Show, Hide, and Help. And in the Data Import menu, you can see just some of the sources of social media network data that we provide. Twitter, Flickr, YouTube, Facebook, what else? Flickr, Email, that's personal email or we also have an Add in to talk to exchange servers, all these ways of getting data, you can even download it from the Graph Gallery. You could open it from your own workbooks, maybe you already have data. You have spreadsheet. So long as your spreadsheet has person 1 or person 2, well, not or, and. So basically, if you have a spreadsheet with two columns with name and name, you have a network. Open that spreadsheet. Open the NodeXL template in the same copy of Excel. And then go to our Import from open workbook and you now have the ability to pull data in from your worksheet, pull it in to NodeXL and analyze it. But in this case, I use this, the Twitter search network importer. We have three Twitter importers with the various kinds of networks you can pull out of Twitter. That's a User's Network. You give me a name, I'll tell you who follows them and who they follow and how those people follow each other if they do. There is the List Network. You give me a list of names or you give me the name of a Twitter list. So if you're a Twitter user, you may be familiar with the idea that in Twitter you can create a list of users. If you give me the name of the list of users, I will show you the connections among them, or if you deliver a list up to 10,000 names, we will give you the network of connections among them. But the one that I use was--the one that I use was a map of SSW or SSW 2012 from Twitter from the search network. So what we do is we throw a query, any query that you can use in a Twitter search box, so that means, ands, ors, nots, and minus sings, quotes, all that kind of stuff works. We throw that at Twitter search. Twitter throws back up to 1,500 tweets. Those tweets have authors. We iterate over the author names. We ask Twitter about each author. We say, "Who follows that author and who does that author follow?" And then we ask NodeXL to run through that list and say, "Of all the people who actually tweeted about X, who follows each other?" And what else could we say? Well, we also click the Replies to mentions and tweet edges. A reply is when I say your name at the beginning of a tweet, at your name, blah, blah, blah. That's a reply. A mention is blah, blah, blah at your name. That's a mention of you and we differentiate those edges. So we pull in those edges. We also say, "Give me the data that Twitter has and the information about URLs and we've just added a feature that does what we refer to as the "double unwrap." So, when you now put a URL into Twitter, Twitter replaces that with letter referred to as the tco URLs, T dot co slash some hash. That's what all URLs in Twitter now look like. But if you resolve the tco URL, what do you get? You'll get a bitly URL. [Laughter] So you don't really want the bitly URL do you? So we double unwrap and we take the bitly URL, we also unwrap that. We now actually have the URLs that we're actually mentioning. You saw that just a few seconds ago in that list of top URLs. Many of those URLs had been wrapped into a bitly URL and then wrapped into a tco URL, we double unwrap them. It does take a little bit of time. So I did this and it took a few minutes. Mostly it did is to transit time of going to Twitter and bringing back all these data and then iterating all those names. And of course, you may have other challenges. You may have the challenge that Twitter at some point is going to go, it's enough for you. No more for an hour. And it could very well be that you can make about one of these graphs a day on one copy of NodeXL using Twitter. You can make as many of them as you want using YouTube or for that matter, even Facebook as long as you're getting the right kind of Facebook. You cannot get anybody's Facebook ego-network other than your own. But you certainly can get a lot of the connections on a Facebook fan page very easily. So we support fan pages and ego-networks, we don't yet support groups. It's coming. So, we got a data set and that data set takes the form initially of just an edge list. And here's the edge list. And I will sort this and do a little tweaking here and you'll see that what we have is a group of people, and smilex3md, that would be Alan Neustadl had a mention of Trent M Kays and

that came in the form of a re-tweet of Trent. And you were saying that you're coming to our Webshop and Alan re-twitted you on the 17th of August at 10:18 p.m. universal time code. So GMT zero. We have all of the URLs, that's the URL that points to the tweet, and if there were URLs in the tweet, you'll see that we populate the separate column with the URLs in the tweet. We also do that for the hashtags. We also give you the longitude and latitude of the tweet if it's present. In here we see that Viege [phonetic], who's Viege? Are you here? Maybe somebody else. And I have a feeling based on those lat/longs that's not somebody here, were definitely not at negative one. So, we could take this and walk it over. It's where? All right. Okay, so let's go to maps, and it's kind of interesting thing that you can take any of these tweets and you just type the lat/long in and you hit this. And we should then promptly zoom directly to this guy's--[inaudible]. [Laughter] And I'll tell you, it's kind of frightening because where--watch where it's going to put the pin. When I look at the tweets that I have done using this, it can tell the difference between the tweets like Twitter from the kitchen, the bedroom, and the bathroom. [Laughter] I don't want you to know which ones are which. So--[Laughter] so, yeah, when you zoom down, you end up really, really seeing. It's not just in the street or in front of the street. It's what part of the house he was in. It looks like he was in his front yard. So here we've gone from a tweet. [Inaudible Remark] Oh, maybe he was on the bus at that time. Oh, look at that. Oh there, he--wave. [Laughter] So partially, this is to terrify you. [Laughter] The good news is less than about 3 or 4 percent of all tweets have lat/longs. The bad news is all tweets have lat/longs. They're just--it's just the case that the lat/long stopped at the phone company and it didn't make its way all the way to Twitter. So, everything you've done looks like this if you're in Leicester. But [laughter] you probably--it'll show you right here. So we can do this with any set of tweets and then we will get from NodeXL these other worksheets. We get lists of all the people and then this is stuff that was summarized in that caption but all note that we actually write that caption for you. This is the graph summary tool, press this button, maybe you want to copy it to the clipboard, you're going to write an email, you're going to begin a report, you now paste that into Word or Google docs and well, it's not a finished report but you are about 60 percent of the way there now. You now add that, which only you can add, which is inside story meaning and that's something that NodeXL is probably not going to do for you anytime soon. So, how did I process this into this image? Did it just come down looking like this? I hope not, that's better. The answer is no, it came down looking like a black and white image with no photos. It was just the arcs. The arcs didn't have thickness, they didn't have color. They didn't have all of the things that you might do to decorate a graph. And so, we have a tool in NodeXL called autofill columns and this allows you to map data attributes to display attributes. And it is dynamic in the sense that if we go into the drop down, you're going to see a row here for every column in your workbook, in your edges worksheet, you're going to see a bunch of columns. If you add a column and in fact, have a look at this let's see if I can--you see this column, the add your own column, column. This is intended to convey to the user, you know, this is Excel, you can add columns. You can add formula, you can sort. You can filter. You can do anything that you know how to do in Excel. You still know how to do it, only you're now doing it in a enhanced version of Excel. And so, you could add a column and it could be shoe size, it could be IQ, it could be income, it could be zip code, it could be area code, it could be anything. And you could then filter or decorate your graph on the basis of that data. And so, you would do that by coming to the menu, going to autofill columns, and you would say, "Hey, you know, I want my edge width to be based on the edge weight, and the edge weight was calculated by this thing that rolled up all the replies in all dimensions and counted them and then added that to every edge." Or we might come over here and go to the vertices tab and on the vertices tab, it's going to say, "Well, what drives the size of the vertex?" And so PJ, this is going to be for you, here we go. This is vertex size, it's currently set to followers, that's my default preference, we're going to change it to between the center alley. And then, I'm going to hit the autofill button. And we're now going to see that the graph is going to--it's going to populate all the data and it's going to re-render the graph and the graph is now going to use size. And the data that drives the size indicator is not going to be how many followers you have. It's going to be, essentially a reflection of your position within the graph. It's going to tell you whether or not you are essentially--Jenny or Jenny. Or Jenny [laughter] or PJ.

>> Yeah, that fixes everything

>> I think that does, yeah. So where are you? I'm sorry.

>> Over on the right there.

>> Oh yeah, there you are. I'm sorry with the hair. That's right, yeah, okay. So, there you go and there's his tweet and notice that we have a little tooltip there. And so, we've changed the size and was that easy? I have to tell you, I'm on the support list for all the other very, very fine quality network tools that are out there and one of the most recurrent questions asked by users in the support for--are, how do I change the size or the color of the node? I'm just going to argue with those guys. Gentlemen, in user interface--I'm talking to the camera. In user interface design, you should be thinking about what your user is most likely to do and that should be one of the simplest things that you can do with that tool. Most users are going to eventually say, can I change the size and color of the node? How would I do that manually? Well, you probably want to do it in the autofill thing, but what we've done is made it really easy to change attributes of these nodes. All you have to do is find the row for, let's say, Nancy, there's Nancy's row and the current shape for Nancy is image and the size is ten and the opacity is 90 and that's the image file. And she has a label and it's her name and there's a tooltip and there's this little description there and if we wanted to change the color, that would be easy. I would just type something like red, no, not there. I would type it there, red, enter. If I now refresh the graph, her node is going to turn red. I'll refresh the graph. So, that's supposed to be easy and we've tried to make it as easy as possible. And so, you can now get data from a variety of sources. Maybe data that you already have that you can import or other data that's up on the web in various social media repositories, you can get that data. You can analyze that data. In the automate feature, we will stack up in the right order all the things that you need to do to get your graph and run it through a process that takes it from a raw state to cooked. And that means things like calculate your matrix and that's great, you probably know that you should calculate network matrix. But what you might not know is that it's probably a bad idea to do that until you've already calculated your clusters. And so what we have done is essentially built a track and you can choose not to do some things on the track but you can't change the order of the things on the track. And these are the right--well, you will assert these are the right order to do them. And in essence, we can't stop you from running with scissors but we can give you the scissors with the rounded tips. And so it's going to be very, very hard to hurt yourself all that badly with this tool, it's designed to do what we think is the right thing. Now, you may dispute that, that's fine, of course, we have open code, we'd loved to see fourth versions or if you want to suggest different features, we very much welcome your contributions on the message board and I should mention to the software developers, we need help. Lots of work items, we need work items done so let me know if you want to pick up or work item. So, this is the order, we add up the number of edges that resemble each other, so how many times did a reply--we recount that. Then we group by cluster, we do this thing where we put them into boxes and you can see here that we have used this feature, that is I think an innovation, and also unique to NodeXL and that is this mash up between the information visualization known as Treemap and network visualizations. And of course, Treemaps were designed and developed here at the University of Maryland. Ben, who did that work?

>> I think it was me.

>> Yeah, it was you. Yes, so Ben invented Treemaps in the early 90s and initially, you know, to apply to things like how do I reclaim a wasted file storage on what were then limited hard drives. You had what--hard drives with what? 5k back then or 10k.

>> There was this megabyte Mac in our lab that was shared by 14 users. And I couldn't figure out who has bulging space. So I had this concept and it was build by graduate student, Brian Johnson and the rest is history, it's just on--very long.

>> And so you'll find Treemaps in lots of places, smartmoney.com has something called the Map of the Market. The Hive Group sells a version of Treemap, you'll find Treemaps in a lot places. There's now a Treemap for d3.js, there's a lot of Treemap out there. And, anytime you want to see a complex hierarchy in a single page, you probably want a Treemap. What we did was we sort of--we adapted Treemap, we created a single level Treemap and made each group or cluster in a network, we gave it a region proportionate to its population. So how many vertices you've got, how many nodes are in your group, that determines how much real estate you get and then we layout each cluster in its region, as if there was no other part of the graph in splendid isolation. That means that each cluster that would otherwise impact the other clusters that would--maybe they would interpenetrate, they would--in some ways, it would be like when galaxies collide, right? And there's--your sun goes spinning out of orbit or the moon moves out of orbit and then we have a very long TV show about living on the moon and traveling through space. But what we've done is we've isolated each cluster, we lay it out and it turns out, it's better. Not just because we say so but because Cody Dunne says so, and who's Cody? Cody is a graduate student of Ben and he's been building tools that measure the fitness of a graph, that tell us how good is your graph or more to the point, how awful is your graph? Because it is the case that people have criticized network visualization, words like ball of spaghetti, bug splatter, hair ball. These are all the words used pejoratively to say network visualization, what is it good for? Absolutely nothing, say it again. But it's not true, bad network visualization is good for nothing but with just a few techniques and there's more to come. And hopefully, over the next day or so, we'll show you what Cody has up his sleeve, something that our team calls network motif simplification in which a lot of the clutter of graphs are reduced by simplifying certain motifs or common patterns that recur in networks. We take them, we remove them, we put a glyph that says, "I used to be 14 people who all hung off of one guy." Something we called a fan or maybe, "I'm tired of a group of people who only connect A and B, we share the quality that we only have two connections, we connected same two people. It might as well get rid of all of us, replace us with a whole bridge.

>> Would you humor me and show that combined edges please.

>> I can, I guess if you insist.

>> I just wanted more discussions on the phone calls regularly and one of the features that our super programmer Tony added was to--instead of having all those edges go back and forth and make the screen so busy, we combine them. Now, you're doing bundle.

>> You didn't want bundling?

>> Yeah.

>> You do--you're doing tight bundling, I'm saying combine. We need to combine feature, combine edges.

>> Is there an edge for every time someone repeats you, repeats with you?

>> Yeah, yeah.

>> And so--and then the--in the curve, that just those edges won't necessarily perfectly overlap with each other.

>> No, they perfectly overlap with each other.

>> Yeah.

>> So they perfectly overlap with each other.

>> Yeah.

>> No, the combine edges feature is a long standing discussion between us that it puts one big gray edge between each cluster. So you can see the numbers are connected between clusters and remove a great deal of the--it's in the group menu.

>> I have done the wrong thing. I'll have to undo that feature and then redo that. Yeah, I--we will.

>> Okay.

>> Okay. But so, you know, what we're playing with are ways of thinking outside the spaghetti bowl, outside the bug splatter, to say it's only a bug splatter because essentially, you didn't take the camera and bring it into focus. And so I am a little frustrated by those who dismissed network visualization. I say you just have to finish the job that, it's a matter of bringing the image into focus. And once you do, then there is some clarity, maybe not on the--

>> I think you need to refresh the graph on this.

>> Oh, right, okay. Well, maybe I should do--no, I don't seem to have control of it. It's a slow virtual machine and we've just done most of the computationally-intensive thing and I apologize. But yeah, so we are playing with these alternative ways of representing, you know, nothing says that every edge must be shown as an arc between A and D. There are ways of aggregating them more glyphically-representing them at a higher level. And so we are looking at that as a way--here we go. I will go to--so that is the bundled edge, see how they all got sort of swooshed down to a com? So it's like what you'll see when good network engineers run cable, right? They bundle the cable very neatly. They're very, very precise about their zip ties. So it is essentially the zip tied version of that, but that's not what Ben is asking for, he's looking for this. Let's go here and go to edges, go to other. You want--now, where is this? Tell me.

>> What is it?

>> I don't know.

>> It's changed.

>> Oh dear, how embarrassing.

>> Is that the thing where, when you put them in the group reviews peer groups?

>> Yeah, it's in the groups. Go to the groups menu.

>> So you collapse all groups and then the edges will view the bundled graphs.

>> Yeah.

>> You don't need to collapse the groups, just go the groups. That's it.

>> I'm not sure Ben.

>> No, no, no, no, no, no.

>> What would you like me to do?

>> Do not do it.

>> I'm lost.

>> Okay.

>> All right, I've got to rehearse that.

>> We'll talk about that later. Okay, we know our tool, I'm telling you. It does networks now?

>> What?

>> It does networks now? That's really cool.

>> Didn't you guys literally write a book on this?

>> Yes. [Laughter]

>> We should find the book.

>> I know, I've heard that--

>> The book is on version 113. We're up to 221 so if you want to help us write the second edition, these features I'm talking about like the Group-In-a-Box are newer than the book, so they're not in the book. And we have to go back in this, that's yet another job. I mean, it's one of these wonderful things where there's a great audience and great interest within those powerful tools. But it takes a lot more energy than we all have when we have students who are after a PhD, writing a user manual for these bold features is so hard and so we--

>> So there is another volume coming?

>> Yeah, yeah.

>> That's what I'm going to do with my summer vacation.

>> Yeah.

>> Well, you know, writing the book was a major ingredient of the success of the tool. And it's just--if you don't have documentation, your tool is of no use to almost anybody. So it's the castor oil of software development, right. You know, writing the documentation, nobody really wants to take it but it's supposed to be good for you. So just to illustrate, this is the before and after so this is "hey, let's make the big splatter graph. This is what NodeXL will do and almost all network tools will do if you just hand it a graph." This is

what happens after the application of Group-In-a-Box. And we believe that the combination of several of these techniques like other little techniques like many network analysis tools, visualization tools, when you feed them isolated vertices, we'll do almost exactly the wrong thing with them. So [inaudible], the layout algorithm for networks will create what I call the asteroid belt which all of these isolates get pushed away from everything else and so they go to the edge of the screen. They kind of form an oval if you have a rectangular screen. Haroll Corin [phonetic] says all isolates share the same structural location. Therefore, I will plot them right on top of each other. We say to these algorithms, you know, you have so many talents. You're so good at so many things, why should we ask you to be good at everything? And so, this is another example of the "don't run with scissors" approach. We take the isolates away and we do something simple. That's known as grid, very computationally-intensive. Every, you know, if we just grid the isolates and we order the isolates by some size attribute and then we grid them and what you get is sort of kind of a histogram or at least a gridogram. And then we box, we Group-In-a-Box these things and you'll find that each of these clusters, it actually turns out that there was a hub and a spoke kind of structure there and that was not visible when we saw it in this form. We really couldn't disambiguate the clusters. Now, we can disambiguate the clusters and this took zero human seconds for that transformation. It took some compute seconds, you just saw that, but it does take zero human seconds. Cody has just helped us implement network motif simplification. This is an example from the book. This is Robert Ackland's Web Graph built with NodeXL, with data from VOSON, the Virtual Observatory for the Study of Online Networks. It's a service from Professor Robert Ackland at the Australian National University in Canberra. And VOSON is essentially a web graph crawler and it brings you back--you give it either a bunch of URLs and it finds out they're connected to each other. And then spiders out from there to some configured radius or you give it a search term and it goes to Yahoo Search and then it gets a bunch of URLs and finds some connections amongst those. And you could use this for all sorts of well-motivated research questions like, how do the banks in Lebanon all connect to each other? How did their web sites connect to each other? Well, there are only what, 18 banks in Lebanon all of which are looking for money right now. Apparently, we--somebody gave them a virus. I can't imagine what that would be and, you know, it would be interesting to look at the patterns of connections amongst all that. Or maybe regulatory agencies or maybe it's a different sector of the economy, maybe it's military, maybe--you can do all sorts of things with how do websites connect to websites and we come out within like that. And we like these graphs, these are pretty graphs and a lot of graphs have this sort of there's a hub or a network head and then all of these sort of flanges coming off. Those are single tips, those are people who only have one connection and in this case, it's a webpage that only connects to that hub but doesn't connect to any of the others. So this is, let's say 10 sites all connect to one site but those 10 sites do not connect to each other. What Cody has done, he said, "Well, why don't we just remove all of that?" And this is what it looks like when you've simplified the graph. It's a way of visualizing with a lot of the clutter gone. And so we think that with development, this is going to be an improvement again like Group-In-a-Box, like the gridding of isolates, relative--in this case, not that simple but the straightforward methods to improve the quality of the graph. So we've been working at a lot of graphs, this is another GOP graph. We see a lot of this kind of thing. We see a lot of pulverization. We see a lot of divergence of URLs. People are not using the same words or roles. And when we go and we build lots of these images, we think something is emerging. Categories are emerging and we can also apply this to other domains. This was applied to the 2007 United States Senate voting records and there, you can see clusters emerge if you say, "I only want to see a link between any two senators if they vote in agreement with one another." Two-thirds at the time are more, you get the two clear divisions in US politics, the east coast and west coast.

>> No, [laughter].

>> No, and of some interest here, worth noting, there is Mr. McCain, this is 2007. Remember, there was some other guy who was in the White House, I can't remember exactly but I've kind of blotted out the whole thing. This is Senator Obama, there's Senator McCain and who are these three people? Specter,

Snow, Collins, what are those three people? Why are they there? 'Cause they're high between these, why are they between? 'Cause they actually have edges into this cluster, why do they have edges into this cluster? Because they are traitors to their party, [laughter]. Specter was immediately voted out of office at--well, 2009, he lost his job. But in--I'm sorry-- [Inaudible Remark] Maybe, I mean okay, we'll discuss that. I believe it was Collins who just retired or it's Snow?

>> Snow.

>> Snow just returned. Collins is the last one standing. So people often ask, is between the centrality good? It all depends. [Laughter] I like to say, between the centrality often means that you have arrows in the back from both sides. [Laughter]

>> But wait a minute, the irony is that--

>> There's the punch line.

>> Okay.

>> There's the punch line, this was 2007. What happened in 2009? Some other, other guy became the president and he had this idea that hey, in really tough economic times, maybe it's the role of government to assist the people of its own nation. It's a radical idea, I appreciate that. But only three Republicans voted for the Obama stimulus. Who were they? Specter, Snow, Collins but this was two years prior to that. So in 2009, we could potentially have predicted who of the Republicans would go with the Democrats and vote for the stimulus. Well, you might calculate the voting record graph and we would see that Specter, Snow and Collins are the only three that are likely. And that two out of three are now gone. So, there is some value to looking at the network structure of all sorts of things, not just social media. You could consider this a kind of antisocial media to the US senate. So, there is a book, there was another book, a fine quality book with my now sadly departed adviser Peter Kollock. And we have applied these tools to a variety of topics. And this one was used as a featured example in a recent publication at a website which was Noshir Contractors conference, he hosted it at Northwestern in early June. And Jana Diesner ran a workshop on Words and Networks, how do you use the words in networks and the structure of networks to inform our understanding of a graph in general but a social media graph in particular? So, this is people talking about what surprises me to be a controversial topic. It's the word contraception or contraceptives, and it turns out that these two groups are strong positives on contraception. They talk about contraception in case you missed it, birth control Griswold. Anybody know what Griswold is? Yeah, Griswold versus Connecticut, 45 years ago. In fact, this day--these days in June, this was the 45th anniversary of Griswold versus Connecticut. What was Griswold? That was the court case that went to the Supreme Court that essentially made contraception legal in the United States of America. So, it's hard to imagine here folks, imagine a time, it's illegal. Not only was it illegal, it was illegal to talk about it in the US male, it was illegal to share information about women's health, reproductive health. Okay, but that 45 years ago, you would think "Hey, things are good now, right? We have a lot legal." And that's true, these people are very happy with, you know, a power of "her" but they're also talking about this word, war on women. Here, up here, this group, they're very tightly tied to each other so maybe they're similar, maybe they're different because at the hub, this is all the people who essentially are around one news outlet, these are the people who are talking to each other, there's the Gates Foundation. And so the algorithm decided that they were separate clusters but it also notes that it has a lot of connective tissue. And here, they're talking about contraception and birth control and the war on women too and repro health watch and P2, that's Progressives 2.0, insanity in women's health. Okay but who is G4? Notice a few things about G4, very little connective tissue to the other groups. They're not that connected and who are they? They're people who don't like birth control. This is--it is these people who make this a controversy, rather than, it's like saying, I don't like aspirin. I

mean, why would you not want it? It seems to me anyway, maybe others have a different opinion but these are the people Teacup, Pro-Life, Catholic Contraception, Tea Party, Pro-Choice, Abortion. These are the people who for whatever reason believe that it's a really bad thing. And they are the ones who make it a controversy by opposing it. And now, we can see and I think I made this point Monday afternoon, that when you look at a bucket of tweets or blog post or messages from any social media source. And you do content analysis on that collection of tweets. You are potentially making a very serious error if you don't recognize that these tweets are coming from socially defined subgroups. So, I'll give you an example. Right now, there's a bit of politics wherein a presidential elections sees and I believe there are 92 days left before the election. And lots of people are getting into the business trying to prognosticate the election on the basis of social media and here at election.twitter.com, the company called Topsy is reporting horse race results on who is ahead and behind in the Twitter battle. Now, I could try to explain how they come up with this number, I don't think I will succeed. It's allegedly the percentage of messages in Twitter that are more negative than the messages about this candidate. And if you understand that, you can explain it to me later but the thing I'll note about it is that one, it's insanely volatile, it goes up and down a lot. At one point, Mr. Obama was at 74 and Romney at 27, they seem to be converging 18-15. But I ask you this question, what does this really report? So, let's just say and as I showed you the GOP graph and--is that a question or that's an answer?

[Inaudible Remark]

If that's true, that's unethical.

>> Oh of course. But it would be--[laughter]

>> I'm talking to you.

>> It doesn't mean they can't do it.

>> And I'll be with you in a second but that's a deep--if that's true, I'm deeply concerned. I'm more concerned than merely methodologically flawed, if you're saying that they're actually manipulating at that. I'm not sure if that's a substantiated allegation but that would be very serious.

>> Correct and I don't think that though you can necessarily [inaudible] approve but that might be true. But it does--it is at least some analysis I'm sure they can work out, since they are kind of promoting this as the analysis.

>> A couple of the caveats and I'll be right there with you. I think it's eight to ten percent of US population are Twitter users. Skews in all sorts of SES directions towards, you know, why yes, I do have a Honda Civic. What do you know? And yes, I am thinking about a Prius. I do have an iPhone. You know, I do have a Bachelor's and yes, I do have a graduate degree so, you know, there is a big skew but my argument is this. If a bunch of GOP-AOK people say, "Today, we're going to tweet a lot of negative about Mr. Obama." Or if a bunch of GOP not so much people say, "Today is the day, we're going after Romney." And that moves these numbers, does that mean anything? If the people who are against you are against you, then that's not news. If the people who are against you got noisier, that's not even that much news but that's being reported as an actual change in the aggregate and I will argue that at the last few hours and days before the election, there is a group of people in any democracy who will vote for the person they think will win making this very dangerous. So, I'll be right with you but I got to go back there.

>> I know because afterwards, they say they voted for the winner. The number of people would say they voted for the winner after an exit poll jumps by 10 percent.

>> I voted for Mondale. Okay, be right with you, I'm going to start back there and we'll come around. I think we've touched a nerve, ma'am.

>> Why don't we decide before by chance and I was trying to find--I think what is not involved is--it's just has one--

[Inaudible Remark]

>> And second, I was trying to find a book that say [inaudible] because on some days, it's like 72 to 20, what does that number mean?

>> Put it this way. If--

>> And maybe you'd know because you would look that it's more left but I couldn't find out what does the numbers mean?

>> So, I think there's--what Jenny may critic, I'm not that sure. I believe it's telling you percentage of messages about--from Twitter that are more negative than the set of messages about Obama or Obama-mentioning tweets as compared to all other tweets. I've read the articles from Topsy, I don't-- [Inaudible Remark] Not only that but it also depends on some magic pixie dust called sentiment analysis.

>> Second, I was looking at the IBM, the IOWA and the kind of market [inaudible] office enterprise, which also has digital tool and literate and computer savvy find that indeed that it needs training and found it to be very much different today, well, [inaudible] left that was 60 percent Obama, 40 percent [inaudible]. So I just figured, you know, which kind is always represented these--

>> I trust the IOM market more than I do this I have to say. At least because I'm going to identify this I think fatal methodological flow, says who, right? If you simply get a lot more pro people to be more pro, that's not news and if you got anti people to be more anti, that's not news either. So, let me see if I can get some of the other comments here, ma'am.

>> Okay, 'cause the thing is, is that in terms of elections in America, how much would we make does matter. How much you want to spend on political eyes, where you target them and when you target them, doesn't necessarily change anyone's mind, it gets buds out of season for election. So what question you're really asking is, does Twitter matter in the same way that other guys at local advertisings do. So Twitter does matter and, you know, we're going to make 50,000 more tweets in the right places and the right, you know, anytime when people want to do or they may in fact have actual political impact.

>> Okay.

>> And I don't know if the answer--I don't know if Twitter's being not actually [inaudible] possible or hurting people but making noise that's always been made by making it louder in the right time actually does [inaudible].

>> That may be but I will still argue that that number, that time series plot doesn't tell me what I think it's trying to tell me. It's telling me how--what's the volume but not the origin of that noise. And so--but I think you're right that, you know, on the eve of the election, if you look at this and it shows an enormous spread and then your candidate is not being favored, you may not vote for that candidate. So, let's see. I've got a comment here, we'll come back there in a moment. Yes sir?

>> Do you have any idea whether it's more important, the value that's being taught or the second derivative?

>> It's really variable isn't it? It's really, really variable but let me show you another example. Let's see, I think that you can see it pretty clearly in two different images. I made a map of the word Obama Care and then I made a map of the word ACA and what really caught my eye was the idea that I'll just do Obama Care. If you said Obama Care, if we look at the Obama Care map, what we're going to see is that two-thirds of the people who are opposed to Obama Care. And so, this cluster and this cluster, these are pretty much negative [inaudible], the hashtag, Obama doesn't care, these people don't like Obama Care. These people use the word Obama Care and another thing to observe is if you look inside the profile images, you'll see that the only media outlets are in this cluster, Talking Points Memo, Daily Feast, I think that's a CNN outlet. So, there is mass media in this cluster, there are no mass media outlets in these clusters. So this is the group of defenders with if you use the word Obama Care, there's a really good chance that you don't like it. However, if you were to use the word ACA, there we go and that of course is the Affordable Care Act. What you get is almost no one who doesn't like it, right? So here, if you're going to use the word ACA, you're not even really going to find--let me get--you're not going to find a cluster that has the word TCOT as its highest frequency hashtag. And so, just the words you used to talk about the phenomena is going to determine what company you keep and depending on the word you use, you may find yourself in enemy territory or among right thinking people who share your view, whatever that view maybe. And so, how you measure this is really fraud. It's dangerous to say, "Hey, send me on ACAs, it's pretty much positive." So with that, let me just now--

>> I found it.

>> You found what?

>> The combined edges. Do you want to go back here?

>> Okay, [laughter] persistence.

>> I wanted to--

>> Persistence and composure.

>> We thought of a lot of arguments.

>> All right, so yeah, where is it?

>> It's under layout.

>> Oh yeah.

>> Go to layout.

>> The hidden layout option, it's in there.

>> And then over there into group edges, hit instead of showing them, combine them..

>> Oh.

>> So the beauty of this is--this graph is not a great example. Now we refresh.

>> Yeah, it's doing its thing.

>> The point is it just strips away of lot of the extra clutter and makes for a much cleaner graph. In this case, there are only two major groups that we're only going to see one, you know, major edge between them. But it allow--it hides what I think is a lot of the clutter then you can actually read the names on the labels. Why is this taking its time?

>> 1 gig virtual machine running Windows on the Mac.

>> Right, okay.

>> Not the ideal demo machine.

>> Okay. I mean this sort of popped up. And it just sort of--I just read at the graph here and it just really cleans up. So we can see all the names. Nancy Baym is the biggest between some [inaudible], Jenny Corn, right over here, PJ Rey, myself, and Marc Smith and then Webshop UMD just lead the way. Come on, you can do it.

>> Okay, so, I hope that this will have given you, now I'm going to leave it, it's going to come back but I'm going to do, you know, closing. It's time for graphs.

>> Yeah, yeah, yeah, it's time [laughs].

>> And beer and graphs.

>> Right.

>> And anybody who wants to hit crack the crabs and I'll--you can open them and I'll eat them.

>> It's a lot of work.

>> So, it's the tool for the rest of us. It's a tool that I hope those who are algorithmic will decide to contribute to and use to help deliver their results to the rest of us. For the rest of us, I think, it's the tool that gives us the super powers of a second year computer science graduate student without actually matriculating. It's grad student in a box. [Laughter] And so our goal is to be that Firefox of GraphML, the thing that let's you think link with a much lower overhead, a shallower learning curve. And I hope that you will also see that it is possible to generate publishable findings and results by bringing--and there you go. That was it.

>> So what you're seeing now, well, we've got to set kind of light but it reduced all the connections between to this thin gray line here. Here we go. It's coming a little better. So you're getting--so we've now--and it gave more space over here. We probably want to resize this a little so we can see it.

>> No, yeah, it's going to be a [inaudible].

>> Oh yeah.

>> Okay, do it.

>> Oh dear, too late.

>> Too late.

>> Oh dear. In any case--

>> You see that line was back there. [Laughter] So, how do you progress if you wanted to join us? Ask your questions on the message board. Send your mail to me or Ben. Catch us while were here for the rest of the week if you have a data set. Somebody gave me a data set. I've got your graph. And, you know, we'd like to help you get to social media insight or network insight broadly but certainly, social media insight with a lot less effort. And we don't assume that giving that it's a reduced amount of effort that then, you know, it's time to watch Breaking Bad for 14 hours. It's then, you know, time to spend that time. Instead telling the story of your graph or maybe repeating the same analysis over a hundred time slices of your graph which becomes practical because you can now automate the collection of this data. You can--as a matter of fact, maybe I could show you very quickly. So yeah, here, I've got a virtual machine up in the Amazon Cloud and that is what NodeXL looks like when it wakes up using task scheduler every 15 minutes, it wakes up and says it's time to do the Romney graph. It's time to do the Webshop graph. It's time to do the--you name a topic that's relevant to your research, graph, you can schedule this. It's going to do that and when it finishes, it's going to write a file to the disc and that's going to be the set of connections amongst all these people. Then it's going to open automatically a copy of NodeXL. And then it's going to hand that data to NodeXL and it's going to say do the thing you do with it. Do that recipe thing. And then it's going to generate a report, put an image on my hard disc and I'm going to see you've got a directory filled with these images not unlike the front page of graph gallery. And then you are able to say what's new. What's happening? What clusters have formed? Which ones disappeared? What's going on in my various topics of interest? And you're now doing this at 60,000 feed with the ability to have, of course, dive right down into that guy's living room in England. So contacts and detail and overview, and in a package that we hope will help you achieve your research goals. And that I think is our story.

>> Woo!

[Applause]

Questions? Yes, you do [inaudible].

>> I heard Marc many times. It's always great, fun, and it's wonderful. Any questions or things to solve? Yeah, let's take a few and then we're going to sort of get--

>> There's crabs and beer between us and crabs and beer. Yes sir, take your shot.

>> Do you know anything about what the general--with the political--?

>> Ah, yes, divided. [Laughter and Inaudible Remark] You know, I would say that the left and the right both very aggressively and actively use Twitter. And I have a tool that allows me to look at individual spaces in Twitter. You would have to go to the Twitter people and say, okay, you know, given a 130 million tweets per day, what's the breakdown? I can't answer that question. I don't have 130 million tweets a day. What I do have is the ability to take a picture of today's set of tweets on a topic. And when I look at, you know, topics that I guess would be right wing topics, they are plenty populated with active discussion. When I look at left wing topics, plenty populated with lots of discussion. And when I look at topics that are at the boundary

between the two, lots of polarization. So I would say that left and right have both very aggressively moved into Twitter. Now, does it matter and this is to your point, does it matter? Is anybody changing anybody's mind in this place? I doubt it. I got to say, part of the downside of the research is that I'm not reading the content from the part of the world that I usually, you know, I don't want to hear it, I don't want--I don't watch Fox News, I don't read National Review, I don't want to know what they know. But now I am. I'm reading all these stuff. And it has had no effect on my political views, if anything. Well, I'll say, it has entrenched my political views.

>> I might say you should ask Dr. Jen Golbeck who did study two, three years ago, she read every tweet written by a member of Congress. [Laughter] Some 6,000.

>> All three of them.

>> And there was an overwhelming strength on the Republican side. This Republicans had organized early on to push the members of Congress to tweet. And so there was a strong, you know, strong and heavy dominance on the Republican side. I think that's balanced out and it'd be interesting what Jen Golbeck has to say about that now. I'm not sure we can say anything else. There was early on the TCOT top, as I learned from Mark, the top conservatives on Twitter but then that was T lot top liberals on Twitter. So, you can look for those groups and see the sizes of that. I would say from what I've seen in the graphs which we've looked at, there's the Republican side is more closely woven together than the Democrat side. And that I don't know about the magnitude but the woven this tight.

>> I always find that I should--what Libby have the next comment.

>> Yes, right.

>> I will answer each of those questions while [inaudible].

>> All right [laughs].

>> Every single one.

>> Good advertising.

>> Congress says replies have had 530 tweets since Mark said 15 number tweets so in the last 45 minutes. Congress generally had 6,000 tweets a day. So Jenny studied from three years ago.

>> Yeah.

>> While useful is--

>> Very old, right.

>> Yeah.

>> But you may have--

>> It doesn't capture a day in--

>> Maybe doing--

>> Perfect teaser, perfect teaser, tomorrow, Libby.

>> Sorry, we'll look Congress not the public Congress tomorrow.

>> Okay, yeah. So these are very different. This is people talking about politics not politicians telling you about politics. But I will underscore that point. Whenever I see a polarized discussion, birth control, contraception, and I'm going to find that there is a group and it's isolated from the other groups and it danced. And those are conservatives. And they also lack any mass media outlets in their group. Whereas--and I say the color is that what we find in the liberal groups and, you know, they're not Liberals folks, they are right of center, right? I mean, nobody here is actually a--[laughter] or advocating the nationalization of anything. Let's just, you know, put this in the right context. There is no Liberals on these maps. But what we find is that the so-called Liberals, they have a lot lower density and they have a lot more different hubs. And so what I like to say is that the problem with the left is that it tolerates diversity.