Triadic Configurations in Limited Choice Sociometric Networks:

## Empirical and Theoretical Results

Katherine Faust<br>Department of Sociology and Institute for Mathematical Behavioral Sciences University of California, Irvine

Corresponding author: Katherine Faust, Department of Sociology and Institute for Mathematical Behavioral Sciences, University of California, Irvine, Irvine, CA 926975100 USA; email: kfaust@uci.edu; telephone: 949 824-9383; fax: 949 824-4717.

Acknowledgements: I am grateful to Carter Butts and A. Kimball Romney for conversations about this work, to three anonymous reviewers for comments on an earlier version of this paper, and to Jim Moody for suggesting triad analyses of the Adolescent Health networks.

Keywords: triads, triad census, network density, sociometric data, meta analysis, social network

## Triadic Configurations in Limited Choice Sociometric Networks:

## Empirical and Theoretical Results


#### Abstract

Previous research demonstrated that information contained in triad censuses from heterogeneous collections of social networks occupies a high dimensional space. Regions of this space, and locations of triad censuses within it, are largely defined by lower order network properties: network density and dyad distributions (Faust 2006, 2007). The current paper extends comparative work on triad censuses by addressing three related issues. First, it determines and interprets the space occupied by triad censuses for 128 friendship networks gathered using a limited choice sociometric protocol. Second, it constructs a theoretical space for triad censuses expected given lower order graph properties and examines the dimensionality and shape of this space. Third, it brings together these lines of investigation to determine where the empirical triad censuses reside within the theoretical space. Results show that the empirical triad censuses are almost perfectly represented in one dimension (explaining 99\% of the data) and that network density explains over $96 \%$ of the variance in locations on this dimension. In contrast, the theoretical space for triad censuses is at least fourdimensional, with distinctive regions defined by network density and dyad distributions. Within this theoretical space, the empirical triad censuses occupy a restricted region that closely tracks triad censuses expected given network density. Results differ markedly from prior findings that the space occupied by triad censuses from heterogeneous social networks is of high dimensionality. Results also reinforce observations about constraints that network size and density place on graph level indices.


## 1. Background

Triads, configurations of triples of actors and the ties between them, are fundamental to many social network methods and theories. Triads are important in social structural investigation since they link local network patterns to theoretically important global structures (Davis 1967, 1970; Davis and Leinhardt 1972; Holland and Leinhardt 1970, 1976, 1979; Johnsen 1985, 1986, 1989a, 1989b, 1998). Triadic patterns and processes provide the basis for many sociological insights, such as forbidden triads in the strength of weak ties argument (Granovetter 1973), structural holes (Burt 1992), network closure (Coleman 1988), brokerage (Fernandez and Gould 1994), tertius strategies (Simmel 1950; Burt, 1992; Obstfeldt 2005), coalition formation (Caplow 1959), and trust (Burt and Kenz 1995). Triads have been widely used to study social network structure, and triadic configurations often are included in statistical models of social networks (Snijders, Pattison, Robins, and Handcock 2006).

The triad census, introduced by Holland and Leinhardt more than three decades ago (Holland and Leinhardt 1970), is a standard means for studying triadic configurations in social networks, yet its formal properties remain understudied.

Recent research on triad censuses has demonstrated that, in aggregate, information contained in triad censuses from heterogeneous collections of social networks is of high dimensionality, requiring at least four dimensions for adequate representation (Faust 2006, 2007). That line of investigation studied triad censuses in various types of social relations (dominance, outcomes of agonistic encounters, expressions of positive and negative affect, co-observation, and choices of work partners, for example) and animal species (humans, baboons, chimpanzees, macaques, cows, red
deer, several bird species, dolphins, and more) using two independent samples of social networks. Evidence for high dimensionality comes from correspondence analysis or singular value decomposition of sets of triad censuses, which require three or four dimensions to account for a substantial percent of the data (Faust 2006, 2007). Within these high dimensional spaces, empirical triad censuses are located in distinct regions that are largely described by network density and the distributions of mutual, asymmetric, and null dyads in the network. Substantively different kinds of social relations reside in different regions of the high dimensional space, for example, contrasting networks of victories in agonistic encounters (in which dyads are primarily asymmetric) with networks of co-observation (in which all dyads are either mutual or null).

The current paper turns from examining contrasts among triad censuses from heterogeneous social relations to studying triadic patterns in networks measured using the same sociometric question in similar social settings. The following analyses use a collection of 128 social networks all measured with the same limited choice sociometric question about friendships between students in American high schools. The general goals of this paper are to determine the space spanned by triad censuses from a single kind of social relation, to construct a theoretical space of probable triad censuses from networks with known lower order network properties, and to locate the empirical triad censuses within this theoretical space. Where do triad censuses for friendships among American high school students reside within this theoretical space of triad census possibilities?

These objectives are subtly, yet fundamentally, different from the objectives of investigations that seek to identify statistically important triadic tendencies in social networks. Examples of such contrasting lines of research include determining whether there are significant triadic effects in a statistical models social networks or quantifying the extent to which observed triad frequencies depart from expectation under a particular conditional distribution. In contrast, the current paper aims to directly study triad censuses for specific social relation and to compare these triad censuses to a theoretical space of possibilities.

## 2. Density, dyads, and triads

Network density, the dyad census, and the triad census are especially important network properties used in the following analyses. For a dichotomous directional relation on $g$ actors, where $x_{i j}$ records the tie from actor $i$ to actor $j$, network density is defined as:

$$
\begin{equation*}
\Delta=\frac{\sum_{i=1}^{g} \sum_{\substack{j=1 \\ j \neq i}}^{g} x_{i j}}{g(g-1)} . \tag{1}
\end{equation*}
$$

If outdegrees are fixed so $x_{i+}=d$ for all actors, $i$, then density can be expressed as:

$$
\begin{equation*}
\Delta=\frac{d}{g-1} \tag{2}
\end{equation*}
$$

This clearly shows the formal relationships between degree, network size, and network density: for fixed degrees, as network size increases, network density decreases.

Dyadic and triadic features of networks are succinctly summarized in censuses of subgraphs of two or three nodes. A dyad consists of a pair of nodes and the state of the arcs between them. For a dichotomous directional relation there are $\binom{g}{2}=\frac{g(g-1)}{2}$ dyads,
each of which must be in one of three isomorphism classes: mutual ( $M$ ), asymmetric ignoring arc direction $(A)$, or null $(N)$. For a given network, its dyad census is a count of the number of dyads in each isomorphism class. These counts are often referred to as $M A N$.

Figure 1 here
A triad is a subgraph of three nodes and the arcs between them. For a directional dichotomous relation there are $\binom{g}{3}=\frac{g(g-1)(g-2)}{6}$ triads, each of which is isomorphic with one of sixteen isomorphism classes. These isomorphism classes are shown in Figure 1 with standard labeling giving the number of mutual, asymmetric, and null dyads along with a letter (T, C, D, or U) to indicate directionality, when there is more than one triad isomorphism class with the same MAN count (Holland and Leinhardt 1970). For a given network, its triad census records the number of triads in each of the16 isomorphism classes, and is summarized in a 16 element vector $\mathbf{t}=\left(c_{1}, c_{2} \ldots c_{16}\right)$, where $c_{k}$ denotes the number of triads in isomorphism class $k$.

A number of graph properties can be derived from the triad census counts for a given network (Holland and Leinhardt 1976; Wasserman and Faust 1994). These properties include network size $(g)$, the number of $\operatorname{arcs}\left(x_{++}\right)$, network density $\left(\frac{x_{++}}{g(g-1)}\right)$, and the dyad census MAN counts, among others. Since network density and the dyad census can be derived from the triad census, but not the reverse, these properties are said to be lower order than the triad census.

There is a long tradition of using the triad census to investigate structural properties of social networks, especially networks of positive interpersonal sentiments
(Brewer and Webster 1999; Davis 1970; Davis and Leinhardt 1972; Hallinan 1974a, 1974b; Holland, and Leinhardt 1972; Leinhardt 1972). Much of this research has been concerned with characterizing common triadic tendencies in these social relations and linking triadic patterns to theoretically important network structures, such as structural balance, clusterability, ranked clusters, and transitivity. It is therefore fitting to continue this line of investigation, focusing on triad censuses for a collection of friendship networks.

## 3. Sociometric data

Empirical data are from The National Longitudinal Study of Adolescent Health in-school questionnaires from Wave I, conducted in 1994-5 (Harris et al. 2003). Given the widespread use of these data it is worth describing them in some detail. The study sampled high schools so that
"... high schools selected are representative of US schools with respect to region of country, urbanicity, size, type, and ethnicity. Eligible high schools included an 11th grade and enrolled more than 30 students. ... Participating high schools helped to identify feeder schools-that is, schools that included a 7th grade and sent at least five graduates to that high school. From among the feeder schools, one was selected with probability proportional to the number of students it contributed to the high school" (Harris et al. 2003).

Sociometric data were collected using an in-class questionnaire. Each student who was present on the day of the survey was given a roster of students in the high school and its feeder school and asked to list, using assigned code numbers, their closest male and female friends. The exact questions were:

- "List your closest male friends. List your best male friend first, then your next best friend, and so on. Girls may include boys who are friends and boyfriends."
- "List your closest female friends. List your best female friend first, then your next best friend, and so on. Boys may include girls who are friends and girlfriends" (Harris et al. 2003).

Space was provided to list, by code number, up to five male friends and five female friends. Students also were allowed to list friends who were not on the roster, did not attend the schools, or were not enrolled in school, though these responses are not included in the in-school sociometric data. A reproduction of the social network section of the questionnaire, showing the exact response format, is in Bearman, Moody, Stovel, and Thaljin (2004, page 205).

Given the response format of the questionnaire, each student is limited to nominate no more than five male and no more than five female friends. Since this imposes an upper limit to the number of choices, and students were not required to name exactly five male and exactly five female friends, the format used in the Adolescent Health questionnaire is a limited choice sociometric protocol.

The following analyses use data from 84 linked junior and senior high schools in which response rates were at least $50 \%$ (Moody 2005). Of the 84 senior high schools, 44 were linked to a separate junior high or middle school, giving a sample of 128 schools. The sample includes $\mathrm{N}=75,810$ individuals. As can be seen in Table 1, schools range in size from 25 to 2,250 students and network density ranges from 0.0012 to 0.3467 . For fixed nodal degree, network density necessarily falls with network size (equations 1 and 2) resulting, for this sample of schools, in a squared Pearson correlation of $r^{2}=.907$
between network density and the reciprocal of school size. Schools also differ in their proportions of mutual, asymmetric, and null dyads.

Table 1 here
Sociometric data from the Adolescent Health project have been used to study a variety of topics, including: adolescent delinquency and violence (Haynie 2001, 2002; Haynie and Payne 2006; Payne and Cornwell 2007), racial and ethnic segregation (Joyner and Kao 2000; Kao and Joyner 2006; Moody 2001b; Mouw and Entwisle 2006; Quillian and Campbell 2003), effects of residential segregation on friendship (Mouw and Entwisle 2006), residential mobility and friendship formation (Haynie, South and Bose 2006; South and Haynie 2004), suicide intentions (Bearman and Moody 2004), obesity and friendship (Strauss and Pollack 2003), grade based subgrouping (Handcock, Raftery and Tantrum 2007), friendship reciprocity (Vaquera and Kao 2008), parental effects on friendship (Knoester, Haynie, and Stephens 2006), and to illustrate new social network methods (Goodreau 2007; Moody 2001a; Handcock, Raftery and Tantrum 2007).

## 4. Analysis and results

Analysis proceeds in three stages. First, triad censuses for the 128 friendship networks are found and singular value decomposition is used to produce a low rank approximation for the collection of censuses. This result is then interpreted using network density. Second, a theoretical space for triad censuses is constructed using singular value decomposition of censuses that are expected given network density and the proportions of mutual, asymmetric, and null dyads from the dyad census. Finally, triad censuses for the 128 empirical networks are projected into this theoretical space and the result is interpreted.

### 4.1 Triad censuses from limited choice sociometric networks

Triad censuses for the 128 empirical networks were found (using PAJEK 1.02, Batagelj and Mrvar 2004) and aggregated into a single matrix. This $128 \times 16$ matrix has networks indexing rows and triad isomorphism classes indexing columns. After transforming to row proportions, the matrix, $\mathbf{T}$, is analyzed using singular value decomposition (SVD) to determine the aggregate space spanned by triad censuses for the 128 networks.

Singular value decomposition of a $p$-by- $q$ matrix, $\mathbf{T}$, is defined as the matrix decomposition, $\underset{p \times q}{\mathbf{T}}=\underset{p \times q}{\mathbf{U}} \mathbf{q \times q} \mathbf{V}_{q \times q}$, where $\mathbf{U}$ is a $p-$ by $-q$ matrix of left singular vectors, $\mathbf{V}$ is a $q$-by- $q$ matrix of right singular vectors, and $\mathbf{D}$ is a $q$-by- $q$ diagonal matrix of singular values, in non-increasing order, $\left\{\lambda_{l}\right\}$ (Ben-Israel and Greville 1974; Digby and Kempton 1987). SVD of $\mathbf{T}$ for the 128 Adolescent Health triad censuses gives left singular vectors in the rows of $\underset{128 \times 16}{\mathbf{U}}$, pertaining to the 128 networks in the rows of $\mathbf{T}$, and right singular vectors in the rows of $\underset{16 \times 16}{\mathbf{V}}$, pertaining to the 16 triad isomorphism classes in the columns of $\mathbf{T}$. When displayed, $\mathbf{U}$ and $\mathbf{V}$ are rescaled so that $\mathbf{U}^{\prime} \mathbf{U}=\mathbf{D}^{2}$ and $\mathbf{V}^{\prime} \mathbf{V}=\mathbf{D}^{2}$, where $\mathbf{D}^{2}$ is a diagonal matrix of squared singular values.

A matrix of rank $W$ requires $W$ sets of singular values and singular vector pairs to reproduce the original data. The $128 \times 16$ matrix $\mathbf{T}$ could have rank equal to 15 , due to the constraint that the 16 triad census counts must sum to $\binom{g}{3}$. The quality of fit of a lower dimensional solution, $w \leq W$, is given by $100 \times \frac{\sum_{l=1}^{w} \lambda_{l}^{2}}{\sum_{l=1}^{W} \lambda_{l}^{2}}$, and is equal to the percent
of the sum-of-squares of $\mathbf{T}$ that is explained by $w$ sets of singular vectors and singular values (Ben-Israel and Greville 1974).

Singular values from SVD of $\mathbf{T}$ for the 128 empirical networks are presented in Table 2 and the first left and right singular vectors are displayed in Figures 2 and 3, respectively. These results show that one singular value and set of singular vector pairs accounts for $99.42 \%$ of the total sum-of-squares of $\mathbf{T}$. It is worth reflecting on this result. Although satisfactory representation of $\mathbf{T}$ could have required a 15 -dimensional solution, the fact that one singular value and pair of singular vectors accounts for over $99 \%$ of the data clearly demonstrates that these triad censuses are adequately summarized using a one-dimensional approximation of $\mathbf{T}$. This result contrasts strikingly with prior findings that required at least four dimensions to account for triad censuses from heterogeneous collections of social networks.

Figure 2 here
Figure 3 here
Table 2 here
In Figure 3, displaying the first right singular vector, $\mathbf{v}$, for triad isomorphism classes, it is apparent that the main contrast is between the 003 (all null) triad and the others. The large distinction between 003 and the other 15 triads on this dimension indicates that its percentage distribution across the 128 networks is markedly different from the other triads. Notably, aggregating across networks, $98.42 \%$ of all triads are type 003, whereas the range for the other triads is from $0.000007 \%$ for type 030 C to $1.19 \%$ for type 012. This suggests an interpretation of this dimension related to network size and density, which prior research has found to be important network features constraining
graph-level indices, including the triad census (Anderson et al. 1999; Faust 2006, 2007; Friedkin 1981). Focusing on the first left singular vector, $\mathbf{u}$, this vector has a squared Pearson correlation of $r^{2}=0.228$ with network size, and $r^{2}=0.962$ with network density. (For the first right singular vector, $\mathbf{v}, \mathrm{r}^{2}=0.305$ with the number of arcs in the triad, though the relationship clearly is not linear, since variability on this vector primarily separates the 003 triad from the others.) I return to the effect of network density on the triad census and dimensionality of the singular value decomposition in section 5, below.

These results demonstrate that triad censuses for the 128 Adolescent Health friendship networks are well fit in a single dimension that is essentially identical to network density. The contrast between the low dimensionality of this result and the high dimensionality required to represent heterogeneous collections of social networks (Faust 2006 , 2007) suggests that triad censuses from limited choice sociometric data are distinctively constrained in comparison to triad censuses from a broader range of social relations. To pursue this further, the following analyses construct a theoretical space spanned by possible triad censuses and then locate the 128 empirical triad censuses within this space.

### 4.2 A theoretical space for triad censuses

Consider the full range of likely triad census outcomes - that is, the probable distributions of observations across the sixteen triad isomorphism classes. This constitutes a "theoretical space" or universe for the triad census. Given prior findings on the triad census, it seems reasonable to construct a theoretical space that takes into account network density and the dyad census. To determine this theoretical space of triad
censuses, two random graph approaches to triad census probabilities are used. Each of these gives an expected distribution across the sixteen triad isomorphism classes, taking into account specific lower order graph properties.

First, the Bernoulli random graph model gives triad probabilities as functions of network density. This model assumes a random graph with arc probabilities equal to $\Delta$. Then, the probability that three nodes form a particular triadic configuration is a function of the number of arcs in the triad and the number of ways that the arcs can be arranged to give the particular configuration (Skvoretz, Fararo, and Agneessens 2004). Equations for these calculations are presented in Table 3. Triad probabilities from the Bernoulli random graph model were generated by varying $\Delta$ from 0.0 to 1.0 in steps of .01 , yielding triad censuses probabilities for 101 values of $\Delta$.

Second, the uniform graph distribution conditional on the dyad census, $U \mid M A N$, is used to calculate a second set of triad probabilities (Holland and Leinhardt 1970, 1976). This approach has as its sample space all graphs with a given distribution of mutual, asymmetric, and null dyads. The probability of a particular triadic configuration is then a function of the dyad proportions on which the distribution is conditioned, the number of mutual, asymmetric, and null dyads in the triad, and the number of ways that the dyads can be arranged to give the particular triadic configuration. Equations for these calculations are in Table 3. To find triad probabilities using the $U \mid M A N$ distribution, the proportions of mutual, asymmetric, and null dyads were varied from 0.0 to 1.0 , in steps of .05. All sets of dyad proportions summing to 1.0 were used to calculate triad census probabilities, generating probabilities for 231 triad censuses.

Table 3 here

Together, the Bernoulli and $U \mid M A N$ triad census probabilities provide the skeleton for a theoretical space of triad census expectations. To determine the dimensionality and shape of this theoretical space, the triad probabilities from the two approaches were combined into a matrix, $\widetilde{\mathbf{T}}$, with $231+101=332$ rows and 16 columns, and then analyzed using singular value decomposition. ${ }^{1}$

SVD of $\widetilde{\mathbf{T}}$, defined as $\widetilde{\mathbf{T}}=\widetilde{\mathbf{U}} \widetilde{\mathbf{V}} \widetilde{\mathbf{V}}^{\prime}$, yields left and right singular vectors $\widetilde{\mathbf{U}}$ and $\widetilde{\mathbf{V}}$ defining the space of triad censuses expected given network density and dyad census proportions. Singular values for the first 13 dimensions of this space are presented in Table 2. The first three singular value/ singular vector pairs account for $75.55 \%$ of the sum-of-squares of $\widetilde{\mathbf{T}}$, establishing that the theoretical space of triad censuses is of relatively high dimensionality.

Figure 4 here
Figure 5 here
The first three left singular vectors, $\widetilde{\mathbf{U}}$, of the theoretical space are displayed in Figure 4, and the first three right singular vectors, $\widetilde{\mathbf{V}}$, are in Figure 5. In two dimensions, the theoretical space is roughly the shape of an elongated triangle, as seen in both Figures 4 and 5. Focusing first on the space for networks (Figure 4) the corners of the triangle are anchored by triad censuses from networks with extreme dyad census distributions. The upper-right corner of the triangle is occupied by triad censuses from extremely dense networks and consequently many 300 triads. The lower-right corner of the triangle is

1 Other distributions that might be used include the uniform distribution conditional on the indegrees or outdegrees (Wasserman 1977), which provides more conditioning than the Bernoulli distribution, and the uniform distribution conditional on the indegrees, outdegrees, and number of mutual dyads (Snijders 1991), which provides more conditioning than $U \mid M A N$ but cannot be calculated directly.
occupied by triad censuses from networks that have extremely low density, and thus many 003 triads. The parabolic "spine", curving from upper to lower right, shows triad censuses from the Bernoulli random graph model and traces decreasing network density, thus moving from 300 to 003 triads. In the far left-center of the triangle are triad censuses from networks that primarily have asymmetric dyads. In these triad censuses two isomorphism classes predominate: 030 T and 030 C . The proportion of asymmetric dyads tracks, from right to left, parabolic curves opening to the right, with higher proportions of asymmetric dyads in parabolas to the left of the figure. The edge of the triangle running from middle left to upper right contains triad censuses from networks with high proportions of mutual dyads. Similarly, the edge running from middle left to lower right has triad censuses from networks with high proportions of null dyads.

The third dimension of the space for networks is similar to the first dimension and is related to asymmetry in the network, as can be seen in Figure 4. When interpreting this dimension it is worth noting that network density is related to asymmetry since highest asymmetry is possible in networks with $\Delta=0.5$.

The first three right singular vectors, $\widetilde{\mathbf{V}}$, for triad isomorphism classes are displayed in Figure 5. This configuration corresponds to the triangular space for networks (Figure 4). The first and third dimensions are related to asymmetry in the triads and the second dimension contrasts the high density from low density triads.

This theoretical space covers the full range of expectations for the triad census conditional on dyad census proportions, and also includes triad census expectations based on network density, ranging from 0 to 1 . The space can be thought of as a skeleton for triad expectations around which triad censuses from empirical networks can be located.

Since regions of the theoretical space are described by distinctive triadic configurations they can be useful in characterizing triadic patterns in particular empirical networks.

### 4.3 Projection of empirical triad censuses into the theoretical space

To determine where triad censuses for the 128 empirical networks are located in the theoretical space, they are projected into the theoretical space, $\widetilde{\mathbf{U}}$, as supplementary points using the equation: $\hat{\mathbf{U}}=\mathbf{T} \widetilde{\mathbf{V}} \widetilde{\mathbf{D}}^{-1}$ (Lebart, Morineau, and Warwick 1984). $\hat{\mathbf{U}}$ gives the locations for the triad censuses from the 128 empirical networks in the $w$-dimensional theoretical space defined by $\widetilde{\mathbf{V}}$.

## Figure 6 here

Figure 6 presents the first three right singular vectors, $\widetilde{\mathbf{U}}$, of the theoretical space and the projected points, $\hat{\mathbf{U}}$, for the 128 empirical triad censuses. This figure shows that triad censuses from the 128 empirical sociometric networks occupy a very limited region in the theoretical space, closely tracking the low end of the parabolic density spine defined by triad censuses from the Bernoulli random graph model. Triad censuses from the lowest density networks are at the tip of the parabolic spine, in the extreme lower right corner of the triangular space for the first two dimensions. The densest networks show a slight departure from the spine in the direction of symmetry rather than asymmetry. This is consistent with a tendency for friendship choices to be mutual.

Fit of the 128 empirical triad censuses in the theoretical space is assessed using canonical redundancy (Lambert, Wildt, and Durand 1988; Stewart and Love 1968). For two sets of variables, $\mathbf{X}$ and $\mathbf{Y}$, canonical redundancy, $\mathbf{R}_{\mathbf{Y} \cdot \mathbf{x}}^{2}$, is the proportion of variance in linear combinations of $\mathbf{Y}$ explained by linear combinations of $\mathbf{X}$. In the current application it is the proportion of variance in linear combinations of the empirical triad
censuses, T, explained by linear combinations of their projections in the theoretical space, $\hat{\mathbf{U}} . \quad \mathbf{R}_{\text {T• } \hat{\mathbf{U}}}^{2}$ is calculated as a function of matrices of correlations between columns of $\hat{\mathbf{U}}$ and columns of $\mathbf{T}, \mathbf{R}_{\hat{\mathbf{U}} \mathbf{T}}$ and $\mathbf{R}_{\mathrm{T} \hat{\mathbf{U}}}$, and correlations between columns of $\hat{\mathbf{U}}, \mathbf{R}_{\hat{\mathbf{U}} \hat{\mathbf{U}}}$, as:

$$
\mathbf{R}_{\mathrm{T} \cdot \hat{\mathbf{U}}}^{2}=\frac{1}{k} \operatorname{trace}\left(\mathbf{R}_{\hat{\mathbf{U}} \hat{\mathbf{U}}}^{-1} \mathbf{R}_{\hat{\mathbf{U}} \mathbf{T}} \mathbf{R}_{\mathrm{T} \hat{\mathbf{U}}}\right)
$$

where $k$ is the number of variables in set $\mathbf{T}$ and $\mathbf{R}_{\hat{\mathbf{U}} \hat{\mathbf{U}}}^{-1}$ is the inverse of $\mathbf{R}_{\hat{\mathbf{U}} \hat{\mathbf{U}}}$.
At least three and as many as five dimensions are required to fit $\mathbf{T}$ in the theoretical space (accounting for $95.7 \%$ and $99.5 \%$ of the variance, respectively), as seen in Table 4. This contrasts with the low dimensionality of $\mathbf{T}$ determined by the SVD presented above (Table 2), where a single dimension adequately reproduced the 128 empirical triad censuses.

Clearly, triad censuses from the from the Adolescent Health friendship networks occupy a restricted range in the theoretical space of outcomes that could have been observed empirically. The basis for this result and its implications are discussed in more detail in the following section.

Table 4 here

## 5. Network density and dimensionality of the Adolescent Health triad censuses

Both singular value decomposition of the 128 triad censuses from the Adolescent Health networks and projection of these censuses into the theoretical space give low dimensional solutions that are closely related to network density. To more fully understand these results, it is worth considering the relationships among network density, triad census distributions, and dimensionality of the space of triad censuses in greater
detail. This section reviews the general results presented above and then elaborates how network density affects the triad census, and, in turn, how this leads to low dimensional singular value decomposition solutions. ${ }^{2}$

First, a theoretical space of triad census probabilities is of high dimensionality: at least three- and possibly as high at 12-dimensional (see Table 2). The first two dimensions of this theoretical space form a triangle with a parabolic density spine running from empty to complete networks. Corners of the space are occupied by networks with extreme dyad distributions. The third dimension of the space contrasts triad censuses from networks with high symmetry from those with high asymmetry. The fact that density and dyadic features describe this space is not unexpected given that it was constructed from triad census probabilities calculated using the Bernoulli and $U \mid M A N$ random graph models. Nevertheless, it is important to recognize that previous research has found that triad censuses from a variety of empirical social networks closely resemble those expected from the $U \mid M A N$ distribution (Faust 2007; Holland and Leinhardt 1979), and density alone is not sufficient to explain those triad distributions (Faust 2006, 2007). In addition, a space of high dimensionality is required to represent collections of triad censuses from heterogeneous social relations (Faust 2006, 2007). Therefore, the high dimensional theoretical space presented in this paper is likely to be a reasonable characterization of triad censuses from many empirical social networks.

In contrast to this theoretical result, triad censuses from the 128 limited choice sociometric networks of the Adolescent Health study occupy a one-dimensional space

2 I am grateful to a reviewer of an earlier version of this paper for calling my attention to the formal relationships between network density and the singular value decomposition result.
that is almost completely described by network density. This single dimension accounts for more than $99 \%$ of the sum-of-squares of the data and network density explains $96 \%$ of its variance. Furthermore, when triad censuses from the Adolescent Health friendship networks are projected into the theoretical space, they occupy a very limited region of the space, closely tracking censuses that would be expected given low network density. This low dimensionality demonstrates that triad censuses from the limited choice sociometric networks are considerably more constrained than the high dimensional potential that might be realized, both in a theoretical space of triad census probabilities and in comparison to triad censuses from a variety of empirical social relations studied in previous research (Faust 2007).

A further observation provides the formal basis for more thorough interpretation of these findings about dimensionality. Specifically, the one-dimensional representation of triad censuses from the Adolescent Health friendship networks and their locations within the theoretical space both occur because network density constrains possible triadic outcomes. It is worth examining in some detail how network density is related to the triad census and, in turn, how this affects the singular value decomposition.

Recall that network density is exactly related to mean in/outdegree, $\bar{d}$, and

$$
\sum_{i=1}^{g} x_{i+}
$$

network size, $\Delta=\frac{\frac{g}{g-1}}{g-1}=\frac{\bar{d}}{g-1}$, (also see equations 1 and 2). For fixed degrees, as network size increases, density necessarily falls as the reciprocal of network size. Fixed and limited choice sociometric data collection protocols restrict individual outdegrees. Thus, networks with nearly identical mean degree but of different sizes must have
densities that are largely a function of network size. With regard to the empirical example in this paper, the limited choice sociometric design of the Adolescent Health study allowed students to name up to ten friends, though the mean number of choices is $4.2(n=75,810)$. Moreover, even though there is considerable variability in school size, the mean number of choices does not vary much across the 128 schools, as can be seen in Table 1. As a consequence, network density is quite low in most schools $\left(25^{\text {th }}\right.$ percentile $=0.0053,75^{\text {th }}$ percentile $\left.=0.0155\right)$ and is well predicted from school size $\left(r^{2}=.907\right.$, as reported above).

Triad probabilities are affected by network density (see equations in Table 3), as are possible empirical outcomes. Some rough calculations using the mean density of $\Delta=0.0193$ and the Bernoulli random graph model illustrate this point. At a density of $\Delta=0.0193$, the probability of an all null 003 triad is $\operatorname{Pr}(003)=(1-0.0193)^{6}=0.8896$ and the probability of an all mutual 300 triad is $\operatorname{Pr}(300)=(0.0193)^{6}=0.00000000005$. Although triad proportions in empirical networks with density $\Delta=0.0193$ might vary from these probabilities, triadic outcomes are severely constrained - this density is simply too low to construct large numbers of the kinds of triads that contain many arcs, such as the 300 triad. Aggregating across triads from all 128 schools in the Adolescent Health data, the relative frequency of the 003 triad is 0.9842 and the relative frequency of the 300 triad is 0.0000007 , illustrating the effect of density on triadic outcomes. This result also shows departure from expectation toward both the all mutual 300 triad and the all null 003 triad, as compared with the Bernoulli random graph model. This pattern is consistent with triadic closure and clustering of friendships.

The prevalence of 003 triads in the Adolescent health networks, in turn, affects dimensionality of the singular value decomposition. To see this effect, suppose that the networks-by-triads array, $\mathbf{T}$, records the relative frequency of the 003 triad in the first column. This means that most rows of $\mathbf{T}$ will have a value in their first column that is around 0.98 (the marginal proportion) and other values that are quite small - together summing to 0.02 . Now, consider the one-dimensional singular value decomposition of T, shown in Figures 2 and 3 and in Table 2. As seen in its definition, singular value decomposition reconstructs a matrix as a product of three matrices, $\mathbf{U}, \mathbf{D}$, and $\mathbf{V}$, in this case: $\mathbf{T}=\mathbf{U D V}^{\prime}$. A one-dimensional singular value decomposition uses the first left singular vector $\mathbf{u}$, with entries $\left\{u_{i}\right\}$, the first right singular vector $\mathbf{v}$, with entries $\left\{v_{j}\right\}$, and the first singular value $\lambda_{1}$. The entry in cell $(i, j)$ of $\mathbf{T}$ is approximated as a function of the first singular value $\lambda_{1}$, the entry in row $i$ of $\mathbf{u}$, and the entry in row $j$ of $\mathbf{v}$ : $T_{i j} \approx u_{i} \lambda_{1} v_{j}$. As a consequence, to reconstruct the Adolescent Health triad proportions in one dimension, the value in the first left singular vector, $\mathbf{v}$, pertaining to the 003 triad must be quite large relative to the other values to approximate the entries in $\mathbf{T}$ that are around 0.98 . It follows that there must be a large difference between the value in $\mathbf{v}$ for 003 and the values for the other triads. This is shown graphically in the gap in triad locations in Figure 3. Since so much of the information in these data distinguishes the 003 triad from the others, a one-dimensional SVD solution approximates the data quite well to capture this distinction.

The one-dimensional space of networks (the first left singular vector, $\mathbf{u}$ ) shows a complementary density effect. Since values in the first right singular vector, $\mathbf{v}$, are comparatively large for the 003 triad, to provide a reasonable approximation to $\mathbf{T}$ a
value in the first left singular vector, $\mathbf{u}$, must primarily accommodate the proportion of 003 triads a network. In the Adolescent Health data a network will have a low value of $\mathbf{u}$ when its 003 triads are less common than is typical, thus offsetting the large value in $\mathbf{v}$ for the 003 triad. Such networks, seen on the left side of Figure 2, are ones with relatively high density and therefore a low percent of 003 triads compared to other networks. To summarize these two results, the one-dimensional singular value decomposition for the 128 triad censuses is an expected outcome of the pervasiveness of 003 triads in these networks, which in turn is an expected result of low density networks.

A related, though slightly different, effect of density is found when the empirical triad censuses are projected into the multidimensional theoretical space. Here, it is useful to think of the theoretical space as representing a universe of expectations for triadic outcomes, given network density and the dyad census MAN. This theoretical space covers triad censuses through the full range of network density (from empty to complete networks) and the full range of legitimate combinations of mutual, asymmetric, and null dyads in the dyad census. Whether triad censuses exist empirically in all regions of this theoretical space is an open question. However, in Figure 6 it can be seen that triad censuses from the Adolescent Health study closely track the low end of the density spine in the theoretical space and do not extend into higher density regions of the space. This indicates that, in this theoretical universe, these triad censuses do not deviate much from what would be expected given their low density and do not appear in extensive regions of the space of possible triad censuses.

## 6. Discussion

Are the results presented in this paper "artifacts" of the affect of network density on triad census distributions, a consequence of using a limited choice sociometric data collection protocol, or due to the nature friendship relations in American high schools? It seems likely that the answer is "yes" to all of these possibilities.

As illustrated above, density constrains possible triad census outcomes. In turn, fixed and limited choice sociometric data collection protocols inescapably limit network density through the relationship of density with mean outdegree and network size. As a consequence, fixed and limited choice protocols preclude observing triad censuses from some (perhaps considerable) portion of the theoretically possible universe. Nevertheless, one might question whether these unobserved, and unobservable, triadic outcomes are indeed socially realistic possibilities given what we know about friendship relationships among adolescents. People can only maintain a limited number of friendships, regardless of the sizes of the groups to which they belong. So, it is reasonable to expect that large groups will have low density, as is found in the large schools in the Adolescent Health data.

Whether the limited choice data collection protocol used in the Adolescent Health study under-represented the "true" number of friends that students have is an important issue to consider. Unfortunately, this question cannot be adequately addressed using the data at hand, since we cannot know how students might have responded had they been presented with a different network data collection protocol. However, a couple of observations might be relevant to the answer. On average, students named 4.2 friends, out of a possible maximum of ten. This could understate the affect of the fixed choice
response format, since there is considerable gender homophily in these networks. The mean number of same-sex friends is 2.77 for females and 2.27 for males, but $17.81 \%$ of females named 5 or more same-sex friends, as did $13.10 \%$ of males ${ }^{3}$. It appears that there is some censoring due to the limited choice format, though the magnitude of the effect might be relatively small.

What, if anything, do these results imply about the nature of friendships among adolescents in U.S. high schools? Three observations seem warranted. First, the number of friends that students have is not unlimited, though the Adolescent Health data do not allow us to determine what the actual number of friends is. Second, friendships tend to be mutual, as seen in the departure of networks (especially high density networks) from the Bernoulli random graph model in the direction of symmetry rather than asymmetry (Figure 6). Finally, and not surprisingly, friendships tend to be clustered, as implied by the higher than expected proportions of 003 and 300 triads. Therefore, in the theoretical space of triads, the Adolescent Health networks are located in the low density region of the space, close to what would be expected given their density, but tending toward symmetry rather than asymmetry.

In summary, the triad census, like other graph-level measures, is heavily constrained by lower order graph features, especially network density. This is consistent with previous observations about limitations that network size and density place on graph-level indices (Anderson, Butts, and Carley 1999; Butts 2006; Faust 2006; Friedkin 1981). The constraint that density places on the triad census becomes especially severe in large networks with low mean degree. We might expect that data collection protocols

[^0]that artificially fix or limit outdegrees will exacerbate the problem. This conclusion has already been well articulated by others (Hallinan 1974a, 1974b; Holland and Leinhardt 1973) but could benefit from further research.

## References

Anderson, Brigham S., Carter Butts, and Kathleen Carley. 1999. The interaction of size and density with graph-level indices. Social Networks 21 (3), 239-267.

Batagelj, Vladimir and Andrej Mrvar, 2004. Pajek 1.02, Ljubljana: University of Ljubljana.

Bearman, Peter S. and James Moody, 2004. Suicide and friendships among American adolescents. American Journal of Public Health, vol. 94 (1): 89-95.

Bearman, Peter S., James Moody, and Katherine Stovel, 2004. Chains of affection: The structure of adolescent romantic and sexual networks, American Journal of Sociology, 110:44-91

Bearman, Peter S., James Moody, Katherine Stovel, and Lisa Thaljin 2004. Social and sexual networks: The National Longitudinal Study of Adolescent Health. Chapter 8 pages 201 to 220 in Morris, M. (ed.) Network Epidemiology: A Handbook for Survey Design and Data Collection, Oxford: Oxford University Press.

Ben-Israel, A. and T. Greville. 1974. Generalized Inverses: Theory and Applications. New York: Wiley.

Brewer, Devon D. and Cynthia M. Webster. 1999. Forgetting of friends and its effects on measuring friendship networks. Social Networks 21:361-373.

Burt, Ronald S. 1992. Structural Holes: The Social Structure of Competition. Cambridge: Harvard University Press.

Burt, Ronald S. and Marc Kenz. 1995. Kinds of third-party effects on trust. Rationality and Society 7(3):255-292.

Butts, Carter T. 2006. Exact bounds for degree centralization. Social Networks 28:283296.

Caplow, Theodore. 1959. Further developments of a theory of coalitions in the triad. American Journal of Sociology, 64(5):488-493.

Coleman, James. 1988. Social capital in the creation of human capital. American Journal of Sociology 94:S95-S120.

Davis, James A. 1967. Clustering and structural balance in graphs. Human Relations 20:181-187.

Davis, James A. 1970. Clustering and hierarchy in interpersonal relations: Testing Two graph theoretical models on 742 Sociomatrices. American Sociological Review 35(5), 843-851.

Davis, James A. and Samuel Leinhardt. 1972. The structure of positive interpersonal relations in small groups. Pp. 218-251 in Sociological Theories in Progress, Volume 2 edited by Joseph Berger, Morris Zeldith Jr. and Bo Anderson. Boston: Houghton Mifflin.

Digby, P. G. N. and R. A. Kempton. 1987. Multivariate Analysis of Ecological Communities. London: Chapman and Hall.

Faust, Katherine. 2006. Comparing social networks: Size, density, and local structure. Metodološki Zvezki (Advances in Methodology and Statistics) 3(2), 185-216.

Faust, Katherine. 2007. Very local structure in social networks. Sociological Methodology, 37:209-256

Feld, Scott L. and Richard Elmore. 1982. Patterns of sociometric choices: Transitivity reconsidered. Social Psychology Quarterly, 45(2):77-85.

Fernandez, Roberto M. and Roger V. Gould. 1994. A dilemma of state power: Brokerage and influence in the national health policy domain. American Journal of Sociology, 99(6):1455-1491.

Friedkin, Noah E. 1981. The development of structure in random networks: An analysis of the effects of increasing density on five measures of structure. Social Networks, 3:41-52.

Goodreau, Steven M. 2007. Advances in exponential random graph (p*) models applied to a large social network. Social Networks, 29:231-248.

Granovetter, Mark S. 1973. The strength of weak ties. American Journal of Sociology 78:1360-1380.

Hallinan, Maureen. 1972. Comment on Holland and Leinhardt. American Journal of Sociology, 77(6):1201-1205.

Hallinan, Maureen T. 1974a. The Structure of Positive Sentiment. New York: Elsevier.
Hallinan, Maureen T. 1974b. Structural model of sentiment relations. American Journal of Sociology 80:364-378.

Handcock, Mark S., Adrian E. Raftery, and Jeremy M. Tantrum. 2007. Model-based clustering for social networks. Journal of the Royal Statistical Society A 107, Part 2: 1-22.

Harris, Kathleen Mullan, Francesca Florey, Joyce Tabor, Peter S. Bearman, Jo Jones, and J. Richard Udry. 2003. The National Longitudinal Study of Adolescent Health:

Research Design [WWW document]. URL:
http://www.cpc.unc.edu/projects/addhealth/design.
Haynie, Dana. L. 2001. Delinquent peers revisited: Does network structure matter? American Journal of Sociology, 106(4), 1013-1057.

Haynie, Dana. L. 2002. Friendship networks and adolescent delinquency: The Relative nature of peer delinquency. Journal of Quantitative Criminology, 18(2), 99-134.

Haynie, Dana. L. and Danielle Payne 2006. Race, friendship networks, and violent delinquency. Criminology, 44(4):475-805.

Haynie, Dana L., Scott J. South, and Sunita Bose 2006. The company you keep: Adolescent mobility and peer behavior. Sociological Inquiry 76(3):397-426.

Holland, Paul W. and Samuel Leinhardt. 1970. A method for detecting structure in sociometric data. American Journal of Sociology, 76(3), 492-513.

Holland, Paul W. and Samuel Leinhardt. 1973. The structural implications of measurement error in sociometry. Journal of Mathematical Socology, 3: 85-111.

Holland, Paul W. and Samuel Leinhardt. 1976. Local structure in social networks. Sociological Methodology 1976, 7:1-45.

Holland, Paul W. and Samuel Leinhardt. 1979. Structural sociometry. Pp. 63-83 in Perspectives on Social Network Research edited by Paul W. Holland and Samuel Leinhardt. New York: Academic Press.

Johnsen, Eugene C. 1985. Network macrostructure models for the Davis-Leinhardt set of empirical sociomatrices. Social Networks 7:203-224.

Johnsen, Eugene C. 1986. Structure and process: Agreement models for friendship formation. Social Networks 8:257-306.

Johnsen, Eugene C. 1989a. Agreement-friendship processes related to empirical social macrostructures. Pages 239-279, in The Small World edited by Manfred Kochen. Norwood, NJ: Ablex.

Johnsen, Eugene C. 1989b. The micro-macro connection: Exact structure and process. Pp. 169-201, in Applications of Combinatorics and Graph Theory to the Biological and Social Sciences edited by Fred Roberts. New York: SpringerVerlag.

Johnsen, Eugene C. 1998. Structures and processes of solidarity: An initial formalization. Pp. 263-302, in The Problem of Solidarity: Theories and Models. edited by Patrick

Joyner, Kara and Grace Kao. 2000. School racial composition and adolescent racial homophily. Social Science Quarterly 81(3): 810-825.

Kao, Grace and Kara Joyner, 2006. Do Hispanic and Asian adolescents practice panethnicity friendship choices? Social Science Quarterly, 87 (5): 972-992

Kao, Grace and Elizabeth Vaquera 2006. The salience of racial and ethnic identification in friendship choices among Hispanic adolescents. The Hispanic Journal of Behavioral Sciences. 28: 23-47.

Knoester, Chris, Dana L Haynie, and Crystal M. Stephens 2006. Parenting practices and adolescents' friendship networks. Journal of Marriage and the Family 68 (5): 1247-1260.

Lambert, Zarrel V., Albert R. Wildt, and Richard M. Durand. 1988. Redundancy analysis: An alternative to canonical correlation and multivariate multiple regression in exploring interset associations. Psychological Bulletin 104:282-289.

Lebart, Ludovic, Alain Morineau, and Kenneth M. Warwick. 1984. Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices. New York: Wiley.

Leinhardt, Samuel. 1972. Developmental change in the sentiment structure of children's groups. American Sociological Review, 37:202-212.

Marsden, Peter V. 1990. Network data and measurement. Annual Review of Sociology 16:435-463.

Moody, James 2001a. Peer influence groups: Identifying dense clusters in large networks. Social Networks, 23: 261-283.

Moody, James. 2001b. Race, school integration, and friendship segregation in America. American Journal of Sociology. 107(3):679-716..

Moody, James and Douglas R. White, 2003. Structural cohesion and embeddedness: A hierarchical conception of social groups, American Sociological Review, vol. 68 pp. 103-127.

Moody, James, 2005. National Longitudinal Study of Adolescent Health Add Health Network Structure Files unpublished Carolina Population Center University of North Carolina at Chapel Hill February 2005

Mouw, Ted and Barbara Entwisle, 2006. Residential segregation and racial friendship in schools. American Journal of Sociology 112(2):394-441.

Obstfeld David, 2005. Social networks, the tertius lungens and orientation involvement in innovation. Administrative Science Quarterly 50 (1): 100-130

Payne, Danielle C. and Benjamin Cornwell, 2007. Reconsidering peer influences on delinquency: Do less proximate contacts matter? Journal of Quantitative Criminology 23 (2): 127-149.

Quillian, Lincoln and Mary E. Campbell. 2003. Beyond black and white: The present and future of multiracial friendship segregation. American Sociological Review, 68:540-566.

Simmel, Georg. 1950. The Sociology of Georg Simmel (translated, edited and with an introduction by Kurt H. Wolff). New York: The Free Press.

Skvoretz, John, Thomas J. Fararo, and Filip Agneessens. 2004. Advances in biased net theory: Definitions, derivations, and estimations. Social Networks 26:113-139.

Snijders, T.A.B. 1991. Enumeration and simulation methods for 0-1 matrices with given marginals. Psychometrika, 56 (3), 397-417.

Snijders, T.A.B., Pattison, P., Robins, G.L., and Handock, M. 2006. New specifications for exponential random graph models. Sociological Methodology. 99-153

South, Scott J. and Dana L. Haynie 2004. Friendship networks of mobile adolescents. Social Forces 83(1):315-350.

Strauss, Richard S. and Harold A. Pollack 2003. Social marginalization of overweight children Archives of Pediatric Adolescent Medicine 157:746-752.

Stewart, Douglas and William Love. 1968. A general canonical correlation index. Psychological Bulletin 70:160-163.

Vaquera, Elizabeth and Grace Kao. 2008. Do you like me as much as I like you? Friendship reciprocity among adolescents. Social Science Research 37:55-72.

Wasserman, Stanley S. 1977. Random directed graph distributions and the triad census
in social networks. Journal of Mathematical Sociology, 5, 61-86.
Wasserman, Stanley and Katherine Faust. 1994. Social Network Analysis: Methods and Applications. New York: Cambridge University Press.

## List of Tables

Table 1. Descriptive statistics for 128 networks
Table 2. Singular values and percent sum-of-squares from SVD of triad censuses, 128 limited choice sociometric networks and 332 theoretical networks

Table 3. Formulas for triad census probabilities
Table 4. Canonical redundancy, fit of 128 networks in the theoretical space of one to five dimensions

## List of Figures

Figure 1. Sixteen triad isomorphism classes
Figure 2. First left singular vector from singular value decomposition of triad censuses from 128 limited choice sociometric networks

Figure 3. First right singular vector from singular value decomposition of triad censuses from 128 limited choice sociometric networks

Figure 4. Left singular vectors for the theoretical triad census space from singular value decomposition of 321 triad census probabilities, conditional on network density and dyad census

Figure 5. Right singular vectors for the theoretical triad census space from singular value decomposition of 321 triad census probabilities, conditional on network density and dyad census

Figure 6. Triad censuses from 128 limited choice sociometric networks projected into the first three dimensions of the theoretical triad census space (Figure 4)


Figure 1. Sixteen triad isomorphism classes


Figure 2. First left singular vector from singular value decomposition of triad censuses from 128 limited choice sociometric networks

```
021D
030T
030C
201
120D
120U 021U
120C 021C
210 111D
300 111U 102 012
M
```

Figure 3. First right singular vector from singular value decomposition of triad censuses from 128 limited choice sociometric networks


Figure 4. Left singular vectors for the theoretical triad census space from singular value decomposition of 321 triad census probabilities, conditional on network density and dyad census


Figure 5. Right singular vectors for the theoretical triad census space from singular value decomposition of 321 triad census probabilities, conditional on network density and dyad census


Figure 6. Triad censuses from 128 limited choice sociometric networks projected into the first three dimensions of the theoretical triad census space (Figure 4)

Table 1. Descriptive statistics for 128 networks

|  | Network <br> size | Density | Mean <br> number <br> of ties | Proportion <br> mutual | Proportion <br> asymmetric | Proportion <br> null |
| :--- | ---: | ---: | ---: | ---: | ---: | ---: |
| Mean | 592.266 | 0.019 | 4.319 | 0.009 | 0.021 | 0.970 |
| Std. Deviation | 440.576 | 0.041 | 1.086 | 0.022 | 0.038 | 0.060 |
| Minimum | 25 | 0.001 | 1.590 | 0.000 | 0.002 | 0.500 |
| Maximum | 2250 | 0.347 | 8.667 | 0.193 | 0.307 | 0.998 |
| Percentiles |  |  |  |  |  |  |
|  | 25 | 309.500 | 0.005 | 3.659 | 0.002 | 0.007 |
|  | 50 | 463.000 | 0.009 | 4.467 | 0.003 | 0.010 |
|  | 75 | 736.750 | 0.016 | 4.996 | 0.006 | 0.019 |
| N |  | 128 | 128 | 128 | 128 | 128 |

Table 2. Singular values and percent sum-of-squares from SVD of triad censuses, 128 limited choice sociometric networks and 332 theoretical networks

| Dimension | 128 Empirical triad censuses |  |  | 332 Theoretical triad censuses |  |  |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
|  | Singular value | Singular value squared | Percent | Singular value | Singular value squared | Percent |
| 1 | 10.540 | 111.092 | 99.416 | 5.397 | 29.128 | 35.058 |
| 2 | 0.793 | 0.629 | 0.563 | 4.829 | 23.319 | 28.067 |
| 3 | 0.138 | 0.019 | 0.017 | 3.213 | 10.323 | 12.425 |
| 4 | 0.067 | 0.005 | 0.004 | 2.693 | 7.252 | 8.729 |
| 5 | 0.017 | 0.000 | 0.000 | 2.301 | 5.295 | 6.373 |
| 6 | 0.010 | 0.000 | 0.000 | 2.160 | 4.666 | 5.616 |
| 7 | 0.006 | 0.000 | 0.000 | 1.109 | 1.230 | 1.480 |
| 8 | 0.003 | 0.000 | 0.000 | 0.938 | 0.880 | 1.059 |
| 9 | 0.001 | 0.000 | 0.000 | 0.682 | 0.465 | 0.560 |
| 10 | 0.001 | 0.000 | 0.000 | 0.615 | 0.378 | 0.455 |
| 11 |  |  |  | 0.369 | 0.136 | 0.164 |
| 12 |  |  |  | 0.111 | 0.012 | 0.015 |

Table 3. Formulas for triad census probabilities

| Triad | Density $^{1}$ | Dyad census $^{2}$ |
| :--- | :---: | :---: |
| 003 | $(1-\Delta)^{6}$ | $N^{(3)}$ |
| 012 | $6 \Delta(1-\Delta)^{5}$ | $3 A N^{(2)}$ |
| 102 | $3 \Delta^{2}(1-\Delta)^{4}$ | $3 M N^{(2)}$ |
| 021 D | $3 \Delta^{2}(1-\Delta)^{4}$ | $\frac{3}{4} N A^{(2)}$ |
| 021 U | $3 \Delta^{2}(1-\Delta)^{4}$ | $\frac{3}{4} N A^{(2)}$ |
| 021 C | $6 \Delta^{2}(1-\Delta)^{4}$ | $\frac{3}{2} N A^{(2)}$ |
| 111 D | $6 \Delta^{3}(1-\Delta)^{3}$ | $3 M A N$ |
| 111 U | $6 \Delta^{3}(1-\Delta)^{3}$ | $3 M A N$ |
| 030 T | $6 \Delta^{3}(1-\Delta)^{3}$ | $\frac{3}{4} A^{(3)}$ |
| 030 C | $2 \Delta^{3}(1-\Delta)^{3}$ | $\frac{1}{4} A^{(3)}$ |
| 201 | $3 \Delta^{4}(1-\Delta)^{2}$ | $3 N M^{(2)}$ |
| 120 D | $3 \Delta^{4}(1-\Delta)^{2}$ | $\frac{3}{4} M A^{(2)}$ |
| 120 U | $3 \Delta^{4}(1-\Delta)^{2}$ | $\frac{3}{4} M A^{(2)}$ |
| 120 C | $6 \Delta^{4}(1-\Delta)^{2}$ | $\frac{3}{2} M A^{(2)}$ |
| 210 | $6 \Delta^{5}(1-\Delta)$ | $3 A M^{(2)}$ |
| 300 | $\Delta^{6}$ | $M^{(3)}$ |

[^1]${ }^{2}$ Numerators for probability, uniform given dyad census (MAN). The denominator is, $\binom{g}{2}^{(3)}$ using descending factorial notation where $z^{(k)}=z(z-1) \cdots(z-k+1)$ (Holland and Leinhardt 1970, 1976).

Table 4. Canonical redundancy, fit of 128 networks in the theoretical space of one to five dimensions

| Dimensions | Canonical <br> redundancy |
| :--- | :--- |
| 1 | 0.831 |
| 2 | 0.864 |
| 3 | 0.957 |
| 4 | 0.981 |
| 5 | 0.995 |


[^0]:    ${ }^{3}$ Students named more than five same-sex friends by putting code numbers for friends of one sex in response blanks intended for friends of the opposite sex.

[^1]:    ${ }^{1}$ Probability in Bernoulli digraph (Skvoretz et al. 2004).

