

CHAPTER 8

Estimating Acquaintanceship Volume*

Linton C. Freeman and Claire R. Thompson

INTRODUCTION TO THE PROBLEM

In their pioneering article on networks, Pool and Kochen (1978) pointed to a whole new area for research by students of social structure. They focussed on networks of acquaintanceship, where a pair of people is considered to be linked if they have met and would recognize each other if they met again.

Pool and Kochen stressed the fundamental nature of the acquaintanceship relation. They observed that, whenever people face problems that they cannot solve alone, they turn to others they already know—to their acquaintances. Acquaintanceship networks, therefore, provide people with sets of others from whom they can and do draw help, support, information, advice, and the like in time of need.

Pool and Kochen called attention both to the universality of such acquaintanceship networks among humans and to their importance in providing contacts who can help when help is needed. In particular, they stressed the importance of such personal acquaintanceship networks in the process of political influence. But later workers have shown that such networks are also important in providing general social support (Wellman, 1981), in the diffusion of innovations (Coleman, Katz, & Menzel, 1966), and in the process of search for such diverse things as a job (Granovetter, 1974), a place to live (Freeman & Sunshine, 1976), and an abortion (Lee, 1969). Thus, acquaintanceship networks turn out to provide an essential mechanism by means of which the individual is connected to the larger world.

Pool and Kochen, however, did more than simply sensitize the social science community to the importance of studying acquaintanceship networks. They went on to provide an explicit agenda for research on such networks. Indeed, Pool and

* The authors wish to thank A. Kimball Romney, Stephen P. Borgatti, Bruce MacEvoy, and Samuel Gilmore for their helpful comments on an earlier draft of this paper.

Kochen (1978, pp. 6–7) set down eight features of acquaintanceship networks that they saw as fundamental to the specification of their form:

1. For any individual we should like to know how many other people he knows, i.e. his acquaintance volume.
2. For a population we want to know the distribution of acquaintance volumes, the mean and the range between extremes.
3. We want to know what kinds of people they are who have many contacts and whether these people are also the influentials.
4. We want to know how the lines of contact are stratified; what is the structure of the network.
5. How great is the probability that two persons chosen at random from the population will know each other?
6. How great is the chance that they will have a friend in common?
7. How great is the chance that the shortest chain between them requires two intermediaries?
8. How far are people aware of the available lines of contact?

Although many of these questions are rather imprecise (4 and 8 in particular), their authors' intent is clear: The first seven all point to the issue of specifying important parameters of the overall form of acquaintanceship networks. And the eighth proposes that the relation between that form and people's perceptions be investigated. Answers to these questions would specify some of the most basic parameters for any network study. We take the position that estimating them, or something close to them, is absolutely essential to progress in structural research.

In this study we focus on only one of these questions: the first, and perhaps the most basic. We will try to estimate people's acquaintanceship volume. In the next section we begin by reviewing earlier attempts at estimating this parameter. Following that, we will describe our own procedures for gathering data. Finally, we shall report our results.

RESEARCH ON ACQUAINTANCESHIP

In their original draft, written in the 1950s, Pool and Kochen (1978) reported two pilot experiments designed to estimate acquaintanceship volume. One was based on a diary of interpersonal contacts, and the other used phone books as an aid in recalling acquaintances. In both cases, there was only one experimental subject, Pool himself.

The diary method required that Pool carry a book and that he record each of his interpersonal contacts each day. A contact was recorded whenever Pool exchanged words with any person he had previously met and whose name he knew. Exchanges on the phone or by mail were included. Multiple contacts with

the same person on a single day were recorded only once. Records were made for 100 days.

Data were tabulated in terms of the cumulative number of names that appeared on the list as time passed. On the first day, the names of all of the people contacted were new to the list. But on the second day, some of the people seen had been listed on the first day, so there were fewer new names. As time passed, fewer and fewer new names were introduced each day. Growth had not stopped entirely, but very few names were coming in on the 99th and 100th days.

This growth curve, based on this declining rate of the appearance of new names, was then extrapolated to estimate the number of names it would include had the data been collected for 20 years. For Pool, the estimate was that he had about 3500 acquaintances.

The phone book method embodied quite a different approach. It was focused, not on contact frequency, but on people's ability to remember their acquaintances. Unaided recall of a long list of this sort is probably impossible. People simply can't sit down and come up with exhaustive lists of hundreds (or thousands) of others they know. What was needed was a device to jiggle the subject's memory. In this case, the device was a phone book.

Pool selected phone books from two large metropolitan areas: Chicago and Manhattan. He randomly selected 30 pages from each book. Then, as an experimental subject, he looked over each of the selected pages. Whenever he could recall an acquaintance with the same family name as someone on the sample page, he recorded a "hit." Then the proportion of pages that generated "hits" was multiplied by the total number of pages in the book to estimate the total number of names that would have been generated if he had looked at every page. For Pool, the Chicago book produced an estimate of 3100 acquaintances and the Manhattan book generated an estimate of 4250.

Other attempts to estimate acquaintanceship volume have been few and far between. One of Pool's students, Michael Gurevich (1961), used the diary method to collect data on 18 subjects. Each subject kept a diary of contacts, just as Pool had earlier. Acquaintanceship volume over a 20-year period was estimated for each of these subjects. Estimates varied from 122 to 5053; their mean was 2130.

Another MIT student, Howard Rosenthal (1960) analyzed 86 days of Franklin Delano Roosevelt's presidential appointment book as if it had been a diary record of Roosevelt's contacts. This led to a 20-year estimate of 22,500 for the President.

Since the late 1950s, the only studies bearing directly on this problem were reported by Killworth and Bernard (1978-1979) and Killworth, Bernard, and McCarty (1984). These investigators used a method completely different from either of those developed by Pool and Kochen. This new method is based on a hypothetical situation.

In both of these studies, each subject was presented with a long list of (actually fictitious) others who serve as targets for a hypothetical personal communication. Each of the targets was characterized by a name, an address, and an occupation. Subjects were asked to imagine that they had an important message that they needed to get to a target through a chain of acquaintances. Their job, then, was to choose from among their acquaintances the one who would be most likely to move the message along toward that target.

This same process was repeated for each of the targets in the list. Thus, subjects generated names of acquaintances whom they chose as facilitators for their communication. Some acquaintances were used only once as facilitators, but the names of others came up again and again. As the lists grew, new names were introduced at a diminishing rate. This listing, therefore is similar to that generated by Pool and Kochen's diary method. And the same kind of extrapolation method was used to estimate the full size of each subject's acquaintanceship volume.

The first of these studies was conducted in Morgantown, West Virginia. Fifty-eight subjects were run through 1267 target descriptions. They produced from 43 to 1131 names, with a mean of 210. From these figures, the estimated network size was 250.

The second study took place in Gainesville, Florida and involved 40 subjects. In this case, 500 target names were used. And, from this second experiment, the estimated acquaintanceship volume again was about 250.

This, then, is the state of current knowledge about acquaintanceship volume. Estimates of acquaintanceship volume provided by this research vary widely. Undoubtedly, individuals do vary in the number of acquaintances they have. But, from the small samples used in these studies, it is difficult to establish any generalizations about either averages or the range of variation. The present study is designed as a first step in providing such generalizations. Our procedures for collecting data are described in the next section.

COLLECTING ACQUAINTANCESHIP DATA

For our study, we wanted a reasonably large sample of a kind of behavior that would allow us to use straightforward procedures of estimation. The diary method was, on the face of it, too cumbersome and too expensive to permit a large sample. And the hypothetical message method introduced by Killworth and Bernard (1978-1979) seemed to us to be estimating a somewhat different parameter. (We will address that question in the discussion at the end of this paper.) So we chose to use a variant of Pool and Kochen's phone book approach.

The phone book approach, as defined by Pool and Kochen (1978), had a major design flaw resulting from the choice of the page as the stimulus unit. The variability in the number of different names on different pages introduced an

uncontrolled source of sampling variation. Several entire pages, for example, were undoubtedly devoted entirely to the name *Smith*, while a name like *Tchoch* might get only one line in one column on one page. Moreover, in any large directory, some whole pages will be made up of commercial listings (*American* this and that, for example). This means that, depending on the luck of the draw, one subject might be exposed to several hundred more usable stimuli than another. Under such circumstances, estimating parameters becomes difficult.

Even more important is the potential bias introduced by this procedure. In general, subjects are more likely to be acquainted with others who have relatively common names. At the same time, common names take up more space in the book. Sampling pages, therefore, is more likely to present subjects with common names than with uncommon ones. Common names, then, are more likely to be known by subjects and they are more likely to be presented as stimuli. This combination means that the page sampling procedure is biased toward presenting the subjects with names that are "good," those that yield recall. In consequence, the page sampling procedure, by overrepresenting the more common names, is biased toward overestimating acquaintanceship volume.

For the present research, we elected to use surnames from the phone book—not pages—as stimulus materials. And since we planned to use Orange County students as subjects, we reasoned that the richest stimulus materials would come from the local Orange County phone book.

The Orange County phone book is a large one. It contains 1416 four-column pages. Each column has 92 entries. The entire book, then, lists $(1416)(4)(92) = 512,088$ phone numbers. Surnames in the directory were sampled by generating a series of random numbers. We programmed a computer to generate a random number between 1 and 1416 to choose a page, a second random number between 1 and 4 to select a column, and a third between 1 and 109 to pick a name. Then the name printed on the chosen page, column, and line had to meet two criteria to be included: (a) it had to be the listing of an individual, not an organization; and (b) to prevent overrepresentation of more common names, the surname in question had to be the first listing of that surname.

If both these conditions were not met, a new set of random numbers was generated and the whole procedure was repeated. When an acceptable surname was found, it was stored and we generated new numbers and began the search for another. In this way, 305 surnames were chosen. These names are a random sample of the surnames listed in the Orange County phone directory (see the Appendix for the listing). They are the stimulus materials for this study.

Subjects for this research were 247 students—primarily undergraduates—from the University of California, Irvine. Most were females—168, to 79 males. They ranged in age from 17 to 59. The mean age was 20.8. Ethnically, 148 were Caucasian, 75 were Asian, and the remaining 24 were an assortment of primarily Hispanics, Blacks and Middle Easterners.

A computer was programmed to run the whole data gathering procedure. The

program began with preliminary questions on each subject's age, gender, ethnic background, and parents' occupations. It 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.

RESULTS

In response to the 305 family names provided as stimuli, subjects produced from 0 to 58 first names of others they classified as family members, friends, or acquaintances. The average was 15.

To use these numbers to estimate acquaintanceship volume, we needed to know the total number of different surnames in the Orange County phone book. We estimated this number by randomly selecting 100 columns in the book and counting the number of different surnames in each. The mean was 19.8 per column. Given four columns per page and 1416 pages, this suggests that the entire book contains $(19.8)(4)(1416) = 112,147.2$ distinct surnames. This number was used in making estimates.

A given subject, let us say, produced N names that were generated by exposure to 305 stimuli. Given a surname from the book, then, the expected number of names generated is simply $N/305$. And, since the 305 names were randomly sampled from all the names in the book, the estimated number of names that would have been produced if the subject had been presented with all those names is $(N/305)(112,147.2)$.

The observed distribution of acquaintanceship estimated by this procedure is shown in Figure 1. The mean of the estimate is 5520 acquaintances and the standard error is 271. This suggests that the true mean number of acquaintances for these subjects is very likely to fall somewhere between 4707 and 6333.

Of the names remembered, 22.2% were described as belonging to friends, 2.6% were classified as family members, and the remaining 75.1% were simply called acquaintances. Fifty-six percent were names of males, and 44% were females.

Neither the age nor the gender of the subjects affected the number of names they remembered. Age, in fact, was not related to responses in any way. Subjects' gender, however, did influence the kinds of others they recalled. Although

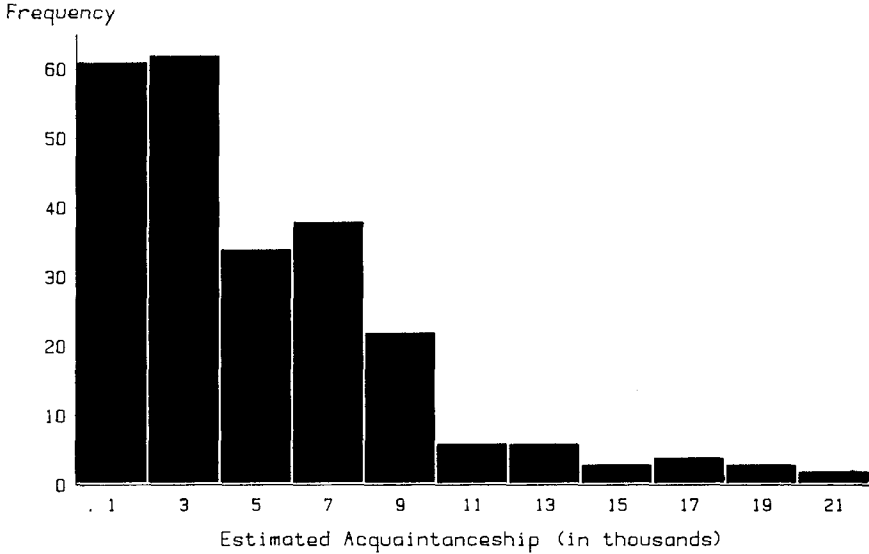


Figure 1. Estimated Acquaintanceship (in thousands)

both male and female subjects remembered more males than females, male subjects remembered significantly more male targets than did females. Moreover, although both male and female subjects characterized those they remembered as kin about equally often, female subjects were more likely to recall others as being acquaintances, while male subjects remembered significantly more others as friends. Figure 2 shows that both male and female subjects remembered about the same proportion of female friends and that they did not differ significantly in their memory for male acquaintances. But males remembered significantly more male friends, and females remembered significantly more female acquaintances.

There were also significant differences in the number of names generated

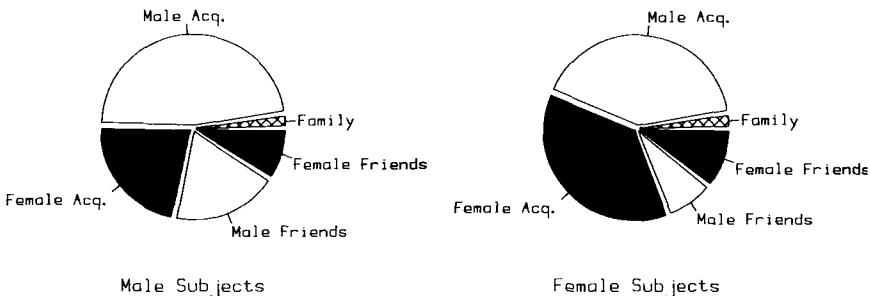


Figure 2. Male Subjects Female Subjects

according to the ethnicity of subjects. Caucasian subjects produced an estimated 6655 acquaintances on average, while Asians produced only 3808. The remaining assortment of subjects with other ethnic backgrounds produced an estimated 3864 acquaintances on average, but their number (24) was too small to take this estimate seriously.

SUMMARY AND DISCUSSION OF RESULTS

This study was designed explicitly to follow up on issues raised in the classic paper by Pool and Kochen (1978). Among other important suggestions they made was a proposal that acquaintanceship volume be estimated. We set out to do exactly that.

To do so, we modified a procedure—using a phone book to stimulate memory—that was described in that same classic paper. Our modifications simplified the procedure and allowed us to specify narrower error bounds for the estimate. The result was that we were able to run 274 subjects and produce an estimated acquaintanceship volume of 5520.

Unfortunately, there are few earlier estimates to which we can compare this value. Those produced by Pool and Kochen were of the same order of magnitude, but smaller. They were, however, based on a sample of 1. That produced by Rosenthal (1960) also involved a sample of 1 (President Roosevelt), and its estimate was almost four times larger than ours. This is not entirely unreasonable if we recall that our subjects were college students and FDR was the President of the United States.

Our current estimate is considerably larger than the 2131.5 estimated by Gurevich (1961). We suspect that this difference is an instrument effect. Gurevich's diary method recorded only those others with whom the subject was currently in contact, but our procedure tapped into a person's memory for all those he or she had ever met.

If we are concerned with acquaintanceship volume, both of these estimates have to be lower bounds. Any person's acquaintances are, by definition, those others whom he or she has met and gotten to know well enough to recognize. For most human beings in the modern world the process of making new acquaintances in this way is a lifelong process. It begins shortly after birth and continues throughout life. If we actually watched people over their entire lives, we could record all their meetings and count the number of their acquaintances at any moment. This, of course, is impossible.

The diary method as used by Pool and Kochen and by Gurevich provides an estimate of the subset of a subject's total acquaintances that are active in a sample time period. In a modern, mobile society where people come and go, past acquaintances may be inactive, but they are, nonetheless acquaintances. The set

of active acquaintances, therefore, would be expected to be smaller than the total acquaintanceship set.

Our adaptation of the phone book method should also produce estimates that are smaller than the total set. Even though all of us continue to enlarge our sets of contacts throughout our lives, it is unlikely that any one of us recalls everyone that he or she has ever met. On the other hand, we can all certainly recall having met people whom we have not seen for years and who would not turn up on any list of contacts with active acquaintances. This method should, therefore, in general produce a larger estimate than the diary method.

Our current estimate is also larger than that provided by Killworth and Bernard (1978–1979) and Killworth, Bernard, and McCarty (1984). It is larger, in fact, by an order of magnitude. We would guess that this difference is again, and even more dramatically, an instrument effect.

Killworth, Bernard, and McCarty's method of hypothetical situations is focused neither on people's active acquaintanceship contacts nor on their recall of acquaintances. Instead, it is concerned with people's actual choice of acquaintances to use in solving a communication problem. Thus, their approach not only involves acquaintanceship contacts and memory of those contacts, but adds the extra dimension of strategy of individual decision making.

Their extremely low estimate of the number of acquaintances is probably a consequence of their having added that extra dimension. Some subjects, for example, may remember acquaintances whom they are unwilling to contact because of unhappy previous experiences. Some may recall possible intermediaries but discard them because they have no idea of how to contact them. Other possible intermediaries may not be chosen because they do not seem likely candidates for the job of moving a message along to Mary Evans in Newfoundland.

Overall, this experiment generated names simply by asking subjects to repeat the same task again and again with varying targets. Wise strategists would confine their nominations to those acquaintances whom they know how to contact and whom they think of as somehow "well connected" to the world at large. Indeed, there is no reason to believe that a narrowly restricted stimulus of this sort would exhaust the subjects' ability to generate the names of acquaintances. It seems reasonable to get a result that suggests that the set of acquaintances generated by this restricted task would be considerably smaller than that produced by more direct methods.

We will end this report with a plea for more research in this area. Over 30 years ago, Pool and Kochen proposed a research agenda. It was a good agenda—the kind that might yield the sort of fundamental social parameters from which real sociological theory might be built. But 30 years has seen little progress toward estimating these parameters. We believe that the field of structural studies will not progress until its practitioners agree upon some goals and move collec-

tively towards their achievement. We hope we have made a small step in that direction, and we encourage others to join us.

REFERENCES

- Coleman, J.S., Katz, E., & Menzel, H. (1966). *Medical innovation*. New York: Bobbs-Merrill.
- Freeman, L.C., & Sunshine, M.H. (1976). Race and intra-urban migration. *Demography*, 13, 571-575.
- Granovetter, M. (1974). *Getting a job: A study of contacts and careers*. Cambridge, MA: Harvard University Press.
- Gurevich, M. (1961). *The social structure of acquaintanceship networks*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.
- Killworth, P.D., & Bernard, H.R. (1978-1979). The reverse small-world experiment. *Social Networks*, 1, 159-192.
- Killworth, P.D., Bernard, H.R., & McCarty, C. (1984). Measuring patterns of acquaintanceship. *Current Anthropology*, 23, 318-397.
- Lee, N.H. (1969). *The search for an abortionist*. Chicago, IL: University of Chicago Press.
- Pool, I. deS. & Kochen, M. (1978). Contacts and influence. *Social Networks*, 1, 5-51.
- Rosenthal, H. (1960). Acquaintances and contacts of Franklin Roosevelt. Unpublished thesis, Massachusetts Institute of Technology.
- Wellman, B. (1981). Applying network analysis to the study of support. In B. Gottlieb (Ed.), *Social networks and social support*. Beverly Hills, CA: Sage.

APPENDIX

SAMPLE OF 305 FAMILY NAMES FROM THE ORANGE COUNTY PHONE BOOK

BURKHARD	HODGES	PUTIGNANI
CAISER	CHAVEZ	CARTHEW
CIKIE	ACOSTA	TAHTI
CARATAN	FAVER	GAY
CALDASSANO	RAVIS	MOKHIARI
CERAZZA	EMPEY	WREN
COLLIARD	SHAYAN	NIENSTEDT
CANDRUFF	MARIACHER	BARBAROSH
CLER	SQUIRRELL	CARRENO
CLERGEORGE	WARRINER	SERRIO
CLAARTMANN-MOE	VONESH	LAUNDRIE
CLVELTRI	AARONS	LEWELLEN
CLOKES	SHIGEKAWA	SONES
CLMOULDS	TABB	BRAVIN
CLAMBRIA	LIEBMANN	HOOLEY

WOMACK	BROZOWSKI	CRONICK
TARAZEWICZ	HENDRICKS	PINELL
CRUME	KOZLIK	HATFIELD
JOYZULIA	RALLYE	KAUFMANN
DAHLROTH	FIELD	KING
MANUKAILEA	DIVIRGILIO	WALAS
MONTENEGRO	EVENSEN	CHOMIK
DENUNZIO	YAMASHIRO	BEHRINGER
REEVE	BORRENSSEN	FRATT
CARDONA	FEINGOLD	NALL
DOUGAL	STANDARD	VAN ORDEN
TARR	SCHERR	KANER
AYLESWORTH	GLASER	COUDRAY
HUMESTON	SARVIS	ROCKMAN
BASOM	CHARM	HERBALIFE
MARCZEWSKI	ALBINI	GATTIS
NEWHALL	KASNER	LEMMON
SANDRITTER	HUSA	GARBERSON
LEONG	MOYAO	SHEETS
ILUSTRISIMO	LYCAN	LIMONICK
FRANKHOUSER	PELTOMAA	TENORIO
NIETZSCHE	DOCKWILLER	ARBOUR
FORREST	KARIM	MORF
BULOW	SHOWLER	DUVE
DOLAN	PRETKO	KITTRIDGE
HERMAN	HOWSER	CHOUERI
GIDLEY	LINDAMOOD	KENNEPOHL
BOURASSA	GERTS	ESSIG
RENNEBERG	RUSSAVAGE	HARJO
RALEY	CHAREUNSY	YARD
ESBENSHADE	MACON	NAHID
MEGGINSON	LOEB	GENNETTE
ST CROIX	SHIPLEY	CARADONNA
JASNER	AVISE	CRAVIERO
GOEBELS	MALLOS	KILSGAARO
VALENTI	MOSCHEL	BRILL
IRISH	SIZEMORE	CADY
BOHLE	PEEL	BOSWELL
WINGARD	BAHARIE	SIGNER
LAUGHMAN	FETTERLING	SITTER
BOHM	MESAROS	HAGER
PISCOPO	BLAUE	BELIK
RICKEY	CONNER	HAMO

HIGHT	BOEHMKE	ALVARADO
GARSOMBKE	KAREL	BODET
FAKOURI	DOUEZ	KIRTLAND
STOLTZE	GALEA	TUSSEY
HEAGEN	MARERRO	SEE
SAXBY	FARNER	LE DU
AMLAW	TAKAI	POST
CALLAHAM	PELLEGRINO	KRUPA
ABOUDAN	STEAK	BIRECREE
RADKEY	HATZOPOULOS	MEMBREZ
TANNENBAUM	SKALBECK	ARNDS
DE GENNARO	OPHASO	DO PHUONG
KUESTER	OVIATT	ACORN
ALONZO	CARMACK	RANGER
ROJKO	ENANY	JOSHI
SURRA	HULET	WILDRICK
KOSKI	KANEASTER	GARDIKAS
CAPASSO	CSABA	TRANI
BORGIO	QUESADA	GODOY
WORKMAN	O'GORDON	DEMAREST
TOYODA	VIRDINIS	NOLAZCO
GRAFICO	OFAKIPOUONO	SHRADER
JEZAK	WESLING	FILLINGANE
MACHT	LISA	ANDRYS
DELIA	PECORA	ALTSTATT
GREYSHOCK	LONGHINI	NAZARIAN
NEIDERT	THRUPP	LA BELLE
GREENLAW	DE GEARE	MOENCH
KROPIDLOWSKI	SKARPHOL	TAVODA
BOLING	GONDA	CARCHIDI
GIESEKE	FUKUNGA	HARAN
LIONELLO	ENTRIKIN	REDINBO
FORTENBERRY	SUNDERMAN	RAUSCH
TRAKTMAN	HENRICHSEN	LILLYWHITE
PHILIPS	DOD	LANDOE
BROZICK	DEMIRJIAN	HALLICY
DOSCHER	ANCLADE	WIGER
MOYNIHAN	PIER	WELTZ
HENRYCK	AVARBUCH	KYSELLA
BLAISDELL	ORDIALES	MERROW
STEINBERGS	CHEVLIN	NOBER
GREENSTONE	LUISI	ARZOLA
SWANN	DELUCA	GLASNER
DES ARMIER	LACARRE	