Data with complex dependencies:

The joint analysis of individual attitudes and networked social systems

Garry Robins

Dean Lusher

Philippa Pattison

Department of Psychology,
University of Melbourne
Australia

Peter Kremer
School of Health and Social Development
Deakin University
Australia

5 December 2007.

Note: This research was assisted by grants from the Australian Football League Research and Development Board and the Australian Research Council. Earlier versions of this paper were presented at the NICTA Workshop on Large Scale Network Analysis, Sydney, Australia (Nov 2005); at the Australasian Mathematical Psychology Conference, Adelaide, Australia (Dec 2005); at the Sunbelt International Social Network conference, Vancouver (May 2006); and at the Age of Networks Speaker Series, University of Illinois, Urbana-Champaign (Nov 2006).

Abstract

The assumption of independence of observations that underpins standard statistical techniques can fail when the observations are ties embedded within a complex relational structure. We show how to analyze the effect of individual-level attitudes on the formation of a social network of relational ties while also recognizing that the relational structure may itself be subject to self-organizing tie formation processes. The analytic question we address is whether (and in what way) attitudes shape the formation of the relational structure over and above other relevant effects. These additional effects may pertain to other individual-level qualities as well as to self-organizing structural processes. Newly developed specifications for exponential random graph (*p**) models permit inferences about competing hypotheses that might explain the emergence of relational patterns. We illustrate this approach using attitudinal and network data from an elite all-male sporting team. We simultaneously estimate structural effects of reciprocity and transitivity, generalized popularity and activity effects, and effects involving player ability and player experience. We show that in this sporting team, even in the presence of these other effects, player attitudes towards women are important in determining who socializes with whom. We infer that attitudes towards women are one organizing principle of this network, reflective of team culture. We conclude by discussing the desirability of bringing some unity to currently differentiated areas of social psychology and social network analysis.

Many of the statistical models most widely used in psychology make a fundamental assumption of independence of observations. Yet, for studies examining a phenomenon of interest in a particular social setting, dependence among observations may not only be expected but may be the object of enquiry. Indeed, it may be argued that a social system without some degree of dependence is not, in fact, a system. The presence of dependencies implies the presence of structure that needs to be understood if correct inferences are to be made.

Such, for instance, is the case with impressive developments in multi-level and spatial modeling (e.g. Bryk & Raudenbush, 1992; Klein & Kozlowski, 2000; Snijders & Bosker, 1999; Raudenbush & Sampson, 1999). In multi-level modeling, for example, the structure of social settings is presumed to comprise individuals nested within a partition of groups, with separate components of variation at individual and group level. It is well-known that by ignoring this structure – specifically, by ignoring the group-level – incorrect inferences may be drawn about the individuals. Certain associations erroneously ascribed to individuals may in fact be due to the grouped nature of the data. In this case, blind application of a statistical model that assumes independent observations may lead to error, whereas more sophisticated multi-level linear approaches can reveal important features about both the individuals and their social context.

Experimental studies in social cognition may sometimes avoid such problems by careful control of contextual variables. For example, the researcher may design a simple dependence structure, by using completely randomized designs that eliminate any unforeseen confounding effects of social context and individual differences. Standard experimental approaches with sufficient control enable confident use of standard statistical models. A classical example in this tradition is the wealth of results and theorizing that has grown from Tajfel's minimal group paradigm (Tajfel, Billig, Bundy, & Flament, 1971). This approach deliberately steps back from the direct study of a social system, but rather concentrates on isolating effects, one from another, with the aim of understanding them separately. The idea is that extraneous dependence among observations is a nuisance that complicates analysis, and control is necessary to eliminate it. The cost is that contextualization is lost.

So if the dependence structure can be deliberately simplified through experimental control, then statistical models that assume independent observations are available for use by researchers. If, on the other hand, the social setting induces dependence among observations, more sophisticated modeling approaches are required. In some cases, the assumption of a relatively simple dependence structure may suffice, such as in the case of the nested multi-level structures amenable to multi-level modeling. But there are occasions where the dependence among observations may take a more complex form. Relational data structures exemplify such forms, where the objects of enquiry are the relationships between (say) individuals. Consider, for example, the social network depicted in Figure 1. Here the individuals are represented by the nodes (the dots) and relationships among them - also called *relational ties* or *network ties* - represented by arcs (the arrows). (We provide more detail on this data below). If we wish to develop a model for this network, then the objects of enquiry are the possible ties among the individuals. Suppose that a relationship between two individuals creates a potential dependence between the ties from those individuals to a third party, in the sense that the latter ties are more likely to co-occur. This pattern of dependence among ties may be replicated across the entire network and cannot be simplified to the nested hierarchical structures suitable for multi-level modeling.

In this article, we suppose that such dependencies among relational observations may exist, and that, in addition, individual-level variables may also contribute to the formation of social relationships among those individuals. Thus we anticipate a situation in which both individual-level constructs and other relational observations create the social context in which any given relational tie is observed. An important research question is then to determine which of these potential effects is at work, recognizing that any given tie may simultaneously serve as social context to other ties as well as be subject to the contextual effects of those other ties. In such a situation, the approach of treating dependence among observations as a nuisance to be eliminated may be counter to the question the researcher is addressing. It is the network structure itself that is the object of enquiry and its underlying dependencies cannot be simply avoided.

In this article we describe a method for analyzing the potential impact of individual constructs – specifically, in our case, individual attitudes – on networked social structures as exemplified by the social network in Figure 1. This method utilizes recent methodological developments in exponential random graph models for social networks. The article is structured as follows. We begin by reviewing some arguments as to why complex forms of dependence might be expected in social systems, and thereby anticipate the type of research questions that could be addressed in those circumstances. We then provide a brief introduction to exponential random graph models, describing new specifications of Snijders, Pattison, Robins and Handcock (2006) that strengthen this class of models considerably. Combining the network model specifications of Snijders et al with the framework for social selection models in Robins, Elliott and Pattison (2001), we propose five important dyadic parameters for both binary and quantitative individual measures in interaction with relational variables. We provide an illustrative empirical example, based on the network in Figure 1, and we argue that there are several possible competing effects to explain the emergence of this network. Fitting these effects simultaneously allows us to make inferences about the effects of the attitudinal variables. We conclude by describing how further extensions of this approach may address issues important to social psychology and to social network theory.

Figure 1 about here.

Why complex social systems?

It is worth noting that some eminent researchers have taken rather strong theoretical positions about the importance of construing social systems in terms of complex dependencies. The social psychologist, Solomon Asch, noted the inevitability of recursive feedback effects in theorizing social systems, implying a complex dependence structure that is both multi-level and cyclic:

> "One could say that all the facts of the system can be expressed as the sum of the actions of individuals. This statement is misleading, however, if one fails to add that the individuals would not be capable of these particular actions unless they were responding to (or envisaging the possibility of) the system." (Asch, 1952, p.252).

An interesting aspect of this conceptualization is its multi-level nature (individual/system) with cross-level dependencies, but with the twist that the higher (system) level is latent. There is no requirement here that the "system" be instantiated in any explicit organizational form, let alone the overt partition of groups required for multi-level modeling.

The social network analyst, Barry Wellman, argued that use of many familiar statistical models implicitly nailed a researcher's colors to the mast of individualism:

> "(Statistical methods) continue to treat individuals as independent units. The very assumption of statistical independence … detaches individuals from social structures and forces analysts to treat them as parts of a disconnected mass. Researchers following this tack can only measure social structure indirectly … They are forced to neglect social properties that are more than the sum of individual acts." (Wellman, 1988, p.38).

Of course, the argument presupposes that there are social properties that are more than the sum of individual acts. It is commonly supposed that such properties do exist, although not all theorists would agree (see Allport's, 1924, radical individualism – all social psychology reduces to "entirely a psychology of individuals", p.4). The question then stands: how would one know empirically whether there were such properties?

The sociologist, Andrew Abbott, presents one of the most radical critiques of decontextualization, even prepared to argue that random sampling is the culprit:

> "… the rapidly advancing discipline of sampling … separated individuals from their social context of friends, acquaintances, and so on, but also deliberately ignored an individual *variable's* context of other variables in the name of achieving 'more complete' knowledge of the variable space. Sampling not only tamed contextual effects to mere interactions, it also thereby produced data sets in which the levels of contextual causation were deliberately

minimized. This would later enable a whole generation of sociologists to act as if interaction were a methodological nuisance rather than the way social reality happens." (Abbott, 1997, p. 1162).

Finally, the organizational theorist, Karl Weick, dryly noted that preoccupation with individual cognition has left researchers "ill-equipped to do much more with the so-called cognitive revolution than apply it to organizational concerns, one brain at a time." (Weick & Roberts, 1993, p.358). Weick is not an adherent to approaches that ignore systematicity in his domain of organizational theory.

We view these firmly expressed arguments as interesting challenges to any social science approach that relied exclusively on models that assume independence of observations. Any method carries with it theoretical assumptions about the nature of the data. For the specific research questions that we address in our empirical example, to make a claim about independence of observations is to make a strong – and indeed implausible – theoretical claim.

So how then can a researcher come to know whether there are systemic properties more than "the sum of individual acts"? Or to put this question in the terms we address in this article: how may a researcher understand which of many potential aspects of social context are at work when observations are both part of the context as well as the phenomenon to be explained? In particular, we ask whether, in the presence of potential dependencies among relational observations, individual attitudes are associated with relational structure.

Exponential random graph (*p**) models for social networks

Models for network data

Social network data includes measures on the relationships (of various sorts) among a given set of individuals, and may also include individual-level variables relating to those individuals. A *network tie* expresses a relationship between two individuals, so a network has many possible network ties, but only some of them will be observed.

Exponential random graph models are a class of statistical models for social networks introduced by Frank and Strauss (1986) with their Markov random graph models. They were elaborated and popularized as *p** models in the 1990s (Wasserman & Pattison, 1996 – see Wasserman & Robins, 2005, for a review, while Robins, Pattison, Kalish & Lusher, 2007, provide a tutorial-level introduction). These are models for observations of possible network ties, and hence differ from most statistical analyses in psychology in that the observations are taken to occur between pairs of individuals, not on individuals singly. As well, there is an explicit recognition that observations are not independent. Contractor, Wasserman and Faust (2006) have emphasized the value of these models in examining a number of theoretically-based cross-level effects simultaneously, including individual and relational effects, and incorporating several social psychological, sociological and social network theories (see also Monge & Contractor, 2003).

Markov random graphs are based on the assumption that two possible network ties are conditionally independent unless they share an individual (that is, a tie from Mark to Jenny is not independent of a tie from Mark to George, because both ties involve Mark). At first glance, this seems to be a rather simple assumption but methods of analysis quickly become complicated (see Robins & Pattison, 2005, for a discussion of dependence assumptions and their implications in social network models). What has become more apparent in recent years is that the Markov dependence assumption is inadequate for much observed social network data (Pattison & Robins, 2002). This has led to the development of more complex dependence assumptions resulting in new specifications for this class of models (Snijders et al, 2006). These new specifications have been shown to be much more successful than Markov models in fitting network data (Robins, Snidjers, Wang, Handcock & Pattison, 2007; see also Goodreau, 2007, for impressive model performance with a large network.)

It is not our intention to go into full detail here in describing the model with these new specifications. Instead we provide an intuitive account, while referring interested readers to the relevant literature, although at the end of this section we provide a formal mathematical description of the model. In essence, these models express the probability of a network as a function of certain types of subgraphs, or patterns, of network ties. Each such pattern has a parameter in the model. A large and positive parameter for a pattern suggests that graphs exhibiting a high numbers of these patterns are more probable, all else being equal, whereas a large and negative parameter suggests that such graphs are less probable. A large positive parameter estimate suggests, therefore, that there is a tendency for that pattern to occur in the network in a way that cannot be explained by the other effects in the model.

What counts as such a pattern, or *configuration* to use a more technical term[1]? Figure 1 is an example of a *directed network* with nodes and arcs (arrows); we focus on directed networks for this article[2]. In Figure 1, a double-headed arrow indicates the presence of two arcs and represents a configuration in which a relationship expressed from a first node to a second is *reciprocated* by a relationship from the second to the first. This situation can arise, for instance, when Harry nominates Sally as a friend, and Sally also nominates Harry. Obviously there is no necessity that Sally should reciprocate the nomination, but in human social networks involving positive affect, reciprocation is often a strong tendency. It makes sense, then, for a reciprocated tie to be a

[1] In recent physics and biological network literatures the term *motif* has been employed. Because *configuration* has been used in the social network literature for over 70 years (Moreno & Jennings, 1938), we prefer to retain it here

[2] A *nondirected network* has nodes and edges, i.e. lines without arrows.

configuration in a model and we might expect its parameter to be large and positive. Transitivity is another example of network-based structural processes, whereby transitive triads are often observed in human social networks (for instance, as exemplified by the old saying, “Friends of friends become friends.”) If we estimate the parameters of such a model for a given dataset, we can obtain evidence whether there are indeed tendencies for reciprocation or transitivity within the dataset.

Figure 2 about here.

Models will often include configurations that are known to be important in social networks. We depict five of these in Figure 2: single arcs (or *density*); reciprocated arcs; *k*-out-stars (or *activity*); *k*-in-stars (or *popularity*); *k*-triangles (*transitivity* or *closure*). We discuss each of these in terms of the interpretation of their associated parameter in the model.

- *arc* or *density* parameter: This parameter reflects the propensity for arcs to occur in the network and is therefore affected by the *density* of the network, i.e., the proportion of observed to possible arcs. Obviously, the more arcs in a network the higher the chance there is of observing, for instance, a reciprocated arc, so in order to assess whether there is a distinct tendency towards reciprocation, there is a need to recognize the impact of density. In that sense, this parameter acts as a baseline effect, analogous to an intercept term in a regression.
- *reciprocation* parameter: This parameter reflects the extent of reciprocation in the network, as discussed above.
- *Activity dispersion* parameter (*k-outstars*): A *k-outstar* is a configuration where a single node expresses *k* arcs to other nodes. Here *k* may range from 0 to $n – 1$, where *n* is the number of nodes in the network. In developing their new specifications, Snijders et al (2006) combined these configurations for different *k* into one statistic by calculating a geometrically weighted sum of configuration counts across *k*, motivated by both parsimony and technical reasons (for details, see Snijders et al, 2006; Handcock & Humter, 2006). A positive parameter estimate indicates that some people in the network have more extensive activity (in terms of expressing arcs) than others, whereas a negative parameter estimate indicates greater homogeneity in activity levels, so that there are few, if any, highly active people, given other effects in the model. (It should be noted, though, that differentially high levels of activity may principally occur within

regions of triangulation, in which case the activity parameter may well be negative in the presence of a positive triangle parameter.)[3]

- *popularity dispersion* parameter (*k-instars*): This is the counterpart of the activity dispersion parameter, except that it relates to structures in which a node attracts multiple network arcs rather than expresses them. In that sense, the parameter relates to dispersion in network popularity in an exactly analogous way that the previous parameter relates to dispersion in network activity[4].
- *multiple transitivity* parameter (*k-triangles*): A *k*-triangle configuration comprises *k* transitive triads, each sharing a single arc as the *base* of the *k*-triangle, as depicted in Figure 2. Counts of *k*-triangles are incorporated into the one statistic in the same way as for *k*-outstars and *k*-instars. A positive parameter is evidence not just for transitivity in the network, but for the formation of denser regions of triangulation built up from multiple transitive triangles[5].

Snijders et al (2006) and Robins, Pattison and Wang (2007) proposed additional parameters that could be included in models for directed social networks. The five we list here are what we would regard as the minimum required for a reasonable representation of network structure for much social network data. Unless a model contains effects for reciprocation, some form of control over degree distributions (i.e., variations in activity and popularity) and a means to capture levels of transitivity, it is unlikely to yield an adequate representation of commonly observed social network data. Multiple transitivity is important because there is compelling empirical evidence that many human social systems exhibit extensive transitivity or *clustering* (for discussion see Snijders et al, 2006). So, while we do not claim that these five parameters are necessarily optimal for modeling all networks, if a researcher is interested in examining individual-level variables within a social structure (as we are), then these five parameters are a reasonable control on the dependencies that might occur within that structure. As a result, we argue that if a model includes individual-level effects in addition to these five parameters and if individual-level parameter estimates appear substantial over and above these five structural effects, then this constitutes evidence for a distinct contribution to the data from the individual-level measures. Moreover, if parameter estimates for reciprocity, activity, popularity and transitivity are substantial in the model, then there is evidence

---

[3] More technically, this parameter is related to the outdegree distribution – see Hunter, 2007, for an extended discussion of different versions of these degree-based parameters (albeit for nondirected graphs).

[4] This parameter controls for the indegree distribution.

[5] The new specifications incorporate an additional parameter λ that specifies the weighting on the sums for the *k*-star and –triangle effects. Hunter and Handcock (2006) show how this parameter may be estimated. In this article we follow the recommendations of Snijders et al (2006) and Robins, Snijders, et al (2007) and set λ=2.

for important endogenous effects that cannot be explained by the individual-level measures, underlining the importance of assessing individual-level effects in the presence of these endogenous effects and indicating that independence among individual-level observations cannot be assumed.

The mathematical form of the model

We present briefly the mathematical form of model. For each pair *i* and *j* of a set *N* of *n* actors, X_{ij} denotes a network tie variable with $X_{ij} = 1$ if there is a network tie from *i* to *j*, and $X_{ij} = 0$ otherwise (usually – but not necessarily – with self-ties excluded, i.e. with X_{ii}.undefined.) Realized values of X_{ij} are denoted by x_{ij} with **X** the matrix of all variables and **x** the matrix of realized ties. The form of the exponential random graph models with the parameters above is as follows:

$$\Pr(\mathbf{X} = \mathbf{x}) = (1/\kappa) \exp\{ \theta_d z_d + \theta_r z_r + \theta_a z_a + \theta_p z_p + \theta_t z_t \} \qquad (1)$$

where:

(i) $\theta_d, \theta_r, \theta_a, \theta_p$ and θ_t are parameters corresponding to density, reciprocity, activity dispersion, popularity dispersion and multiple transitivity effects;

(ii) $z_d(\mathbf{x}), z_r(\mathbf{x}), z_a(\mathbf{x}), z_p(\mathbf{x})$ and $z_t(\mathbf{x})$ are the *network statistics* pertaining to the corresponding parameters; these are either direct counts of the relevant configuration in the observed data, or, for the sake of parsimony, combinations of direct counts (see Snijders et al, 2006, for the detailed calculation of these statistics);

(iii) $\kappa = \kappa(\theta_d, \theta_r, \theta_a, \theta_p, \theta_t)$ is a normalizing quantity to ensure that (1) is a proper probability distribution.

The model represents a probability distribution of graphs on a fixed node set, where the probability of observing a graph is dependent on the presence of the various configurations expressed by the model. One can interpret the structure of a typical graph in this distribution as the result of a cumulation of these particular configurations.

It is worth emphasizing that these are models for the network **x**, comprising $n(n-1)$ possible network ties. In the example below, we have a (directed) network of 38 individuals, implying 38 × 37 = 1406 observations on the presence or absence of ties.

Models for network data and individual-level variables

There are a variety of ways to incorporate individual-level variables into this model, but in this article we incorporate some simple dyadic *social selection* parameters (Robins, Elliott & Pattison, 2001). In social network terminology, individual-level variables are often referred to as measures on the nodes, or as node *attributes.* Social selection processes occur when network relationships are formed as a result of node attributes. For instance, *homophily* is a process whereby individuals who are similar to one another form a relationship (McPherson, Smith-Lovin & Cook, 2001).

With attributes included, the form of equation (1) is not greatly changed, except that in a social selection model, the attributes are assumed to be exogenous, so that we have a probability form for **X**, conditional on a vector of attributes **Y**.

In this article, we use both binary and quantitative measures on the nodes. Information about a binary attribute[6] can be incorporated into visual representations of networks and network configurations by coloring the nodes whereas information about a quantitative attribute can be incorporated using node size. For each type of measure, we present five configurations that may be incorporated into the model, as depicted in Figure 3. As is shown in the Figure, these configurations can be seen as various interactions between attribute variables and the network variables X_{ij}. If we denote by Y_i a binary attribute measure on node *i*, then the network statistic for an attribute-based activity effect is $\sum_{i,j} y_i x_{ij}$. The network statistic for reciprocated attribute homophily is $\sum_{i,j} y_i y_j x_{ij} x_{ji}$. Large positive parameter estimates for parameters corresponding to these configurations provide evidence that individuals who have an attribute score of 1 (i.e. who "have" the attribute) are more active in the network (often called a *sender effect* because the "sender" of the tie has the attribute) and are more likely to have reciprocated relationships with each other, respectively. Because the statistics are in the form of multiple interactions, the latter effect can be inferred as over and above the former, that is, individuals with the attribute are more likely to have reciprocated relationships with each other, even taking into account the fact that they are more active in the network. Similarly, parameters include an attribute-based popularity effect, or *receiver effect*, indicating that people with the attribute tend to be more popular by receiving more ties.

The quantitative attribute statistics are similar, although because of interest in homophily here we use absolute differences between attribute values for pairs of nodes. Homophily would then be evidenced by a strong negative parameter estimate. The corresponding statistic is $\sum_{i,j} \left| y_i - y_j \right| x_{ij} x_{ji}$ where X_i denotes a quantitative variable in this case[7]. The reciprocated activity parameter, based on the sum of attribute values, corresponds to the statistic $\sum_{i,j} (y_i + y_j) x_{ij} x_{ij}$. To incorporate these effects in the model, the attribute-based statistics, weighted by parameters for each effect, are included as additional terms in the summation within the exponential in (1).

We are not suggesting that all five parameters must be included for every attribute, although we take this approach in the application presented below. For instance, sometimes reciprocity may

---

[6] We require the binary variable to be an indicator variable, taking values 0 and 1.

[7] Mathematically simpler multiplicative interactions could be used in direct analogue to the binary case, but the natural interpretation regarding homophily makes the absolute difference statistic appealing.

not be of particular relevance to the attribute under study, so that reciprocated activity and homophily may be excluded from models in the interests of parsimony. We have found that it is usually important to include activity and popularity parameters as attribute-based main effects. Strongly positive activity and popularity parameters alone are evidence for a form of homophily since individuals with the attribute both send and receive more arcs, and hence the probability of an arc from a person with the attribute to another person with the attribute is enhanced. What we cannot ascertain, though, from these two parameters alone is whether there are distinct homophily effects over and above activity and popularity. This may or may not be important to an investigation. For clarity of interpretation, it is also often helpful to include any lower order interactions to higher order interactions included in the model.

Figure 3 about here.

Estimation of parameters; model fit

Monte Carlo Markov Chain Maximum Likelihood Estimation (MCMCMLE) methods have been developed to obtain estimates of parameters and standard errors for exponential random graph models (see Hunter & Handcock, 2006; Snijders, 2002). Robins, Snidjers et al (2007) reviewed the three programs (*siena*, *statnet* and *pnet*) that are currently available for estimation. The results presented here were obtained using *pnet*.

Following Snijders (2002) and the *siena* program, *pnet* implements MCMCMLE and involves updating parameter estimates from simulations. The process converges, in principle, on a set of parameter estimates with the property that the random graph distribution simulated from the estimates has average values of the statistics that are very close to the observed values. Failure to converge on such estimates may indicate an inadequacy in the model for the particular data set. Changing the structural effects in the model may improve the prospect of convergence. Convergence *t* statistics indicate whether convergence has been achieved for each parameter: the statistic locates the observed value of the graph statistic in the distribution of graph statistics corresponding to the parameter estimates. A statistic of less than 0.1 is regarded as indicating good convergence. Issues of convergence (and the wider but associated issue of model degeneracy) involve matters of technical detail that have been discussed elsewhere (Handcock, 2003; Snijders, 2002; Snijders et al, 2006; Robins, Snijders et al, 2007).

Empirical example: Attitudes towards women within an elite all-male sporting team

As an illustrative example, we model data from a larger study of attitudes towards women among professional male athletes from elite sporting teams. This study arose because of a number of widely publicized incidents involving athletes in their after-hours activities, incidents that often centered around negative behaviors towards women or aggressive behaviors among males. Club administrations and the sport's administering authorities were concerned about the effects of these incidents on the reputations of clubs and of the sport itself. The research aim was to investigate whether attitudes towards women – or more generally, attitudes about masculinity – were in some way associated with the social structure of the team. Theoretical arguments suggest the possibility of an association between attitudes about masculinity and social network structure among certain groups of males (Lusher & Robins, in press).

We present results from one team of 38 male athletes. These team members completed a questionnaire as a group after a training session. Members were asked to nominate other players for several network based questions. Here we model the data from the "socializing" network, already presented in Figure 1. For this item, members were asked to nominate other members with whom they socialized after hours. In Figure 1, an arc indicates that the sender of the arc nominated the receiver. As such, this network data is about choices by the respondents and not about whether the socializing actually occurred. Hence there is no need to ensure agreement between pairs of individuals about whether they socialized together.

In a separate question, members were also asked to nominate those whom they considered to be the best players in the team (they could freely nominate as many best players as they wished). For each member, the summed nominations by other members was taken as a quantitative attribute of being one of the best players in the team. Team experience in terms of the number of games played was also used as a quantitative attribute. As a measure of attitudes toward masculinity, we used the Masculine Attitudes Index (MAI) which is an inventory to measure male dominance beliefs incorporating facets of anti-femininity, gay-male homophobia, violence and manliness as sexual success (Lusher, Robins & Dudgeon, 2007). In other analyses, we have used the MAI as a quantitative measure, but for the illustrative purposes of this article we have categorized the scores into high and low categories based on a median split. The high MAI category comprises team members with "more dominative attitudes" (i.e. more anti-feminine, more homophobic, etc…) and is assigned a score of 1 on the binary variable.

<u>Competing explanations for the network data</u>

Figure 4 depicts the socializing network with colors on the nodes representing the high and low MAI categories. It can be seen immediately from the figure that the circled subset of athletes all have more dominative attitudes and form a tightly-knit group whose members socialize together. It might be concluded that the presence of this group in the network constitutes a risk for this particular sporting club. In social situations involving this group, there would seem to be a consensus about attitudes that might exacerbate rather than ameliorate the possibility for an unfortunate incident to occur.

Figure 4 about here.

Yet it is obviously important not to reach conclusions without considering alternative explanations. Other structural or attribute processes could be a reasonable explanation for this network, and the apparent relevance of masculine attitudes may merely be epiphenomenal of these other effects. Figure 5 depicts the network with the other attributes, experience and best player, represented by node size. It can be seen that both these attributes may have a role to play in explaining how this network came into being. There appear to be tendencies for experienced and best players, respectively, to socialize together.

Figure 5 about here.

Finally, there are important structural features, discussed above, that may shape human social networks. We know that most human social networks exhibit reciprocity and regions of multiple transitivity. From Figure 6, it can be seen that these features appear to be present in this network as well. As well, activity and popularity effects in human networks are commonly differentiated across nodes, and it is important to control for this diversity. Incorporation of activity and popularity dispersion effects in the model may be seen as controlling for other latent attributes that, although unmeasured, may be responsible for some individuals being more or less active or popular.

Figure 6 about here.

Accordingly, we fit models with parameters relating to configurations from Figures 2 and 3. We denote as significant those parameters with a ratio of estimate to standard error in excess of two in absolute value (Snijders et al, 2006). We have evidence for an association between masculine attitudes and social structure if any of the attribute parameters involving the MAI are significant, when controlling for the other attribute and structural effects.

<u>Models for the data</u>

We have fitted a number of models to this data. We begin by noting that the best player effects are not significant and do not seem to change the estimates of other parameters greatly. So for the sake of simplicity we present here a model without best player parameters, but involving all the structural effects in Figure 2, and from Figure 3, the effects for experience and MAI category. Convergence was obtained for all parameters. Parameter estimates and standard errors are presented in Table 1.

Table 1 about here.

We provide a brief interpretation for the different parameter estimates in turn. The density and reciprocity effects are not significant. Although it is evident in Figure 1 that many ties in the network are reciprocated, the extent of reciprocity is modeled as a consequence of other effects described below. We see significantly negative effects for activity and popularity dispersion (while the attribute-based activity and popularity parameters are not significant). These effects imply that there is little variation in activity or popularity among these athletes, unless it occurs as a result of other effects. The significant positive multiple transitivity effect is expected given that it is such a common feature of human networks pertaining to social affiliation. The estimate indicates that there are regions of triangulation in the network, even after other modeled effects have been taken into account. In other words, the measured individual-level attributes alone are unable to explain all of the structural effects, suggesting that there may be important endogenous dependencies in the network.

For the experience effects, it is important to take account of the scale of measurement, since experience can range from 0 to hundreds of games played. We see two significant effects, both related to homophily. The non-reciprocated parameter estimate is negative indicating that two players with a small absolute difference in games played are more likely to have an arc between them. In other words, players with similar experience are more likely to be linked than players of different experience. There is also an interesting positive reciprocated homophily effect. Of course, if players have the same experience the absolute difference term is zero, so that the statistic here will be zero. Accordingly, this effect tells us something about players with different levels of experience. The overall conclusion from the two effects is that players of different experience are less likely to be connected, but the effect is reduced for reciprocated arcs. If we suppose that reciprocated arcs represent stronger social connections, then the effect suggests that the processes underpinning strong tie formation are less sensitive to differences in experience than the processes leading to weaker social connections[8].

There is also a significant homophily effect for MAI. Players with more dominative attitudes are more likely to have connections. This effect is over and above experience homophily and the formation of regions of multiple triangulation.

Discussion of results

The results show that in this sporting team, even in the presence of substantial structural effects, player attitudes towards masculinity are important in determining who socializes with whom. We also infer that experience is important to patterns of socializing, but not best players, and that there are strong multiple transitivity effects over and above the effects related to the measured attributes. Our conclusion is that masculine attitudes are one of at least three organizing principles for this network: broadly, that players tend to socialize with players of similar experience, that they tend to socialize in groups (multiple transitivity effects), and that they tend to socialize with players with similar masculine attitudes. That masculine attitudes are important as a basis of social choice may reflect a particular team culture that could bring with it risks for the club at two levels: first, impact on team harmony/team dynamics and, secondly, impacts at the Club level in terms of sponsorship, publicity and membership support in the event that dominative attitudes result in negative incidents.

Conclusions

This paper illustrates one method for analyzing individual-level constructs within a social structure. Such methods are always going to be more complicated than standard general linear model techniques because they necessarily implicate complex dependencies in the data. It is not a simple matter to untangle important organizing principles that might help explain a complex social system. Nevertheless, the broad modeling approach presented here provides a principled approach for doing so and the advent of new software for estimation, simulation and goodness of fit makes such approaches accessible.

We emphasize that the model we have presented is a social selection model where it is supposed that the attributes are important to the formation of network ties. It is also possible that social influence effects may apply, whereby network ties may lead to similar attitudes (Friedkin, 1998; Robins, Pattison & Elliott, 2001). With cross-sectional data, as we have here, it is not

---

[8] We have also fitted models with a variable related to age rather than experience in terms of games played. Our conclusion is that experience provides the stronger effects, although the two variables are reasonably correlated. Accordingly, we present the experience rather than the age results here.

possible to say whether one or the other (or both) process applies, so we make no attributions of causality. What a social selection model does permit, however, is inferences about associations between network relationships and individual attributes. Recent methods for longitudinal data enable differentiation between social selection and social influence processes (Snijders, Steglich & Schweinberger, 2007).

There is often a research gap between social psychology and social network analysis, both theoretically and empirically (Robins & Kashima, in press). There is too little work on how relevant psychological variables may interact with network structural effects within networked social systems. Our own experience in exploring these matters is that, as is the case in our example here, psychological attributes never fully explain network structures, that network structural effects alone never fully explain the social patterning of people with certain psychological predispositions, and indeed that the two interact with one another. The full story is not complete with just individual-level attributes, nor just relational structure. Both are needed.

References

Abbot, A. (1997). Of time and space: The contemporary relevance of the Chicago School. *Social Forces, 75*, 1149-1182.

Asch, S.E. (1952). *Social psychology*. Englewood Cliffs, NJ: Prentice-Hall.

Bryk, A.S., & Raudenbush, S.W. (1992). Hierarchical linear models: Applications and data analysis methods. Newbury Park, CA: Sage.

Contractor, N., Wasserman, S. & Faust, K. (2006). Testing multi-theoretical multilevel hypotheses about organizational networks: An analytic framework and empirical example. *Academy of Management Review, 31*, 681-703.

Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, *81*, 832-842.

Friedkin, N. 1998. *A structural theory of social influence*. New York: Cambridge University Press.

Goodreau, S. (2007). Advances in Exponential Random Graph (*p**) Models Applied to a Large Social Network. *Social Networks, 29*, 213-248.

Handcock, M.S. (2003). Statistical models for social networks: Degeneracy and inference. In Breiger, R., Carley, K., & Pattison, P. (eds.). *Dynamic social network modeling and analysis* (pp. 229-240). Washington DC: National Academies Press.

Hunter, D.R. (2007). Curved exponential family models for social networks. *Social Networks, 29*, 216-230

Hunter, D. & Handcock, M. (2006). Inference in curved exponential family models for networks. *Journal of Computational and Graphical Statistics,* 15, 565-583.

Klein, K. & Kozlowski, S. (2000). *Multi-level theory, research and methods in organizations: Foundations, extensions, and new directions.* San Francisco: Jossey-Bass.

Krackhardt, D. (1988). Predicting with networks: nonparametric regression analysis of dyadic data. *Social Networks*, *10,* 359-381.

Lusher, D., & Robins, G.L. (in press). Hegemonic and other masculinities in local social contexts. *Men and Masculinities.*

Lusher, D., Robins, G.L., & Dudgeon, P. (2007). The masculine attitudes index (MAI): A scale for the measurement of male dominance beliefs in adolescent and adult contexts. Working Paper, University of Melbourne.

McPherson, M., Smith-Lovin, L., & Cook, J. (2001). Birds of a feather: homophily in social networks. *Annual Review of Sociology, 27,* 415-444.

Monge, P., & Contractor, N. (2003) *Theories of communication networks*. New York: Oxford University Press.

Moreno, J. L., & Jennings, H. H. (1938). Statistics of social configurations. *Sociometry, 1*, 342-374.

Pattison, P. E., & Robins, G. L. (2002). Neighbourhood-based models for social networks. *Sociological Methodology*, *32,* 300-337.

Raudenbush, S., & Sampson, R. (1999). 'Ecometrics': Toward a science of assessing ecological settings, with application to the systematic social observation of neighbourhoods. *Sociological Methodology*, *29,* 1-41.

Robins, G.L., Elliott, P., & Pattison, P. (2001). Network models for social selection processes. *Social Networks, 23*, 1-30.

Robins, G.L., & Kashima, Y. (in press). Social psychology and social networks. *Asian Journal of Social Psychology.*

Robins, G.L., & Pattison, P. E. (2005). Interdependencies and social processes: Generalized dependence structures. In Carrington, P., Scott, J., & Wasserman, S. (Eds) *Models and Methods in Social Network Analysis* (pp.192-214). Cambridge University Press.

Robins, G.L., Pattison, P, & Elliott, P. (2001). Network models for social influence processes. *Psychometrika, 66,* 161-190.

Robins, G.L., Pattison, P.E., & Wang, P. (2007). Closure, connectivity and degrees: New specifications for exponential random graph (*p**) models for directed social networks. *Social Networks.* Under review.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (*p**) models for social networks. *Social Networks, 29,* 173-191.

Robins, G.L., Snijders, T.A.B., Wang, P., Handcock, M., & Pattison,P. (2007). Recent developments in exponential random graph (*p**) models for social networks. *Social Networks, 29*, 192-215.

Snijders, T.A.B. (2002). Markov chain Monte Carlo estimation of exponential random graph models*. Journal of Social Structure, 3*, 2.

Snijders, T.A.B., & Bosker, R. (1999). Multilevel analysis: An introduction to basic and advanced multilevel modeling. London: Sage.

Snijders T.A.B., Steglich C., & Schweinberger, M. (2007) Modeling the co-evolution of networks and behavior. In: van Montfort, K., Oud, H., & Satorra A (Eds), *Longitudinal models in the behavioral and related sciences*. Hillsdale, NJ: Lawrence Erlbaum pp.41-71.

Snijders, T.A.B., Pattison, P., Robins, G.L., & Handcock, M. (2006). New specifications for exponential random graph models. *Sociological Methodology, 36*, 99-153.

Tajfel, H., Billig, M.G., Bundy, R.P., & Flament, C. (1971). Social categorization and intergroup behavior. *European Journal of Social Psychology*, *1*, 149-178.

Wasserman, S., & Pattison, P. E. (1996). Logit models and logistic regressions for social networks: I. An introduction to Markov graphs and *p**. *Psychometrika*, *61*, 401-425.

Wasserman, S., & Robins, G. L. (2005). An introduction to random graphs, dependence graphs, and p*. In P. Carrington, J. Scott & S. Wasserman (Eds.) *Models and methods in social network analysis* (pp.148-161). New York: Cambridge University Press.

Weick, K.E., & Roberts, K.H. (1993). Collective mind in organizations: heedful interrelating on flight decks. *Administrative Science Quarterly*, *38,* 357-381.

Wellman, B. (1988). Structural analysis: from metaphor to substance. In B. Wellman & S.D. Berkowitz (Eds.), *Social structures: a network approach* (pp.19-61). Cambridge, UK: Cambridge University Press.

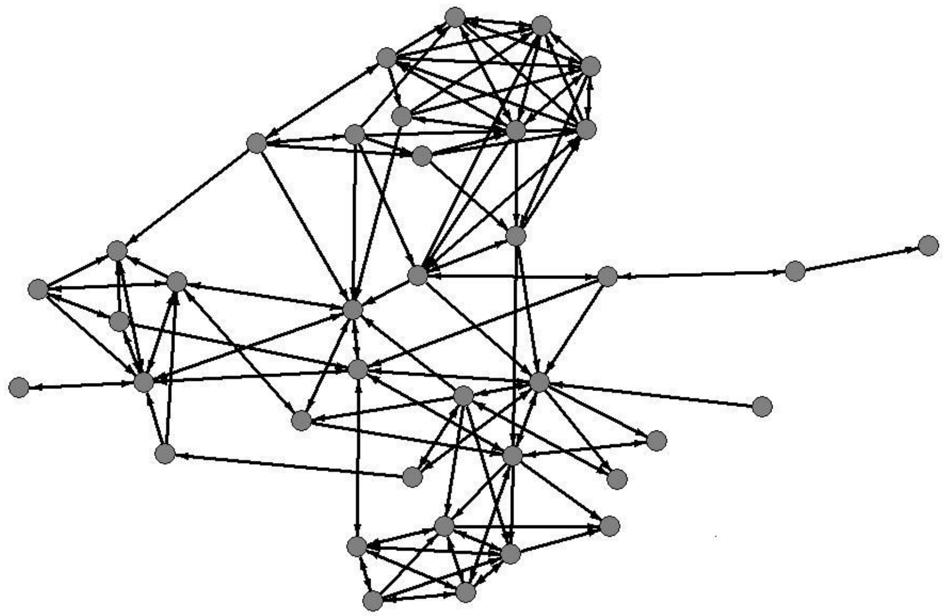

Figure 1

A social network:

After-hours socializing among members of an all-male elite sporting team

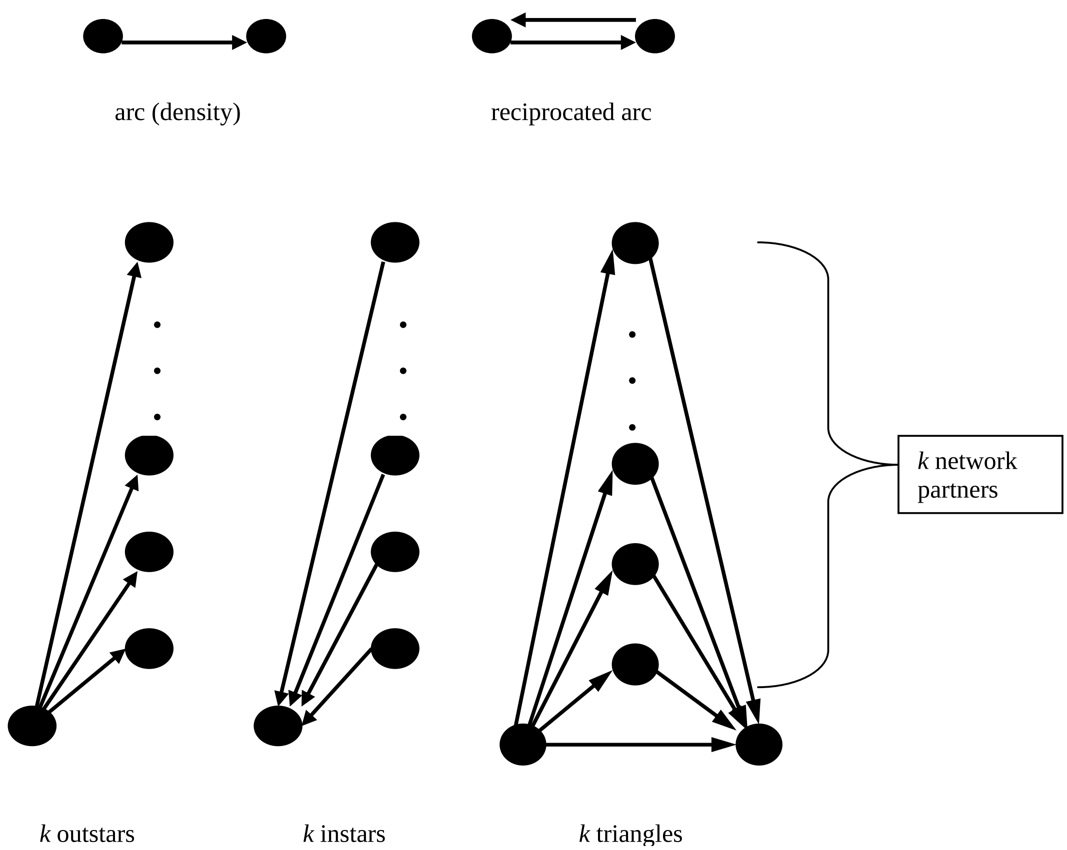

Figure 2

Some configurations for a directed graph

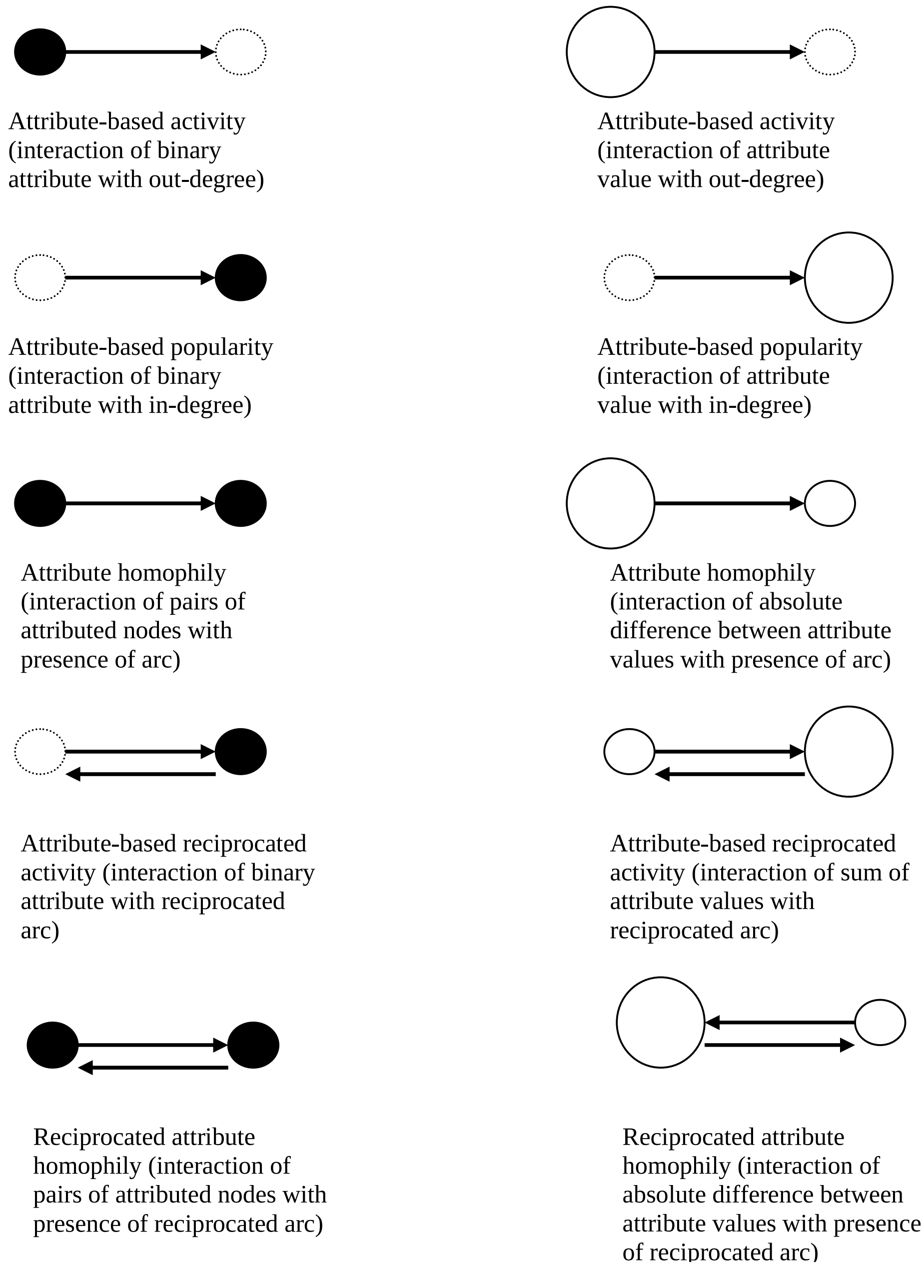

Figure 3

Configurations including attribute variables.

(Binary attributes are represented in left column; quantitative in right column. A dotted border indicates a node irrespective of attribute value; a black node indicates binary attribute value of 1; the size of node for quantitative attributes indicates value of the attribute measure. Attribute-based activity and popularity effects are also called sender and receiver effects.)

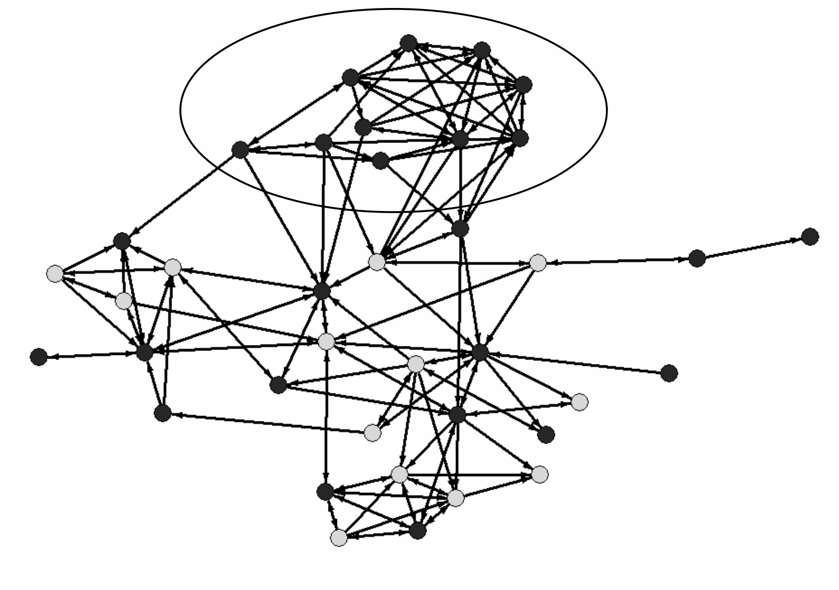

Figure 4

The socializing network of figure 1 with colors on the nodes to indicate masculine attitudes

(Black nodes indicate more dominative masculine attitudes; grey nodes indicate less dominative attitudes; a tightly-knit socializing group of athletes with high MAI scores is circled.)

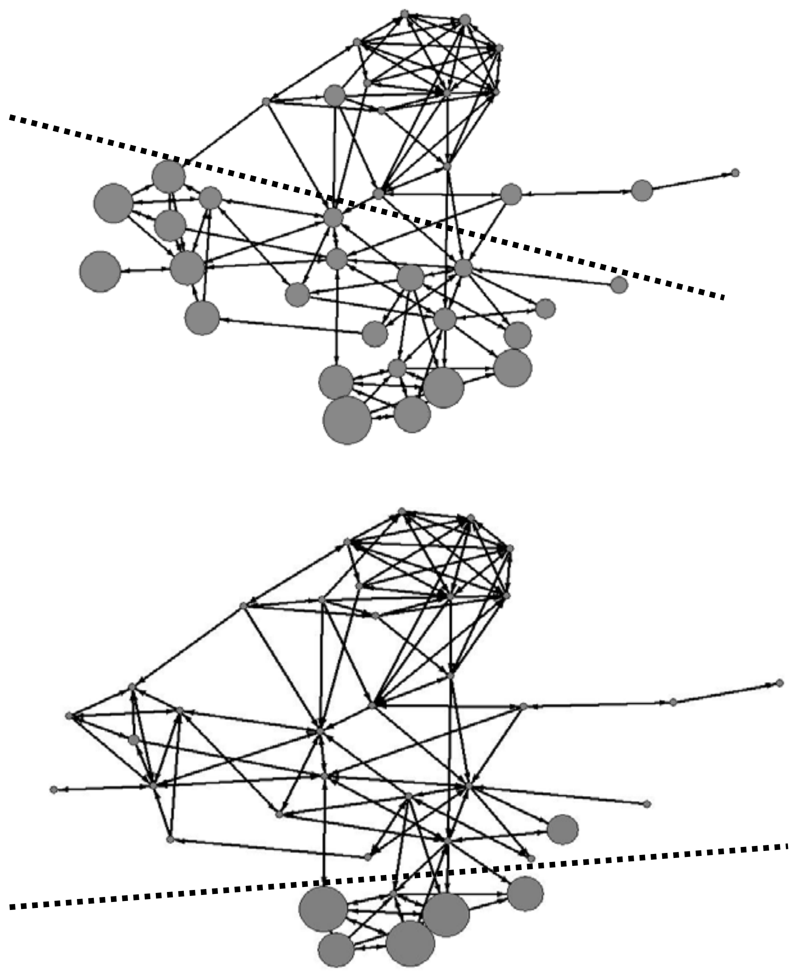

Figure 5

The socializing network with other attribute effects represented

(In top panel, larger nodes represent more experienced players; in bottom panel, larger nodes represent best players; in both panels the dotted lines represent how these attributes approximately partition the network.)

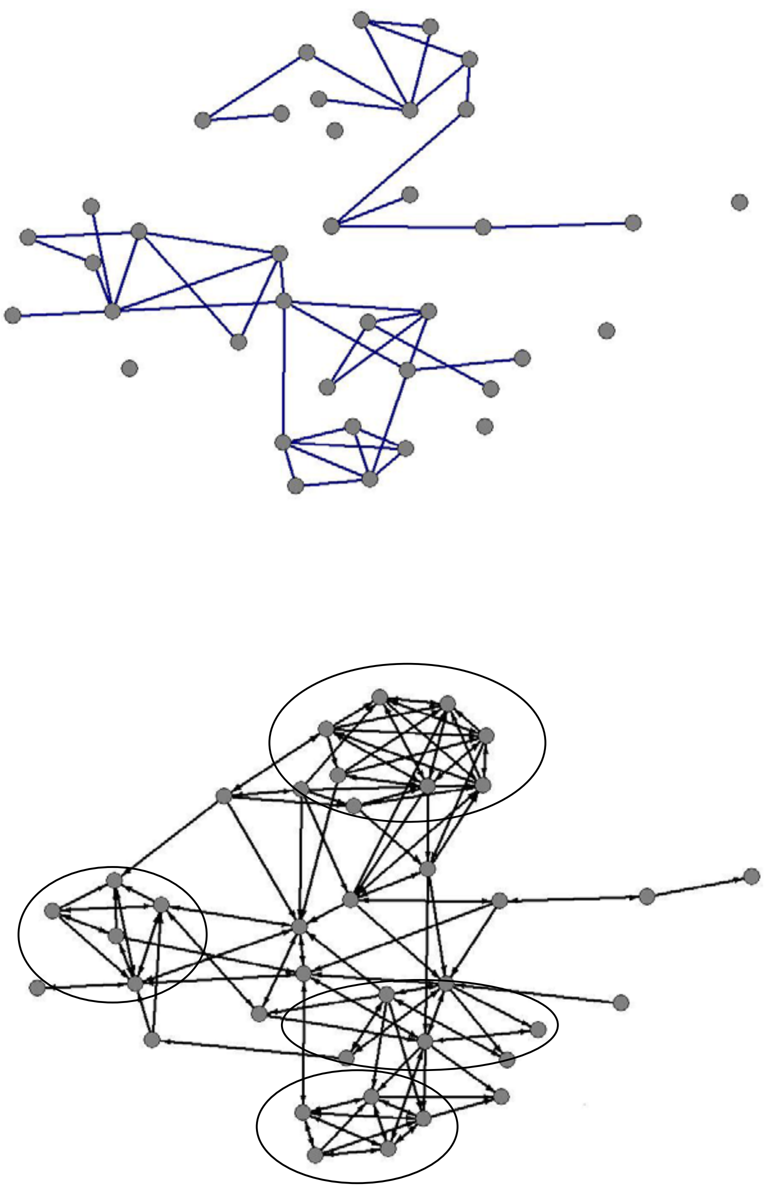

Figure 6

The socializing network with structural features highlighted.

(Top panel depicts mutual edges based on the reciprocated arcs in the network; circles in the bottom panel indicate some regions of the network with multiple triangulation and greater density.)

Parameter	Estimate	Standard error
	Structural effects	
Density	– 0.55	0.62
Reciprocity	0.79	0.57
Activity	– 0.88	0.34 *
Popularity	– 0.70	0.32 *
Triangulation	0.97	0.13 *
	Attribute effects: Experience	
Activity	– 0.001	0.003
Popularity	0.001	0.003
Homophily	– 0.014	0.004 *
Reciprocated activity	0.003	0.003
Reciprocated homophily	0.018	0.009 *
	Attribute effects: MAI	
Activity	– 0.26	0.37
Popularity	– 0.83	0.43
Homophily	1.30	0.64 *
Reciprocated activity	1.22	0.72
Reciprocated homophily	– 1.63	1.00

Table 1

Final model

(asterisks indicate effects where absolute value of estimates exceed twice the standard error.)