CONNECTIONS 24(2): 30-34 http://www.sfu.ca/~insna/Connections-Web/Volume24-2/242-Bernard.pdf

Estimating the Ripple Effect of a Disaster[1]

H. Russell Bernard
Department of Anthropology, University of Florida, USA
Peter D. Killworth
Southampton Oceanography Centre, Southampton, UK
Eugene C. Johnsen
Department of Mathematics, University of California–Santa Barbara, USA
Gene A. Shelley
Georgia State University, USA
Christopher McCarty
Bureau of Business and Economic Research, University of Florida, USA

We apply our network scale-up model to estimate the number of people in the U.S. who ***know someone*** *who experienced the terrorist attacks of September 11, 2001 and the number of people who* ***know someone who knows someone*** *who experienced those attacks.*

INTRODUCTION

On Tuesday, September 11, 2001, Bernard got a call from a reporter asking: "Are you the same person who studied the Mexico City earthquake?" The reporter wanted to know how many people in the U.S. would be affected directly by the attack on the World Trade Center and the Pentagon. What he meant by "directly" was this: How many people know someone who worked at the World Trade Center or at the Pentagon or who were on those four planes on September 11, 2001? And further: How many people know someone who knows someone who experienced the attack? Could this, he asked, be worked out as a kind of six-degrees-of-separation problem?

Yes, it can, but there are at least three ways to define the population of people who experienced the attack. First, there are those who are usually present at the sites of the attacks at some time or other during the week (this includes people who regularly take those flights); this is about 50,000. Second, there are the people who were actually present at the sites at the time of the attacks. This is about 20,000. And third, there are those of the 50,000 who are missing, estimated at the time of this writing (9/22/01) to be 6,333.

[1] We acknowledge with gratitude the support of the U.S. National Science Foundation for our work on the network scale-up method (grant SBR-9710353).

THE MODEL

We used our network scale-up model to work out the numbers. The model is simply ***m* ' *c* = *e* ' *t***, where:

- ***e*** is the size of the event population, here 50,000, 20,000 or 6,333
- ***t*** is the size of the population within which the members of ***e*** are embedded, which in this case is the total U.S. population;
- ***c*** is the average size of a personal social network among members of the ***t*** which, from our recent work, we estimate to be 290 for the U.S., and
- ***m*** is the average number of people in ***e*** whom people in ***t*** know.

When we first developed our model, it was to estimate the size of hard-to-count populations, but the reporter who called Bernard made clear that in this case, the numbers of interest to his readers were (a) the number of people who *know someone* in ***e*** and (b) the number of people who *know someone* who *knows someone* in ***e***. We'll call these n_1 and n_2.

Our working definition of "A knows B" means that A knows B by sight or by name; that A can contact B in person by telephone or by mail; and that A has had contact with B in the past two years. Over the last 15 years, we've done several national studies to determine ***c*** for the U.S. We call a representative sample of people across the country and ask them how many people they know in 29 populations whose sizes we know. Examples of those populations are: people named Nicole, American Indians, people who hold a commercial pilot's license, and people on dialysis.

From their answers, we use a back estimation technique to estimate ***c*** for each person—that is, the number of people they must know in order to produce as closely as possible the patterns of answers to our questions about how many people they know in those 29 known-size populations. (For a review of the model, its variants and its tests, see Bernard et al.1989; Bernard et al.1991; Johnsen et al.1995; Killworth and Johnsen et al.1998; Killworth and McCarty et al.1998; McCarty et al.1997; McCarty et al.2001; Shelley et al.1995.)

Across five such surveys, the number for ***c*** is 290, plus or minus just a few people. We plug this estimate for average personal network size into the network scale-up model and estimate the number of people in each of those same 29 populations. Since we know the number of people in those populations, we can see how well the model is doing. We estimate about a dozen of these populations reasonably well, but we seriously underestimate or overestimate others. Our program of research involves testing confounds and trying to improve our estimates for populations, like suicides, that we miss by a lot.

We also use our model for its original purpose, to estimate the size of hard-to-count populations. Our estimates for the number of HIV-positive people, the number of rape victims, and the number of homeless in the U.S. track closely with estimates made by others. Our answers are thus reliable, but not necessarily valid. We might simply be mirroring the incorrect answers produced by others who happen to use quite different methods to measure these same things. One thing we do know: Our estimates of the number of HIV-positives, the number of rape victims, and the number of homeless cost very little to produce, compared to other methods currently used.

Estimating *m*

With appropriate caution, we can use the network scale-up model to answer the reporter's questions: How many people would be one or two links away from people who did or could have experienced the attacks on September 11, 2001? Assuming that 50,000 people could have experienced the attacks at the World Trade Center or the Pentagon or on the hijacked planes that day; that 20,000 actually had experienced the attacks; that about 6,333 are currently missing and presumed to have died; and assuming ***t*** is about 250 million, ***m*** would be, respectively, about 0.058, 0.023, or 0.0073, to two significant figures. (We'll quote estimates from here on out to two significant figures.)

That is, across the U.S., each person knows, respectively, about 0.058, 0.023 or 0.0073 of a person who experienced, in one way or another, the calamity of September 11th. (We use ***t*** = 250 million because that was the approximate population of the U.S. when we developed and tested the model and obtained the figure of ***c*** = 290.) We expect the distribution of answers across the U.S. to the question "How many do you know who experienced the attack on the WTC or the Pentagon on September 11th?" to consist largely of 1s and 0s, so we might treat the above m values as proportions, in which case about one person in 17, 43, or 140, respectively, knows someone who experienced the attack in one way or another.

We know, however, that if someone knows someone in a population they will tend to know others in the population, and this will almost certainly be true for those living in the New York City and Washington, D.C., metropolitan areas. We represent this by a "lead-in" factor, 8, which is the average number of members of a population known by those who know at least one member of the population. In general 8 > 1.0, so including this factor in our calculations should improve our estimates of ***m***.

In previous work (Johnsen et al. 1995), based on General Social Survey data, we obtained lead-in factors for the populations of homicides, suicides, and AIDS victims of about 1.60, 1.26, and 1.75, respectively. The relatively low figure for suicides points to the relative social isolation and stigmatization of those who commit suicide, while the relatively high figure for AIDS reflects the relatively high social cohesiveness of those afflicted with AIDS despite stigmatization. We think that the figure for homicides reflects the fact that this is a population that is neither stigmatized nor uninteresting and that is generally not a cohesive group (though their survivors might be). We think that people who experienced the attack of September 11 are in this category, so we assume a lead-in factor of 1.60 for these populations.

Estimating n_1 and n_2

Using the lead-in factor, we can calculate n_1 and n_2 — that is, the number of people who know people in e and the number of people who know people who know people in ***e***. The effect of the lead-in factor is to decrease n_1 and so we have $n_1 8 = mt = ce$. From this analysis, one person in about 28, 69, or 220, respectively, knows someone who experienced the attack and so, our estimates for the three possible populations affected by the attack are 9.1 million, 3.6 million, and 1.1 million.

To answer the reporter's next question, we take the analysis another step, using the three populations of n_1 members from the first step. Thus, the number of people who know someone who knows someone who experienced the attack on September 11th is about $n_2 = c^2 e / 8^2$.

We justify the use of the same 8 here by assuming that these three populations are of the same type as the ***e*** and the homicides. Thus, at two steps removed in the network chain from the disaster, about 1.6 billion, 660 million, or 210 million, respectively, know someone who knows someone who experienced the horror of September 11th, 2001. (The formulas for n_1 and n_2 are approximate, but obvious corrections to them have little effect on the results.)

The first two figures exceed the population of the U.S., so practically everyone in the U.S. knows someone who knows someone who was in the 50,000 or 20,000 who experienced the attack. The third figure says that about 83% of the U.S. population is two steps removed from the 6,333 who are currently thought to have died in the attack. Without the repeated lead-in factor, the third figure of 210 million becomes 530 million, which indicates that practically everyone in the U.S. knows someone who knows someone in the 6,333.

Here we have analyzed the ripple effect of knowing people and knowing people who know people; however, knowing a person who experienced the attack is not the same as knowing that that person experienced the attack. This flow of information — finding out that people whom we know, or have heard of, experienced the attack — has its own ripple effect, and, as Shelley et al. (1990) found, it can take years to complete.

Estimating *c*

We discussed this by e-mail with Barry Wellman. Wellman pointed out that his database of network connections, including e-mail connections, ran to several thousand, not several hundred. Did our model and our research take into account the fact that the distribution of network size was very broad?

This raises a very interesting question: How many people would there be in an average network if everyone had access to an electronic database of all the people they know? In their pioneering work on network size using phone books as their cuing device, Pool and Kochen (1978 [1959]) came up with an estimate of between about 3100 and 4250 for the size of Pool's social network. Freeman and Thompson (1989) replicated Pool and Kochen's work with the phone book method, using a sample of respondents, as have we (Killworth et al.1990). These results, using the phone book method, vary between about 1700 and about 5500 for total network size.

Our more recent work on the network scale-up model consistently produces, as we said, an estimate of about $c = 290$, sd = 232. Obviously, different ways of measuring total networks can produce results that differ by at least an order of magnitude. Note, though, that if t is six billion (the Earth's population), the average network size would be about 6,960—more like the value found using the phone book method.

All measures of network size depend on asking people questions involving recall. Seymour Sudman (1985) made clear that aided recall produced far more data about social networks than does unaided recall. We tested several quite different methods for getting at the total network size (Killworth and Bernard 1978; Bernard, Killworth, and McCarty 1982; Killworth, Bernard, and McCarty 1984; Bernard et al. 1988, 1990; Killworth et al. 1990; McCarty et al. 2001) and it is certainly the case that different measures produce different numbers. In fact, giving a national representative sample of Americans a list of categories to cue them about people in their networks (name your blood relatives, kin by marriage, people you know at work, neighbors, people you know at church, etc.) produces an average network size of 437, sd = 415 (McCarty et al. 2001).

Thus, our estimate of 290 for the average network size is a minimum. The number would surely be larger were everyone to have a database of their network handy when we come calling to ask them how many people they know in each of 29 populations. Aware of this, we nonetheless use the figure of 290 because (a) it is very stable, and (b) it produces results (estimates of the size of populations) that are demonstrably correct in many cases, probably correct in others, and incorrect in others. In other words, it's the best estimate we've got at the moment.

From the evidence so far, one way to improve our results (defined as estimating correctly more of the 29 known-size populations) is to improve our estimate of m, the number of people whom each person knows in each of the known-size populations. This, of course, requires basic research on the obstacles to knowing specific things about people in our networks, and this is what we've been doing in recent years.

REFERENCES

Bernard, H.R., P.D. Killworth, and C. McCarty. 1982. INDEX: An informant-defined experiment in social structure. *Social Forces* 61: 99–133.

Bernard, H.R., P.D. Killworth, M.J. Evans, C. McCarty and G.A. Shelley. 1988. Studying social relations cross-culturally. *Ethnology* 27: 155–179.

Bernard, H.R., E.C. Johnsen, P.D. Killworth, and S. Robinson. 1989. Estimating the size of an average personal network and of an event population. In: *The Small World*, M. Kochen, Ed., 159–175. Norwood, NJ: Ablex Publishing.

Bernard, H.R., E.C. Johnsen, P.D. Killworth, C. McCarty, S. Robinson and G.A. Shelley. 1990. Comparing four different methods for measuring personal social networks. *Social Networks* 12: 179–215.

Bernard, H.R., E.C. Johnsen, P.D. Killworth, and S. Robinson. 1991. Estimating the size of an average personal network and of an event population: Some empirical results. *Social Science Research* 20: 109–121.

Freeman, L.C. and C.R. Thompson. 1989. Estimating acquaintance volume. In: *The Small World*, M. Kochen, Ed., 147–158. Norwood, NJ: Ablex Publishing.

Johnsen, E.C., H.R. Bernard, P.D. Killworth, G.A. Shelley and C. McCarty. 1995. A social network approach to corroborating the number of AIDS/HIV+ victims in the U.S. *Social Networks* 17: 167–187.

Killworth, P.D. and H.R. Bernard. 1978. The reverse small world experiment. *Social Networks* 1: 159–192.

Killworth, P.D., H.R. Bernard and C. McCarty. 1984. Measuring patterns of acquaintanceship. *Current Anthropology* 25: 381–397.

Killworth, P.D., E.C. Johnsen, H.R. Bernard, G.A. Shelley, and C. McCarty 1990. Estimating the size of personal networks. *Social Networks* 12: 289-312.

Killworth, P.D., E.C. Johnsen, C. McCarty, G.A. Shelley, and H.R. Bernard. 1998. A social network approach to estimating seroprevalence in the United States. *Social Networks* 20: 23–50.

Killworth, P.D., C. McCarty, H.R. Bernard, G.A. Shelley, and E.C. Johnsen 1998. Estimation of seroprevalence, rape and homelessness in the United States using a social network approach. *Evaluation Review* 22: 289–308.

McCarty, C., H.R. Bernard, P.D. Killworth, E.C. Johnsen, and G.A. Shelley. 1997. Eliciting representative samples of personal networks. *Social Networks* 19: 303–323.

McCarty, C., P.D. Killworth, H.R. Bernard, E.C. Johnsen, and G.A. Shelley. 2001. Comparing two methods for estimating network size. *Human Organization* 60: 28–39.

Pool, Ithiel de S. and M. Kochen. 1978. Contacts and influence. *Social Networks* 1: 5–51.

Shelley, G. A., H. R. Bernard and P.D. Killworth 1990. Information flow in social networks. *Journal of Quantitative Anthropology* 2:201–225.

Shelley, G.A., H.R. Bernard, P.D. Killworth, E.C. Johnsen, and C. McCarty. 1995. Who knows your HIV status? What HIV+ patients and their network members know about each other. *Social Networks* 17: 189–217.

Sudman, S. 1985. Experiments in the measurement of the size of social networks. *Social Networks* 7: 27–151.