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.

>> Noshir Contractor: So I hope you had as exerting a lunchtime break as I did. I'm going to mention what I did at lunch. I didn't see Esther. Esther? Oh, there she is, okay. So Esther Hargati [assumed spelling] and David Hoffic [phonetic] and myself went on a geocaching. So how many of you know what "geocaching is? Oh my God, and successful as well. It was a lot of fun. But it was -- it kept us down to the wire to come back here [inaudible]. And had a nice workspace for the walk that we took there, et cetera. And I'm going to mention two things about this, one [inaudible]. The first thing I'm going to mention is that as many of you know that people -- very often you have people -- is Nancy here?

>> Nancy: Yes, I'm here.

>> Noshir Contractor: Yes. Sorry, there's more than one Nancy. [Inaudible]. So one of the things I've discovered is that people like us who are in technologies and who have a lot of interest in doing these things as hobby often find a way of taking our hobbies and then justifying a different search and getting involved with it, et cetera. And Esther is an example of that. You know, she got into some geocaching [inaudible] and then now justified it as a professional activity because she has written an article on geocaching, which I think was titled "Cache Me If You Can." Is that the title of the article? And so and I was thinking about that, and I said, "Who else do I know who does this?" And then it came to mind that Nancy Beam [phonetic] is another person who does this. I knew Nancy from when she was a grad student at Illinois. I was there for 20 years before moving to Northwestern 5 years ago, and Nancy was really obsessed, if you don't mind me saying this, Nancy, about keeping track of rats; and by rats I mean radio and television soap operas. And she got so involved in it that at some point she had to justify to her mom and everyone else around her, her colleagues, why she was spending all this time looking at -- you know, reading about it online, et cetera, and the next thing you know Nancy's dissertation is on rats and how people react to doing discussion [inaudible] about it. So there's a lesson there for so many of you that you could take technology addictions and convert them into a professional activity. Okay. So I'm going start talking about, Some Assembly Required" [inaudible] have been on sabbatical this past year. I've been working on a book with the same title -- not the same subtitle I don't know [phonetic], but the same title, "Some Assembly Required." And it's about how we assemble teams from networks and how these teams in turn then influence the networks [inaudible]. So I'm going to share a little bit with you about those ideas here. I'm going to go through it pretty fast so I can hopefully spend some -- save some time for questions. So many of you have -- some of you senior scholars in the room have seen this so bear with me; may have heard how do you get -- oh there is it, yes. So the lovegety; how many of you knew about the lovegety other than [inaudible]? Okay; almost nobody, oh a couple people do, okay. So the lovegety is this little toy that was very popular 10 years ago in Japan. [Inaudible] all the kids with bags, little lovegety on their belts, and it was basically

[inaudible] the tamagotchi, some of you remember that? It's kind of like that but a big difference. The lovegety is a device which you program with the kind of music, movies that you like, and the kind of music and movies that you might want in a love partner. And then you wear this. So this is kind of going back to this morning's discussion about chemistry.com and eHarmony, et cetera. But this was a hardware solution that you programs in that little thing there, and you walk around the streets in Japan and anytime your lovegety starts flashing or beeping or buzzing, you look around to see who around you is also flashing or beeping or buzzing, and then you know that you've made sort of a love connection that comes out there [phonetic]. You might wonder why am I going to be talking about romancing issues, not by no means [inaudible] studying it or necessarily practicing it, but I'm going to move on next to another one, which is even though it's a little more darker, and that is "Sniff [phonetic]." And this is called "Sniff" because it's social networking in [phonetic], so it's the dogs, and you see -- that's what you see out there. You can see these little dogs. And it's a little device that you put on the collar of the dog. And you put in on the collar of the dog and it -- with each time it meets another dog that has the same thing like they do here, the two dogs, little USB infrared thingamajiggy exchange business card information about the dogs. So, you know, and at the end of the day you take the collar off, and then you plug this into your laptop, and you go on the web and you see network of all the dogs that your dog likes to hang out with. And then you can click on any of those dogs and find out who the owner is and then -- this was developed in the media lab at MIT, and so the article in Wide Magazine, which featured the story said you can set up dog dates and, of course for those of you who know, why they like, they geek speak [phonetic]. They went even one step too far for them I think. Oh, this makes them move from social networking to social pet-working. [Laughter] Okay. So why am I starting with two examples? I actually am talking about it because the two examples talk about something that is becoming increasingly common, desire and an aspiration to use technology, not just to communicate and collaborate with people, but to find the right people to communicate and collaborate. That was true in finding romantic partners, it's true in finding dog dates, and as David Faruchi [assumed spelling] wrote in the New York Times earlier this year -- this is a beautiful article if you want to take a look at it. All of you who live in Watson, Jeopardy, IBM, yes. So David Faruchi was the head of the team that came up with Watson. And he wrote this article in the New York Times earlier and he was saying that we may think that, you know, building Watson was a challenge technologically. But the bigger challenge for me was building the team that built Watson, which was the title of this article in the New York Times. And so we clearly see that there is a need for us to do better at building teams. Some of us are successful, like David, and others not so successful. And I'm going to have to do this with my power [phonetic]. Is this on a USB [inaudible]? Okay. So in fact there's a lot of [inaudible] and articles published in science by my colleagues Brian Nutsy [phonetic] and collaborators, which shows that more and more work is being done in teams. So they looked at all the articles published about [phonetic] science over the last 30 years, and they looked at this about [phonetic] 20 million articles [inaudible]. And what they found was that increasingly articles that are being published are written by teams, that it's multiple authors, in all areas, including the social sciences, even

the humanities, [inaudible]. Not only is it being in use, and not only every [inaudible], but finding number two is that teams that get higher [inaudible] than individual arts. And the cynic in all of us says, "Well, of course, each of the author cites their own articles." [Inaudible] citation, and you still have this high impact of articles done by teams. Finding number three, articles done by teams for people from different disciplines have an even higher impact than those that are done in the single discipline. And finding number four, articles that are done by teams from different disciplines and different universities have an even higher impact than those that are done from different disciplines but from the same university. And so you look at these findings and you go, "Well, I guess we now know everything we need to do. We are already here and it's [inaudible] set of workshops, so all you've got to do for the next week out here is find other people that you want to work with, make sure they are from different disciplines, from a different university, and you're all good go to." Well, not quite so fast. It does [inaudible] -- oops. [Laughter] This is a reminder for the right thing [inaudible]. So move this -- so this is the move to virtual team science, meaning distributed locations as well. But while this was happening, Jonathan Cummings, who is at Duke, and his collaborator, Sarah Kesla, [phonetic], [inaudible], have done a series of studies looking at an Assar projects [phonetic]. And what they found was that in general the more interdisciplinary an Assar project, and the more distributed -- geographically distributed it was, the less likely it was to be successful. So you think this is a paradox and it's actually not a paradox, because what it's really saying is that most of the time interdisciplinary geographically distributed work fails. But when it succeeds -- on the few occasions when it succeeds, it succeeds spectacularly. And that's the bias that you see in the published materials in the world of science. And so part of what our goal -- and my goal here is to try to understand how can you make the teams more like the stuff that Ryan Lutsi [phonetic] and Ben Jones found in the web of science, and less like the projects that were dominating in the dataset that Jonathan Cummings and Sarah Kesla had. So what I'm going to argue here is four take-aways, and that is understanding and enabling TMSP, which is my abbreviation -- and I'll get to that, our abbreviation for "technologically media social participation," that understanding this and enabling it, and that's the intervention I know. Ben uses the word "intervention." I think of understanding and enabling as the two phrases, that you're really well-poised to make a lot of improvement in this area for four reasons, because the first time we have good social science theories that already tell us a lot, but leave new questions still to be addressed and researched. We have methods that are better than what we have had -- [inaudible] we have methods ranging from ethnographic quality and into some pretty advanced interferential network analytic techniques that I'll talk a little bit more about today. The third is that we have data, obviously; we have big data. We also have what I think is -- what I [inaudible] is like it was broad data. This is Jim Hendler's [phonetic] word. He says, "We don't have big data, we have broad data, because you are able to take data from diverse sources and collate them in ways that we were never able to do before." And then finally fourth we have the computational link structure with [inaudible] that allow us to do that. I'm not going to talk much about the fourth one. I'm going to focus on the top three during the [inaudible]. So why do we create -- and this is

a general question, why do they create the links we do? Later on I'm going to ask the question, "Why do we team up with the people we team up with? What is the motivation for doing that?" And I have this picture from last year, taken right outside this building. And it was taken right after that earthquake that we've talked about. And I thought it would be appropriate to talk about this today because what you see in this picture is that people [inaudible]. But what you may not be able to see very clearly is that not only some people are talking to each other, but a lot of the people are on their cells phones and mobile phones and smart phones talking to other people. And so the question is this is a great way to capture at that moment when you had a crisis like the earthquake, who are you going to call? Right; it's that sort of thing. And it tells you the motivations -- see they don't get that. [Laughter] "Ghostbusters." So it's this idea of how -- what is the motivation to do it? So in a book that Beonji [phonetic spelling] and myself wrote, and then Stanley Russerman [phonetic], KD Fost [phonetic] and I wrote a subsequent piece in the Academy Management Review, we looked at what are the motivations for what could be for technologically media and social participation. And we basically categorized the literature in the review into two categories. One is self interests. I'm going to reach out to Mark because I want something that Mark has. So it's my self-interest, and that's why I'm going to reach out. Or of course Mark may not answer my calls or my Skype calls, [inaudible], and so that's because it doesn't always work. Sometimes it works, but in many cases it may not work. The theory of -- it's an economic model, right, I'm trying to maximize my [inaudible]. The second is it's a socially exchange model. So this is more like a market model which says that I want to create a link with -- where's Jenny? [Inaudible]. She had to step out. Oh, there she is. Okay. So I want to get something from Jenny, but I know that Jenny also wants something from me. So we set up a social exchange, and that's another way then to be able to -- why I may [inaudible]. I want something from them, and I know they want something from me. The third one says, "I don't want anything from Ben, and Ben doesn't want anything from me, but we know that getting together we have a better shot at getting something from a third party." So any collective action, any lobbying effort, any setting of standards, all of that will fall under that. So these are different motivations of creating webs [phonetic]. The ones on the right are less strategic than the ones on the left. The one on the right says, "Theories and [inaudible]," that the reason I want to talk to Bonnie Logan [phonetic] is because I look around and everyone wants to talk to Bernie Logan, right? So he's a very popular guy, and so I feel like I'm a stranger in this group, I should go talk to the one person that everyone talks to, and that's why Bernie Logan, they're the rich get richer. People who have more network guys get even more network guys. I only said that to make sure that Bernie was listening and paying attention. [Laughter] [Inaudible]. That's good. I'm glad it's time [phonetic]. [Laughter] Another reason -- another thing that could happen is that people are creating links because of tolerance, because your friends are friends, right? So I know Bernie and then Bernie will introduce me to the next great star in the web world because he knows all these people -- friends with them and knows something about them. Is it necessarily a good thing, not necessarily. If that person is really going to say something interesting, I could just get it from, where, from Bernie, and instead I could spend my time talking to people who are not connected to

people that I'm connected with because they may give me more novel information [phonetic]. [Inaudible] birds of a feather; so this goes back to my story of the geocache. And let's do our morning introductions. When we did the morning introductions, did you notice how we would have these clusters of people from the same institution and one after the other sitting next to each other, right? And why were they hanging out together, because it's comfortable to talk to people like us. And I'm only saying -- and having said this now I really would request to you to mess with the next time where they have people like Bernie. Remember Bernie said this morning, "Well, so and so when he gave a lecture here three days ago he was sitting there, and he was sitting there, and she was sitting there?" I think we need to shuffle things up and not sit next to people you know because there is a [inaudible] while it's good news and jumps of being able to talk to people like yourself. It's bad news because that's not were creative ideas come from, and you're not here to talk to people like yourself. Having said that, I plead guilty that I'm the one day I'm here today because I have to leave this evening sadly -- on the one day that I'm here I take the whole lunch break and hang out with two people that I know pretty well, Esther, who's my colleague at Northwestern, and David, who was my colleague at Northwestern as a graduate student before he went to greener pastures and better climate. So it's easy to fall into the habit of [inaudible], but it's not always the sort of best thing to do; though I had a pretty good time necessarily, David, just proximity says if you try to [inaudible] people who are close-by to you. So having sat next to each other you're obviously more likely to talk to each other. And you can ask anyone in technology you could talk to anyone, anytime, anyplace, and that's all true, but the fact of the matter is that even with technologies today we forget that we use technologies inordinately more to actually communicate with people who are close-by to us. The biggest applications of using things like Skype et cetera, are to talk to people in the same city, in the same building, on the same floor, in the same dorm rooms. There's lots of different research that shows that. And while technology allows us to talk to people [inaudible] inordinately just to control a lot of our approximate communication. So all of these things are still relevant today as we move forward. The nice thing is that each of these theories that I've talked about have unique structural signatures. By that I mean if you have -- if people are operating on the basis of self-interest you will see certain kinds of structural configurations in the network that are more likely to happen than by random [inaudible]. This is the easiest one to show, it's an exchange, if there's a solid line from A to F, there's [inaudible] from F to A, social exchange, as opposed to a link from B to C. So what is nice about this is that if each of these theories has a unique structural signature, then when you get an observed network -- like some of the networks we saw this morning, in those mangled networks we can use techniques that you can think of them as statistical microscopes, if you may, or statistical MRI's, and that allow you to be able to -- oops, sorry, what am I doing here; that allow you to -- hold on a second here, all right? These are slides [inaudible]. So I'm a little confused. Oops. I seem to have -- that happens, I'm sorry, [inaudible]. So that's --

>> Male: [Inaudible]?

>> Noshir Contractor: Beg your pardon? [Laughter]

>> Male: [Inaudible]?

>> Noshir Contractor: I didn't --

>> Male: [Inaudible].

>> Noshir Contractor: The rose, no, not that. I'm just going -- okay so just -- this is the MRI. So you could look at those signatures and be able to identify which of those theories may be operating in certain conditions, right? So in other words you can -- I can do a network analysis using Facebook and mygen.dev now, not mygen.web; is that right?

>> Male 2: Yes; mygen.dev is the one that's not relevant as of last week.

>> Noshir Contractor: Right. So when you do that, I can look at the newly spaghetti network of this and say is this network has been driven 15% by self-interest, 20% by amophaly [phonetic], 54% by social exchange, and so on and so forth. So I'm able to understand what drives this particular network in ways that I wouldn't have been able to do before. This is talks about statistically testing this. I'm not going to talk much more about it, but there are some nice statistical models that have come up. Where is Zach? Zach's advisor, Carter Butts [phonetic] at UC-Irvine is one of the leaders in setting up this really interesting statistical techniques called "Pstar" [phonetic], or exponential random graph models that allow you to test for those signatures and be able to analyze data in some very creative [inaudible]. Now, of course we have theory and we have methods. The challenge has always been of apparently testing these theories, and of course [inaudible]. The Hubble Telescope, these are slides that David Lazar [phonetic], who's at Northeastern shared with me, and it's a wonderful set of way of describing our [inaudible] sets. If you're an astronomer and you have a Hubble Telescope you know you spend -- you have all the data you want and you spent $2.5 billion, and you're able to -- you know, you're in data heaven as a result; literally. And if you're a particle physicist, then you've got the [inaudible] idea of the CERN particle accelerator. It's a billion dollars a year and that gives you all the data that you want. And that's the way in which do it. But what if you're a website? What if you're a scientist like someone in this room, et cetera? Well, this is what we have. It's priceless, sort of, because we obviously have all kinds of privacy issues, et cetera. But it is in fact the place for a lot of where we are going to be rely on extensively to be able to analyze the kinds of questions that we want to address in this area. So as I look at it from my lens, the web today has helped us create this multidimensional network, where some of the notes are people, some of the notes are documents, some of the notes are datasets, some of the notes are analytic tools, and we have links amongst all of these people; people-to-people coauthoring, people -- documents. Documents have key webs [phonetic] and they links to create documents, and key webs. So you create this incredible fishnet web. And you want to be able to pick up any one note and see how it's connected to other notes. Or you might be wanting to pick up any particular -- and the note that I might pick up here for me might be different from the neighborhood that will show up for you because where we are in the network is

different. So we will think about having recommender systems along these lines allow us to say that if [inaudible] says he's interested in topic A, it might recommend person X, document Y, and dataset C. But if David Hoffard [phonetic] picks up the network and asks the same topic, he'll get a different set of recommendations based from where each of us are in the network and given the motivations that we talked about there. So of course what has helped a lot with this recently is this notion of linked open data, right? We had a lot of the data that we have today on the web. You know, we have web pages that are linked, but we don't have data that has been linked until relatively recently. And that's a large part of what [inaudible]. And so people like Jim Handler who used to be in Maryland, who is one of Ben's colleagues here, now at RPI, has been championing the idea of linked open data, where data that is made available is linked to other datasets and it creates this incredible way in which not just having pages, but you have deep data that is linked with one another and makes it into a broad dataset, not just a big dataset. So that's a large part of what is happening, but it first started out you see that a lot of the data that had started with DD media [phonetic] -- and some of you would know about database, DD media, and then you have all these datasets, your online activities, publications, life sciences, geography, and music. As you move you see a couple of years later and it's really exploded. So this is something that we should think about as a way of understanding the network of data that has been created that accompanies the network of people that we have been typically studying on the web. So obviously this was -- I don't expect many of you have seen this, but these are all the kinds of activities that we do on the web. This was something based on 60 seconds on the web in the amount of time we spent 168 million emails since our 700,000 search queries and so on and so forth. It gives you the scope of what the different kinds of activities in which we interact with the web. So on the basis of this -- I mean, I talked with David Lazar, David Lazar and his colleagues and then myself included here, we wrote an article in "Science" a few years ago that talked about the fact that we have this new thing called "computational social science." And computational social science is an approach that complements what are the other forms of social science. So it doesn't take away from experiments, it doesn't take away from large-scaled surveys or survey-based data, but it's a new quiver in our -- a new arrow in our quiver activity to look at it. And I think it's particularly useful -- and this is a separate talk that I'm actually going to give at Oxford in a couple of weeks, on can big data motivate new theories and data and new methods? So it gives us an opportunity to actually think about new theories that we go beyond what we had looked at previously. So what I want to talk about substantively then -- and that is a backdrop, is to say that if I want to look at teams, and I can think of four levels in which this was analyzed. I will say that last year when I did this talk I talked about some things that were similar to this, but I only had three levels then so this year I added a fourth level based on some research that we've done in the past year. So when you look at teams, people who study teams, some of them look at it as opposition. So they look at it as a collection of individuals; right. So it just people are the people and they say, "Well, how does, you know, gender diversity in the team affect something," for example. So that's a compositional variable. People who do network stuff move to the relational level. That's the second one there. And in the

relational level you start asking questions, "How about the connections between people in the team going to influence the performance of the team?" And the third category is to say you can look at the people and then you can look at the task they are doing, and create a bimodal network. That is to say, "I'm going to see how these people relate to these tasks and to what extent are teams assembled around certain artifacts," like Wikipedia pages, for example. And then the fourth one is to what extent [inaudible], because we recognize the world we have teams that overlap, people are multiple teams in the same time and they come in and out of teams. To what extent are your teams in external connection, not internal connection as we saw out here, but external guys with other teams, in the ecosystemal [phonetic] teams, how does that influence your performance of your team? Okay? So these are -- I'm going to get some quick examples that run through all of these approaches. I'm going to get several examples here. The first one will be from a project we've been doing on Virtual Worlds. Virtual Worlds is people who online games who come together. I know it sounds very frivolous. At least I know Jenny thinks it sounds very frivolous. She's given me a hard time about this on a couple of occasions because, you know, "Why don't you study important societal problems rather than study -- seeing people who play silly games online?" [Inaudible] --

>> Male: Have you been written up in a congressional report yet?

>> Noshir Contractor: Beg your pardon?

>> Male: Have you been written up in a congressional report?

>> Noshir Contractor: Three of them.

>> Male: There we go. [Laughter] [Inaudible].

>> Noshir Contractor: And I say that with pride because some of my best colleagues like Nicole Nelson [inaudible], and many others were also written up in those reports. And I know why you are sensitive to that. [Laughter] [Inaudible]. But actually it's jokes aside, and I -- Jenny's right, but I think that there are some really interesting insights that we can get about the serious issues that you work by looking at those. And the question, of course, pivots on, "Do we think that people's behaviors are --" how many of you know about "World and Warcraft" [inaudible]? Oh, I love this crowd; no explanation required. So, you know, what can we learn from people's playing it, and that really translates into serious work. Well, [inaudible] shows that in many cases it translates quite well when we have done that kind of work. And we are not alone in this respect. There was an article that was published in "Harvard Business Review" that by Byron Reeves and Don Malone and Tony Briscol where they make the case that actually studying this it's important, not only because it has the kinds of expertise that you would expect of -- you know, the findings that were general last year [phonetic], but the second reason is because maybe you know how young colleagues today are playing those games and learning leadership skills there. And so for those of you who have played on one of these games, you know that if you get together in gills [phonetic] and gills go on these big quests -- that where people from gills go on these big quests or

raids, there could be 70, 80 people, and it's often coordinated by some 13-year-old guy somewhere. And so they are learning their leadership skills, so being able to understand this is good in and of itself, and also because it translates into other cases. So I want to take two cases. We have data from Sony Online Entertainment that I'm going to talk about here. It's a thing called "EverQuest II." It's like "World of Warcraft." And the basic idea is we were interested in asking two questions, "How does the virtual team composition influence the crew [inaudible]," and the second is, "What are the motivations for creating this?" So the first one is in that slight [inaudible] it's at that composition level. So this particular one was a project that I did with my now former student, Mook Shaw [phonetic], who just got a job at [inaudible] in communication and computer science there. And so we were looking at combat groups in EverQuest II. These are groups that come together to go kill something, kill a monster, right? And we looked at one week of data. This is service site data, so it's [inaudible], and this is dream data for people who do group research, right, 8,400 players, 46,000 plus groups, 9 million combats. That's the kind of data we can dream of when you study group behavior. And what did we find? We basically hypothesize that the diversity of the group will impact performance, and that approved members [inaudible], that is the group member has played with a lot of other different types of people in other groups. You know, that might affect it. So this is sort of like an external group connection, because they bring experience from all those groups. Well, what did we find? Well, first of all I should say that performance is very nicely measured in these games. They had 4 measures of how many points you gain, how many non-play characters or monsters you kill, how much gain in level for you, and how many times you died in the game, because as you know, in these games you can get killed and then you come back after a while. So it's a negative measure of performance, if you will. And what did we find? Well, we found that diversity in fact does help the group to achieve more. So if you have a group that has people from different class characters, so these are people who could be fighters or maids [phonetic] or priest, and as you have more diverse groups you improve experience points, you improve non-fear character, and you improve your level game. So it does make a difference the compositional diversity in the group. Interestingly, their cosmopolitan also makes a difference, but for a different one, for debt. In other words, people group people -- there have in groups where people have played with a lot of other different types of people, so people are cosmopolitan. It doesn't help you gain positive performance, but it helps protect you against negative performance indicators, like being killed. So that's an example of how we can think about compositional level as affecting it. Running quickly to relational level; and you'll be looking at the links between people and seeing how those links might affect it. Here this was a project I did with Yuan Wong [phonetic], who is a [inaudible], Brian, who's [inaudible] as a post-doctorate at Northeastern, and Jeff Treen [assumed spelling], who just took a job with TT Austin. You can see there's a lot of churn once the people who are leaving Northwestern at this point, just as we welcome you people like somebody who's joining Northwestern. So in this case we had network data from the game to see how these people in a technologically media way were coming together to participate socially. So the data for this was just 3,000 players, very small dataset taken from across the -- from the US server. And we were asking questions about

this particular set of network. So this is a partnership network. Who do you play with in the game? Who do you join with? So this is where you come together and said, "Why do I team up with someone [inaudible]? What is the motivation for teaming up with them?" This is the IMing network. This is who you mail with where you mail -- it's different from IM because you can actually mail things and [inaudible] artifacts within the game. And this one is the train network. When you look at the -- the train network is when you can buy and sell goods in the game. So as you know, you can buy things to make yourself better weapons and get yourself better armory and so on and so forth, provide medicine to heal yourself. And so going back to the partnership network you can notice that the middle is what is called "one large component," sometimes reflected in a giant component. Those who took Mark's [inaudible] workshop know all about components and networks. And these are all -- it basically says anyone's connected to anyone else. The black here refers to male and the red refers to female. And so everybody's connected to somebody. You've played with somebody who played with somebody else who played with somebody else. But on the perimeter you also have these people who only played with 2 or 3 others. Can you guess how that happens? Why is it that they've only played with a couple of other blacks or a couple of other reds; and that is it, they didn't play with anyone else in the game. Who do you think they are? Any thoughts?

>> Male: Fathers and sons?

>> Noshir Contractor: Yes. It turns out it's family members. It could be siblings or it could be romantic partners, and they just -- they go online to play only with their closest offline friends; right? So that's a large part of what we see out there. And that's an important finding because -- an insight that we'll talk about a minute more, because we have this notion that when you have technology you could communicate with anyone, anytime, anyplace, and yet the predominant evidence shows that you're communicating online with the same people that you communicate offline with. So when you look at these data then, the train has a similar kind of story, which is quite interesting. In the train dataset you see that there's this very heavy and intense knitted thing inside, right? Who are these people who are constantly traders, et cetera? Well, it turns out that in that sort of netting knitting world mess there, there are some very interesting insights that Brian Keaton [assumed spelling], in particular in my lab, along with some colleagues at University of Minnesota have been looking at. And they found that sometimes the people who trade are trading illegally. Well, what do you mean by illegal trade? How many of you know about "Gold Farming?" Some of you have got "Gold Farming" not as many as "World Warcraft." Well, Gold Farming is the same but you play it as illegal activity that's mostly in China, as it turns out, but as you start with the level one character, and you start playing in one and you start playing a new level two character, and you play the character doing repetitive tasks going and getting things, and getting your level moved up so that by the end of the day you've moved from a level one to a level 17 character. That's a high-level character. And now you sell that level 17 character to some lazy Westerner who didn't have the time but has the money to buy a level 17 character so that all of a sudden the game is populated by people who don't know what they are doing but they have a level 17 character. Right?

So of course the game and the regular players don't like it, and so Sony doesn't like it. It's the part of the reason why we were able to get these data from Sony is they asked us, "Can you help us figure out how to get rid of these people? Can you tell us something about the networks?" Well, these are smart people, and so it's not like, you know, if you just say, "Oh, if I see a stranger who's selling things to a lot of different desperate people I'll just knock them out and close out their accounts." No, no, no. They know that's going to happen so they engage in very sort surreptitious behaviors. And so it turns out that there are some real interesting structural signatures associated with how a gold farmer will start the character and sell the level 17 character, not initially sell it, gift it to a friend who then leaves it in someone's house in the game, who then sells it for a minor price to an accomplice or a fellow accomplice, who then will sell it to the final person, et cetera. So they have these very complex patterns, and we even study -- we've compared the structural signatures of people who are gold farmers and found that they have a lot of similarity with the structural signatures of drug trafficking dealers on a dataset that was collected in Canada. So what you see here is that there are some really interesting ideas that you can look for in offline behavior that are being mapped quite well in these online works. So what did we find here? We found that going back to the question of who do you team up with, selectivity is very important. You don't play around at [inaudible] people. You play with friends are friends, so that was the balance theory. You play with people the same age and having the same online game experience. And you play with people who are -- you're 22.6 times more likely to play with someone within 50 kilometers and between 50 and 800 kilometers. Proximity matters, okay; people are playing with people close-by because most of the people they are playing with close-by are their offline friends, et cetera. We also found that time zones matter, that you're 1.25 times more likely to play someone in your same time zone than the next time zone. And oddly we found that [inaudible] for gender did not matter. In other words, you are not more likely to play with the same gender people. So this is where what we do [inaudible] on a team one of our buddies, friends, Tracy Kennedy, and so Tracy's on our team and we were [inaudible] finding. And so we had a bunch of people --

>> Male: Tracy was also web shopping [phonetic]. [Laughter]

>> Noshir Contractor: Okay.

>> Male: She sat here.

>> Noshir Contractor: And she sat there.

>> Male: She sat right behind Mark and that's where I sat.

>> Noshir Contractor: And there you go.

>> Male: She sat on Dave's lap I think.

>> Noshir Contractor: I'm not going there. [Laughter] Okay. So Tracy and her team of ethnographers were able to find things out. They said, "Well, it turns out that --" oh and we did -- she asked us the question, "Well,

is it males who don't like to play with males or females who don't like to play with females?" And we found out males like to play with males. This was, again, part of the analysis, but females were not particularly likely to play with other females. And so then we found out the reason for that. And it turns out that there are many females who like this game and they love playing the game, but a vast majority of women who were talking with and observing, we asked them why they play the game, it turns out that they play the game not because they like the game, but it's the only way they can hang out with their significant other male partners who love the game. So it's the football widow idea, right, that we are part -- and the football season and all the women will get disenchanted with it. And so in fact that was what was happening out here is that many of the women were -- the reason we didn't find this effective is that a lot of women just play the game to hang out with their significant others. And so that's a nice example of how ethnography can work well with the computational techniques and the kinds we've talked about. The third one I'm going to look at is multi-modal network levels. So here the idea [inaudible] saying, "How do you --" we are leaving the realm of virtual games now, and saying, "How can you look at teams like this and see how they're more likely to join certain tasks that is there out there?" So this was work again that I did with -- this is Brian's dissertation, by the way, and it is work we did with our colleague at Northwestern, [inaudible]. And here what we looked at is how are the teams that come together to -- in Wikipedia -- in particular he's interested in breaking news at Wikipedia. So that when you look at the self-organization of teams that come together in the event of a disaster to come and edit a page. You could think of that as a team that comes together. And we wanted to understand a little bit more about how those teams came together. So there's really interesting findings about the earthquake that was there in Japan. And if you look at the Wikipedia, the Japanese Wikipedia article was up within 11 minutes of the beginning of the earthquake. The English Wikipedia article was up within 33 minutes, within 90 minutes of the earthquake. Front page of Wikipedia had been vetted by 12 editors. And the English article had 220 editors. This is -- yes, this was within 19 minutes, right, and 82 editors. Articles were also available in 12 other languages in Wikipedia, and the New York Times had one wired story, and one story bought. So it gives you a sense of the difference in which technologically social participation can make a difference in this day and age here. So when he looked at here -- I'm not going to talk much more about his findings other than there were a few pretty pictures, because this is his station and I want him to be able to preserve the conduct with that. But it shows -- it's interesting to see how you can begin to look at the longitudinal nature of the kinds of revisions. But these are the articles that were created, new articles, new revisions, old articles that were revised, articles about earthquakes and now added this piece of information on, et cetera. And you see -- and old users, that is new people who have never written a Wikipedia article and came online. And you see some very interesting changes in it. The site that was here is coinciding with when a nuclear disaster happened and there was a new group of people who came on, people who would not have been interested but were experts in nuclear materials, et cetera. So and I know we had this stuff there. And then we looked at the co-authorship by [inaudible]. So this is a multi-modal input. The blue notes are people, the red notes are Wikipedia pages. The Wikipedia pages are

larger. This is done in -- this is done in NodeXL, right? This is NodeXL. I'm an early and often user of NodeXL as marketing as well. And so what we had here was -- and what we did was we had -- it was a picture that would show us how -- there's some really interesting insights that you have here. So that's an article, that's an editor, but this was a big article on earthquake and tsunami. These were pages about towns and cities that were affected by the tsunami. This was about the nuclear accidents. You see they're very different types of people who are writing these different articles in different areas. And we -- and also you have -- you know, these are all people who are single article contributors. They only wrote one article and that was it. And then you had some generalists who are contributing to many topics who are writing on a lot of different Wikipedia pages. And then you had specialists who were only contributing to the nuclear crisis idea, for example. This got picked up and one thing that Mark did was -- no, I'm sorry, that Brian did, was he took these pictures dynamically and created an incredibly powerful video of an animated version of the NodeXL graphics that got picked up in Fortune and Forbes Magazine. And so as a grad student I went to several conferences for the next two months where people said, "Do you know Brian Keaton? Do you know the famous Brian Keaton?" And I was very proud to say yes I did know Brian. I'm going to go to the last one here, which is the ecosystem level. The ecosystem level is talking about not staying inside the team. Most the work we've done is looked inside the team to understand how teams come together and why they are -- why they have a certain performance. In this case we wanted to see the effect of the ecosystem which teams go. The basis is actually taking a bug and making the feature [phonetic]. Most research on team says, "I'm going to assume that people belong to one team, and I'm not interested in overlapping teams. That contaminates my data. When I do experimental research in the lab I set it up so that people are by definition on different teams, and in the real world when I have people who have overlapping teams, I think of that as noise or I try to remove that from the dataset, because of the dependence problem when you're doing statistical analysis and how you know whether the person -- how that person influenced two different teams." We think it's time to turn that around, be realistic of the fact that in the real world we have teams that overlap. And it's you are thinking of this as a bug we need to think about making this into a feature and see how that might explain why some teams do better than other teams. So this is work that I'm doing with Alina Logiani [assumed spelling], who's a PhD student in technology and social behavior. Dochio [assumed spelling] and Dorothy are post doc students at Georgia Tech. And I forgot one name who is a part of this, and it's mentioned here, it's Leslie Church, who's a faculty at Georgia Tech, and runs the lab where Dorothy and Dochio are working. So here we wanted to assume -- look at the Assembly assigned to the research teams [phonetic]. So what we've been doing is we've looked at -- we had for a short while I was on the advisory subcommittee at [inaudible] to computing and information sciences and engineering directory. And we -- and for a short period of time I had highly secure access to data about funded and unfunded NSF [phonetic] projects. And here's what we found [inaudible] I'm going to just say this already and then talk about the ecosystems. What our first analysis was how can we distinguish why people come together to write a proposal? What are the reasons why you join someone to write a proposal? And what are the reasons why when you join someone that -- well some

proposals are funded and some are not funded. So you could think of that as a performance measure; that at least the project was funded as opposed to unfunded. And what did we find? We basically found that people -- and we have several factors, but two factors that are network related that I'll touch on is that people who have coauthored with each other were more likely to submit [inaudible]. This kind of makes sense. You already have a history of working together. We also found that people who site each other a lot are more likely to submit reports. Well, if I know what Ben does and Ben knows what I do, we've cited each other, we are more likely to know what each other does if we cite each other. And then we're more likely to submit a proposal. However, then when we wanted to distinguish between successful and unsuccessful proposals, what we found is that in fact submitting a proposal with someone that you have coauthored then is actually going to make it 4.2 times more likely to be funded. So sending a proposal with somebody who have coauthored it definitely is a good thing. However, submitting a proposal with someone you cite and who cites you reduces you're likelihood of getting funded by 1.6 times. You're less likely to get funding if you submit a proposal with someone you've cited. You think, "Why this is possible?" Well, it says that what happens is that when two people don't cite each other that much it's because they represent two different intellectual communities. They are both citing their respective intellectual communities. And what they are doing is by their collaboration they're bringing together incuminent [phonetic] ways of different pockets of expertise thereby increasing my data, coming up with a [inaudible] proposal which is meant more to get it funded [phonetic]. By the way, we've [inaudible] this now with some [inaudible] proposals as well, and so this is a kind of interesting finding that you can get by looking inside the team. [Inaudible] to end with is how do you look outside the team? And what we are saying is that you can think of the team -- as we have a full focal researchers, each of them are circling around the team. So now we want to represent the team and not representing it by a bunch of lines, because that's the problem. If you represent it by a bunch of lines, you don't know whether the three lines between the three of us represents one team or it represents a team of two players, a team of two here and a team of two here. Right? So that's why you circle it, and that's what is called a hypo-graph [phonetic], as opposed to a regular graph. And a hyper-edge is different from a regular edge. The regular edge, as Mark told you, is an edge between two notes. A hyper-edge can have connections with more than two notes, could be there, four, five and so on. So here you see these hyper-edges around these things, and the goal is to say, "To what extent would I predict that these four people, based upon their overlapping team memberships with other people, are likely to come together, and then are they likely to be successful in the setting?" So it says, "Teams are not assembled in a vacuum. They're part of an ecosystem of teams." And there's just very little work that I have come across that looks at ecosystems of teams to see how that might explain the assembly of a new team. So we have all kinds of questions here. What are the ecosystem characteristics that lead to team assembly? Are there a few key teams that dominate the intellectual discourse? In other words, how do we know that a particular area people are likely to submit proposals, to go back to the other [inaudible]? What is involved with the existing set of teams that tells us, that signals us what's going to happen? And the second is are they -- do they need to be part of a coherent neighborhood? Are

proposals more likely to be written in areas where teams are already having a lot of overlapping ties, that is people are connecting to others who are already working together; so one member of one team is also another team. And then how is the immediate local neighborhood, that is the team that is submitting a proposal, is their immediate local neighborhood pretty coherent or is it -- are they highly embedded? And is it three of us work on a team here are we more likely to work on our team, if each of us belonged to teams that has a lot of overlap with others of us who work on teams? So that will make us highly embedded. Is that more or less likely to increase the creation of teams? And what we found here is -- I'm just going to jump in pretty quickly, what we found here is that there are some really interesting results that come around this. I'm not going to talk about statistics here. I have to talk about each one [phonetic]. We found that you're more likely to see proposals written or come together, or papers for that matter -- and this we actually based on papers written, that papers were more likely to be written in areas -- no, I'm sorry, this was [inaudible]. And so what we found that more proposals come in a field where there's already some amount of general coherence. And as you have some amount of connections between the people, there's a lot of overlap amongst the teams, the more overlap there is amongst the teams in a certain disciplinary area, the more likely it is that people are writing proposals. So it kind of makes sense because it says there's some intellectual coherence within this area, and that there's a tipping point at which point it's more likely for teams to begin to emerge, for more people to work, because they think it's a little to do with area. I think Jessica's our example of what we do ourselves. I think the area of TMSP falls into that category. We have enough of overlap amongst the various people -- and [inaudible] has an interesting network graphics of people who have collaborated with other people within this area. So we now know that through workshops like this, this field is a level of coherence where it's more likely that people will be submitting proposals, or will be writing articles itself. And that said, the second part of it is also interesting. And that is you're more likely to submit a proposal and assemble into a team in a field that is coherent, but you're more likely to be successful in the proposal being funded if your immediate neighborhood is actually less embedded than the larger area. So the larger area you want to be fairly connected so that you will have a chance to feel legitimized about doing what you're doing. But your local neighborhood actually should be less connected so that you are brokering between groups that have not been connected together. And if you are highly embedded then you are less likely to come up with a good idea because what you'll find is that you're surrounded by people who are in highly incestuous team relationships. So you want to be in an area that's somewhat connected, but your immediate neighborhood should be less connected in order to average in the outside [phonetic]. So, again, these are the kinds of findings that you have not thought about or theorized about before, because we've not had access to these kinds of data before. I want to end by saying that we've talked about understanding, and it's now time to talk about how could we take what we have learned in our research and help to enable TMSP? I was in a hurry, sorry, coming out so I haven't had a chance to fire up my browser so it will take a minute to do this. But I want to talk about how we build recommender systems based upon the social network theories and research results that we have [inaudible]. So let me

go back to it. So here is our research project where our goal was to try to take what we have done and build better networking tools for scientists to come together in research teams. So this is a website that we developed for people at Northwestern. It's for people who do work in the area of clinical translational science. I need to move this back one. And it's a recommender system. But it's a recommender system which is based upon taking data from each person, seeing all the articles they wrote -- it's linked-open data so that makes it easier to do the work with this, and it's a research networking tool. And just like in Facebook you get recommendations on who you want to be friends with, or on Amazon you say if you're interested [inaudible] some other people would be interested. What we do out here is we try to put a recommender system for people who are interested in certain topic and want to form teams. So here you look at the top -- and I'm almost done, and I think we'll be on time. These are all the researchers who were part of this particular Center on Clinical and Translation Science. This is an NIH effort. So let's say I'm [inaudible], and let's say I'm interested in some topic -- I've already put the word "brain," but if I didn't you'd get auto complete on all the key words in the articles that any of the people in the network wrote. So let's say we stick with brain, and then I say, "Okay." What becomes interesting is how the recommendations are made. And you recognize some of this. You can ask for a recommendation by motivation. So self-interest is most qualified. Show me the most qualified person [inaudible]. Or show me a friend of a friend, or show me an exchange, the social exchange mechanism, or birds of a feather, the [inaudible] mechanism, or globalizing the collective action mechanism. So we could say, "Use one or each of these criteria based on where I'm the network, and amongst all the people that are experts on brain, rank all of them based on that criteria." Or you could use the "I'm feeling lucky" approach here. And the "I'm feeling lucky approach" is based on a statistical model, which remember, a model set is a little bit of this, a little bit of self-interest, a little bit of [inaudible] and so on. So we run that model and come up with the best results based on a combination of these [inaudible] motivations that might influence this. So this is now almost done with. You can make a recommendation. I'm taking it your basis of most qualified, and so I chose the most qualified here, and it shows me the names of the people. You will recognize someone like Mark Williams, Shi Wong Chen [assumed spelling]. He's our most qualified because either they wrote a lot of publications, or the ones they wrote were cited a lot. You can then also click on "wide." This might take a minute because I don't know whether I launched the Java before. There it is. So it puts me in the center there, and shows me the people that recommended and sized by the amount of citations they get I can double-click on this node and then it will bring the information about that person. The nice thing is the information it's bringing me about that person here, his coauthors, his citations and so on and so forth, it's based on linked open data. So we are not keeping the data centrally. Wherever the data might be, because there are links to these people it's able to bring all that data in. And so that's the way in which we work that recommender system. And of course, you know, the same thing, you can say, "Well, it says that I cite him, but I'm not sure where I cited him." You'll click on that link and it will bring up -- I'm going to do it in another window somewhere. It will bring up the citation [inaudible]. Here we go. So there it is. It says, "[Inaudible] cites Mark Williams," and it

gives you the articles in which he was cited. So what is nice is it tells you -- it's able to take the knowledge we have for the kinds of theories we have. And this is a recommendation that is built for scientific research, but it's the same kind of situation that could be built in a lot of other context. And we're thinking about doing this one. We have a project we just [inaudible] to look at how to build up mentoring ties between practitioners, activists, advocates and scientists women who are doing -- are activists and advocates and scientists in the area of sustainability. So this way we're building a recommender system along the same lines as [inaudible]. So, again, to go back to the [inaudible], I hope that what I've tried to do during this presentation here is -- oops, sorry, is to say that we have been able in good point in times in terms of helping assemble the next Watson team, if we may. And I'm not going to go into the details of these slides here, but as I was hoping you've seen here is that we do have theories that were created and new ones like the hypo-graph [phonetic] that we had an opportunity to create. We have some really interesting combination of methods that we could use in this area. Obviously we have fantastic data, and I really recommend thinking more about the kind of work. Linked-open data is one area. Many of you know that Google came out recently with what they call "the knowledge troth," or "the knowledge net," which is very similar idea. In fact, some of the people who have been studying semantic web technology say, "Well, Google -- the good news is -- the bad news is Google stole our idea and gave it a new label. The good news is that finally it's getting some visibility." They were talking about this quite unsuccessfully in the general public. And then finally we have also obviously all of this stuff meets really high computational in structure in order to be able to [inaudible]. I will stop with that and acknowledge our funders; and thank you again.

[Applause]

>> Female: Great talk. We've got about five minutes for questions, and maybe Dave would like to set up -- well [inaudible] some suggestions. Who's going to be first?

>> Noshir Contractor: You could wake us now.

>> Female: Yes; endless amount of content, really wonderful example of how network research ends up in product that it's really helpful to people and can be used across a variety of different disciplines. Oh, it's close to lunch [inaudible].

>> Female 2: Wait, I have a question about the diversity, when we were talking about individual networks -- or individual consumer network and the "World and Warcraft" composition of the -- as gills. And you were talking about how diversity [inaudible] performance. Are we talking about user diversity or avatar diversity? I mean, I know that's probably [inaudible] --

>> Noshir Contractor: Okay. This is -- no, no, no, no, no, it's a very good question. In this case I clarified it afterwards so it was avatar diversity.

>> Female 2: Avatar diversity.

>> Noshir Contractor: So in other words what we were saying is that there are people -- those who may not be that familiar with the game, you have sort of 4 roles. You're going to have maids or a scout or a priest or so a fighter is somebody who goes on offense. You can have sculptures [phonetic] good on defense, you can have majors and magician who helps give you medicine, and the priest who heals you. And so having diversity of those characters makes it -- there's a whole bottle of work of literature in computer science, as it turns out, on what they call "team chemistry." And it's compositional work. And they look -- and some of you may be familiar with Shane Badyer's [phonetic] effect. Have you heard of the Badyer effect; anyone in this room? It's a great example. You know Shane Badyer, right, who used to play basketball at Duke, did very well, went into the pros, was heralded was going to be one of these fantastic players, didn't do so well, people wondered why. And then there was this one guy who was a great fan of his who ended up going on and doing some additional statistics and found that while Shane's stats were not great, what he found is that the stats of everyone on the team were much higher when Shane was on the court than when he was on the bench. So these -- or least that's what the Badyer effect, in other words he's not good but he makes everyone around him obviously much better. And it's not just by assists, in fact his assists score is not that high. And this got, again, another resurgence because this year guess what happened, shortly before Miami granted the finals, he ended up actually moving to Miami. Miami couldn't win all these years without him, and then Shane shows up and now they got the NBA championship this year. So those are the kinds of compositional issues. And so we've been looking at it even in our net piler [phonetic] research. And said proposals [phonetic] found women are not more likely to join teams overall [inaudible]. But when women join teams they are three times more likely to be a successful proposal than an unsuccessful proposal, which is -- which can -- there are two or three explanations for that. But one is well I [inaudible]. You want to ask something.

>> Female 2: But, I mean, part of what makes it easier seems to me that these are tangible ways of being able to establish what the strengths and the weaknesses of each member of the team is. And there's no real -- I don't know how to say it, like substantive way, you know, quantitative way to make those differentiations when you're talking about, "Would you like to write an article with me? Will you please give your resume of, you know, your writing skills, your research ability and your --"

>> Noshir Contractor: Oh, you have so much to learn.

>> Female 2: You know.

>> Noshir Contractor: You have so much to learn. We do that all the time. So consciously or unconsciously we are very cognizant by it. You see, what is that [inaudible] movie about --

>> Male: [Inaudible].

>> Noshir Contractor: "Money Law." [phonetic] So I actually would -- I would say that maybe we don't do it consciously, but subconsciously we

are constantly evaluating and making judgment calls, and there's nothing wrong with that, to make judgment calls about who we team up with based upon those tangible attributes. So I would say that we do it. We do it.

>> Female 2: We do it, we just don't have a quantitative --

>> Noshir Contractor: And we're getting there to do that as well. And that's part of what I [inaudible].

>> Male 2: Thank you for a wonderful presentation. [Inaudible] the scholars in your recommendations this time are predetermined, right?

>> Noshir Contractor: Yes. They are part of a -- the boundary of the network -- if they're asking how do you decide who's in the network, these are all members of a particular center at Northwestern called "NEW CATS," the new Northwestern University Clinical and Translational Science Center.

>> Male 2: Oh. Is there any way friend them [inaudible] and their name on the system [inaudible]?

>> Noshir Contractor: Yes, yes; good question. So last year I spent a week at Web Shop -- at this web shop, and I was telling earlier that I don't know how much you folks learn from the faculty but I can tell you the faculty learned loads from you because you have so many insights out here. So I can't believe they pay us to learn from you, and it's because of all these great opportunities. I bring this up now because today I said I'm not going to be able to stay. I'm leaving tonight, but I'm going -- the reason I'm leaving tonight is that tomorrow I have a presentation that is about that topic. So we have now developed -- and I didn't show it today here, but we've developed a networking tool that allows you to go beyond your home institution, that if everyone is sharing their data about this kind of information in linked-open data. So we are presenting tomorrow at a conference on linked-open data and research networking in Miami. And it's a paper which says you can put it in, we've showed -- we demoed the tree. You can put in the key word like "red," and it will show you people in your own institution, Northwestern, as well as Florida and Cornell, both of which have now put their data in this linked-open data format. So now the idea is that because of the standardization we can scale this interconnectivity to any institution that shares its data in the linked-open data [inaudible]. There are some political problems because some of these citation data isn't linked data, but it's not open data because it's provided by commercial companies like Elsevier [phonetic] and Townsend Writers' so there is some gated community issues that come up in that case. But technologically we are sharing that at this talk that I'm giving tomorrow.

>> Female: So thank you for a great presentation.

>> Noshir Contractor: Thank you. [Applause]