ELSEVIER

Social Networks 17 (1995) 167–187

SOCIAL NETWORKS

A social network approach to corroborating the number of AIDS/HIV+ victims in the US [☆]

Eugene C. Johnsen [a,*], H. Russell Bernard [b], Peter D. Killworth [c], Gene Ann Shelley [d], Christopher McCarty [e]

[a] *Department of Mathematics, University of California, Santa Barbara, CA 93106, USA*
[b] *Department of Anthropology, University of Florida, Gainesville, FL 32610, USA*
[c] *NERC Oceanography Unit, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK*
[d] *Department of Anthropology, Georgia State University, Atlanta, GA 30303, USA*
[e] *Bureau of Economic and Business Research, University of Florida, Gainesville, FL 32610, USA*

Abstract

Accurate estimates of the sizes of certain subpopulations are needed to inform important public policy decisions in the US. Laumann et al. (1989, 1993) have attempted to assess the accuracy of the reported data on the incidence of AIDS in the US, collected by the Centers for Disease Control and published in the AIDS Weekly Surveillance Reports (AWSR) and HIV/AIDS Surveillance Reports (HASR), by comparing these data with response data from the 1988, 1989, 1990 and 1991 General Social Surveys (GSSs). To establish reference comparison subpopulations, they did a similar assessment of the reported numbers of homicides during previous 12-month periods, published in the Unified Crime Report (UCR) and the Vital Statistics of the United States (VSUS), comparing these data with other response data from these same GSSs. The GSS data were compared with the AWSR, HASR, UCR and VSUS figures by sex, race, ethnicity, age, and region of the US. Their results for homicide victims are reasonably similar to the UCR and VSUS figures for these categories, while their results for AIDS victims are reasonably similar to the AWSR and HASR figures only for sex and age.

There is then the question of whether reported total figures for the incidence of homicides and AIDS (as well as suicides during a previous 12-month period) are reasonably accurate. There is concern that there is significant undercounting of the total incidence of

* Corresponding author.

☆ Supported by NSF grant SES 8803577. Different parts of this paper were presented at the Social Networks Conferences, Sunbelt X, February 1990 in San Diego, and Sunbelt XIV, February 1994 in New Orleans.

SSDI 0378-8733(95)00261-8

AIDS. Here we investigate this issue using a simple mathematical model of social networks, some elementary social theory, recent estimates of average personal network size in the US by the authors, and response data from the 1988, 1989, 1990, 1991 and 1993 GSSs. Assuming the average personal network sizes of homicide victims, suicide victims and AIDS victims are about the same as those for the general US adult population, we test as null hypotheses the propositions that the incidence of homicides reported in the UCR, of suicides reported in the VSUS, and of AIDS reported in the AWSR and HASR are accurate. Our analysis does not detect a significant inaccuracy in the reported figures for homicides. For suicides and AIDS, however, we encounter discrepancies of varying degrees, suggesting that either the reported figures are overcounts, AIDS and suicide victims have significantly smaller average personal network sizes than the general US adult population (smaller for AIDS victims than for suicides), or else the GSS respondents significantly underreported their personal knowledge of suicide and AIDS victims, either because they did not know that some of the people they know committed suicide or have AIDS, or because they falsely state their lack of knowledge of such victims, perhaps because of the stigma attached to such associations. These issues are illuminated by the results from an earlier survey of personal network size in Gainesville, Florida, and a previous national survey among surgeons, which suggest that lack of knowledge of personal network members who have AIDS seems to play a major role in explaining the discrepancies with the reported figures for AIDS.

1. Introduction

To begin to deal with a serious social problem such as homelessness, rape, AIDS, HIV infection, homicide and suicide it is important to know its true extent, namely, the size of the subpopulation associated with the problem. It is usually the case that some estimate of the subpopulation size is available, but it is often of questionable accuracy. Here we are concerned primarily with the accuracy of the estimate for the subpopulation of AIDS/HIV + victims and incidentally with those for homicide and suicide victims.

Laumann et al. (1989, 1993) have used data from the General Social Surveys (GSSs) of 1988, 1989, 1990 and 1991 (National Opinion Research Center, 1993) on respondents' knowledge of AIDS victims in order to assess the accuracy of the data collected by the Centers for Disease Control and Prevention (CDC) in Atlanta for the incidence of AIDS in the US population, particularly with regard to the distribution of the disease across various demographic subgroups. The latter AIDS data are from the CDC reports of March 1988, 1989, 1990 and 1991 (see References). As a benchmark comparison these authors use the 1988, 1989 and 1990 GSS data on respondents' knowledge of homicides within the previous 12 months to assess the data reported in the *Uniform Crime Report* (UCR) and the *Vital Statistics of the United States* (VSUS). Their homicide data are from the 1986, 1988 and 1989 UCR (Federal Bureau of Investigation, 1986, 1988, 1989) and 1987 and 1989 VSUS. The GSS is conducted between February and April of the years in which it is administered, with randomly selected respondents age 18 and older, one from each of approximately 1500 households. The demographic and socioeconomic characteristics of the GSS respondents compare favorably with those of the general US population (Bureau of the Census, 1983, 1989). The response rate for the years of interest here run around 75%, which is well within the range of response rates for the GSSs going back to 1972 (Laumann et al., 1989, 1993).

In the GSSs of 1988, 1989 and 1990 respondents were asked to report on their acquaintance with victims of homicide during the previous 12 months, and in the GSSs of 1990 and 1991 they were asked this for victims of suicide during the previous 12 months. In the GSSs of 1988, 1989, 1990, 1991 and 1993 (National Opinion Research Center, 1993) they were asked to report on their acquaintance with anyone, living or dead, who had contracted the disease called AIDS. For all three victim groups the responses included the number of members in each group with whom respondents were acquainted and various demographic characteristics of the three best known members. These data were compared with the corresponding AWSR, UCR and VSUS figures classified by sex, race, ethnicity, age, and region of the US. Laumann et al. report that in the GSS data the relative incidence of homicide victims is fairly similar to that in the UCR and VSUS with respect to sex, race, age and region, and of suicide victims with respect to sex and region but not age and race, whereas the relative incidence of AIDS victims is fairly similar to that in the AWSR only for sex and age. They also report that these data suggest underreporting of suicides for Blacks, overreporting of AIDS for Blacks and Hispanics and in the East, and underreporting of AIDS for Whites and in the Midwest. It also shows evidence that lower-status respondents are more likely to know a homicide victim than higher-status respondents, while higher-status respondents are more likely to know an AIDS victim than lower-status respondents. Several critical issues about the authors' first study are discussed in a letter of Berkelman et al. (1989) and a response from Laumann and Gagnon (1989).

Putting aside the issue of the *relative* distribution of these event populations among various demographic subgroups of the US population, there is the question of whether the official total figures for the sizes of these event populations in the US are reasonably accurate, or whether they represent fairly significant undercounting of these populations (there is little reason to believe that overcounting is a problem). Here we investigate this issue for homicide, suicide and AIDS victims. The presence of the AIDS questions (but not those for homicides or suicides) in the 1993 GSS allows us to examine the changes in the general populations' knowledge of AIDS victims over a 5–6-year span, a period in which there is a marked increase in the incidence of AIDS. Our approach takes the form of a set of hypotheses tests of the reported data for homicides, suicides and AIDS against the data from the GSS, using a simple mathematical model of social networks, some elementary social theory, recent estimates by the authors of average personal network size in the US, and the respondent data from the 1988, 1989, 1990, 1991 and 1993 GSSs (there was no GSS in 1992). We treat homicide victims as a benchmark population against which to compare the results for AIDS victims and suicide victims. We conclude with a discussion of some of the issues and difficulties that arise in employing such model, theory and data.

2. Hypotheses tests – a simple mathematical model and some elementary social theory

We consider a population T of size t having a subpopulation E of size $e \equiv e_E$ associated with some attribute or event. We call E an *event population*. For a member u of T we let K denote the people in T with whom u is acquainted and let k denote the number of people in K. The members of K we call the *personal (acquaintance) network* of u and k we call its *size*. Theoretically, the value of k could vary from 0 to $t-1$; however, it will usually be greater than zero and almost certainly no more than 1 800 000 (consider an 85-year social life in which an individual meets 5 new acquaintances per hour for 12 hours per day, 350 days per year). The average personal network size of all the members of T we denote by $c \equiv c_T$. Except when otherwise stated, we assume that an acquaintance tie is sufficiently established so as to be *symmetric*, that is, the tie is such that if v is an acquaintance of u then it is reasonable to assume that u is an acquaintance of v. Let c_E denote the average personal network size in T of the members of E and μ_E denote the average number of members of E acquainted with a member of T. Then, counting all the acquaintance ties between the members of E and the members of T in two different ways and equating the counts, we obtain

$$c_E e = \mu_E t \tag{1}$$

so that

$$\mu_E = \frac{c_E e}{t} \tag{2}$$

The situation of particular interest to us here is when the size of E is in question or is difficult to evaluate or estimate directly. Suppose we know c_E within a certain range of values, know t rather accurately, and have a reported value of e, $e^{(r)}$, to assess. By substituting these values into Eq. (2) we obtain a value of μ_E against which we may be able to test an empirical GSS estimate m_E in order to determine from a 95% confidence interval whether the value of μ_E is supported by this estimate. Now, suppose that E is a relatively rare subpopulation of T, $e \ll t$, and that the number of members of E known to a person in T has a distribution over T with mean μ_E and standard deviation σ_E. If we take a large random sample S of s people from T, sufficiently large to justify the assumption that the distribution of μ_E is virtually normal (here $s \geq 1370$), and ask each of them to recall all of the members of E with whom he or she is acquainted (it is possible that some members of S may be in E), then the average number of members of E known to a member of this sample, m_E, is distributed with mean μ_E and standard deviation $\bar{\sigma}_E = \sigma_E/\sqrt{s}$. Suppose the total number of acquaintance ties elicited by this request from all the members of S is τ. The average number of ties from a member of S to the members of E is then $m_E = \tau/s$. Thus, to test the value of $e^{(r)}$ on the right-hand side of Eq. (2) we set up the null hypothesis

$$H_0: \qquad \mu_E = \mu_0 \equiv \frac{c_E e^{(r)}}{t} \tag{3}$$

versus the alternative hypothesis

$$H_1: \quad \mu_E \neq \mu_0 \tag{4}$$

to see whether m_E lies in the 95% interval $\mu_0 \pm 1.96\ \bar{\sigma}_E$, where $\bar{\sigma}_E$ is estimated by s.d.(m_E).

Under the assumption that ties are symmetric, τ/e is the average number of members of S known by a member of E, and assuming that S is representative of T we may scale this up to T to derive an estimate for c_E in terms of m_E,

$$\sim c_E = (t/s)(\tau/e) = (t/e)m_E \tag{5}$$

when a value of e is given. Alternatively, we can use m_E to derive an estimate for e from Eq. (2),

$$\sim e = t(\tau/sc_E) = (t/c_E)m_E \tag{6}$$

provided we are given a value for c_E.

We bring this model to bear on the three subpopulations E = H ≡ homicide victims in the last 12 months, E = S ≡ suicide victims in the last 12 months, and E = A ≡ cumulative AIDS victims, in the US adult population T. For each of these event populations we test the reported value $e^{(r)}$ of the event population E by means of Eqs. (3) and (4) for known values of t and various restricted estimates of c previously obtained by the authors (Killworth et al., 1990) which we shall assume represent c_E.

3. The questions and data from the GSS

In the GSSs conducted in February–April of 1988, 1989, 1990, 1991 and 1993 the following questions were asked of approximately 1370–1598 randomly selected adult household respondents:

Homicides – 1988, 1989, 1990

H.0: "Within the past 12 months, how many people have you known personally that were victims of homicide?"

Suicides – 1990, 1991

S.0: "Within the past 12 months, how many people have you known personally that have committed suicide?"

AIDS – 1988, 1989, 1990, 1991, 1993

A.0: "How many people have you known personally, either living or dead, who came down with the disease called AIDS?"

The following questions were also asked of respondents, questions 3–6 first being asked in 1989:

Homicides

H.1: "Think about the person you know best who was a victim of homicide. Please tell me the letter of the category on the card which best describes your relationship to that person."

H.2: "We would like to know a few things about that person.
- A. Was that person male or female?
- B. How old was that person? Was (he/she)... (read list).
- C. What was that person's race? Was it black, white, hispanic or other?
- D. What state did that person live in?"

H.3: Same questions as H.1 and H.2 for the person second best known to the respondent.

H.4:

H.5: Same questions as H.1 and H.2 for the person third best known to the respondent.

H.6:

Suicides

S.1: "Think about the person you knew best who has committed suicide. Please tell me the letter of the category on the card which best describes your relationship to that person."

S.2: "We would like to know a few things about that person.
- A. Was that person male or female?
- B. How old was that person? Was (he/she)...
- C. What was that person's race? Was it black, white, hispanic, or other?
- D. What state did that person live in?"

S.3. Same questions as S.1 and S.2 for the person second best known to the respondent.

S.4:

S.5: Same questions as S.1 and S.2 for the person third best known to the respondent.

S.6:

AIDS

A.1: "Think about the person you have known best, living or dead, who came down with AIDS. Please tell me the letter of the category on the card which best describes your relationship to that person."

A.2: "We would like to know a few things about that person.
- A. Is that person currently living, or has that person died?
- B. (Is/was) that person male or female?
- C. How old (is/was) that person? (Is/was) (he/she)... (read list).

Table 1
Reported values of subpopulation sizes and corresponding average personal network sizes

	1988	1989	1990	1991
t (US total)	246048000	248251000	250410000	252502000
e_H (homicides)	20675	21500	23438	24703
$\sim c_H\ (t\tau_H / se_H)$	2052	1859	1560	–
e_S (suicides)	30407	30232	30906	30628
$\sim c_S\ (t\tau_S / se_S)$	–	–	952	1089
e_A (AIDS)	62200	85992	121311	162408
$\sim c_A\ (t\tau_A / se_A)$	663	425	405	401

D. What (is/was) that person's race? (Is/was) it black, white, hispanic or other?
E. What state did that person live in?"

A.3 Same questions as A.1 and A.2 for the person second best known to the respondent.

A.4:

A.5: Same questions as A.1 and A.2 for the person third best known to the respondent.

A.6:

Overall, about 94%–96% of all victims reported by respondents were classified as husband/wife, partner or lover, son or daughter, other relative, friend, neighbor, co-worker, acquaintance or patient, indicating that respondents' personal relationships to the reported victims were sufficiently strong to warrant the assumption that these relationships were approximately symmetric.

In Table 1 we give, for the years 1988, 1989, 1990 and 1991, the projected values of t, the reported values $e^{(r)} = e_H$ for E = H, $e^{(r)} = e_S$ for E = S, and $e^{(r)} = e_A$ for E = A, and the derived values of $\sim c_H$, $\sim c_S$ and $\sim c_A$ (to the nearest integer) for the years for which there were GSS responses to the above questions about acquaintanceship with homicide, suicide and AIDS victims (m_H, m_S and m_A values appear in Table 2). For simplicity we shall assume that persons less than 18 years of age are known by respondents because of their parents, and that they know a substantial part of the (symmetric) acquaintance networks of their parents (activated by their parents' prompting when necessary). Although the US population sampled in the GSS is of age 18 or older, we shall assume the samples to be representative of the total population, including those less than 18 years of age, by virtue of the above assumptions. Thus, the values for t we shall use are those for the total population of the US (Census projections, Bureau of the Census, 1989, Table 4, pp. 38, 40, 42 and 44). The numbers of homicide victims per year are from the 1988, 1989, 1990 and 1991 *Uniform Crime Reports* (Federal Bureau of Investigation, 1988, 1989, 1990, 1991), the numbers of suicide victims per year are from the 1993 *Statistical Abstract of the United States* (Bureau of the Census, 1993), and

Table 2
Average number of subpopulation members known per person $\pm 1.96 \times$ standard deviation of average from GSS

	1988	1989	1990	1991
m_H (homicides)	0.17241	0.16102	0.14599	–
$\pm 1.96 \times$ s.d.(m_H)	$\pm$0.03140	$\pm$0.02701	$\pm$0.02894	–
μ_H ($c = 2100$)	0.17646	0.18187	0.19656	–
($c = 1700$)	0.14285	0.14723	0.15912	–
($c = 1500$)	0.12604	0.12991	0.14040	–
($c = 1300$)	0.10924	0.11259	0.12168	–
m_S (suicides)	–	–	0.11752	0.14531
$\pm 1.96 \times$ s.d.(m_S)	–	–	$\pm$0.02084	$\pm$0.02302
μ_S ($c = 2100$)	–	–	0.25919	0.25473
($c = 1700$)	–	–	0.20982	0.20621
($c = 1500$)	–	–	0.18513	0.18195
($c = 1300$)	–	–	0.16045	0.15769
m_A (AIDS)	0.16757	0.14723	0.19621	0.25777
$\pm 1.96 \times$ s.d.(m_A)	$\pm$0.03561	$\pm$0.02825	$\pm$0.03463	$\pm$0.03701
μ_A ($c = 2100$)	0.53087	0.72742	1.01734	1.35071
($c = 1700$)	0.42975	0.58887	0.82356	1.09343
($c = 1500$)	0.37919	0.51959	0.72667	0.96479
($c = 1300$)	0.32864	0.45031	0.62978	0.83615

the cumulative numbers of AIDS victims are from a March 1988 *AIDS Weekly Surveillance Report*, and the March 1989, 1990 and 1991 *HIV/AIDS Surveillance Reports* and exclude victims outside the 50 states and District of Columbia.

In Table 2 are given, for 1988, 1989, 1990 and 1991, the values m_H, m_S, m_A, $\pm 1.96 \times$ s.d.(m_H), $\pm 1.96 \times$ s.d.(m_S) and $\pm 1.96 \times$ s.d.(m_A) derived from the corresponding GSS data, and the values μ_H, μ_S and μ_A for the values c_H, c_S and c_A posited as the high (2100), middle (1700), low (1300) and mid-low (1500) values of the range for c previously obtained by the authors (Killworth et al., 1990).

Finally, in Table 3 are the results of the hypotheses tests given by Eqs. (3) and (4) for the data in Table 2. Here the symbol "=" indicates that H_0 is accepted (the reported value of $e_E = e_H$, e_S or e_A is accepted), "<" indicates that H_0 is rejected on the low side (the reported value e_E is rejected as too high, i.e. $m_E < \mu_E - 1.96\text{s.d.}(m_E)$) and $(1/2)\mu_E \leq m_E$, "≪" indicates that H_0 is rejected on the low side and $(1.3)\mu_E \leq m_E < (1/2)\mu_E$, "⋘" indicates that H_0 is rejected on the low side and $m_E < (1/3)\mu_E$, and ">" indicates that H_0 is rejected on the high side (the reported value of e_E is rejected as too low, i.e. $\mu_E + 1.96\text{s.d.}(m_E) < m_E$).

The acceptance of H_0 for the reported values of e_H for all 3 years that the homicides questions were part of the GSS indicates that homicide victims may be a reasonable benchmark subpopulation for comparative purposes (as discussed by Laumann et al., 1989, 1993) and, in the process, suggests that $c \approx 1700$ may be a reasonable working estimate for the US adult average personal network size for

Table 3
Comparison of event population sizes derived from GSS data with reported event population sizes (UCR, VSUS, CDC) by hypothesis testing

	1988	1989	1990	1991
Homicides				
$c = 2100$	=	=	<	–
$c = 1700$	=	=	=	–
$c = 1500$	>	>	=	–
$c = 1300$	>	>	=	–
Suicides				
$c = 2100$	–	–	≪	<
$c = 1700$	–	–	<	<
$c = 1500$	–	–	<	<
$c = 1300$	–	–	<	=
AIDS				
$c = 2100$	⋘	⋘	⋘	⋘
$c = 1700$	≪	⋘	⋘	⋘
$c = 1500$	≪	⋘	⋘	⋘
$c = 1300$	<	⋘	⋘	⋘

this period of time. The rejection of H_0 and the level of the rejection for the reported values of e_S and e_A (except e_S for $c = 1300$) calls for further interpretation. After describing in the next section the method we used for estimating the average personal acquaintance network size for the US adult population, and presenting estimates of this value for two US communities on opposite sides of the country, we shall discuss and try to explain these results in Section 5.

4. Estimates of c for the US: two American communities

Freeman and Thompson (1989) describe a method for estimating the average number of persons a person knows in a population universe. They obtained a random sample of 305 surnames from the 1984 Orange County, California, phone book and presented these surnames by programmed computer to a sample of 247 student subjects (primarily undergraduates) from the University of California, Irvine. As described by Freeman and Thompson (p. 152), after some preliminary questions on each subject's age, gender, ethnic background, and parent's occupation.

'It (the computer program) then went on to define acquaintanceship, to describe the task, and to encourage the subject to take it seriously. Subjects were then presented with a surname. They were asked to try to remember if they had ever been acquainted with a person having that surname. Subjects who could not recall such an acquaintance were presented with the next surname. But those who could recall an acquaintance were asked to type in the person's first name and to identify whether that person was a member of their family, a close friend, or an acquaintance. They were then asked whether they knew anyone else with that surname, and the process was repeated until they ran dry. This process was repeated until all 305 names had been presented to each subject.'

If a given subject listed a total of N first names, then the expected number of first names generated per surname is $N/305$. Since the 305 names were a random sample of all the surnames in the 1984 Orange County phone book, Freeman and Thompson reasoned that the estimated number of distinct surnames of 112 147.2 in the book (a proxy for the US) would produce $(N/305)(112\,147.2)$ first names if the subject had been presented with all those names. With the average value of N being $\tilde{N} \approx 15$ they obtained an estimate of the mean number of acquaintances of 5520, with a standard error of 271. Of the names remembered, 22.2% were described as belonging to friends, 2.6% to family members, and the remaining 75.1% to acquaintances. We note that this estimation of c is essentially the same as that given in Eq. (5), with the proportion $112\,147.2/305 = t/e$ and the average value of N over the 247 subjects $\tilde{N} = m_{\mathrm{E}} = \tau/s$.

The reader should notice that in order for this computation to produce an accurate (but not necessarily precise) estimate of c for the US, scaling $\tilde{N}$ from the sample surnames to all surnames in the phone book requires that the sample surnames have the same number of listings in the book per surname as do all the surnames in the book, i.e. that on average the sample surnames have the same 'popularity' as all the surnames in the book, which is implied by $112\,147.2/305 = t/e$. It seems clear that if the sample surnames are on average more popular than those in the book (the universe) they will trigger in the minds of the subjects proportionately more first names per surname than they should, and similarly if less popular, proportionately fewer than they should. In addition, the estimation of the number of distinct surnames in the book by finding the average number of distinct surnames per column or page requires that an estimate of the average number of 'dud' lines occupied by empty space, large bold face advertisements, company listings, multiple phone listings and multiple lines for single phone listings be taken out of the count.

Using a later repeated count in the 1984 Orange County phone book by Alaina Michaelson and taking into account these corrections, we re-estimated c for Orange County as 2025 (Killworth et al., 1990). Thus, the correction to the original Freeman and Thompson estimate is substantial, our estimate of c being 37% of theirs. In a recent study by the authors (Bernard et al., 1990; Killworth et al., 1990) the Freeman–Thompson method with 98 subjects yielded the estimate $c \approx 1391$ for Jacksonville, Florida. Taking the Orange County and Jacksonville values as independent estimates of c for the US, we obtain the average $c \approx 1708$, which is virtually dead on 1702, the mean of the four estimates obtained by different methods for Jacksonville, having a spread of about ± 400 (Killworth et al. 1990) and a value which fits all 3 years of reported homicide data in Table 3. We also note that all the values for $\sim c_{\mathrm{H}}$ shown in Table 1 fall comfortably within the range 1700 ± 400, suggesting that homicide victims have personal networks similar to those of the total population.

The acceptance of hypothesis H_0 for homicides for all 3 years, using the earlier estimated value of $c = 1700$, indicates that if a respondent has an acquaintance who becomes a homicide victim then the respondent will with a high probability know this fact, and probably fairly rapidly. Even with a potentially large error in the value 1700, this is consistent with the belief in the accuracy of the reported annual counts of homicides. The validity and accuracy of the official statistics for homicides is agreed, by most researchers, to be the most reliably reported of the FBI's index of seven serious crimes (Gove et al., 1985).

5. Reconciling the results with the reported figures

The first and most direct interpretation of the results of the hypothesis test given in Eqs. (3) and (4) is that all the values of e_A and almost all the values of e_S reported are too high, that is, that suicides and AIDS victims are overcounted, and in particular that suicides are overcounted by a factor of about 1.6 and AIDS victims by a factor of about 3.7. It would appear, however, that the only ways AIDS victims could be overcounted relative to the GSS results would be that either

(a) a sizable proportion of AIDS victims generate multiple diagnoses, perhaps to ensure qualification for disability benefits, all of which are reported to the CDC, or
(b) there is a sizable multiple reporting of AIDS victims or sizable reporting of fictitious victims by health agencies, perhaps to qualify for more support from funding sources.

An average of about 3.7 reports per actual victim, however, seems highly unlikely. False positive diagnoses should not account for any significant part of an overcount either, since the social knowledge that such members were diagnosed with AIDS would get out into their networks just as it would for true positives, and be detected in the same way in a random survey such as the GSS. In the case of suicides, multiple and fictitious reporting seem almost impossible.

A second interpretation is that the average acquaintance network size for suicide victims is about 62% of that of the general population, while that for AIDS victims is only about 27%. From a social psychological perspective this possibility for suicide victims may have some merit, since suicide may be an indicator of a more withdrawn person having a smaller than average long-term acquaintance network. For AIDS victims this possibility is problematical.

With $m_A = 0.47309$ (1993 GSS), $e = 279\,854$ (CDC, 1993) and $t = 256\,466\,000$ (Bureau of the Census, 1989), we obtain the estimate $\sim c_A = 434$ for 1993, which is about the same as the estimates of $\sim c_A$ for 1988–1991 in Table 1. Assuming that respondents are likely to know when an acquaintance has AIDS, these figures imply that AIDS victims have an average acquaintance network size a little less than 1/3 the average for the total population. Before they had AIDS, victims would presumably have been developing acquaintance networks similar in size to those of the general population. Even though contracting AIDS may result in a significantly smaller network of individuals with whom a victim later interacts, this should not appreciably change the total number of the victim's acquaintances, many of them from the past. The assumption that a respondent is highly likely to know when an acquaintance has AIDS is thus highly suspect.

Finally, a third interpretation is that the GSS respondents are underreporting the members of their acquaintance networks who are suicides or who have AIDS. This could be because

(c) respondents did not know that members of their acquaintance networks have AIDS or were victims of suicide, probably because the victims or their close network alters severely controlled this information because of the stigma attached, or
(d) respondents knew but were not acknowledging that members of their acquaintance networks were victims of AIDS or suicide, probably because of the 'stigma of association' attached to knowing an AIDS or suicide victim.

These seem plausible, with (c) appearing more so than (d). In a national survey of surgeons by Shelley and Howard (1992), 38.5% and 44.6% of the sample said they had a friend, patient or acquaintance who had died of AIDS or been diagnosed as infected with HIV, respectively.
From an earlier model of the authors (Eq. (7) in the next section), with estimates of 92,435 deaths due to AIDS (CDC, 1993, Table 9; using a much later table of data, on the assumption that at the time of the survey surgeons had received information about AIDS victims in their personal networks more completely and sooner than had members of the general population) and total US population size $t = 249{,}690{,}000$ (Bureau of the Census, 1989, Table 4) at the time of the survey (estimated as March 1, 1990), we obtain an estimate of c for surgeons of 1313 relative to AIDS deaths. This figure is in the previously obtained range 1700 ± 400 for c and, when used in Eq. (7), with the estimated probability $p \approx 0.446$, produces the size estimate of 112,286 for the HIV + subpopulation at that time. It is not clear to what extent these surgeon-respondents included those known to have been diagnosed with AIDS, and still living, among those known to be HIV + . On average the survey respondents knew 0.669 friends or acquaintances who had died of AIDS, which by Eq. (5) yields the estimate $\sim c \approx 1807$ for surgeons, which is again in the range 1700 ± 400.
There was no 'stigma of association' which could be attached to individual surgeons here as the responses were anonymous and mailed in to the investigators. Surgeons are respondents for whom we expect medical information in their personal networks to be relatively unsuppressed. For the general US population, the relatively constant values of $\sim c_{\mathrm{A}}$ for 1988–1993 in the face of an apparent lessening of the 'stigma of association' in public during this period suggests that (c) is the effect which continues here.

Thus, we assume that AIDS victims have the same size distribution for their acquaintance networks as the general population. This implies that on average the AIDS victims in a respondent's acquaintance network whom the respondent knows have AIDS are a little less than one-third of all AIDS victims in that network. What could explain that? One possibility is that the one-third whom the respondent knows have AIDS are in the more intimate and smaller subnetwork consisting of his or her intimates – sympathetic friends, relatives and lovers – which we will call the *close personal network*. Since having AIDS is a highly stigmatizing attribute it is not a piece of information given out casually to others either by the victim or by close friends and relatives of the victim. In particular, it is not given out casually to mere acquaintances. One assumes that it would only be given out to members of the close personal network.

Freeman and Thompson (1989) found that 22.2% of all persons named by their subjects were friends and 2.6% were family. In their average acquaintance network of size 2025 (corrected) this yields $(2025)(0.222) \approx 450$ for friends alone and $(2025)(0.248) \approx 502$ for friends and family together. From Eq. (6) these figures yield the respective estimates 67 934 and 60 897 for the size of the AIDS population A in the US in 1988, which are fairly close to the reported figure. Unfortunately, the relatives, friends and acquaintance data from the Freeman–Thompson module in our Jacksonville survey (Bernard et al., 1990) were lost. We have a proxy, however, in the corresponding data collected in the Reverse-Small-World module in this survey. Here friends alone and friends and relatives together were about 36.4% and 51.3%, respectively, of all acquaintances named by the respondent, yielding the values $(1391)(0.364) \approx 506$ for friends alone and $(1391)(0.513) \approx 714$ for friends and relations together in 1988. From Eq. (6) these yield the respective estimates 60 416 and 42 816 for e_A. If we apply these "effective" figures for c_A to the AIDS data for 1989, 1990, 1991 and 1993 we again obtain rather pronounced low estimates for e_A.

According to our analysis to this point, we do not find evidence that the CDC figures are serious undercounts of the US AIDS population. However, this method of evoking a plausible "effective" subnetwork to deal with knowledge about the AIDS population, while reasonable a priori, may still run into a problem. We note from the responses to the GSS questions H.1, H.3, H.5, S.1, S.3, S.5, A.1, A.3, and A.5 that, excluding patients, respondents' classifications of the first best, second best and third best known AIDS victims as a close relative or friend are 43.7%, 41.8% and 39.2% of those three layers of their total networks, respectively, whereas for homicide victims they are 55.1%, 45.4% and 46.2%, respectively, and for suicide victims 53.3%, 41.4% and 64.7%, respectively. Thus, the expected closer personal relationship to respondents of reported alters who are AIDS victims compared to those who are homicide or suicide victims is not evident in the data. From the respective estimates 663, 425, 405, 401 and 434 of $\sim c_A$ for 1988, 1989, 1990, 1991 and 1993, we see that over a 5–6-year period there is no evidence that personal networks are opening up to relatively easier availability of information about AIDS victims, even though the population of AIDS victims and the probability of knowing an AIDS victim have increased over this period (see Tables 1 and 7).

6. Discussion

The above analysis gives independent confirmation of the UCR figures for annual numbers of homicide victims without the need for further explanation, but this is not true of the CDC figures for the total number of AIDS victims. The social context of this analysis for homicide victims seems straightforward, but the context for AIDS victims suggests that we need more information about peoples' acquaintance networks, such as the characteristics of their friend and family subnetworks and the ways that different types of information pass through their total acquaintance networks. Here we look for a further explanation, in the hope of developing some reasonable hypotheses. For this the results from a previously unreported acquaintanceship study in Gainesville, Florida, and a previously reported communication study in Jacksonville, Florida, are relevant.

In 1987–88 we surveyed a random sample of 372 respondents from Gainesville, Florida, who were asked to state whether they knew (i) a parent of twins, (ii) a parent of triplets, (iii) a police officer, (iv) a surgeon, and (v) a midwife. For these subpopulations we have known values of the size of e. Here we apply an equal likelihood probability model (Bernard et al., 1989), analogous to the one which gives the estimate (Eq. (5)), to the survey data obtained. Letting p_{E} be the probability that a member of the random sample S knows someone in the event population E,

$$\alpha_{\mathrm{E}} = \ln(1 - p_{\mathrm{E}})/\ln(1 - e/t) \quad (7)$$

yields a lower bound estimate for $c = c_{\mathrm{T}} \approx (t/e)m_{\mathrm{E}}$ (Bernard et al., 1989, Theorem A.2). For each event population the value of α_{E} and its range is given in Tables 4–6 based on the value of p_{E} and its 95% confidence interval, for T = total population of the US, Florida, and Gainesville, respectively. The data in these tables show the figures derived from this model when the total population and the event population, as well as the responses of the subject, are restricted to these three geographic locations. A random sample of subjects from Gainesville may not be a sufficiently representative sample for the US, and thus we include the data for Florida and Gainesville as well.

Since the α values are lower bounds for c_{T} for each total population T, we should try to obtain an estimated correction to them. Here we base our correction on what we call the *lead-in factor* $\lambda_{\mathrm{E}} \geq 1$, defined as the number of members of E

Table 4
Event populations in the US (1988) (t = 246048000)

Subpopulation E	e	Range of p_{E}	α_{E}	Range of α_{E}
Parents of twins	2400000	0.488 ± 0.051	68	59–79
Parents of triplets	12800	0.054 ± 0.023	1067	605–1540
Police officers	1140000	0.500 ± 0.051	149	128–172
Surgeons	128000	0.407 ± 0.050	1004	849–1174
Midwives	2800	0.105 ± 0.031	9748	6756–12846

Table 5
Event populations in Florida (1988) (t = 12306000)

Subpopulation E	e	Range of p_E	α_E	Range of α_E
Parents of twins	52390	0.334 ± 0.048	95	79–113
Parents of triplets	580	0.038 ± 0.019	822	407–1245
Police officers	33400	0.449 ± 0.051	219	187–255
Surgeons	2670	0.348 ± 0.048	1971	1644–2324
Midwives	261	0.083 ± 0.028	4085	2667–5547

a member of S will know on average if the member reports that he/she knows some member of E. It is easy to show, for n_0 = number in S who know no member of E, that

$$\lambda_E = \mu_E[s/(s-n_0)] = \mu_E/p_E \tag{8}$$

For the three event populations H, S and A above, we are able to obtain empirical values of λ from the GSS data, which are given in Table 7. It is clear from these values that the lead-in factor depends on the event population and gives an indication of how well an event population can be known once one knows a member of that subpopulation. Note, in particular, the relatively low value of λ_S, suggesting that socially there is less support for knowing further members of E if one knows some member of E when E = suicides than when E = homicides or AIDS victims. Thus, in a certain perceptible social sense suicides are relatively more isolated from each other than are homicides or AIDS victims.

For rare-event populations such as midwives, one might expect an approximately Poisson distribution of the number of its members in the acquaintance networks of the members of the total population and thus a value of λ very close to 1. If we make a strong assumption that other nonstigmatized event populations have lead-in factors like those for homicide victims we can produce, from Eqs. (5) and (8) with $\lambda \approx 1.6$, corrected estimates for $c = c_T \approx c_E$ from these subpopulations as

$$c \approx \sim c_E = (t/e)\lambda_E p_E \approx (1.6)(t/e)p_E \tag{9}$$

We show these estimates for the population T = US in Table 8 using the 95% confidence intervals for p_E given in Table 4. With this correction the estimates of c using E = parents of triplets and E = surgeons start becoming consistent with the

Table 6
Event populations in Gainesville (1988) (t = 85000)

Subpopulation E	e	Range of p_E	α_E	Range of α_E
Parents of twins	356	0.175 ± 0.039	46	35–57
Parents of triplets	4	0.024 ± 0.016	516	171–867
Police officers	206	0.296 ± 0.046	145	119–172
Surgeons	67	0.256 ± 0.044	375	302–452
Midwives	13	0.067 ± 0.025	453	281–631

Table 7
Probabilities of knowing a member and lead-in factors from the GSSs for the homicides, suicides and AIDS populations

	1988	1989	1990	1991	1992	1993	All years pooled
Homicides							
p_H	0.104	0.104	0.092	–	–	–	0.100
λ_H	1.656	1.553	1.587	–	–	–	1.599
Suicides							
p_S	–	–	0.096	0.112	–	–	0.105
λ_S	–	–	1.220	1.294	–	–	1.262
AIDS							
p_A	0.093	0.093	0.128	0.157	–	0.240	0.144
λ_A	1.797	1.580	1.528	1.639	–	1.969	1.751

previously discussed estimates $c_H = 2052$, 1859 and 1560 in Table 1 and the range 1700 ± 400.

When we restrict the data to Gainesville, we should obtain the most accuracy and consistency on such acquaintance networks since the sample is now random in the total population. The results in Table 6 appear to bear this out. We note that at the Gainesville level the confidence intervals for knowing surgeons, midwives and parents of triplets are nested, one within the next, and that the confidence interval for knowing police officers just overlaps that for knowing parents of triplets. At the Florida level the five confidence intervals are completely disjoint and at the US level there is almost no overlap of confidence intervals, except that the interval for surgeons is contained in that for parents of triplets.

The c values for Gainesville relative to parents of triplets, surgeons and midwives would all need to be scaled up by a factor of roughly 3.8 in order to attain $c \approx 1700$ at the US level. Because of the strong effect of geographical proximity on the composition of personal networks, it is reasonable to expect significant attenuation in the proportional rise of personal network size in moving from the Gainesville to the US level (population rise factor ≈ 2900), but we cannot say yet whether the factor 3.8 is at all reasonable. Preliminary p values of 0.49, 0.57 and 0.50, based on 542 respondents from a similar study conducted at the national level by Fairlee Winfield, are reasonably consistent with the p values we obtain with respect to parents of twins, police officers, and surgeons, respectively

Table 8
Estimates of c from four event populations in the US (1988) ($t = 246\,048\,000$, $\lambda = 1.6$)

Subpopulation E	e	Range of p_E	c	Range of c
Parents of twins	2400000	0.488 ± 0.051	80	72–88
Parents of triplets	12800	0.054 ± 0.023	1661	953–2368
Police officers	1140000	0.500 ± 0.051	173	155–190
Surgeons	128000	0.407 ± 0.050	1252	1098–1406

(the only event populations in her study which are identical to ours), so it appears that one can obtain reasonably reliable values for p even when sampling from localities of the US.

The average percentages of relatives, friends, and acquaintances in the personal networks of our respondents for the five event populations at the US level are 15.5%, 41.7%, and 42.9%, respectively. The averages for these three types of alters, when considered over all five event populations, range from 6.0 to 31.5% for relatives, from 39.2 to 45.2% for friends, and from 29.3 to 51.0% for acquaintances. Only the parents of twins population shows three extreme values in these ranges, the maximum for relatives and the minimum for friends and acquaintances, and only the surgeon population shows two, the minimum for relatives and the maximum for acquaintances. The police officer population shows the remaining extreme percentage, the maximum for friends. The extreme positions occupied by all three averages for twins suggests that parents of twins in acquaintance networks are rather biassed towards being relatives rather than friends or acquaintances. This appears to be a reason for the uniformly and unusually low values of α based on this event population. An argument similar to the one reconciling the AIDS data and results, assuming that information on twins is largely constrained to the subnetwork of relatives and close friends, would produce much improved estimates for c. The subnetwork of relatives and close friends typically occupies a relatively small part of a personal acquaintance network, and it appears that this is the locus of information and communication about the occurrence of twins. Further, there is a rather significant gender difference in the way parents of twins are remembered by respondents. Male respondents remember only slightly more female parents than male parents, whereas female respondents remember almost six times more female parents than male. Thus, a further interpretation of what is happening here is that information about parents of twins tends to remain concentrated among females, probably most intensely among female relatives.

This highly disproportionate remembering of parents of twins by gender does not occur for the other event populations, in particular for parents of triplets. Thus, the substantive grounds for expecting the argument for parents of twins to carry over to parents of triplets are absent. Perhaps the best explanation for this is that there is far more visibility and general news interest associated with triplets than with twins, so that even the farther regions of a personal acquaintance network have a reasonable chance of receiving this news in a short time. Note that for news about surgeons the favored part of a personal network is the mere acquaintance subnetwork, which is typically the largest part. This suggests the 'strength of weak (acquaintance) ties' for the transmission of information about surgeons, a group of persons occupying a highly instrumental position in personal networks (cf. Granovetter, 1973). Since the size of the friends subnetwork is typically between those for the relatives subnetwork and the mere acquaintance subnetwork, and since police officer's networks appear biassed toward friends, it is not hard to construct a similar plausibility argument to explain why the α value for police officers is between that for parents of twins and those for the remaining three subpopulations.

The above interpretations are further supported by another survey and analysis on communication to and from 21 informants in Jacksonville with a variety of news contents (Shelley et al., 1990). With regard to the news generally, it was found that the travel time of the news on average is about 4–30 times longer to or from a friend than a relative, and about 2–2.5 times longer to or from an acquaintance than a friend. It was also found that news travelled about 2.5–13 times faster between women than between men. Within these data, one informant, who went to his 25 year college class reunion during the time of the study, is an outlier. The ranges in the above comparisons are the result of both excluding and including him. He significantly distorts some of the data that is obtained when he is not included. However, this kind of activity, in which someone receives information from old friends and acquaintances at a reunion long after an event, is probably not rare in the US population.

We may wonder whether there is any significance in the extremely large values of α relative to knowing midwives in the US and in Florida, as shown in Tables 4 and 5. One possibility is that midwives are both rare and "exotic", which tends to make them have news interest disproportionately higher than that for other subpopulations. Another is that the assumption that respondents' acquaintance ties are symmetric may not apply very well to midwives. It is not hard to believe that if a very large number of remembrance ties are claimed to a midwife, many of these ties may not be reciprocated. This phenomenon may have occurred in our earlier Mexico City study (Bernard et al., 1991), where knowing a priest led to an average network size of at least 2254. Of course, there is the possibility that priests in Mexico City just might have acquaintance network sizes well above the average for the general Mexico City population, about 570 ± 460 in our recent study (Killworth et al., 1990), but another plausible reason is that there is a lot of directionality to remembering a priest, and that the remembering may be unreciprocated. When this is the case our model does not hold under the assumptions made in this paper (although it can still hold under other assumptions on the distribution of E in personal networks). However, another observation which needs to be made is that the rarer the event subpopulation E the greater the change in the estimated value of c from reporting knowing one more member of E. For rare subpopulations ($e \ll t$) this change is approximately t/se, where s is the size of the respondent sample. At the US level of the Gainesville survey, with $s = 372$, the increase in c is about 236 per additionally reported member of E = midwives, but only 52 for E = parents of triplets and 5.2 for E = surgeons. Thus, a small discrepancy in reporting midwives has a much larger effect on the estimated value of c than the same discrepancy in reporting parents of triplets or surgeons.

7. Further work

We have mentioned several items regarding personal networks about which we need much more information. After our analyses above and in Bernard et al. (1990) and Killworth et al. (1990), we are more optimistic about elaborating the structure of personal networks and the ways in which they operate. This becomes an exercise in theory building.

We need to determine certain details about personal networks – the characteristics of their members (the alters of ego), the different relations they have to ego, and the uses to which different members are put for communication and accomplishing tasks. Anthropologically, we need an elaboration of ego's network of friends and acquaintances like that of ego's kin. The personal network at different affective levels, as well as any formal organizational networks overlaying ego's personal network, are clearly important, as well as the kinds of activities which transpire in these networks.

We also need a better grasp of how a person knows and is related to others. How do people discern best friends, close friends, 'just' friends and 'mere' acquaintances, as well as the different levels of kinship? How are these different subnetworks used instrumentally? How does communication occur in these networks with regard to contents of messages, and at what speeds? How does the sensitivity and interest of news and the reluctance or eagerness of transmitters and relayers affect the details of its transmission through a network? What are the basic processes which occur within networks and what are their implications? For example, is a high lead-in factor for acquaintanceship with the members of a certain group related to high cohesiveness or gregariousness in that group?

As we close in on some of the fundamental quantities connected with human networks, such as the number of persons a person knows according to a reasonably precise definition of knowing, and the distribution of these quantities in various networks, we will be able to apply them with more rigor in the further investigation of the more detailed structures and processes occurring within these networks. Certainly a clear and major purpose of this work is the application of what is developed to the accurate and precise estimation of the sizes of important but hard-to-count event populations.

Acknowledgements

Behind this effort are the labors of several other people whose contributions were essential: Alaina Michaelson for conscientiously re-counting the 1984 Orange County phone book, and Timothy Brazill, Jamie Jacobs, and Chellyn Rinehart for data collection in the Gainesville, Florida, survey. We also want to express our thanks to Noah Friedkin for his careful reading of an earlier version of this paper and his useful comments, questions and suggestions.

References

Berkelman, Ruth, James Curran, William Darrow, Tim Dondero and Meade Morgan
1989 "Monitoring the US AIDS epidemic." Letter in *Science 245*: 908.

Bernard, H. Russell, Eugene C. Johnsen, Peter D. Killworth, Christopher McCarty, Scott Robinson and Gene A. Shelley
1990 "Comparing four different methods for measuring personal social networks." *Social Networks 12*: 179–215.
Bernard, H. Russell, Eugene C. Johnsen, Peter D. Killworth and Scott Robinson
1989 "Estimating the size of an average personal network and of an event subpopulation." In: Manfred Kochen (Editor) *The Small World*, Norwood, NJ: Ablex, pp. 159–175.
1991 "Estimating the size of an average personal network and of an event subpopulation: Some empirical results." *Social Science Research 20*: 109–121.
Bureau of the Census
1983 *1980 Census of Population*, Vol. 1. Characteristics of the Population; Chapter D, Detailed Population Characteristics; Part 1, United States Summary; Section A; United States. US Department of Commerce, Washington, DC, US Government Printing Office, March 1994.
1989 "Projections of the population of the United States, by age, sex, and race: 1988 to 2080." *Current Population Reports*, Population Estimates and Projections, Series P-25, No. 1018, US Department of Commerce, US Government Printing Office, Washington, DC, Table 4, p. 38ff.
1993 *Statistical Abstract of the United States: 1993*. 113th Edition. US Department of Commerce, Economics and Statistics Administration, Washington, DC, 1993.
Centers for Disease Control
1988 *AIDS Weekly Surveillance Report* US Department of Health and Human Services, Public Health Service, Atlanta, GA, 21 March.
1989 *HIV / AIDS Surveillance Report*, First Quarter Edition. US Department of Health and Human Services, Public Health Service, Atlanta, GA, March.
1990 *HIV / AIDS Surveillance Report*, First Quarter Edition. US Department of Health and Human Services, Public Health Service, Atlanta, GA, March.
1991 *HIV / AIDS Surveillance Report*, First Quarter Edition. US Department of Health and Human Services, Public Health Service, Atlanta, GA, March.
1993 *HIV / AIDS Surveillance Report*, First Quarter Edition. US Department of Health and Human Services, Public Health Service, Atlanta, GA, May.
Federal Bureau of Investigation
1986 *Uniform Crime Reports for the United States*. US Department of Justice, US Government Printing Office, Washington, DC.
1988 *Uniform Crime Reports for the United States*. US Department of Justice, US Government Printing Office, Washington, DC.
1989 *Uniform Crime Reports for the United States*. US Department of Justice, US Government Printing Office, Washington, DC.
1990 *Uniform Crime Reports for the United States*. US Department of Justice, US Government Printing Office, Washington, DC.
1991 *Uniform Crime Reports for the United States*. US Department of Justice, US Government Printing Office, Washington, DC.
Freeman, Linton C. and Claire R. Thompson
1989 "Estimating acquaintanceship volume." In: Manfred Kochen (Editor) *The Small World*. Norwood, NJ: Ablex, pp. 147–158.
Gove, Walter R., Michael Hughes and Michael Geerken
1985 "Are uniform crime reports a valid indicator of the index crimes? An affirmative answer with minor qualifications." *Criminology 23*: 451–501.
Granovetter, Mark S.
1973 "The strength of weak ties." *American Journal of Sociology 78*: 1360–1380.
Killworth, Peter D., Eugene C. Johnsen, H. Russell Bernard, Gene A. Shelley and Christopher McCarty
1990 "Estimating the size of personal networks." *Social Networks 12*: 289–312.
Laumann, E.O. and J.H. Gagnon
1989 Response to Berkelman et al. *Science 245*: 908–910.

Laumann, E.O., J.H. Gagnon, S. Michaels, R.T. Michael and J.S. Coleman
1989 "Monitoring the AIDS epidemic in the United States: A network approach." *Science 244*: 1186–1189.

Laumann, Edward O., John H. Gagnon, Stuart Michaels, Robert T. Michael and L. Philip Schumm
1993 "Monitoring AIDS and other rare population events: A network approach." *Journal of Health and Social Behavior 34*: 7–22.

National Opinion Research Center
1993 *General Social Surveys, 1972–1993: Cumulative Codebook*. Chicago, IL: University of Chicago, Press.

Shelley, Gene Ann and Richard J. Howard
1992 "A national survey of surgeons' attitudes about patients with human immunodeficiency virus infections and acquired immunodeficiency syndrome." *Archives of Surgery 127*: 206–211.

Shelley, Gene Ann, H. Russell Bernard and Peter D. Killworth
1990 "Information flow in social networks." *Journal of Quantitative Anthropology 2*: 201–225.