A Method for Systematic Reduction of the Number of Questions in a Network Matrix Questionnaire

ROBERT T. TROTTER, II, ANNE M. BOWEN and HEATHER J. HURLBERT

Anthropology Department
Campus Box 15200
Northern Arizona University
Flagstaff, AZ 86011, USA

ABSTRACT: This article describes a systematic method for reducing the number of questions in a network instrument, without losing key elements of the analytical structure present in the larger questionnaire. Our goal in this process has been to conserve the instrument's theoretical structure and to preserve its analytical power, while reducing unnecessary redundancy. The process results in a revised data collection instrument which creates less of a burden on the networks from which we collect data, and a reduced number of analyses that must be completed on the data set, without sacrificing richness of data or strength of interpretation.

INTRODUCTION

The Flagstaff Multicultural AIDS Prevention Program (FMAPP)is a NIDA funded project[1] that conducts HIV prevention research with active drug users. The primary goal of the project is to provide outreach and intervention to members of hidden drug using populations who are at risk for HIV infection through injection drug use or sexual activities tied to drug use. We employ psychosocial and network interventions to improve our clients knowledge, skills, and motivation to reduce their HIV infection risks. The project conducts ethnographic network mapping, ego-centered data collection, and reciprocal relational data collection on drug using networks, as one group in the overall cadre of techniques used to accomplish its goals.

We apply network approaches in three arenas: outreach, intervention, and program evaluation. The project's outreach activities identify and recruit active drug users using a network based approach (Trotter et al. 1993). This has proven a successful system for finding, recruiting, and managing follow-up processes for hidden populations (Klovdahl 1985; Klovdahl et al. 1994). Our intervention research goals are addressed by randomly assigning networks to one of two prevention strategies. The first is primarily didactic. The second includes the use of both psychosocial theories (Bowen and Trotter in press), and theories from network approaches (Trotter, Bowen,

Journal of Quantitative Anthropology **6**, 35–47, 1996.

Potter 1993). The basic premises behind the network intervention are, 1) that changing group norms can cause protective behavioral changes within a group which might otherwise prevent change at the individual level (peer pressure), and 2) that cost effective diffusion of protective information to individuals not directly recruited into the project can be accomplished through a network intervention approach. The analysis of project outcomes are based on ethnographic observations and interviews, psychosocial quantitative measures of intervention outcomes, and network analysis of risk, communication, social relationships, and changes over time.

II. DEVELOPMENT OF THE FMAPP SOCIAL NETWORK INSTRUMENT

We have created a relational network matrix questionnaire which allows us to conduct social network analysis on drug using groups recruited into the project. The matrix questions are constructed to provide information on 1) risk relationships within the network, 2) social conditions and network characteristics that would assist in reducing risks (communication patterns, influence, cliques and factions, strong versus weak relationships, etc.), and 3) variables that allow us to measure (either directly or by proxy) changes in network relationships and structures through time. The change measures were selected to indicate a reduction in risk taking behavior (e.g. elimination of weak ties, such as casual sexual partners, reduction in needle sharing, etc.) The current instrument went through a two stage development; pretest and revision, then application to newly recruited drug networks.

The groups we are studying are at risk for HIV infection from injection drug use (direct and indirect sharing of infected injection paraphernalia), or from sexual relationships connected with drug use (sex for drugs, sex with injection drug users, high risk casual sexual partners, etc.). The items included in the pretest questionnaire asked about social relationships, drug use relationships, and issues of trust. Each question in the pretest instrument was generated from ethnographic data, confirmed by both research respondents and project personnel. Each question was matched with a reciprocal question. For example, we asked both, "How often are you willing to share needles with ____?" and "How often is ____ willing to share needles with you?" Each question was measured on a six point scale that is verbally anchored at each level of the scale. We pretested the questionnaire on three drug using groups and found that the reciprocal questions were redundant. The respondents reported virtually identical relationships in both directions for these questions. Therefore we kept only one question out of each reciprocal pair in the final questionnaire.

We computed basic exploratory network analysis measures (distance, flow, influence, structural equivalence, cliques, etc.) for the three pretest networks and found that the questions were distinguishing between social

relationships and drug relationships in ways that were both theoretically and pragmatically important (Trotter et al. 1993; Trotter Bowen and Potter 1993; Trotter and Potter 1992). However, we were convinced of the theoretical and practical importance of adding questions which would measure additional intimacy relationships in the group. These supplemental questions would help us measure the likelihood of diffusion of HIV prevention information through "intimate communication" channels. If people are not willing to talk about AIDS, sexual relationships, or other intimate subjects, then the diffusion of information from the project, or others like it, would potentially be blocked within some networks.

The result of the pretest exercise was the construction of a 27 item social network instrument designed to measure conditions relevant to HIV infection risks, and linked to our measures of psychosocial behavior change and our ethnographic measures of risk and risk reduction. The 27 network questions follow in Table I. The questions allow our respondents to differentiate their relationships with one another in four areas: measures of social relationships, judgements about trust, drug use relationships, and communication intimacy. After having the opportunity to administer the questionnaire, we found that some questions could potentially be interpreted as having more than one area of social meaning. Therefore, we conducted a classification exercise using expert assignment of the questions to four categories: social relationships, drug relationships, trust measures, and intimacy relationships. We asked our project staff (outreach, intervention, and interviewing) to make their best judgement of the classification of each question. Individuals were allowed to identify two potential classifications of each question, if they felt a question went in both categories equally strongly. The results produced the classification found in Table I. This classification is useful for understanding both the theoretical construction of the questionnaire, and the factor analysis results presented below. The four originally conceived categories (trust, social relations, drug relations, and intimacy) are the core of the final classification. However, the expert ratings also indicated that trust and intimacy are closely related to one another and that some questions combine these elements in varying degrees. Therefore, some of the questions should be considered as measuring both relationships simultaneously.

As we began to administer and analyze the data from the questionnaire, we encountered a pragmatic problem with the instrument which resulted in the development of this paper. When we used the instrument on larger drug networks, it became a potentially serious burden for our respondents. One of our drug networks contains more than 40 individuals. Thus the matrix each individual filled out had no fewer than 1080 (27×40) judgements which had to be made about network relationships. We feel this condition produced potential respondent irritation and fatigue, with the likelihood of a consequent loss of precision in the data. The size of the

TABLE I
Expert classification of network question

TRUST

T1 How willing would you be to lend or give money to _____?
T2 How much would you trust _____?
T3 How honest is _____ with you?
T4 How often does _____ tell important things to you?
T5 How often would you go to _____ for help?
T6 How comfortable would _____ be to share works with you?*
T7 How often does _____ tell their problems to you?**
T8 If you had AIDS, how willing would you be to tell _____?**

SOCIAL Relationship measures

S1 How much contact do you have with _____
S2 How close a friend is _____?
S3 How much do you hang around with _____
S4 How often do you talk to _____?
S5 How often would you do fun things with _____

DRUG Relationship measures

D1 How often do you use drugs with _____?
D2 How often do you go to _____ for drugs?
D3 How often are you willing to lend or give drugs to _____?
D4 How willing are you to trust _____ to score drugs for you?**
D5 How often do you party with _____?
D6 How often would you ask _____ for help in a drug emergency?***

INTIMACY measures

I1 How comfortable would you feel discussing condoms with _____?****
I2 How comfortable would you feel discussing AIDS with _____?****
I3 How comfortable would you feel discussing a one night stand with _____?*
I4 How comfortable would you feel discussing an affair with _____?****
I5 How comfortable would you feel discussing casual sex partners with _____?
I6 How comfortable would you feel discussing unwanted sex with _____?****
I7 How willing would you be to talk about sex with _____?****
I8 T5 How often do you tell intimate things to _____?****

* This question was originally constructed as a drug question, but the expert raters classified it as a trust question.

** These questions were classified as measuring trust, but at least 25% of the experts also judged the question measures intimacy.

*** These questions were classified as drug questions, but at least 25% of the raters also classified them as measuring trust.

**** These questions were classified as measuring intimacy, but at least 25% of the raters identified them as a trust measure as well.

instrument also presented analytical problems, since we had identified a minimum of 10 different types of exploratory analyses to run on each question (plus possible combinations of questions). This plan would produce at least 270 analyses for each network. The burden of interpreting this

level of data, even for a single network, is significant and for the estimated 10 networks we will survey in the next year is horrendous (270 × 10 = 2700 analyses).

While each of the 27 questions on the instrument may represent a different network relationship (Knoke and Kuklinski 1982), we felt there might be sets of questions which produced virtually identical or overlapping network results. We decided to reduce the total questions analyzed (and used in the future) by systematically identifying the ones which measure the same relationship. After considerable searching, we were unable to find a technique of this sort in the literature which dealt with large data sets like ours. Most researchers have been content to ask fewer questions, or to report a smaller number of questions. Talking the problem over with other network researchers produced a number of qualitative suggestions for eliminating questions, but no known systematic techniques.[3] Therefore we are proposing the following procedure.

III. METHODS

We are presenting the full reciprocal network data on 8 drug using networks for this paper, two from our pretest and 6 from the full 27 question network questionnaire. The networks range in size from 7 to 41 individuals, and are composed of active cocaine and heroin users in a small town (pop. est. 45,000). The data for each matrix was collected by bringing the entire network together in the same room, where they filled out the network questionnaires and received a network based HIV risk reduction session targeted at the group.[4]

The responses to the questions were entered as 27 individual person by person network matrices, using UCINET IV's data entry capabilities. For this analysis, the matrices were then combined into a single file containing all 27 distinct matrices for a single network, using the UCINET MERGE routine. This multiple matrix file was reshaped to form a single question by subject matrix file, using the UCINET RESHAPE routine. This process produced a file in which the "X" axis (columns) of the matrix was the nuraber of subjects squared, and the "Y" axis (rows) was the set of questions. The reason for the subjects squared condition is that the original question matrices were syrnmetrical (person by person) matrices. The reshaped network matrix was then transposed, using the UCINET TRANSPOSE routine, so the "X" axis represented the questions and the "Y" axis the subjects, in order to fit the data format expectations of SPSS for Windows.

The UCINET IV question-by-subject files were saved as ASCII files and exported into SPSS for Windows. In SPSS, the files were factor analyzed, using the FACTOR routine (Principle Components Analysis with

a VARIMAX rotation). This produced eight factor solutions: six from networks administered the full question sets, and two which were administered the pretest question set which includes only one of the intimacy questions. Factors with an eingenvalue of 1.0 or greater were considered significant. SPSS was also used to test the reliability of each factor solution, for each network, using Cronbach's Alpha.

The factor solutions for each network were compared using General Set Theory, to identify sets of co-occurring questions across the network factor solutions. In order to be constructed as a set, questions had to co-occur in a single factor in at least four of the eight network factor solutions. This analysis produced six question sets containing a total of 19 questions. There were 8 remaining questions which did not meet our set theory "rules of inclusion" for co-occurrence, and consequently were not assigned to a set.

The question sets and single questions were assessed for the strength of the relationship they demonstrated within factor solutions by comparing their correlations to factors. We identified the questions in each set that made the strongest overall contribution to various factors, and which could be judged as the best representative question to keep in the questionnaire without destroying either the theoretical underpinnings for the matrix, or the factors found in the empirical administration of the questions. This dual condition of meeting the a *priori* and empirical research conditions was cross checked by comparing Cronbach's Alpha for the original factor with the Alpha for the reduced set of questions.

DATA

Table II identifies the factor groupings and correlation scores for the questions administered to each network. These data represent the information used to determine the interrelated question sets in the questionnaire.

The data were analyzed according to the Set Theory rules established in the methods section. Table III illustrates the question sets which were identified by the set theory analysis, based on the potential associations established within the structure of the original classification.

The Set Theory analysis also identified the following questions, Table IV, which did not meet the co-occurrence rules and therefore could not be classified into sets. These questions appear to measure net in the questions sets.

The next analytical stage was to assess the individual contribution of each question. Table V displays the factor correlation values for each question, organized according to the question sets or individual questions identified by the Set Theory analysis. We selected the question within each set which displayed the highest overall values, or highest average values, and met

TABLE II
Factor analysis solutions for each network

n1	n4	n6*	n9	n11*	n15	n17**	n18
Factor 1	**Factor 1**	**Factor 1**	**Factor 1**	**Factor 1**	**Factor 1**	**Factor 1**	**Factor 1**
T3 (0.894)	T4 (0.852)	S1 (0.896)	I2 (0.993)	DS (0.938)	T4 (0.956)	T3 (0.897)	I6 (0.913)
I4 (0.870)	S5 (0.832)	S4 (0.896)	I5 (0.914)	T3 (0.909)	S3 (0.953)	S2 (0.870)	D2 (0.908)
T2 (0.802)	I8 (0.825)	S2 (0.885)	I1 (0.883)	D2 (0.903)	S1 (0.942)	S1 (0.869)	I2 (0.901)
D5 (0.773)	T2 (0.767)	T1 (0.878)	T7 (0.837)	D1 (0.892)	54 (0.939)	S3 (0.868)	T5 (0.890)
T5 (0.764)	I6 (0.744)	T2 (0.878)	D6 (0.827)	T6 (0.863)	T5 (0.938)	D5 (0.865)	T4 (0.890)
S1 (0.712)	T1 (0.737)	T3 (0.817)	S2 (0.753)	T2 (0.818)	S5 (0.918)	T2 (0.864)	S2 (0.888)
T1 (0.705)	S4 (0.684)	T5 (0.791)	D5 (0.740)	D3 (0.795)	I8 (0.913)	D4 (0.753)	D3 (0.866)
	T3 (0.663)	T8 (0.783)	I4 (0.682)	S4 (0.791)	T8 (0.905)	T4 (0.730)	D4 (0.866)
Factor 2	T5 (0.642)	S5 (0.774)	I3 (0.674)	T1 (0.789)	T1 (0.895)	T8 (0.727)	I1 (0.840)
	S1 (0.S95)	T7 (0.771)	SS (0.649)	53 (0.782)	T2 (0.853)	D1 (0.709)	I8 (0.892)
D4 (0.884)		T4 (0.771)	T1 (0.614)	S1 (0.778)	T3 (0.864)	T1 (0.670)	S5 (0.827)
D3 (0.772)	**Factor 2**	I8 (0.687)	T8 (0.614)	D4 (0.748)		T7 (0.601)	D1 (0.780)
S3 (0.705)		S3 (0.668)	T3 (0.501)	52 (0.746)	**Factor 2**	T5 (0.576)	I7 (0.786)
D1 (0.705)	D6 (0.792)			T5 (0.714)			D5 (0.774)
S4 (0.694)	D3 (0.782)	**Factor 2**	**Factor 2**		T7 (0.961)	**Factor 2**	I4 (0.772)
D2 (0.681)	D5 (0.746)			**Factor 2**	I7 (0.959)		S3 (0.769)
T8 (0.612)	D4 (0.700)	D1 (0.815)	T5 (0.774)		S2 (0.918)	I7 (0.874)	T3 (0.748)
	T6 (0.612)	D3 (0.777)	T2 (0.767)	T7 (0.857)	I6 (0.750)	I1 (0.856)	T1 (0.731)
Factor 3		D4 (0.764)	S3 (0.761)	T4 (0.815)	I4 (0.693)	I6 (0.837)	S1 (0.731)
	Factor 3	D2 (0.720)	T6 (0.748)	I8 (0.759)	I5 (0.644)	I5 (0.770)	S4 (0.720)
I2 (0.895)		T6 (0.628)	I8 (0.743)	S5 (0.732)		I2 (0.709)	I3 (0.707)
I1 (0.890)	I4 (0.835)	D5 (0.626)	S1 (0.727)	D6 (0.723)	**Factor 3**	S4 (0.668)	T7 (0.707)
S5 (0.795)	I5 (0.832)	D6 (0.599)	S4 (0.674)	T8 (0.701)		D3 (0.660)	T2 (0.707)
D6 (0.776)	I3 (0.829)		D4 (0.601)		I1 (0.946)	I4 (0.658)	
T6 (0.556)	S2 (0.706)				I2 (0.945)	D6 (0.641)	**Factor 2**
	I1 (0.665)		**Factor 3**		I3 (0.928)	S5 (0.557)	
Factor 4	I2 (0.607)						T6 (0.781)
			D2 (0.906)		**Factor 4**	**Factor 3**	D6 (0.760)
S2 (0.911)	**Factor 4**		D1 (0.787)				T8 (0.691)
I7 (0.879)			D3 (0.695)		D5 (0.796)	D2 (0.733)	I5 (0.661)
I6 (0.806)	T8 (0.794)		I7 (0.583)		D4 (0.774)	I3 (0.069)	
T7 (0.787)	S3 (0.665)				D6 (0.758)		
I5 (0.782)	I7 (0.465)		**Factor 4**			**Factor 4**	
					Factor 5		
Factor 5	**Factor**		I6 (0.766)			T6 (0.827)	
			T4 (0.707)		D1 (0.667)		
I8 (0.922)	D1 (0.826)				T6 (0.517)		
I3 (0.864)	T7 (0.826)						
T4 (0.692)	D2 (0.690)				**Factor 6**		
					D2 (0.924)		
					D3 (0.656)		

* This network data was collected during the pretest and does not include questions I1, I2, I3, I4, I5, I6, I7.

** question I8 had no variance in this network and was eliminated from the factor analysis.

TABLE III
Set theory associations of questions

n1	n4	n6	n9	n11	n15	n17*	n18
SET 1							
F1	**F1**	**F1**	**F2**	**F1**	**F1**	**F1**	**F1**
S1	S1	S1	S1	S1	S1	S1	S1
–	–	S3	S3	S3	S3	S3	S3
–	S4	S4	S4	S4	S4	–	S4
T5	T5	T5	T5	T5	T5	T5	T5
SET 2							
F5	**F1**	**F1**		**F2**	**F1**		**F1**
–	S5	S5		S5	S5		S5
T4	T4	T4		T4	T4	–	T4
I8	I8	I8		I8	I8	–	I8
SET 3							
F1	**F1**	**F1**	**F1**	**F1**	**F1**	**F1**	**F1**
T1	T1	T1	T1	T1	T1	T1	T1
T2	T2	T2	T2	T2	T2	T2	T2
T3	T3	T3	T3	T3	T3	T3	T3
SET 4							
F2	**F2**	**F2**		**F1**	**F4**	**F1**	**F1**
D1	–	D1		D1	–	D1	D1
D4	D4	D4		D4	D4	D4	D4
	D5	D5		D5	D5	D5	D5
SET 5							
F3	**F3**		**F1**		**F3**	**F3**	**F1**
I1	I1		I1		I1	I1	I1
I2	I2		I2		I2	I2	I2
	I3		I3		I3	–	I3
SET 6							
	F3		**F1**		**F2**	**F2**	**F1**
	I4		I4		I4	I4	I4
	I5		I5		I5	I5	I5

Set Rule: complete co-occurrence must be present in at least four factor solutions.
* I8 answers had no variance and was eliminated from the factor analysis.
** Pretest network instrument: I1, I2, I3, I4, I5, I6, and I7 are not included.

our programmatic needs. We also identified the individual questions from Table IV which contributed strongly and met programmatic needs. Each selected question is identified in bold letters in the table. They are the primary candidates which appear to best maintain the original theoretical

TABLE IV
Questions that do not cluster into sets

S2	How close a friend is ___
T6	How comfortable would _____ be to share works with you?
T7	How often does _____ tell their problems to you?
D2	How often do you go to _____ for drugs?
D3	How often are you willing to lend or give drugs to _____?
D6	How often would you ask _____ for help in a drug emergency?
I6	How comfortable would you feel discussing unwanted sex with _____?
I7	How willing would you be to talk about sex with _____?

TABLE V
Comparison of factor values for each question set and individual questions

	n1	n4	n6	n9	n11	n15	n17	n18
S1	0.712	0.595	0.896	0.727	0.778	0.945	0.864	0.731
S3	**0.705**	**0.607**	**0.668**	**0.760**	**0.782**	**0.958**	**0.867**	**0.769**
S4	0.694	0.684	0.896	0.687	0.791	0.922	0.688	0.719
T5	0.764	0.642	0.791	0.774	0.714	0.941	0.576	0.576
T1	0.705	0.737	0.878	0.614	0.788	0.858	0.669	0.730
T2	0.802	0.767	0.878	0.767	0.818	0.818	0.864	0.706
T3	**0.894**	**0.663**	**0.817**	**0.501**	**0.909**	**0.864**	**0.897**	**0.747**
T8	0.612	0.794	0.783	0.614	0.700	0.863	0.727	0.691
S5	0.795	0.832	0.774	0.649	0.732	0.905	0.557	0.827
T4	**0.692**	**0.852**	**0.771**	**0.707**	**0.814**	**0.970**	**0.730**	**0.889**
I8	0.922	0.825	0.687	0.743	0.759	0.916	*	0.829
D1	**0.705**	**0.826**	**0.815**	**0.787**	**0.891**	**0.613**	**0.709**	**0.800**
D4	0.884	0.700	0.720	0.601	0.748	0.852	0.753	0.865
D5	0.773	0.746	0.626	0.740	0.938	0.585	0.865	0.774
I1	0.890	0.665	–	0.883	–	0.676	0.856	0.804
I2	**0.895**	**0.607**	–	**0.933**	–	**0.697**	**0.709**	**0.901**
I3	0.864	0.828	–	0.674	–	0.677	0.069	0.707
I4	**0.870**	**0.835**	–	**0.682**	–	**0.846**	**0.658**	**0.773**
I5	0.782	0.831	–	0.914	–	0.822	0.770	0.661
S2	**0.911**	**0.707**	**0.885**	**0.753**	**0.746**	**0.749**	**0.870**	**0.887**
T6	**0.556**	**0.611**	**0.628**	**0.748**	**0.864**	**0.475**	**0.826**	**0.781**
T7	**0.787**	**0.825**	**0.771**	**0.837**	**0.856**	**0.835**	**0.601**	**0.707**
D2	**0.681**	**0.690**	**0.720**	**0.905**	**0.903**	**0.268**	**0.732**	**0.908**
D3	0.772	0.782	0.777	0.695	0.795	0.524	0.659	0.865
D6	0.776	0.792	0.599	0.827	0.723	0.102	0.641	0.760
I6	**0.806**	**0.744**	–	**0.765**	–	**0.844**	**0.837**	**0.912**
I7	0.879	0.465	–	0.583	–	0.797	0.873	0.786

I8 answers had no variance and were eliminated from the factor analysis.

structure of the questionnaire, while maintaining the best representation of the empirical data for each network.

This data allowed us to produce the following reduced question set which meets both our theoretical and empirical conditions for creating a sound reduced question set for the analysis of existing data and for future data collection.

As a final test of the applicability of this method, we calculated the Cronbach Alpha factor reliability scores for the questions in the original factor solutions, for each network. Then we calculated the Cronbach Alphas for only the questions that remained in each factor, using the reduced question set. Table VII presents the results of those analyses.

The rule we applied to the comparison of the factor reliability solutions for the original question set and the reduced question set were as follows. We felt that the reduced question set should be constructed in such a way that there would be at least one question remaining that represents each of the original factors, when comparing the original and the reduced question sets. Second, the reliability measure for each solution containing two or more questions should not be significantly lowered when using the reduced question set to compute the Cronbach's alpha reliability measure for that factor. The cutoff point we established was an alpha reliability of no lower than 0.700 in the reduced question set. The first rule was met in all but one factor of one network. Factor 4 of network 15 does

TABLE VI
Reduced questions for network questionnaire

TRUST

T3 How honest is _____ with you?
T4 How often does _____ tell important things to you?
T6 How comfortable would _____ be to share works with you?
T7 How often does _____ tell their problems to you?
T8 If you had AIDS, how willing would you be to tell _____?

SOCIAL Relationship measures

S2 How close a friend is _____?
S3 How much do you hang around with

DRUG Relationshio mearsures

D1 How often do you use drugs with _____?
D2 How often do you go to _____ for drugs?

Combined INTIMACY/TRUST measures

I2 How comfortable would you feel discussing AIDS with _____
I4 How comfortable would you feel discussing an affair with _____
I6 How comfortable would you feel discussing unwanted sex with _____

TABLE VII
Comparison of Cronbach's alpha for original factors and the reduced question set

All questions	N1	N4	N6	N9	N11	N15	N17	N18
Factor 1	0.907	0.953	0.977	0.973	0.986	0.979	0.981	0.934
Factor 2	0.937	0.930	0.896	0.952	0.925	0.934	0.960	0.896
Factor 3	0.927	0.901		0.901		0.966	0.548	
Factor 4	0.934	0.732		0.864		0.761	*	
Factor 5	0.853	0.946				0.418		
Factor 6						0.642		
Reduced questions	**N1**	**N4**	**N6**	**N9**	**N11**	**N15**	**N17**	**N18**
Factor 1	0.971	0.816	0.942	0.922	0.963	0.933	0.967	0.977
Factor 2	0.885	*	0.702	0.852	0.921	0.875	0.855	0.555
Factor 3	0.800	0.840		0.912		*	*	
Factor 4	0.855	0.647		0.864		**	*	
Factor 5	*	0.946				0.418		
Factor 6						*		

* Factor only contains one question. This makes reliability testing impossible.
** no remaining question exists for this factor.

not contain a question from the reduced set. The second rule was met in all cases except factor 4 of network 4, which had a alpha reliability of 0.647 for the reduced set, compared with a reliability of 0.732 for the original questions,[5] and Factor 5 of Network 15. In the latter case, the original factor was a weak factor to begin with, below our 0.700 cutoff (0.418). The reduced question set factor was identical to the original question set factor. Since they were identical, this was not taken to be a violation of the rule. The results of this comparison provide strong support for the overall question selection procedure and minimize the number of qualitative or unsystematic judgements that would otherwise need to be made in reducing the number of questions in a network questionnaire.

V. SUMMARY AND CONCLUSIONS

This paper identifies a systematic and relatively simple method for reducing the number of questions which need to be administered and analyzed in a complex network data analysis process. The procedure allows for the preservation of the original theoretical construction of the network questionnaire and also preserves most of the original complexity and underlying structure of the empirical data. This process allows us to significantly reduce the burden placed on informants who participate in large network data collection process. It simultaneously allows for a more thorough exami-

nation and presentation of the complexity of network data, both for exploratory analysis and for final network comarison, by reducing the number of overall computations that have to be completed for each network and for between network comparisons.

NOTES

[1] This project is supported by the Community Research Branch of NIDA (Grant # U01-DA07295).

[2] This question has one meaning for our respondents, "who do you use drugs with in a group situation," and another meaning for some of our staff, who thought it meant how often do you go to parties with someone (regardless of drug use). We kept this question in the drug classification, rather than the social classification, since the intended use of the question is a proxy for "how often do you use drug with _____?, which is the meaning assumed by the majority of our clients. However, it may also have lead to an ambiguous interpretation in that group as well.

[3] We would like to thank both Steven Borgatti and Richard Rothenberg for their advice and creative suggestions in this discussing the existing literature and potential qualitative solutions to this problem.

[4] Our questionnaire does not directly ask about sexual relationships in the group, to avoid potential ethical and logistical problems. We wanted to avoid possible individual or group conflict that could have arisen if someone saw someone else's answers on questions, about a sexual relationship that they were not aware of, or which might anger them. The questions we developed as measures for intimacy and communication about sexual matters are the best proxy measures we could find for the approximation of potential risky relationships in this area, without creating this type of problem.

[5] These results assume that the factors in which there is only one question remaining demonstrated sufficient reliability in the original factor to be a reliable solution, since it is not possible to compare a question to itself for this procedure.

REFERENCES

Bowen, A. M. and R. T. Trotter, II 1995 HIV Risk in Drug Users and Crack Smokers: Predicting Stage of Change for Consom Use. Journal of Consulting and Clinical Psychology (with A. M. Bowen).

Klovdahl, A. (1985) Social Networks and the Spread of Infectious Diseases: The AIDS Example. Soc. Sci. and Med. 2(11).

Klovdahl, A., J. J. Potterat, D. E. Woodhouse, J. B. Muth, S. Q. Muth and W. W. Darrow 1994 Social Networks and Infectious Disease: The Colorado Springs Study. Soc. Sci. and Med. 38(1): 79–88.

Knode, D. and J. H. Kuklinske 1982 Network Analysis. Newbury Park, CA: Sage Publications.

Trotter, R. T. II, A. M. Bowen and J. M. Potter 1995 Network Models for HIV Outreachand Prevention Programs for Drug Users. NIDA Monograph Series. Washington, DC: National Institute on Drug Abuse (with A. M. Bowen and J. M. Potter). NIDA Monograph Series no. 151, pp. 144–180.

Trotter, R. T. II, A. M. Bowen and J. M. Potter 1993 Network Models for HIV Outreach and Prevention Programs for Drug Users. NIDA Technical Review on Network Analysis. Washington, D.C.: National Institute on Drug Abuse, Community Research Branch.

Trotter, R. T. II, J. M. Potter, A. M. Bowen, D. Jiron, K. Crawley, E. Moore and L. J. Price 1993 Ethnographic Network Analysis: Tools for HIV and Drug Intervention. NIDA Monograph Series. Washington, D.C.: National Institute on Drug Abuse.

Trotter, R. T. and J. M. Potter 1992 Network Approaches to Outreach Prevention for HIV Risk Reduction for Drug Users. Paper presented at the American Anthropological Association Annual Meeting, San Francisco. December.