

*A Coalition to Protect Research
Congressional Briefing*

“Six Degrees of Separation:

**USING SOCIAL NETWORK RESEARCH TO INFORM PUBLIC
HEALTH AND NATIONAL SECURITY”**

Major Brian Reed, Military Fellow
The Center for Research on Military Organization
The University of Maryland

Katherine Stovel, Ph.D.
Associate Professor of Sociology
The University of Washington

Duncan Watts, Ph.D.
Associate Professor of Sociology
Columbia University

Sally Hillsman, Ph.D. (Moderator)
Executive Officer
The American Sociological Association

June 10, 2005
1539 Longworth House Office Building

*Transcript by:
Federal News Service
Washington, D.C.*

ANGELA SHARPE: Good morning. Welcome to today's briefing, "Six Degrees of Separation: Using Social Network Research to Inform Public Health and National Security." I am Angela Sharpe, the Deputy Director for Health Policy with the Consortium for Social Science Association, and Co-Chair of the Coalition to Protect Research along with Karen Studwell of the American Psychological Association. The Coalition to Protect Research was formed in the fall of 2003, and our membership includes 60 scientific and public health associations representing thousands of scientists, practitioners, patients, and individuals that are committed to promoting public health through research.

I would like to thank the co-sponsors of today's event, the American Academy of Political and Social Science, American Psychological Association, the American Sociological Association, the Association of American Medical Colleges, the Association of Population Centers, the Consortium of Social Science Associations, the Federation of Behavioral, Psychological, and Cognitive Sciences, the Institute for the Advancement of Social Work Research, The Mautner Project, the National Lesbian Health Organization, the Population Association of America, the Reproductive Health Technologies Project, the Society for Research in Child Development, and The AIDS Institute.

Now it is my pleasure to introduce today's moderator, Dr. Sally Hillsman, Executive Officer of the American Sociological Association. Prior to coming to ASA, Dr. Hillsman was responsible for developing and managing the External Behavioral and Social Science Research Program and the intramural research at the National Institute of Justice, the research & development arm of the U.S. Department of Justice. Before joining NIJ, she was Vice President for Research and Technology at the National Center for State Courts in Williamsburg, Virginia, and prior to that, Associate Director of the Vera Institute of Justice in New York City. She earned her doctoral degree in sociology from Columbia University. Please join me in welcoming Dr. Hillsman. (Applause.)

SALLY HILLSMAN: Thank you very much, Angela. It's a great pleasure for the American Sociological Association to join our colleagues, particularly our colleagues from the Consortium of Social Science Associations and the Coalition to Protect Research in sponsoring this briefing this morning. We are pleased that so many of you were able to come and that we have some representatives of the press here as well, because we think the subject of today's briefing is going to be extremely interesting and, we hope, productive for many of you who are here. All the sponsors of this briefing are eager to bring to this audience research from a longstanding and well-developed area of social science research, but an area whose importance is, I think, only now becoming increasingly visible to those outside the academe.

Social network analysis, at first glance, even to the most informed non-social scientist, might appear a somewhat boring or perhaps soporific, academic exercise. But I assure you that it is anything but that. Its significance is increasingly evident as a central tool for improving health, for preventing disease, and potentially for countering terrorism. The intent of today's presentations and the resources that we put in your packets to help you clarify how this science of social network analysis might make significant contributions to your own work and the areas about which you are most concerned.

To provide a little context for the three presentations we have this morning, let me borrow, perhaps a little crudely, from the social network analysis to convey the premise of this briefing in ways that are familiar to the different networks of people who are in this room. The title of this session comes from the 1990 play by John Guare. It was later adapted into a movie and it was entitled "Six Degrees of Separation." The play's premise was that each of us is connected by virtue of circumstance or acquaintances through social linkages that involve six or fewer people. That is, as one of his main characters in the play says, everybody on this planet is separated by only six other people. I am bound to everyone on this planet by a trail of six people. It's a profound thought, how every person is a new door opening other worlds.

For those of you more familiar with the sciences than the arts, you might recall the origins of this fascinating science of social network. The social scientist, Herbert Simon, said in his 1991 autobiography that, and I quote, "The purpose of science is to find meaningful simplicity in the midst of disorderly complexity." Our extraordinarily complex social world, Simon understood, is like other parts of the natural world. While it is composed of a seemingly endless set of highly diverse parts, we can, for the most part, develop fairly simple theories to understand these complex systems. Mathematics, of course, is central to revealing these simpler patterns and rules that govern these complexities of social behavior.

In one of the early experiments in social network analysis, the social psychologist Stanley Milgram, who was at Harvard in 1967, studied what was then called the Small World Phenomenon. He asked what the probability was that any two people in the world would know each other. He set out to learn by experiment how many intermediate connections or acquaintances were needed on average to find a common social connection between any two people. He was essentially asking, what is the mathematical structure of society? Since then, research on social networks has helped understand the dynamics of disease epidemics, aiding significantly in vaccine strategies. Small World Theory has been applied to analyzing the spread of news and rumors, aiding our understanding of how the social world responds to crises. It also has been used to analyze fashion. Maybe we should leave that for another briefing.

Obviously, we each feel we are at the center of a unique universe, and in fact, we are at the center of a socially- defined universe or socially- defined universes of co-workers, families, relatives, and friends. But in today's electronically and recorded world, many of these complex connections are more transparent and measurable than in a less-wired time. The computerized recording of our lives can unveil the networks in

which we participate, wittingly or unwittingly. The May 22nd New York Times, for example, published a network analysis of Enron employees, derived from e-mail correspondence without relying at all on the content of the messages – only the network transactions, not the content. The resulting pattern of connections within this corporate chatter is being used to understand and document the financial misdeeds of this now-fallen mega-corporation.

We are delighted you could all join us this morning to hear our engaging speakers and, at the conclusion of all of their remarks, I hope you will all join us in continuing this conversation with questions and with discussion. Let me tell you a little bit about all our speakers and the order in which they are going to speak and then I am going to turn the podium over to them.

Major Brian Reed was commissioned as an infantry officer in 1989 from the United States Military Academy and has served in a variety of command and staff positions. He received his Master's degree in Sociology in 1998 from the University of Maryland, College Park and taught at West Point in the Department of Behavioral Sciences and Leadership. Major Reed served as the Operations Officer in the 1st Brigade, 4th Infantry Division while the unit was deployed in Iraq to support Operation Iraqi Freedom I. Presently, he is a military fellow at the Center for Research on Military Organization at the University of Maryland and is completing his doctoral dissertation on social network analysis and resistance networks in fourth-generation warfare.

Katherine Stovel is associate professor of sociology at the University of Washington. She received her baccalaureate degree in Political Science from Stanford and an M.A. and Ph.D. in Sociology from the University of North Carolina – Chapel Hill. She uses social network analysis to study adolescent sexuality and sexually transmitted disease potential. She also studies how networks are used by employers and workers in the job-finding process.

Duncan Watts is associate professor of sociology at Columbia University, and an external faculty member of the Santa Fe Institute. His research on the structure of social networks and the dynamics of network processes has appeared in journals as diverse as Nature, Science, and the American Journal of Sociology. He holds a Bachelors in Science and in Physics from the University of New South Wales, and a Ph.D. in Theoretical and Applied Mechanics from Cornell. Last month, he was a key participant in the National Academy of Sciences' computational analysis of social networks symposium, in which the panelists explored computational, mathematical, and social science methods to better understand such things as web communities, economics and game theory, and the effects of connectivity on the marketplace.

We have a diverse group of speakers, all of whom I hope will engage your interest in the vast potential of social network analysis. If you would join me in welcoming our speakers, I will turn the podium over to Major Reed. (Applause.)

BRIAN REED: Well, before I begin, I would first like to thank the American Sociological Association for extending to me this invitation to come and speak today. It is quite an honor, and especially to be in the company of two distinguished scholars and researchers who I am on the panel with today. I would also like to thank all of you for taking time out of your busy schedules to be here this morning. My talk today is just a snapshot of a larger project that I am presently working on, that of social networks and resistance networks.

The network perspective has proved fruitful in a wide range of social and behavioral science disciplines. The diversity of contemporary research is a result of sociologists linking network concepts with a variety of technical and substantive concerns. Many topics that have traditionally interested social scientists can be thought of in relational or social network analytical terms. My research is centered on the role of social networks in providing enough information to infer an organization in the absence of a formal structure, but most especially those organizations characterized by secrecy and stealth. With this project, I hope to merge the academic with the operational by taking one idea, trying to provide a clear definition for it, and then applying this idea to other situations. In the end, it is my hope that the results of this project will be helpful in drawing policy prescriptions for a greater and more thorough understanding of resistance networks by offering a sociological application via a network theory approach.

Arguably, the present nature of warfare is characterized by decline and territorial control in the rise of novel actors and relationships attempting to have their voice heard on the world stage. Such actors are intent not on just short-term gains and goals, but with the inclination to make a long-term difference. The essence of contemporary warfare is the symmetry and asymmetry of will and means. The enemy of contemporary conflicts is network-based, flexible, and trans-national in scope. And it is my position that the tools and concepts inherent in a network analysis approach allow for a more complete understanding of how such network-based enemy systems behave, and how that behavior is affected by their connectivity. Of note, however, is that asymmetry is not something new. The Bolshevik Revolution in Russia was an insurgency structured as a network of cells, and in that tradition, so was the Viet Cong in Vietnam.

Given that the term “social network” refers to the set of actors and the ties among them, and that the network analyst seeks to model these relationships in order to depict the structure of a group, how does one depict such a structure if the group does not want to be known? Most network analyses are based on obvious sources of data regarding the nodes of the network, frequently, by simply asking the network participants with whom they are connected or with whom they associate. However, we cannot do this with, for example, the resistance in Iraq or al Qaeda in Afghanistan.

So specifically, when dealing with resistance networks, how do you discover a network that focuses on covertness and coerciveness? The intelligence and information work that we did in Iraq is known as the intelligence preparation of the battlefield, and from it, one can infer the relationship between this process and sociological research. The intelligence background and link diagrams that we built were rooted in the concepts

of network analysis. The intelligence preparation of the battlefield, or IPB, is a vital command and staff function for successful counter-insurgency operations. It begins with a commander's estimate of how the enemy is fighting, as well as the enemy's strengths and weaknesses. It is complemented by a staff estimate. It becomes a hypothesis confirmed or denied by daily reconnaissance, surveillance, reporting, and command assessments. It thus evolves.

The purpose is to assist the commander in identifying targets, objectives, and friendly tactics. Its value is that it allows commanders to direct operations against the enemy. An IPB at brigade level will have many components based on the threat the commander sees. Link diagramming of various enemy cells is a key component to the IPB process. Terrain analysis is absolutely vital, as is an understanding of the diverse population. The enemy and the people are not monolithic. The understanding of the terrain and the population at-large give the commander an understanding of how to choose the time and the location of future operations, as well as tactics. This supported offense of operations that kept the enemy reacting to our will.

When examining a resistance network, the task of interpretation is to identify significant patterns of motivation, interest, and need, around which characteristic types of action are organized. Whether interpretation is focused on a single personality or a resistance elite, the problem is one of isolating a set of interdependent variables forming a unique system. When we have identified the system, we can tell something about the way the behavior is summoned and constrained. This helps us make predictions concerning certain likely responses in future action. On the ground, this translates into the painstaking work of building link diagrams showing how everyone in the network is connected and why. It involves detailing relationships by family, clan, prior associations and club memberships, business transactions, et cetera. It involves an ongoing analysis of the ties and links between network members.

To date, my research has led me to three main propositions concerning resistance networks as framed within the context of social network analysis. First is that resistance networks often do not behave like normal social networks. Resistance networks are both coercive and covert. In a normal social network, strong ties reveal a cluster of network players. It is easy to see who is in the group and who is not. In a resistance network, because of their low frequency of activation, strong ties may appear to be weak ties. The less active the network, the more difficult it is to discover. Resistance networks are characterized by incompleteness, that is, the inevitability of missing nodes and links. Resistance networks are characterized by ambiguous boundaries, and thus, the difficulty of deciding who to include and who not to include. Resistance networks are dynamic. These networks are not static and they are always changing.

Second, when analyzing a resistance network with the intention of eliminating one key node, the best solution for network identification may be to discover possible suspects and then map their ego networks, see whom else they lead to, where they overlap, powers of function of one's position in the network of communication, where this position is assessed not only simply in terms of structural centrality, but also in terms

of the power of the people with whom one is connected. People who are well placed in a communication network also tend to be the central players in terms of power and influence. Disruption occurs by focusing on the connections.

Lastly, a network analysis approach to understanding resistance networks allows for predictions concerning certain likely future behavior of the network, thereby allowing for a focus on both prevention and prosecution of coercive behavior. There is a clear division of labor within the network, for example, financiers, logisticians, operators, and decision makers. Control over resources and the importance of the node in the network is derived from the division of labor. This, in turn, gives some positions or groups more power or control over critical tasks than others. Power, then, comes from the control over resources. Disruption occurs when the power node is eliminated. Additionally, a resistance network lacks a protective bureaucracy, thereby omitting some checks and balances to guarantee secrecy and covertness and thus making it potentially more vulnerable to infiltration.

So in conclusion, I believe that there are two points that are worth highlighting. First, a requirement on today's battlefield is to make an assessment of the political and social architecture of the operating environment from both the friendly and the enemy perspective. And two, to successfully accomplish this requires more than a reading of field manuals, but also an understanding of the complexity of the nature of warfare. At least as it applies to understanding how the enemy operates, I believe that there is a value added to a network approach in to analyzing what that enemy looks like and how that enemy fights. Resistance networks often do not behave like normal social networks, but by asking what kind of social network is a resistance network, one is afforded a window into that network, thereby understanding what that network looks like, how it is connected, and how to best destabilize it. Thank you.

KATHERINE STOVEL: Thanks, it's really nice to be here, and I'd like to thank Angela and the other members of COSSA for organizing this briefing about what I think is a really rich and exciting area of social science research. I teach graduate and actually now an undergraduate seminar on social network analysis and this is a real challenge for me because there are so many substantive areas that we cover in this broad field, and it seems in some ways like every week we're kind of building on substantive and theoretical knowledge in a variety of different parts of social science. And so you're seeing a little of that diversity here today. I'm here to talk about, what in these classes I refer to as, the "sex and bugs week." And I'm going to be talking about the romantic and sexual networks of adolescents.

Social network analysis is embedded in the public health realm in broader questions of social and peer influence on a variety of health outcomes. And in this context, we often think of these influences, which are of great concern to parents and teachers, as coming in two forms. One is a concern with composition of a peer group, so what kinds of kids are particular adolescents hanging around with and do they get ideas or do they get influenced in their behavior from those friends or do they start to find each other simply because of shared interests? And this leads to lots of work on the

composition of peer groups. What I'm going to be focusing on more today is the actual structure of peer groups and the way in which the density or the expansiveness of networks can have an impact on certain kinds of behaviors that we have specific public health interest in.

But to give you a sense of the breadth, these kinds of social and peer influences of both forms, the compositional and the structural, have been shown to be important for effecting rates of drug and alcohol use among adolescents, eating behaviors, both over- and under-eating, exercise behaviors, suicidality, and what I'll be focusing on here today, the romantic and sexual behavior of adolescents.

Now, the particular problem that I'm going to focus on today and the work that I'm going to describe is sort of organized with respect to is the prevalence, the distribution, and the risk of sexually transmitted diseases, particularly among adolescents. And the focus on adolescents is not at all accidental, because the rate of a variety of sexually transmitted infection is extremely high among adolescents. And this is some data from the CDC that shows the age- and sex-specific distribution prevalence rates of chlamydia in the United States in 2003, which is the last year for which we have clearly reliable information. And you can see that there is a big – the largest prevalence among women is among adolescents 15-19, followed by slightly older women 20-24. We have many fewer cases of reported chlamydia among men, but here we also see a boost among younger people. A similar pattern with respect to gonorrhea, more girls, particularly the young girls have the highest prevalence rates of anyone of any age groups, and there's also a lot of gonorrhea among young adult women 20-24. Similar pattern, lots and lots of disease among young people in these two bacterial STDs.

We turn to HIV, which is of course of great concern, and we see a slightly different pattern [referring to chart]. These are reported cases of HIV infection for males and females. These are cumulative rates through 2003.

Young people 15-24 have lower incidence of detected HIV infection, but we do suspect that there's a lot of undetected HIV infection among these people. So many people who actually contract HIV probably contract it when they're adolescents or young adults.

So this is the background; this is the puzzle that we want to understand. Why do we have these high prevalence rates with particular STDs that are of concern from a public health perspective? To make sense of this, I first want to give you a really, really brief sort of EPI-101, fundamentals of disease epidemiology. And I know Duncan's going to talk to you a little bit about this as well. But for infectious disease epidemiology, really the key thing to understand is the reproductive rate, which is the number of infections that a particular infection will cause or trigger over the course of the infectivity of that host.

And this typically has been understood as the product of a family of parameters, the first of which is the transmission rate for the disease itself, and this is a biological

characteristic. How infectious is the disease; what is the likelihood of actually becoming infected with the disease, given contact with that disease or with an infected host? The second family of parameters is known as C, which refers to contact, and this is what's the probability of contact between someone who is infected, or an infectious agent, and an uninfected but susceptible agent. And the final family of parameters is the duration of infectivity. This is another characteristic that is typically thought to be controlled biologically for specific diseases. And this is, how long is the disease actually infectious? For a disease like HIV, it appears that the duration of infectivity is as long as one actually has the infection, whereas for something like the common cold, the duration of infectivity is much shorter, and by the time you know that you have symptoms, you're probably no longer infectious. You're just at the tail end of that infectious rate.

Now, this family of models is sort of the mathematical background for the familiar S-shaped curve that we observe with lots and lots of infections. But what's important for us – for me, as a sociologist, certainly – is understanding this contact parameter better. In a lot of infectious disease epidemiology for many, many years, contact was assumed to be random. That is, individuals who were infected were randomly bumping into those who were uninfected. Now, there's been some refinement on that, but still, we suspect that for certain kinds of diseases, assumption of random interaction is very poor.

And so, as a sociologist, we might begin to think that we have some intuition at least about who interacts with whom, which would create some biases in these contact parameters. Is it random or is there some kind of selection with respect to various attributes that people have? People choose different types of people purposefully as friends. They may be in environments in which there is some sorting on these attributes. Or, in the context of sexually transmitted disease diffusion, we might be interested in people who find one another who share certain kinds of behavioral practices or levels of activity. So beginning to think more seriously about the structure of contact should be able to inform these models and certainly over the last ten to fifteen years, there's been a lot of progress in better refinement of this contact parameter.

Now, Brian is a rare network person who didn't show pictures and didn't show any kind of dots and lines. But I am going to succumb. So I'm going to give you a brief – just an overview of the notation that I'm going to use and then just sort of comment among a lot of network scholars, where we often talk about nodes, which in a picture we might represent with dots. And these nodes could represent people; they could represent airports; they could represent companies; they could represent informal organizations that could be connected through transactions in time. And we connect these through lines, which represent relationships that exist between nodes. And again, relations are very dependent on the kind – the relations we're interested in are dependent on the context we're focusing on – so friendship, sexual intercourse, air travel can all be different kinds of relations that could be represented in a network. Now what's of particular importance here is variation in these structures, and so we can just imagine a little universe – or this isn't a universe – but a sample of types of network structures that could emerge when we

combine nodes with relations. So we could have a star; we could have people organize into a clique or a chain or a random kind of collection of people.

Now, if we take some of these ideas from network epidemiology and we try and translate them into a network framework, we can represent some of the traditional thinking about why disease might incubate and produce sort of a reservoir or an endemic pool of disease or even an epidemic under some circumstances. One of the core ideas in infectious disease epidemiology is that we might have some kind of preferential mixing among people who have a disease or are likely to get a disease if they interact frequently, and so we can see that repeated interaction will produce a disease and we recognized this in a network sense as a core. This is a group of people here surrounded by these blinking dots who all have many, in this case, sexual partners. They interact with one another and then particularly for diseases that might allow for re-infection, this is an important way of thinking about why we get kind of the persistent presence of a disease in a small population.

Now, we don't really know always if there's a core. We might imagine other ways that people could be arranged. We call this network structure a romantic cascade. It looks very different structurally from that core. Here we have a setting where, instead of having enough people who have lots of partners interacting with each other, we have them interacting also with people who have few partners, and it produces this kind of chain. This romantic cascade is characterized by two structural features. One, this long chain and the other is the absence of cycles. It's the cycles that create these redundant clusters or cores. Now, this kind of a structure has a very different disease potential. We could imagine this as the architecture over which disease would travel, and we can see that there is a possibility of a great deal of transmission, assuming that the disease is relatively easy to contract and is infectious over a long period of time. However, this is a very fragile structure because of the absence of cycles.

Now, we actually for a very long time had very little information about what these patterns of actual interaction looked like among people, so scholars could sit around and imagine all sorts of different configurations and play what-if games – what if we had a star; what if we had a clique; what if we had a chain – but we didn't actually know what real people did when their sexual partnerships were linked together into a structure. And luckily, thanks to the efforts of a lot of different people, we now have the ADD Health study, which provides really rich data about exactly this question.

ADD Health is a multi-, multi-everything study actually. (Chuckles.) It studies all kinds of different things, but it really was designed with social networks in mind. It was designed with the idea that there are many, many different influences on adolescent health and that we need to collect rich information about all of these different influences in order to get a more complete picture about how adolescent health outcomes are shaped and influenced. So it's a very – it's a contextual study where information about adolescents is collected both from the adolescents, from friends who attend the same school, from family members, and from characteristics of the school as well. This study was funded by 17 federal agencies, with lead funding from NICHD.

In the work I'm going to talk about now, we focused on one of the schools that's in the ADD Health study, what we called Jefferson High. Jefferson is a mid-sized school in a Midwestern town. This is a school in which every kid in the school participated in the study, so we have the opportunity to get information of all the kids reporting on each other, including information about their romantic and sexual relationships. We have data from 800 students approximately. It's a predominantly white school and we have really remarkably comprehensive information about their romantic and sexual relationships.

The data I'm going to show you is measured in wave one of the ADD Health study, which is the data collected in 1995, which amazingly is ten years ago. It's kind of scary. And so one of the things that we did first – well, first we did, although it took awhile – was actually try and map this. And this really was one of the first times we had complete information or relatively complete information about the romantic and sexual network of a population that interacts with one another and really not so much with a whole lot of other people, so this is fairly complete. And here we see – we learn a lot of things about the structure of interaction and begin to think, well, is there a core or is there something else going on here?

In this community we see that 126 people are involved in dyadic relationships and they have no other partners – that's those 63 dyads. A lot of people are embedded in fairly short chains of romantic relationships. Most of these are where they alternate boy-girl-boy-girl. But these are all connected and from a disease perspective, these kids are relatively low-risk, because they are in these fragmentary chains. The vast majority of kids, over half of the kids, in Jefferson, who have romantic partners, are imbedded in this large component – 288 people – which really does not look at all like a core, it's much more like a chain.

Now, – this structure is somewhat surprising and we have lots of questions about what forms it, because the absence of small cycles is somewhat curious. And so we have thought about this; why do we get this spanning tree, which is what this is technically called? And we believe that there is actually a local proscription against these kinds of cycles that we explain by thinking back to Bob and Carol and Ted and Alice, from the old movie, who were married couples. Eventually Carol and Ted had a relationship and the question is, will Bob and Alice get together or not? And if they do, you get a cycle. And if they don't, you get a chain. And what we observed is that actually, in these adolescent contexts, where kids are all watching each other, they almost never complete these cycles. It always remains as a chain and there's sort of an informal rule, it seems, among these kids that you don't do that. This could be because they don't want to lose status in the eyes of their peer. It could be because of jealousy dynamic. It could be what we very scientifically sometimes call *the yuck* factor it's just a little close.

And actually, it's sort of funny if you think about your own sexual network if you have one and if you have a sexual relationship, you might have a sexual network. And if you then start talking to someone else that you know – you know, late at night – it may turn out that you discover that you're in the same component. And very frequently, when

this kind of comes up, people just sort of say, eww, yuck. And we think that's sort of the same thing. There's sort of an aversion to this closeness, this closure of these sexual networks.

MR. Paul Gaist : Just real quickly, before you move on, in contemporary time and place, now you have to ask the question, if Carol and Allison get together, will Bob and Ted get together. And I don't say that just flippantly, I actually think that in terms of what are your cycles, you'd actually have to look at them way beyond this heterosexual –

MS. STOVEL: Yes, yes, I focus here – and that's a good point – and we focused on a four-cycle, which is the minimal redundant structure in a heterosexual network. But of course, where there's same-sex partnerships as well, you get a triangle instead of a square and you get redundancy that way as well.

Now the implications of these kinds of findings – and actually beginning to really have data about the structure of these networks, I think, is important. Now, we can start by thinking about a dyad. People know they're in a dyad. They know they have a partner. The question is, what's their risk because of that? And here, I want to focus on this pink dyad – we'll call it a girl – who has one partner. And we might imagine that's a pretty safe thing. But if that partner has partners, the potential exposure for disease may be much, much greater. Imagine that their partner has three former partners. Now, for a highly infectious disease, this person's risk has quadrupled. If we spin it out, we get much, much more exposure. Now, of course, if our pink star has actually been self-infected, they could also be putting that many people at risk themselves. So with this kind of a structure, we begin to get a different picture on the emergence of an explosive epidemic because of this structure of connectivity.

Now, here we have a spanning tree, not a core. Now, just to bring this back to Jefferson, here I've highlighted those adolescents at Jefferson who have only one partner, which is over half of the population. So, many of them are in small components and they are at the moment relatively isolated from disease if it gets introduced into the population. But many of these people are connected to this big component where, if disease enters, there is a great deal of potential exposure.

MR. Gaist: Is that only one partner within a given timeframe or –

MS. STOVEL: This is within the last 18 months. I've elided all of the temporal dynamics. We can talk about that in questions.

Now, I think we can now go back to the interventions that we might think about with respect to disease. And this is something that, in many ways, has been known with disease epidemiology before, and has influenced vaccine patterns as well. But if we have dense clusters and we break a particular relationship, we don't really change the structure at all, in fact the structure is quite robust to that. So in this respect, subtle changes, either by breaking a particular relationship by maybe not having sex with that person or by using a condom, we actually won't do a whole lot to change the overall picture of risk,

both for individuals – we may reduce the individual risk a little bit – but from the disease potential, there’s no change really at all. Even if we remove nodes in the core altogether, we still will have a remarkably robust structure.

In contrast, with a spanning tree or the romantic cascade, if we break a relationship, we actually sever the chain and separate people. And so in this context, very subtle changes to the structure of the network can have massive impact on the disease risk. And here we have – this is contingent at some level on the levels of beta – the infectivity or the duration of infection – of the specific bug. So if beta is very low, then it’s quite likely that the disease simply won’t be transmitted in a particular relationship. If beta is high, then these kinds of interventions are much more important.

This slide just actually duplicates what I’ve said. If we have dense clusters, the recommendation really is to target interventions at the high-risk people and really get them to maybe spend a lot of energy on working on behavior change, or the change in the structure of relationships. When we have long chains, we really have a different message, which is a broadcast message, because everyone is at risk and everyone plays a role in the transmission. And this is a moment where we can say, well condoms are one of the things that can break the transmission chain, as can not having the relationship at all, as can having a low-infectious – the particular diseases that have low infectivity, transmission may not occur.

So just to conclude, I think there are a lot of things that we learn from networks. The most important perhaps is that we really need to rethink individual risk. It’s not just what an individual does – how many partners they have or what kinds of behaviors they engage in – but it’s who they do it with and the macrostructure in which they’re embedded. If they’re in a dyad, no problem. If they’re part of a larger component or if they’re part of a core, it’s a very different story. And I think finally, I would just say that different interventions are really appropriate for different network structures, and until we have a sense of what network structure is in place that is relevant for a particular disease, then we don’t really know which intervention to recommend. So that’s it, thank you.

DUNCAN WATTS: [Please do not quote without permission] Thanks very much for coming. It’s nice to be here. I just want to preface my remarks by noting that last week in the op-ed section of the New York Times, there was an interesting piece by Senator Obama - about calling for a global warning and management system to counter the emerging and ongoing threat of avian influenza virus. Now, for those of you who don’t sort of read the newspaper very quickly and sort of the latest pages of the front section, you might not think that avian influenza, which typically affects countries in Southeast Asia, is a terribly interesting topic. But, one thing that they pointed out in that op-ed piece and which is absolutely true is that it is the single greatest threat to humanity. Alright, so that is not an exaggeration. Terrorism is a sort of minor and inconsequential threat compared to the potential for global death associated with an avian influenza pandemic.

So this is not a trivial matter. And on that note, I would like to point out that the current state of theory and knowledge regarding the dynamics of large-scale epidemics is extremely poor. And there are some reasons for that that I want to go through and then I want to try and suggest a direction that might help us build some sort of framework that is better at understanding these sorts of problems. So there's an obvious question that one can ask and I started thinking about this problem in 2003, when there was another epidemic going on of SARS. And a reporter called me and asked me, how big is this going to be? And I said, well, I don't know, and he said, well, can't you figure it out? And I said, well, no. And he said, well, how about an estimate, and I said, well, I can't even give you an estimate. And he said, what are you talking about? And I said, well you know, the way these things work, it could sort of be, you know, a few hundred people or it could be, you know, 50 million people. And we don't really know which one it's going to be. And he said, well, what is all this social network analysis good for then? And I said, well, that's a good question. (Audio break, tape change) – has a very good understanding of why he doesn't know the answer. So this will help you be an expert.

Okay, so this is a very natural question one can ask. And now, all these questions on dying epidemics like HIV/AIDS pandemic that we are experiencing right now -- also malaria, TB. And somewhat amazingly, we don't have a good way to answer this question. And it is not because people haven't been doing mathematical epidemiology in a serious way for very long. This is a subject that goes back to none other than Daniel Bernoulli in the 18th century when he proposed a very simple model of the smallpox epidemic. And really it has been going sort of seriously since the 1920s following a series of papers by Kermack and McKendrick.

And at this stage, I haven't done a literature survey of mathematical epidemiology, but it is an enormous field. There are many, many thousands of papers that have been written on this subject, literally hundreds of different models have been proposed that deal with all sorts of different variations, different diseases, different demographic influences. Kate mentioned some of them in her talk.

So there is sort of incredible diversity of models and it is kind of bewildering to get immersed in this literature. And what is sort of remarkable is how at a fundamental level, many of these models are very similar, as different as they look. And so this is basically the model that Kate also discussed. This is the model that Kermack and McKendrick introduced in 1928 and that most of these models are still based on. And it is called the SIR model, not, as one of my students once suggested, for a stupid, ignorant, retarded– but for the three states of a disease through which you can imagine individuals in the population cycling.

And the “S” stands for susceptible, the “I” stands for infected, and the “R” stands for recovered or removed. So as “susceptible” person is someone who is not infected but can get infected. An “infected” person is both infected and infectious. You can split that up if you want. It is not always the case that being infected makes you infectious. And recovered is when you either get better and you have immune response that prevents you from being re-infected, or else you're dead, in which case you're not playing any more

role in the dynamics. Either way, the “R”s are taken out of the system. Of course, you can lose immunity over time and this is why we see cyclical outbreaks of influenza, for example, because over the course of the year, the disease mutates slightly, a virus mutates slightly, and you can be infected the next year.

So the way it works is that infections are transmitted when “susceptibles” bump into “infecteds.” And as Kate mentioned, a key assumption that you need to make or something that you really need to know about is how people bump into each other. And the standard model is to assume that individuals bump into each other at random. And so the probability of an “infected” running into a “susceptible” is just proportional to the product of their population sizes.

Friends, this is a tremendously simplifying assumption because now you don’t need to model the disease at the level of individuals; you can instead model it at the level of populations. So this is called the mass-action assumption, which is a term borrowed from chemistry. And if any of you have studied chemistry, you will have seen that assumption in practice. You just imagine people being stirred in a large beaker just like chemicals, and the rate of the disease progression is the rate of the chemical reaction.

Okay, so this is a very powerful assumption. It enables you to write down differential equations instead of having to do very complicated simulations. And it has the consequence that there is really just one parameter that people really pay a lot of attention to, and that is what is called the basic reproduction number, or R_0 . And as Kate mentioned, it is the average of new “infecteds” that are generated by a single infected individual in a susceptible population. And so when that number is greater than one, and each “infected” is infecting more than one other person before the recover – and then so you can imagine there is this – you can see easily that there is going to be some exponential growth of the disease going on. If it is less than one, then the disease burns itself out. you – what, I have got 15 to go? Oh, okay, thanks. (Laughter.) That’s a fast 15 minutes.

Okay, so there is one parameter that people really pay a lot of attention to, and that is what is called the basic reproduction number, or R_0 . And as Kate mentioned, it is the average of new “infecteds” that are generated by a single infected individual in a susceptible population. And so when that number is greater than one, and each “infected”e is infecting more than one other person before they recover – and then so– you can see easily that there is going to be some exponential growth of the disease going on. If it is less than one, then the disease burns itself out.

And so there is critical value: $R_0 = 1$, which is called the epidemic threshold, and, you know, mathematically speaking, public health interventions are all designed to get R_0 below one. Okay, so I’ll have some more things to say about R_0 .

One consequence of this is that standard models have size distributions that are what we called bimodal. All right, so when R_0 is less than one, you never see any epidemic. So that is this peak over here [referring to chart]. So here we have a

distribution. This is the size, from zero to one. So zero is – no one is being affected. One is a fraction of the population, right, so this is 100 percent of the population is getting infected.

And you can see that really only one of two things can happen. Either no one gets infected, or else the disease – there is an epidemic, the disease spreads, and you get a very large fraction of the population infected. And this is just a standard result that comes from almost all of these models. You can go a little further and you can actually predict what size that ought to be. And here we have a plot of the reproduction rate, R_0 , versus the fraction of the population that is infected in its final state [referring to chart]. And you can see when R_0 is less than one, there is only one state of equilibrium where no one gets infected. And then above $R_0 = 1$, you have two stable equilibria or two possible equilibria. But you can predict exactly what they are going to be, right. So if you know the population size in and you know zero, you ought to be able to answer the reporter's question and tell him exactly or approximately even how big the epidemic ought to be.

Another feature of this standard model is that you should see a very – a smooth series of cases. You only see one – what we call the epidemic curve only has one peak. So this is cases per day [referring to chart] that are generated by the epidemic and you see that it starts off very slowly and then it peaks, and then it starts to burn itself out when the susceptible population gets exhausted. And so Kate showed the S-shaped curve. This is just the derivative of that curve. So the S-shaped curve would be the cumulative number of cases. This is the actual number of cases per day, so you have this very striking and simple shape compared to these sorts of curves.

So that is what ought to happen if epidemics progress as the standard models suggest that they should. You should be able to predict the final size and you should know about in the course of the epidemic you are.

Okay, so in reality, things are very different. Epidemics vary dramatically in size. So the first three examples here are all maybe influenza epidemics that occurred in the last century [referring to slide]. And you can see that there are some sort of staggering numbers. In 1918-1919 pandemic, over 500,000 deaths in the U.S; no one knows how many people died worldwide. The estimates range from 40 to 80 million. This is the largest loss of life in the history of the planet. So it is pretty serious.

You know, 40 years later, you get the same kind of virus spreading and you get another disaster, but it's ten 10-times less bad than the 1918-1919 epidemic. You get another one ten years later. It's called the Hong Kong flu because that is where it was first detected, and, again, it is colossal numbers here, but half of what happened ten years before. And then finally in 2003, we had a SARS epidemic – very similar. It spread slightly differently from influenza, but it turns out it has some similar properties. And no many knew how many people were going to die. And as it turns out, it was a bit of a fizzer – only 800 deaths worldwide. And people afterwards said, well, you know, that is ridiculous; more people sort of die falling over in their bathtubs than in this SARS epidemic. Why was everyone making such a fuss about it? Well, we'll get to that.

Okay, so all of these diseases have about the same value of R_0 . And so according to the standard model, they should all behave in a similar way. And yet, they have obviously – you know, orders of magnitude difference in outcomes. So it is hard to be any more quantitative than this because the historical data is – at least I haven't been able to locate good datasets for avian flu. But there is a dataset that we do have, which is extremely comprehensive and it deals with Iceland.

The Icelandic people are extremely good at recording things, it turns out. So they have 100 years of data on all epidemics that have occurred for a variety of childhood diseases. And the two that we have here are measles and whooping cough. And what we are looking at is the equivalent of that distribution that I showed you before that had only two peaks. And as you can see, these distributions do not have two peaks; they have many peaks. They are what we call “multi-modal.”

So, you see, you have a very big peak at the beginning, just like in the other plot, where lots of these epidemics are very – these outbreaks are very small. But then once the epidemic takes off, lots of different things can happen, right. And so you can see a very large range of possibilities associated with different epidemics in Iceland. So this is a very sort of odd-looking distribution I would say. And if you're not used to looking at probability distributions, you might not think it looks odd. But this is a peak with a sort of long flat section is a very strange distribution indeed. Nevertheless, that seems to be what is going on.

The other characteristic – switching scales and diseases here – is that real epidemics don't tend to display one peak, right. In fact, you look at the caseload data for the 2003 SARS epidemic, you see many peaks and valleys. So the disease is taking off and then it's dying back down again, and then it gets somewhere else, and it takes off again, and then you get a big peak, and then it starts to die down again, and then it affects somewhere else. So you can imagine that as the disease is hopping around the world, you get these bursts.

So the result is unpredictability. Their multi-modal size distributions implies that any particular disease can have very different fates depending on something – we don't know exactly. And furthermore, temporarily there is a lot of uncertainty as well. Just because the disease is burning itself out doesn't mean that it can't find itself a new population and start right back up again. And in fact that actually happened in 1918 – that the epidemic was burning itself out, and then after Armistice Day everybody was celebrating and so there were these huge parades in the streets, and the disease kicked right back up again and killed more Americans than had died in the war itself.

Okay, so what makes them unpredictable? So this is where the network literature comes in. The sort of very kind of general-level observation is that large populations exhibit structure. And there are lots of different kinds of network structures that are relevant to the spread of large epidemics. The population is distributed in homogeneous fashion around the world, in countries, even in sort of urban-versus-rural areas. We have

a lot of transportation and infrastructure networks that help spread diseases. And also we have all kinds of networks between individuals – social, organizational, and sexual networks being some obvious examples.

So the result is that the assumptions of the standard model only apply at best in very small contexts. Within this building, for example, you could probably get away with assuming that people bump into each other at random – it's not quite true but it's sort of close enough to true that the model will probably work effectively. But once you get outside this building, you have all sorts of barriers that means that certain people are more likely to interact with each other than they are with others.

So you can't think of large epidemics as singular events. You have to think of them of concatenations of many, many smaller epidemics. And we can see this pattern [referring to chart] – it's a bit hard to read this view graph, but this is the 1957 influenza pandemic and it started about here. And you can see these numbers, these little circles have numbers in them that refer to the different waves of the epidemic.

And you can see that the numbers here are very small and then they start to get – it's sort of like a rock falling into a pond. You see these ripples kind of circulating out, but then you get these big jumps, right? And so it's not just that it is sort of diffusing geographically; it is sort of diffusing geographically and then it has these large jumps that are probably associated with somebody hopping on a plane and starting off a new epidemic in a different part of the world.

So we would like to – if we are going to understand how these big epidemics spread, we need to understand in some way to capture this kind of multi-level structure. And so that is what we have done. And we have just assumed very briefly that within contexts, the random interaction model works. However, each of these contexts is embedded in a – a sort of nested series of larger contexts, so sort of successively larger pools. So you can have your building that you work in but then your building is located in the capital, and then the capital is located in D.C., and D.C. is located in the Northeast, and the Northeast is located in North America, and so on.

And there is a sort of hierarchy of these nested populations. So the infections are all occurring locally, but then people, individuals, can transport themselves around. And so this is just a very, very strip-down way of representing this sort of high-level idea. You can see that – here are the sort of random interactions going on in the small contexts. Then you can lump these contexts together and then you can lump these contexts together, and you can lump all of these contexts together.

So just making this very simple modification to the standard models, you can see that you start to get the right kinds of answers.

So there is two things to notice here. One is that for each value of R_0 , there is a huge range of possibilities for what can actually happen. The other thing to notice is the reverse, that for very different values of R_0 , these distributions look very similar, right.

So R_0 , in other words, doesn't tell you very much at all, which is – doesn't win me friends with my mathematical epidemiology colleagues because R_0 is something that is very near and dear to their hearts.

The same story can be told in time. Here, everything is exactly the same. R_0 is the same, the population is the same, and yet we see very different outcomes, right, and the reason is that these – this one stops and this one keeps going, and the reason why is that the difference is being driven by one or two individuals. So you're talking about a population of hundreds of thousands of millions of people, and yet the macro behavior of the epidemic is being by the behavior of a couple of people. So this is a very, very hard thing to deal with using these sorts of statistical models.

Okay, I'm just going to skip this slide here and just go to the end, which is to say that it is important to think about these processes in a network way because, first of all, it's right. Large populations do exhibit this kind of network structure. And secondly, it has important consequences. You know, it's not just sort of it being finicky and trying to sort of put bells and whistles on perfectly adequate models. The models are fundamentally wrong. And only if we start to understand the importance of population structures and driving the unpredictability of these very large epidemics can we start to think of new kinds of intervention strategies.

And just to conclude, I think that the WHO did something very smart in 2003, which was to issue travel advisories very earlier on, right. They made a big out of it and they scared the hell out of everyone. And afterwards, people were sort of mad at the WHO, especially the mayor of Toronto for causing such a fuss and all of this economic damage for an epidemic that seems inconsequential in retrospect.

But what he did was he – what the WHO did was that by issuing these travel advisories, they dramatically reduced long-range travel, and they cut down on these random transport of individuals. And you just need to think, like, if that epidemic had gone to Harari instead of Toronto, it could have been a whole different story. So I think that now we have a way to sort of understand and quantify the effect of imposing, for example, travel advisories, which is a relatively unobtrusive kind of intervention strategy. And it can be actually extremely effective. So thanks very much.

MS. HILLSMAN: I don't know about you, but as I was listening to all three of these, I was kind of jumping across from one to the other thinking about how the different areas being very different substantively were actually very inter-related. I was thinking a lot about how the terrorism networks could be affected by the kinds of logical work being done in other areas as well. So that was kind of interesting.

Okay, we have some time for questions. Anybody like to – yes, back here in the corner.

Q: Actually, with this slide in your conclusion where you say knowledge of diseases is how R_0 does not predict size or duration of an epidemic – what do you think

about the statement that R_0 is a necessary, but not sufficient element, and that even when you talk about your one or two people driving the resurgent events in the epidemic, that you still need to understand the infectiousness of those two people to predict what is going to happen in your epidemic –

MR. WATTS: That is correct.

MS. HILLSMAN: Could you repeat the question?

MR. WATTS: The question is that I have a statement there saying that R_0 doesn't help predict the size of the duration of the epidemic. The questioner is correct. A necessary condition for an epidemic is R_0 greater than one. That is still true. And so I should amend that statement and say in the event that you have an epidemic – I mean, you can't have an epidemic where R_0 is less than one. So in the event that you have an epidemic, you know, you are already talking about R_0 greater than one. In that case, you're not going to learn much from its particular value.

MS. HILLMAN: Another question and would you introduce yourself when you ask a question please just so we will know who you are? Anyone else going to jump in?
Yes

Q: I have a question for Major Reed. I'm Mary Jo Hoeksema with the Population Association of America and the Association of Population Centers. I was wondering as to what role technology has played in the rise of resistance networks and the identification of them– specifically speaking about the role of the Internet maybe has played. Can you –

MAJ. REED: Yeah, I mean, from the research that I have done and also from some practical experience, I think to some degree it has made it more vulnerable because any time you are using cell phones or telephones or any type of technology, you become more vulnerable. But then on the other hand, it has allowed for speed of communication and things like that. But, you know, we know that and folks that are part of these resistance organizations know that as well. So, you know, I think we're also starting to see a trend towards the more primitive modes of communication in order to reduce the amount of disruption that we might be able to cause or create. Does that answer your question?

Q: Mm-hmm.

MAJ. REED: So it is sort of a doubled-edged sword I guess to a degree.

MS. HILLSMAN: Yes, sir.

Q: I'm David Siegel from the University of Maryland. Also, a question for Major Reed -- and probably an unfair one because I want you to get on the other side of the coin: We have talked about how we might understand insurgencies as networks. But

in your presentation, you very clearly said that operations against insurgencies are military and political. And on the political side, I was wondering if network analysis is useful, or as useful, for understanding the people who are on our side, as the people who are on the other side, people who we are trying to build alliances, who we want to help turn the country over to?

MAJ. REED: Yeah– I mean, I would certainly agree with that statement. I mean, right now, I don't know enough about it to comment in much detail about it. But, you know, I think your question gets at this, you know, as we sort of – I don't want to say stand up these governments because that's really not the right terminology, but in terms of our allies and the folks that we're working with, in terms of understanding who they are associated with, I think there is some merit to it.

MS. HILLSMAN: I'm sure that is a lot of what historians do after the fact when –The question is whether can do it in more contemporaneous and systematic ways.

MAJ. REED: I think we would be surprised to find out who people are associated with.

MS. HILLSMAN: Yes, in the back.

Q: Susan Newcomer from NIH. Again, just kind of an observation. This data – that if you have reliable data, you are several steps ahead. If you have credible data, that is even better – right? Reliable and credible. And that continues to be a challenge. (Inaudible) – and I would imagine when you were reviewing intercepts and insurgencies, even more – (inaudible).

MAJ. REED: Yes, ma'am, it is. It is because – and it's also I think – as I alluded to, it's a lot more involved because you can't simply ask people per se with whom they are associated, and even if they tell you, you can't necessarily expect that to be the truth. So in terms of building these networks, it requires a lot of painstaking details, confirmations and denials, and things like that to really sort of put this together. And then in terms of looking at disruption, figuring out -- as I alluded to in my talk – in terms of who does what within the network. You know, maybe not necessarily is the leadership the most important thing at that particular time to sort of disrupt or neutralize the network, but maybe there are financiers or logisticians, and things like that.

MS. HILLSMAN: Perhaps I could take the position of the chair for just a second and ask our presenters whether or not the collection and structure of data for these kinds of analyses are improving in various areas and whether there are suggestions they have in their own area of expertise as to what things – bodies such of the government, et cetera, could be doing to help improve the datasets in this area, recognizing of course that we have another balancing act here, which are issues about privacy, both privacy and confidentiality, which are obviously somewhat different things. So could you comment on those kinds of data in these areas?

MS. STOVEL: I would love to comment on that. I think data has improved for studying a variety of different kinds of networks. But it is very difficult because the easy way to collect network data – and here I use “easy” carefully. I mean, it’s not that easy but it’s easier than the alternative – is to ask a sample of people about their acquaintances, about their sex partners, or perhaps in some context we could do archival information and get some mapping. And this is what we call egocentric network data when we start with a person and we sort of ask them something.

Now, from a structural perspective, that makes everybody look like a star – a socio-metric star, because all we know is them at the center of a small network. And it is really clear that really understanding the consequences of some of these structures means figuring out how these stars fit together, whether they are all overlapping or whether they are in chain, whether some people have Duncan calls shortcuts, which might be getting on a plane. People who travel frequently might play a really different role.

And so to understand the ways in which these egocentric networks connect to one another requires a slightly different perspective. Most people don’t actually know very much about the second and third steps. Most people don’t know very much about the second and third – the histories of their sexual partner. They might know one step back, but they might not know a lot more than that; there may be social reasons why people don’t like to tell that.

So it is very important to get network data from interacting people so we can begin to see what these maps look like, and that is a much more difficult expensive problem. We are increasingly recognizing how important it is and in some context being able to do it. And then there is a lot of work to think about whether those are generalized patterns, whether these make sense. And then we can maybe use some of – some mathematical modeling to begin to sort of do this generalization.

MR. WATTS: I have two responses to that. One is that with regard to certain kinds of social interaction, there is a tremendous advantage going on now in recording who interacts with whom and possibly even with what consequences. Unfortunately, for publicly funded researchers, that is almost all in the corporate world and it’s proprietary.

So, you know, Google knows an awful lot about you. And if Google ever gets together with eBay –and, you know, a couple of other companies, they can tell what you’re searching for, what you’re clicking on when you perform those searches, who your friends are, what they do, what they e-mail you, the content of those e-mails, and then what you subsequently purchase. So from – in a certain class of activities, the day when somebody can know everything about what a very large number of people are doing does not seem so far off.

When it comes to the diseases and the sorts of things we have been talking about today, it gets much more complicated. To illustrate that, you just have to go through the record of the SARS epidemic – you know, the – it was first recorded in Hong Kong after a single individual got a train. It was spread throughout a hospital because the person

who was sick was misdiagnosed, and so they thought he had congestion in his lungs so they put a ventilator down his throat to get the stuff out of his lungs. So he is spewing up virus particles into the air.

And there was a class of medical students standing next to him and they all got infected. And then the next burst came when somebody went home to their apartment building and it got into the sewage system. And then somebody – a woman left Hong Kong and flew to Toronto. So many of the transmissions were relatives visiting sick people in hospital.

So, you know, how many networks do we have now? You know, we have trains, we have roads, we have airlines, we have sewage and water, wastewater systems, we have hospitals, we have social networks. You know, at some point, the idea, which I think has underpinned a lot of network analysis over the years of sort of a literal map of the network that you can measure properties of, right, and use those properties to predict things I think is sort of doomed, that that is not the successful way; that there has to be another way.

And in fact, this is what we're trying to do, is to sort of think in terms of networks but you don't necessarily have to map everything because I think that in these sorts of problems it's not feasible. So depending on the kind of application area, I think in some cases it is extremely promising, except we would like to get the data. In other cases it is not promising and we need a different theoretical approach all together.

MS. HILLSMAN: I might add here that I think you can see from this part of the discussion why a coalition to protect research is very interesting in this social-network analysis because we're going to have to rely on this kind of method for studying in order to engage in much research that will be for the benefit of not only a public, but many publics. And yet at the same time, the kind of data, the kind of research that is going to need to be done is not only complex, it also does raise the issues that I said before – confidentiality and privacy violations, and so forth and so on.

And so we are going to have to ensure that the cannons of science and research are protective on the one side and are providing the information and the data that we need on the other. And this is going to I think require a fair amount of complicated sort of public thinking; not just thinking within the science community, but thinking within the policy community, thinking within the leadership communities, and thinking – and engaging the broad public on some of these issues. So I think that is one of the lessons of some of this.

Are there other questions? I have taken more time –

Q: My name is David McMillen. I am a member of the – (inaudible) – community network. And I'm not sure this is a fair question but let me try it anyway.

MS. HILLSMAN: It's fair game.

Q: After 9/11, one of the owners of the large data conglomerate private sector outfits said, wait, I have got all of this information on people; I can figure out who these people are. And he went to his study and wrote up some code, and within about eight hours, he ran through his database and identified 319 people based on credit card information and this sort of thing, who, you know, could be the 19 terrorists. And sure enough, 11 of the 19 were in his accumulation of 319 people.

It seems to me this is one of the dangers of rushing research into policy that we face. You know, we don't know what happened to the other 308 people that were identified by that analysis. You know, they may still be rotting in some prison somewhere. But how do we draw that line between what is valuable research and when is it appropriate to move that research into the policy realm, where the consequences are quite big?

MR. WATTS: There is actually sort of a technical distinction that I think might be helpful. I mean, I agree that ultimately this is, you know, sort of something that needs to be – that it's going to be part of some sort of ethical problem. But in the particular example that you pose it's interesting because the kinds of questions that these methods are well suited for are not those kinds of questions, right. The problem with individuals is that they are very complicated, and two people who have very similar histories and backgrounds can end up behaving very differently. You know, even if you happen to be genetically identical, you can have very different life courses depending on any number of variables.

And so trying to use network tools to predict individual behavior is I think just scientifically problematic. And so, you know, you may be able to sort of shortcut the issue by saying, well, that is just not good science. The sorts of things that these tools are good for are sort of average statistical properties of very large numbers of people because when you have very large numbers of people, their idiosyncrasies tend to average out. And so you can make statements about distributions of events, for example, that are very robust, but the individual events are very unpredictable. So earthquakes is an example from the natural world, right? You can't predict the size of and timing of a particular earthquake, but the distribution of earthquakes is very stable.

So you have to ask the right questions and unfortunately, a lot of the questions that people want the answers to, particularly in the policy world are the wrong kinds of questions scientifically. And I think we can help them understand that and they usually react poorly to that advice) -- but I think that it is our responsibility as scientists to give it to them anyway.

MS. HILLSMAN: Yes, sir.

Q: My name is Mike – (inaudible) – and my question is for Dr. STOVEL. I was wondering – the HIV epidemic going on in the U.S., are you all studying the relationship

– (inaudible) – to the actual people in the U.S. to see basically how we can approach – (inaudible)?

MS. HILLSMAN: Would you repeat the question?

MS. STOVEL: Yeah, this is a question about sort of studying the sexual networks of other populations maybe with particular interest toward the HIV epidemic in the United States. And I would say that with respect to the ADD Health study, we have not – the orienting questions at the beginning of the ADD Health study certainly took very seriously HIV risk and the structure of interaction among adolescence that might amplify the HIV risk in some communities or some populations.

We have not specifically focused on that in the work that I have talked about. There has been a lot of research in other places that is trying to understand some of the context and structure of sexual networks and how that might relate to HIV risk. And here, we have really tried to import some information from other contexts, think about what if the network looks like this; what if the network looks like that. Can we actually understand why we get outbreaks in some places and maybe more smaller epidemics and less severe outbreaks in other places?

This work is in many ways hindered by data constraints. There is sort of a simultaneous feeding of increased data where we get a little bit of estimation of contact parameters or the structure of interactions, the structure of migration. Certainly in Southern Africa, it is playing a very big role in epidemics sort of leapfrogging from one place to another. And I think this is a context in which there is a lot of collaboration between people who have expertise in mathematical modeling and people who are in the field trying to give a better sense of what the on-the-ground relationships between persons are that may have some kind of disease transmission potential.

And so I think this is a context where we really do need people working with different kinds of expertise to feed better estimates of social behavior, sexual behavior, whether it's sexual relationships per se or other kinds of things that might bring people into proximity with one another and so on.

I think at the moment, the most intensive research activity in this area is going on in parts of the world outside of the U.S. where HIV is more acute, whether the epidemic is more acute and it's more of a problem – Southern Africa and certain parts of Southeast Asia.

MS. HILLSMAN: We have – hang on just a second – we have run over the – we still have this room but I realize that if some people have to leave, please do, but we'll continue questions for a little while longer if there are still questions, but we recognize that some people have time constraints and we did say we would end at 11:30. So, yes, another question.

Q: Mine is actually – first to thank everyone for presenting really how important this research is in nationwide outcomes. I’m Barbara Solt with the Institute for the Advancement of Social Work Research. I think one of the things that we also need to highlight here for the coalition and the importance of developing research in this area is the need for getting data through networks that are sometimes difficult to reach and unpopular. And I think it is important that we recognize the important truck stops, and prostitutes, and other areas of high risk of high resistance if you will, and the importance of having support for research that reaches into the communities where things are happening so that your model can then be developed in order to form preventive approaches in the future. I just want to make that comment to make –

MS. HILLSMAN: Well, some of you might want to comment on – one of the things that was interesting to me in your presentation, Kathy, was the notion that a whole high school participated is really quite a remarkable research phenomenon, that that entire community participated in the adolescent health survey – is really – was really remarkable. And you’ll see the consequences of that being able to study something that hadn’t happened before. So clearly something happened in that community that enabled all of the students to be willing to participate and families agreeing to that and so forth.

Does anyone want to comment on that – on the issue of difficult and sometimes not very socially popular populations?

MS. STOVEL: Well, I’ll just make one quick comment about that, which is that I think the more we study networks, the more we find both that there are strong similarities, there are some principles that seem to guide people’s interactions with one another. These may be conditioned by infrastructure, by technology, by politics, by other things. But there are lots of similarities that we are beginning to discover, and some of these similarities – you know, the more robust they are, the more we can use them in context, and then we don’t have to keep collecting data about them.

And yet there is important variation as well. And to the extent that we can understand that variation, understand the mechanisms that produce it and so on – maybe in small contexts, we then will be able to begin to think about the consequences of networks in a more robust way. And I think that we are increasingly beginning to do this, to recognize this, to begin to see the relationship between individual motivations and the networks that those produce in the aggregate and the cumulative sense, and then we can begin to think about what the consequences of that might be.

Now, too – I guess I’ll make this a little longer – two other kinds of contexts. One is that the more we understand these, the more people then might do things to change them. So I live in Seattle where we have Microsoft and Amazon that seem to influence lots of people’s behavior. And the more you understand about how they collect data about you, the more you can do certain things to pervert that, to not let them know what you’re doing. And we see this in all kinds of networks.

The more people begin to understand this, the more they may take certain kinds of behaviors to transform the network itself. And so we need to be aware of these reactive things, and this is one of the things that is both so exciting and so difficult about social science in general.

With respect to the ADD Health study in particular, this is a study in which we went into 80 communities and collected a great deal of data. We collected data about friendships from 90,000 kids. Basically all of the kids were enrolled in 140-or-so schools in these 80 communities. That was a truly phenomenal effort and has produced just this really remarkable data that begins to give us more confidence when we make generalizations about what adolescent friendship networks look like.

We don't have the sexual- or romantic-relationship data from all of those 90,000 kids. But we do know a lot more about their interaction patterns than we did before that study, and so now we can begin to understand sort of the contexts – or the variability, and then we can go into some other contexts other places that might be a particular risk or where there are specific social issues that are of concern to people. And if we see a little bit of evidence that one of these network structures seems to be in place, we may not need to do quite as much or quite as basic research, but we can then begin to move forward from that.

MS. STOVEL: Well, ADD Health may have broken the bank. But understanding sort of the range of variability and beginning to discover that, you know, there really is sort of a social vocabulary, at least for interaction among kids in schools, and, you know, there is different ways to combine it but there is sort of a limited number of elements, and that begins to be a really important discovery.

MR. WATTS: Can I just add something to that?

MS. HILLMAN: Yes, please.

MR. WATTS: Just on that expensive component, I agree but I think that I would just like to emphasize there are no easy answers here. And this sort of – often is something I get frustrated with because people will call me up and say what can social network theory do to stop terrorism, to predict diseases, to help us account for the impact, to help us sort of maximize the impact of our advertising campaign. And the answer is that right now, it can't do any of these things.

And, you know, the analogy that I think is appropriate is if in 1953 you read Watson and Crick's seminal Nobel-prize-winning paper on the structure of DNA, and you had written a letter to James Watson and said, Dear Dr. Watson, I read with great interest your paper in – (audio break, tape change) – but, it's going to take us 50 years and several trillion dollars and tens of thousands of people to figure it out. And you know, 50 years later and several trillion dollars later, it's starting to look like it may help

us figure it out, but we're still not there yet. And I think that we are facing the same kind of magnitude of question, when you say how can social network theory help us do X?

And you know, the real shift that has to come is a cultural shift, where we say, this is science like any other kind of science. And if you want answers, you have to pay for them. And you know, you don't get there with just one person sitting and dreaming up his or her theory of the world in their office. And you saw the dates on Kate's papers. You know, ten years from conducting the study to publishing the papers. That's how long it takes when it's just one or two of you doing it. And that's not how you make real progress. You know, you need to industrialize these things. So I think that that's something that, you know, in this particular environment, is important to emphasize.

MS. HILLMAN: On that positive and challenging– it is a positive note and a challenging note – I would like you to join me in thanking our guests and for their research. It's very good. Thank you, guys.