Introduction to KliqueFinder ©

by Kenneth A. Frank

September 29, 2005

Table of Contents

What is KliqueFinder?

A clustering algorithm.

Identifies non-overlapping cohesive subgroups in social network data.

A technique for producing a crystalized sociogram.

A mapping of ties within and between cohesive subgroups.

Initial presentations of KliqueFinder

Frank, K. A. 1993. Identifying Cohesive Subgroups. Unpublished doctoral dissertation, University of Chicago.

Frank. K. A. 1995. Identifying Cohesive Subgroups. *Social Networks* (17): 27-56.

Frank, K. A. 1996. Mapping interactions within and between cohesive subgroups. *Social Networks* 18: 93-119.

Frank, K. A. and Yasumoto, J. 1996. "Embedding Subgroups in the Sociogram: Linking Theory and Image" *Connections* 19 (1): 43-57.

Some Published Examples

Krause, A., Frank, K.A., Mason, D.M., Ulanowicz, R.E. and Taylor, W.M. (2003). ""Compartments exposed in food-web structure." Nature 426:282-285

Frank, K. A. and Yasumoto, J. 1998. "Linking Action to Social Structure within a System: Social Capital Within and Between Subgroups." *American Journal of Sociology* 104 (3): 642-686.

Foster-Fishman, P.G., Salem, D.A., Allen, D.A., & Fahrbach, K. (2001). Facilitating interorganizational exchanges: The contributions of interorganizational alliances. American Journal of Community Psychology., 29(6).

Plank, Stephen (2000). *Finding One's Place*, Chapter 6 "More on Peer Relations: Cohesive Subgroups" pages 80-110. New York: Teachers' College Press.

McFarland, Daniel (1999). "Organized Behavior in Social Systems: A Study of Student Engagement and Resistance in High Schools." Ph.D. dissertation, University of Chicago.

Bidwell, C., & Yasumoto, J., (1999). "The Collegial Focus: teaching Fields: Collegial relationships, and the Instructional Practice in American High Schools." *Sociology of* Education, 72 (4), 234-256.

Quiroz, P.A., Gonzales, N. F., and Frank, K. A. 1996. Carving a Niche in the High School Social Structure: Formal and Informal Constraints on Participation in the Extra Curriculum. In *Sociology of Education and Socialization*, edited by A. Pallas, Volume 11. London: JAI Press.

The index to maximize

For $n(n\text{-}1)$ pairs of actors i and i', define:

$X_{ii'}$ = 1 if actor i is tied to i',
0 otherwise, and

$samegroup_{ii'}$ = 1 if actors i and i' are members of the same subgroup,
0 otherwise.

Then define

$$\log\left(\frac{P[X_{ii'}=1]}{1\text{-}P[X_{ii'}=1]}\right) = \theta_0+\theta_1 samegroup_{ii'} \quad . \qquad \textbf{(1)}$$

The index to maximize is then θ_1. This is equivalent to the logodds (AD/BC) of the following table:

Association Between Common Subgroup Membership and
The Occurrence of Ties Between Actors

		Tie Occurring No $X_{ii'} = 0$	 Yes $X_{ii'} = 1$	
Subgroup Member-ship	Different 0	A	B	$n(n\text{-}1)\text{-}\Sigma_g\, n_g(n_g\text{-}1)$
	Same 1	C	D	$\Sigma_g\, n_g(n_g\text{-}1)$
		$n(n\text{-}1)\text{-}\Sigma_i\Sigma_{i'} X_{ii'}$	$\Sigma_i\Sigma_{i'} X_{ii'}$	$n(n\text{-}1)$

(n_g represents the number of actors in subgroup g)

The Clustering Algorithm

1) Start with subgroup seeds –

those who interact directly and who interact with common others

2) Iteratively reassign actors to subgroups to maximize an index defining cohesiveness

(new subgroups emerge and old ones dissolve during iterations)

3) stop when no one assignment can improve function

Example of Iterations

Each Cell Represents The Number of Transactions (Also Known as Connections or Exchanges or Ties) Between Actor i (Row) and Actor j (Column)

```
 N             Group And Actor Id
 28          |AAAA|BBBBBBBBBBBBBB|CCCCCCCCCC|
             |    |              |          |
             |  12|1111212 1222 2|  22111   |
 Group     Id|1824|71452655916840|7337803269|
-------------+----+--------------+----------+
  1 A       1|A...|1.......1.....|1.....1...|
  1 A       8|.A.1|..............|.....1.1..|
  1 A      12|..A1|..............|1.........|
  1 A      24|.11A|..............|..........|
-------------+----+--------------+----------+
  2 B      17|1...|B.111.1.111.11|....1.1.1.|
  2 B      11|....|.B.......1....|..........|
  2 B      14|....|1.B11.1.1.1...|..........|
  2 B      15|....|1.1B....111...|..........|
  2 B      22|....|1.1.B......1..|..........|
  2 B      16|....|.....B..11....|..1.......|
  2 B      25|....|1.1...B.11....|...1....1.|
  2 B       5|....|.......B.1....|..........|
  2 B      19|1...|1.11.11.B.111.|......1.1.|
  2 B      21|....|11.1.111.B1...|..1....1..|
  2 B      26|....|1.11....11B...|.......1..|
  2 B      28|....|....1...1..B..|.......1..|
  2 B       4|....|1.......1...B.|..........|
  2 B      20|....|1............B|..........|
-------------+----+--------------+----------+
  3 C       7|1.1.|..............|C11.....1.|
  3 C       3|....|..............|1C..11..1.|
  3 C      23|....|.....1...1....|1.C1....11|
  3 C      27|....|......1.......|..1C....1.|
  3 C      18|....|1.............|.1..C.1111|
  3 C      10|.1..|..............|.1...C.1..|
  3 C      13|1...|1.......1.....|....1.C.1.|
  3 C       2|.1..|.........111..|....11.C1.|
  3 C       6|....|1.....1.1.....|11111.11C.|
  3 C       9|....|..............|..1.1....C|
```

Move 1 from cluster 1 to cluster 3.

CHOICES BY GROUPS
Each Cell Represents The Number of Transactions (Also Known as Connections or
Exchanges or Ties) Between Actor i (Row) and Actor j (Column)

```
 N             Group And Actor Id
 28          |AAA|BBBBBBBBBBBBBB|CCCCCCCCCCC|
             |   |              |           |
             | 12|1111212 1222 2|  2211  1  |
 Group     ID|824|71452655916840|13378067329|
------------+---+--------------+-----------+
  1 A      8|A.1|..............|.....1...1.|
  1 A     12|.A1|..............|.......1...|
  1 A     24|11A|..............|...........|
------------+---+--------------+-----------+
  2 B     17|...|B.111.1.111.11|1...1.1.1..|
  2 B     11|...|.B.......1....|...........|
  2 B     14|...|1.B11.1.1.1...|...........|
  2 B     15|...|1.1B....111...|...........|
  2 B     22|...|1.1.B......1..|...........|
  2 B     16|...|.....B..11....|..1........|
  2 B     25|...|1.1...B.11....|...1..1....|
  2 B      5|...|.......B.1....|...........|
  2 B     19|...|1.11.11.B.111.|1.....1.1..|
  2 B     21|...|11.1.111.B1...|..1......1.|
  2 B     26|...|1.11....11B...|.........1.|
  2 B     28|...|....1...1..B..|.........1.|
  2 B      4|...|1.......1...B.|...........|
  2 B     20|...|1............B|...........|
------------+---+--------------+-----------+
  3 C      1|...|1.......1.....|C......11..|
  3 C      3|...|..............|.C..1111...|
  3 C     23|...|.....1...1....|..C1..11..1|
  3 C     27|...|......1.......|..1C..1....|
  3 C     18|...|1.............|.1..C.1.111|
  3 C     10|1..|..............|.1...C...1.|
  3 C      6|...|1.....1.1.....|.1111.C111.|
  3 C      7|.1.|..............|111...1C...|
  3 C     13|...|1.......1.....|1...1.1.C..|
  3 C      2|1..|.........111..|....111..C.|
  3 C      9|...|..............|..1.1.....C|
```

Final solution: Sample Results of Clustering

:**Table 1**

Partitioned Friendships Among the French Financial Elite

```
 N             Group And Actor Id
 28           |AAAAAAAAAA|BBBBBBB|CCCCCCC|DDDD|
              |          |       |       |    |
              |  11222222|   1111|  11122|  12|
 Group      ID|8929014567|2560347|1456828|3713|
------------+----------+-------+-------+----+
  1 A      8|A.....1.11|...1..1|.......|1...|
  1 A      9|.A...1..11|1.....1|.......|....|
  1 A     12|..A...1111|.......|.1.....|1...|
  1 A     19|...A....11|.......|.......|....|
  1 A     20|....A...1.|.......|.......|....|
  1 A     21|.1...A..11|..1....|.......|....|
  1 A     24|1.1...A111|.......|1......|....|
  1 A     25|..1...1A11|.......|.......|1...|
  1 A     26|11111111A1|1.....1|1......|1...|
  1 A     27|1111.1111A|......1|1.....1|.1..|
------------+----------+-------+-------+----+
  2 B      2|.1......1.|B1..1.1|.1.....|....|
  2 B      5|..........|1B1...1|....1..|....|
  2 B      6|.....1....|.1B..11|...1...|....|
  2 B     10|1.........|...B.11|.......|....|
  2 B     13|..........|1...B1.|.......|....|
  2 B     14|..........|..111B1|.......|11..|
  2 B     17|11......11|1111.1B|.1.....|....|
------------+----------+-------+-------+----+
  3 C      1|......1.11|.......|C.....1|....|
  3 C      4|..1.......|1.....1|.C..111|1...|
  3 C     15|..........|.......|..C1.1.|....|
  3 C     16|..........|..1....|..1C...|....|
  3 C     18|..........|.1.....|.1..C1.|....|
  3 C     22|..........|.......|.11.1C.|....|
  3 C     28|.........1|.......|11....C|....|
------------+----------+-------+-------+----+
  4 D      3|1.1....11.|.....1.|.1.....|D111|
  4 D      7|.........1|.....1.|.......|1D..|
  4 D     11|..........|.......|.......|1.D.|
  4 D     23|..........|.......|.......|1..D|
```

Producing a Crystalized Sociogram

Metric multidimensional scaling within each subgroup to locate actors relative to subgroup.

Metric multidimensional scaling between subgroups to locate subgroups relative to one another.

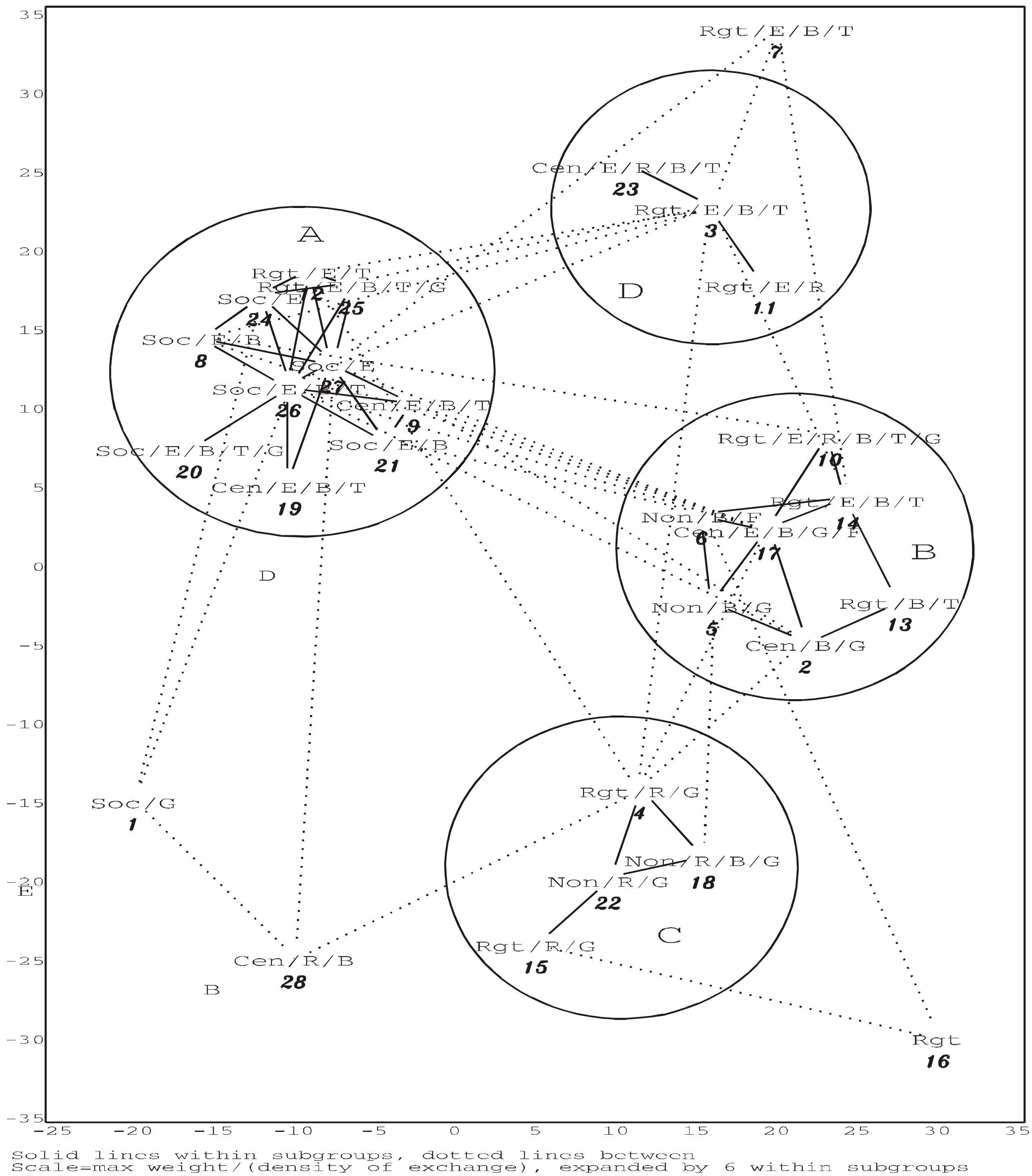
Rotate actors within subgroups to reduce the number of between subgroup interactions which cut across the space of a subgroup.

Combine results to produce a single plot with a common metric.

Sample Result of Crystalized Sociogram: Friendships Among the French Financial Elite

Figure 1

Crystallized Sociogram: Friendships Among the French Financial Elite



Party affiliation: Soc=Socialist, Cen=Central, Rgt=Right, Non=None/unknown),
R if member of Resistance de Socialism.
E if attended Ecole National d'Administration (ENA)
T if was employed in the treasury
B if was employed in the banking industry
G if was employed in Grand Banque
F if was a member of the prestigious Firme de la Finance

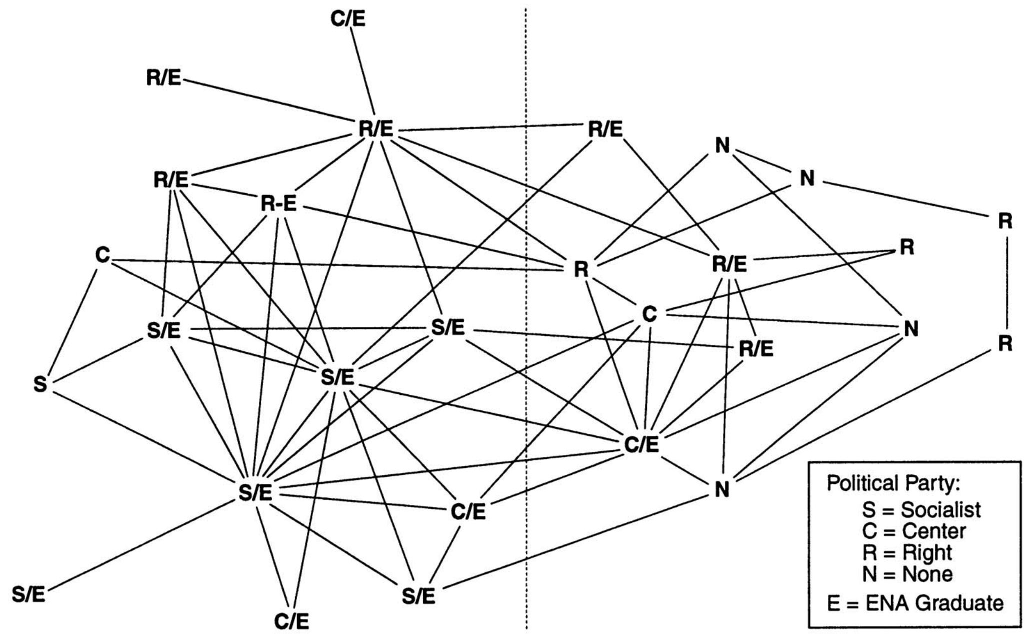


Figure 1. Sociogram Illustrating Friendship Links, Political Party Preference, and ENA Graduation: 28 Inner-Circle Elites in France, 1990

correlated with friendship were then analyzed by creating an adjacency matrix composed of 1s and 0s corresponding to whether or not any given pair of the 28 possessed the same attribute. Since visual presentations are useful in getting a sense of the data, I first turn this matrix into a graph. Then I develop analyses to create a more conventional graph that depicts a "path analysis" of the friendship structure and attributes of friendship.

Figure 1 is a computer-drawn sociogram.[11] Each letter or group of letters indicates persons and represents party preference and ENA graduation. Lines between the letters indicate a direct friendship connection. Distances between persons do not represent the strength of the friendship. Rather, the multidimensional scaling algorithm used initially to locate persons, places them closer together if they have a large number of friendship with the same other persons.

X's friend. Two respondents, however, named almost everyone as their friend. For these two respondents, friendships were counted only if they were reciprocated by other respondents. Even so, these two respondents had more friends than anyone else.

When these procedures were completed, one nonrespondent of the top 28 still had no friendship link to any other in the top 28. de Quillacq added two friends for this person, and one for another of the nonrespondents, who unaccountably was not linked with a respondent whom "everybody" knew was his friend. Sociometric graphs constructed without de Quillacq's additions, however, offer substantially the same "gestalt." de Quillacq insisted that they were incorrect for the nonrespondents, however, so we agreed to use the present version, which is augmented by her participant observation.

[11] Figure 1 was developed in several steps. First, I created a Hubbell Input-Output measure (Hubbell 1965) of similarity between each pair. This method of measuring similarity between pairs in a dichotomous matrix, a relative of Leontief input-output analysis, essentially notes how many paths at what distances lie between any pair of friends; but as pairs become further and further distant, each path counts for less. Second, the Input-Output matrix was subjected to three-dimensional nonmetric multidimensional scaling. Third, lines were drawn between points that were originally directly connected. The resulting graph was then rotated to show the lines

General Advantages of KliqueFinder

1) The crystalized sociogram as an intuitive image

Embedding subgroups in a sociogram generates an image consistent with the intuitions of researchers (e.g., Kadushin, 1995; Laumann, 1970; Moreno, 1934; Nakao and Romney, 1993). These researchers, and many others, have drawn circles around subsets of actors in a sociogram. But there is an objective, theoretical, and statistical basis for the circles drawn by KliqueFinder compared with the ad hoc and esthetic criteria employed by the researcher's hand.

2) The crystalized sociogram facilitates micro and macro analyses

Like the primary group, the subgroups define the micro contexts that affect actors actions and behaviors and filter influences from the larger system. Cohesive subgroup boundaries are sociologically salient because they define regions in which information circulates and influence is concentrated.

The subgroups also define components which are integrated to form the macro level structure. This duality was appreciated by Festinger et al, (1950) who used the physical proximity of apartment complexes to define rumor as spreading within, then between, cohesive subgroups (see Frank, 1996, Figure 1).

3) Other aspects of social structure defined relative to cohesive subgroups.

Actors can be characterized as central to a subgroup, spanning between subgroups, unaffiliated with a subgroup, etc. Structural holes can be defined in terms of gaps between subgroups. Centralization of the network can be defined in terms of ties between subgroups rather than between actors. See Frank (1996), Frank & Yasumoto (1996, 1998).

4) Peer Review Journals

The KliqueFinder clustering algorithm and the crystalized sociogram were first presented in *Social Networks*, a peer review journal. As such, the ideas and their presentation have been formally acknowledged by other social network researchers. Moreover, the process sharpened the presentation of the ideas as well as aspects of the procedures. Applications of KliqueFinder have appeared in multiple peer review journals. This is in large part because the subgroup memberships are related to actors' beliefs and behaviors, helping social networkers transcend the mere description of social structure.

Specific Advantages of the KliqueFinder Algorithm and Index

Relative to other cohesion-based algorithms & criteria

1) No need to prespecify or use ad hoc criteria of connectedness. Compare with *k*-clans, *k*-plexes, etc., in UCINET.

KliqueFinder's flexible index eliminates the need to pre-specify the extent of connectedness defining cohesion, or to select a criterion post-hoc.

2) Non-overlapping subgroups

KliqueFinder's flexible index accommodates variation in pattern of ties and thus allows for non-overlapping subgroups. Compare with *k*-plexes, *k*-cores, algebraic solutions, lambda sets, etc.,

Are subgroups really non-overlapping (Simmel)?

The ego-centric paradox: across many systems and relations, each person is a member of multiple subgroups. But in a bounded system, for a single relation, our conception is of non-overlapping subgroups (ask yourself about the system of closest colleagues among graduate students and professors, or among those involved in a favorite hobby):

Durkheim *The Division of Labor in Society* (1933)
Roethlisberger and Dickson *Management and The Worker* (1941)
Festinger, L., Schachter, S., and Back, K. *Social Pressures in Informal Groups* (1950)
Blau *Inequality and Heterogeneity* (1977)
Burt *Toward a Structural Theory of Action* (1982), *Structural Holes* (1992).

How can we apply theories of subgroups to individual behavior if two actors can simultaneously be the members of the same and different subgroups? Overlapping boundaries fail to establish "an inside and an outside" (page 872, Abbott, 1996) necessary to define a sociological entity.

3) No need to pre-specify the number of groups.

The KliqueFinder algorithm allows for the emergence and dissolution of subgroups as it iterates.

4) Pre-simulated sampling distribution for $\mathbf{2}_1$ (the index of cohesiveness).

Frank (1995) reports a model for $\mathbf{2}_1$ from the application of KliqueFinder to random networks of varying sizes and distributions of ties, but with no tendency for subgroups. Each KliqueFinder output contains a statistical test of the observed value of $\mathbf{2}_1$ in your data using a baseline of the predicted value from simulations. There is also the capacity to generate a sampling distribution unique to a specific data set (see manual for details).

```
PREDICTED THETA (1 base) BASED ON SIMULATIONS.
VALUE BASED ON UNWEIGHTED DATA.

          0.65962

ESTIMATE OF THETA (1 subgroup processes)

          0.43653

THE TOTAL THETA1 IS:

          1.09616
```

Compare the two models:

$$\log\left(\frac{P[X_{ii'}=1]}{1\text{-}P[X_{ii'}=1]}\right)=\theta_0+\theta_{1\ base}\ samegroup_{ii'}\ . \qquad \textbf{(2)}$$

Versus

$$\log\left(\frac{P[X_{ii'}=1]}{1\text{-}P[X_{ii'}=1]}\right)=\theta_0+\theta_{1\ base}\ samegroup_{ii'}+\theta_{1\ subgroup\ processes}\ samegroup_{ii'}\ . \qquad \textbf{(3)}$$

In this case the values are:

$$\log\left(\frac{P[X_{ii'}=1]}{1\text{-}P[X_{ii'}=1]}\right)=\theta_0+.66\ samegroup_{ii'}+.44\ samegroup_{ii'}\ . \qquad \textbf{(4)}$$

```
Approximate LRT BASED ON PREDICTED THETA (1 base)

         17.92620

COMPARE TO CHI-SQUARE ON 1 DF
P-VALUE (LESS THAN OR EQUAL TO):

          0.00100
```

5) The algorithm has been calibrated

Each KliqueFinder output reports the predicted performance of the algorithm given the structure of the data. This prediction is based on application of the algorithm to simulated data with a known solution. Note that performance is quite good when there is evidence of subgroups (see point 4, above).

```
PREDICTED ACCURACY: LOG ODDS OF COMMON SUBGROUP
MEMBERSHIP, + OR - .5734 (FOR A 95% CI)

          1.64921

The Log odds applies to the following table:

                  OBSERVED SUBGROUP
                  DIFFERENT   SAME
                 ___________________
                 |        |        |
       DIFFERENT |   A    |   B    |
KNOWN            |        |        |
SUBGROUP         |--------|--------|
                 |        |        |
            SAME |   C    |   D    |
                 |        |        |
                 -------------------

THE LOGODDS TRANSLATES TO AN ODDS RATIO OF

          5.20288
WHICH INDICATES THE INCREASE IN THE ODDS
THAT KLIQUEFINDER WILL ASSIGN TWO ACTORS TO
THE SAME SUBGROUP IF THEY ARE IN THE SAME
KNOWN SUBGROUP VERSUS IF THEY ARE IN DIFFERENT
KNOWN SUBGROUPS
```

6) Index of cohesiveness is defined by a sociologically meaningful model

The index is proportional to the increase the probability that two actors are tied given that they are members of the same subgroup. This is the same index that would be used to assess the effect of a common affiliation in a p* based model.

7)The index of cohesiveness is linked to similar indexes

Frank (1993, 1995) links $\mathbf{2}_1$ to other statistical indexes of cohesion such as compactness (Hubert 1987) or bias in statistical networks (Fararo, 1981; Skvoretz, 1991). $\mathbf{2}_1$ also can be linked to the classic problem of the constant frame of reference (Bronfenbrenner, 1943; Criswell, 1950; Freeman, 1978; Smith, 1947).

8)Direct extension to weighted or valued data

Assuming the weights define a ratio scale (10 is twice as big as 5), one can redefine the function to be maximized in terms of a weighted log-odds (see Table 2 in Frank, 1996).

Relative to structural equivalence (e.g., Concorr)

1)Cohesive subgroups consistent with longstanding theory:

Nadel (1957) offers a description of the organization in terms of integrated subgroups:

> Sub-groups, like that widest group 'society at large', are made up of people in determinate relationships. And any group is characterized by the kinds of relationships that occur between the people in question, holding them together. Now, inasmuch as subgroups are discrete entities, bounded units, at least certain of these characteristic relationships must be equally bounded, that is, they must come to an end somewhere, their cessation demarcating the boundaries of the group. From this point of view, then, we might describe sub-groups as areas of bounded relationships. But inasmuch as they are also subdivisions of a wider collectivity and not isolated, self-sufficient units, the bounded units themselves must be interrelated (pages 13-14).

Note that the boundary of each subgroup marks the cessation of relations within that subgroup, implying that interactions are concentrated primarily *within* subgroups.

Merton (1957): “sub-groups are structurally constituted by those who develop distinctive social relations among themselves which are *not shared* with other members of the larger group” (page 287, *emphasis added*)

Burt (1982): "relational models [based on direct interactions among actors] have a high face validity stemming from their consistency with classic concepts of communication in face-to-face primary groups as the socializing unit in society" (page 221).

2) KliqueFinder can identify structurally equivalent blocks in social network data

Based on Panning’s algorithm, maximizing sums of squares between blocks, the KliqueFinder algorithm can identify blocks of structurally similar actors, and these blocks can be embedded in a sociogram. But the result is not necessarily productive (see Frank, 1996, Figure 3).

3) KliqueFinder can identify cohesive clusters starting with two-mode data

When the original data matrix is rectangular, such as people participants in activities, KliqueFinder can begin by multiplying XX’ to obtain a square symmetric matrix of commonalities among rows (the people) or X’X to obtain a square symmetric matrix of commonalities among columns (the activities). One can then apply KliqueFinder to identify cohesive subgroups among the rows or columns.

Although individuals in structurally similar blocks may behave similarly (e.g., White et al. ,1976), we typically conceive of organizations and collectives as composed of cohesive subgroups which mediate between individual and higher levels of analysis.

Comparison with Hierarchical Clustering

1) Reporting a hierarchy of subgroups

The conception of a hierarchy of subgroups begins with non-overlapping subgroups. KliqueFinder has the capacity to order subgroups according to the ratio of in to out degree, or other measures of prestige.

2) Nested subgroups produced by hierarchical clustering

The regions in the crystalized sociogram indicate the next level of hierarchy. It is also possible to use the density of ties between subgroups to perform a second layer of clustering of subgroups. But, generally, when clusters from a hierarchical algorithm are interpreted, researchers often use ad hoc criteria to focus on a single level of partitioning. This is necessary because one cannot model the effects of subgroup membership if two actors are simultaneously members of the same and different subgroups.

Uses of KliqueFinder

1) A theoretical basis for cohesive subgroups

Although I have adapted KliqueFinder to perform multiple analyses (e.g., to identify structurally equivalent blocks) it is best <u>not</u> to use KliqueFinder as an all purpose clustering algorithm. Before applying KliqueFinder one should be able to state the theoretical mechanisms through which cohesive subgroups are induced, and through which subgroups have salient effects on actions or beliefs.

2) Best used for social network ties or relations

There are many other algorithms in SAS, SPSS , etc., for identifying clusters in rectangular bimodal data (e.g., a typical data set of people and variables). Many of these procedures have evolved over years, have known properties and weaknesses, and can be linked to statistical and theoretical principles (generally, I am not a fan of most clustering algorithms which impose a structure on fluid data that could better be analyzed with simple principle components). In contrast, KliqueFinder takes advantage of the unique structure of social network models and tables to define $\mathbf{2}_1$, and thus is best used for such data. For bimodal data, one might consider correspondence analysis to plot both row and column elements in a fixed space.

3) Emphasis on the Crystalized Sociogram

Although KliqueFinder reports many statistics of centrality, linkage, etc., the interpretation of the subgroups should be based primarily on the crystalized sociogram. For example, in the crystalized sociogram, one can observe when two subgroups are plotted closely together. This type of information is difficult to discern from statistical output, but graphically suggests caution in interpreting the boundary separating the subgroups.

I refer to the KliqueFinder output primarily to reference the sampling distribution of $\mathbf{2}_1$ and to evaluate the likely performance of the algorithm.

4) Best for symmetric, undirected, ties or relations

The processes occurring in cohesive subgroups, such as information sharing and influence, are most likely to occur through symmetric ties such as communication or friendship. When a tie is directed, such as telling or liking, it may be more worthwhile to employ a structural equivalence approach which capitalizes on commonalities extended or received by pairs of actors.

5) Some programming skill required

Unfortunately, the interface for KliqueFinder is not as user friendly as some other software. Although running a basic analysis requires just two simple commands, the user must typically modify the SAS program, *socgram.sas*, to produce a publishable crystalized sociogram. The sAS program can be modified in Unix or windows, but the user must engage the SAS code (with help from the user's manual).

The basics: All You need to do to run KliqueFinder

Go to pikachu.harvard.edu/wkf and enter kf as the logon. The password (get from me)

Follow directions to install.

1) Under basic specs, do browse to choose the directory in which you want to work. Note that "Browse" cannot create a new directory.

2) setup file
   choose "basic setup" and then "Run setup file"
3) Choose a data file (either your own or one of the examples)
4) choose options as described in manual
5) click "run analysis"
6) look at clusters output, confirm evidence of clusters

to make the picture

7) click "make graph"
8) make modifications based on contents of this manual
9) click run sas
   If SAS doesn't run ...
   Open sas
      1) from the program editor, open the file socgramz.sas in the directory in which you were working above.
      2) find (ctrl-f) "inlist.axdjust"
      3) modify "inlist.xadjust" to include full path: "c:\...\inlist.xadjust";
      4) find (ctrl-f) "socgramx.eps"
      5) modify "socgramx.eps" to include full path "c:\...\socgramx.eps"
      6) run

10) view file "socgramx.eps" using word perfect, adobe, or ghostwriter.

You may have to edit socgramz.sas (as described in this manual) to modify plots.

Conversion of data between UCINET and KliqueFinder

(note these commands are based on a slightly older version of UCINET, but I think they pretty much hold)

A) KliqueFinder ÷ UCINET

% setup ucinet
% kliqfind mydata.list

There is now a file called *ucwn.DAT* that contains information to be read by UCINET. To read this, ftp the file to your c drive, open up UCINET-V, and go to datasets/import/ucinet3.0 . Enter the filename containing the data (probably *ucwn.DAT* unless you renamed it), indicate the data are in integer format, and designate an output file name.

B) UCINET ÷ KliqueFinder

1)run UCINET
2)go to datasets/export/raw
3)indicate data set to export. Consider these defaults:
    diagonals: absent
    decimal places: 0
    fieldwidth: 9
    Guarantee space: yes
    page width: 255
    embed row labels: no
    embed column labels: no
    embed level variables: no
    indicate output data set
4)ftp output data set to edstat2
This data set should be in a format ready to read directly into KliqueFinder

Data sets available

1) French financial elite (look in kliqfind/data/ffe directory)

Jeff Yasumoto and I analyzed these data in the paper: Frank, K. A. and Yasumoto, J. (1998).

The data you have include friendships among the french financial elite (ffe.list), hostile actions (hostile.list) and supportive actions (suppor.list) as well as background characteristics (ffe.sas7bdat).

A table linking the ID numbers in your data and those used in the paper is in *nicetab3.lst,* along with other characteristics of the people.

UCINET data files are in **.##h* and *#.##d* -- you need both the *h* and the *d* extension to call up in UCINET.

2) "Our Hamilton High" --stanne (look in kliqfind/data/stanne)

These are the data I analyzed in Frank (1995), and Frank (1996). They consist of professional discussions among teachers in a high school (*stanne.xlist*). Characteristics of individuals are in a sas file called *background.ssd01*. A table linking ID's in the data to those reported in the article is in *nicetab3.lst.*

UCINET data sets are in *magent.##h*, *magent.##d* (moral agency data) and *stanne.##h* and *stanne.##d* (professional discussions among teachers).

3) Participants in a social capital conference (look in kliqfind/data/soccap)

At a conference at MSU in April, 1998, we had many internationally known researchers in the field of social capital (Etzioni, Putnam, Fukuyama), as well as in education (Valerie Lee, Charles Bidwell, Bo Bileau, Aage Sorensen, Ralph McNeal, Barbara Schneider, Steve Morgan). The basic list consists of who was aware of whom academically (soccap*1.list*). We also developed measures of each person's definition of social capital (*socdef.ssd01*). We have limited background information. I have a 5 page write up on this if you are interested.

KliqueFinder Applications

1) adding in attributes:

run KliqueFinder

data file soccap.list

make graph

use ID from other file? Yes:

sas file name: c:\kliqfind\soccap2 [be sure to include full path]

id variable: chooser

string variable: r1 (social capital is about relationshps)

Save

run sas:

retrieve socgramz.sas from working directory and run

2) Alternate lines

run KliqueFinder

data file nonpr1.list

make graph

save

run sas

retrieve socgramz.sas and run

replace all occurrences of nonpr1.list with nonpr2.list

run