

Small and Other Worlds:

Global network structures from local processes

Garry Robins
Philippa Pattison
Jodie Woolcock
Department of Psychology
University of Melbourne
Australia

American Journal of Sociology
To appear

An early version of this paper was presented at the Sunbelt XXII International Social Network Conference, New Orleans, February 2002. The authors would like to thank Tom Snijders, Andrew Seary and Mark Handcock for helpful suggestions in regard to this research, and for the helpful comments of two reviewers. This research was conducted with support from the Australian Research Council, and was supported in part by grants from the NIH (R01-DA012831 and R01-HD041877).

Abstract

Using a simulation approach based on the Metropolis algorithm, we contrast broad global features of network structure – in particular, small world properties – with the local patterning that could generate the network. It is not difficult to infer local structures emerging from certain simple social processes but, as these localized patterns agglomerate, the global outcomes are often not apparent. In such cases, computational techniques are necessary because analytic solutions are simply not available. In this paper, we show how to simulate a distribution of Markov random graphs based on assumptions about simple local social processes. We examine the resulting global structures by comparison with an appropriate Bernoulli distribution of graphs and provide examples of various stochastic global “worlds” that may result, including small worlds, long path worlds and dense non-clustered worlds with many four-cycles. In the light of these results we suggest a locally-specified social process that may result in small-world global properties. In examining the movement from structure to randomness, we show how parameter scaling relates to a phase transition occurring at a certain scaling (“temperature”) so that a regular structure “melts” into a stochastically-based counterpart. We provide examples of “frozen” deterministic structures, including highly clustered “caveman” graphs, bipartite structures, and global cyclic structures involving structurally equivalent groups.

1. INTRODUCTION

It matters that networks differ. To justify this claim, we need look no further than the compelling description of the rise of the Medici by Padgett and Ansell (1993). They argue that Cosimo de Medici came to power in fifteenth century Florence in large part because the Medici family was at the centre of a star-like structure of marriage and business alliances, a patterning of inter-familial relations that could be efficiently activated by – and only by – the Medici. In contrast, the dense interconnections among Cosimo’s political opponents, the traditional oligarchic families of Florence, hindered a decisive response. In the crucial showdown, the oligarchic opposition to Cosimo splintered into confusion and mistrust. Because of their dense pattern of ties, the oligarchs knew who should take the field, so absences were treated as defection. The Medici clients knew too little to feel betrayed. They stayed; the oligarchs fled.

1.1 Global network structure and local network processes

So, networks differ, but in what sense? It is the global structure of the Florentine network that counts in the description above. The argument rests on the patterning of relations among all families. By *global structure* we refer to features that can only be determined by examining the entire network (or at least large parts of it.) In the Florentine families network, the juxtaposition of large patterns – a star-like structure adjacent to a dense clique-like pattern – cannot necessarily be surmised by considering individual nodes or their immediate network neighborhoods, i.e. more *local* sub-networks.

In this paper, we contrast such broad global features with the local patterning that could give rise to them. Our representation of local structure relies on small network *configurations* that we describe below as subgraphs involving only a few network ties. We construe these configurations as the outcomes of *local social processes*. Network ties emerge, persist and disappear by virtue of actions made locally at the scale of the individual actors in a network (whether they be persons, or families, or companies, or some other social entity.) Actors do not usually cast their gaze across the entire network, possibly because in most cases they can only “see” what is in their local social neighborhood (Pattison & Robins, 2002). On the basis of their localized view, they form strategies and make decisions that intersect with those others who are socially proximate. Combinations of these competing or complementary intentions and actions constitute social processes that see the formation of local patterns of relationships. These local patterns agglomerate to create the global structure. The strength of Padgett and Ansell’s account of the Medici lies in showing how the localized processes – involving individual marriage and business partnerships within and across neighborhoods of medieval Florence – led to the global features that mattered.

1.2 The role of computation and simulation

Because we can construe global structure as an agglomeration of local patterns, an understanding of the global can be greatly enhanced by computational and simulation techniques. It is not inherently difficult to infer the local structural outcome of a simple postulated local social process (e.g. as discussed below, a tendency to *structural* balance leads to localized triadic structures.) What is not apparent, however, are the global outcomes when several of these localized patterns combine, perhaps with different strengths. Indeed, except in a small number of cases (examples are discussed below),

computational techniques are necessary because analytic solutions are simply not available.

The advent of computer technology enables us to traverse this conceptual and methodological gap from individual local patterns to the possibility of various global phenomena. This is potentially a two-way process. We observe certain global properties of interest in a network (e.g. small world properties as discussed below). We can then speculate how these properties might emerge from localized social processes, develop a model embodying such processes, and then use simulation procedures to investigate the global outcomes of this model. If the results mirror the global properties of interest, we can then check whether some or all of the postulated local processes are operative in the observed network.

In other words, we do not see observation, modeling and simulation as inherently different approaches towards understanding, but rather a potentially powerful combination of tools that readily complement each other. In this paper, we concentrate on the results of simulating from some simple but plausible models, but we are mindful of the need to subject these models to stringent empirical testing. For instance, one step that is needed to go beyond this paper is to show that the processes inherent in the models are actually observed in empirical networks.¹

1.3 Which global properties?

Despite the possibility that local patternings shape global properties of the network, it is not always clear which global features are best to examine, or indeed simply how best to describe global network structure. Sometimes, as in the Medici example, important global qualities seem apparent²; but if we consider the range of networks that Faust and Skvoretz (2002) attempted to compare, it is not at all obvious what the optimal criteria are. Even so, the recent flurry of activity on small world networks, following the seminal work of Duncan Watts (Watts & Strogatz, 1998; Watts, 1999a, 1999b), illustrates just how much can be gained by investigating one or two global features that seem of general relevance, in this case the prevalence of short paths in networks in the presence of some randomness. *Average path length* is clearly a global feature of a network. What local processes could make it “short”?

Following Watts’s innovative approach, there has been an upsurge of exciting research into global network phenomena with a focus on properties (described in more detail below) such as path length, clustering and degree distribution (e.g., Amaral, Scala, Barthélémy, & Stanley, 2000; Barrat & Weigt, 2000; Barthélémy & Amaral, 1999; Bohland & Minai, 2001; Comellas, Ozon & Peters, 2000; Kirillova, 2001; Mathias & Gopal, 2001; Newman, 2000; Newman & Watts, 1999; Pandit & Amritkar, 2001; Pastor-Satorras & Vespignani, 2001; Yang, 2001). Much of this research concentrates on large-scale networks, including growing networks such as the internet. One advantage of considering large networks with indeterminate numbers of nodes is that asymptotic results may sometimes be available analytically, at least for simpler models. Strogatz (2001), Albert and Barabási (2002) and Newman (2003) provide excellent reviews of this burgeoning literature.

¹ For at least some observed networks, we have evidence that the models described below are indeed plausible descriptions, not just on the basis of measures of fit, but also because important global properties can be reproduced (Robins, 2003).

² at least, once Padgett and Ansell (1993) have done all the hard exploratory data analysis to reveal the important structural features.

1.4 Locally specified network models

Yet, despite its innovation and rigour, this literature is strongly centered on global network features with at best passing attention to the local processes that might generate most real networks. Few of the models can be locally specified. The original model of Watts (1999a) *starts* with an exogenous global structure (a large circle through all nodes, termed by Watts a *cyclic substrate* and also known more technically as a *one dimensional lattice*) and then introduces some local processes that add to that structure (random edge addition combined with a tendency towards clustering, or triangle formation). In the short time since its introduction, this basic model has been used and elaborated by many others, clearly constituting an important new approach to understanding global features. But with an exogenous global structure as its starting point, it is not locally specified, and so cannot tell us how small worlds might emerge in real networks³.

Another well-known recent model is that of preferential attachment (Barabási & Albert, 1999), where for a growing network a new node is attached to existing nodes with a probability dependent on the degree distribution of existing nodes. This model can be shown to result in a power-law degree distribution with resulting *scale-free* properties. Here we have more of the flavor of a locally specified model, but to determine the probabilities of attachment the new node has to “know” the existing degree distribution across the entire graph. This model has considerable strengths in describing the structure of the internet, which is indeed an expanding network. But we disagree with the claim by Albert and Barabási (2002) that most real world networks describe open systems that grow by the continuous addition of new nodes. Such a claim ignores the large amount of empirical work in a range of literatures – from the anthropological to the organizational – that searches for an understanding of small scale, relatively closed network systems. For instance, the membership of the Florentine elite was relatively stable, in comparison to the changes in their inter-relationships resulting from ongoing social and economic processes.

Of course, simple random graph models, arising originally from the work of Erdős and Renyi (1959 - see also Gilbert, 1959), are much closer to local specification. Two simple random graph models have similar properties. In the first, L edges are added at random to a graph of n nodes, sometimes referred to among social network analysts as the $U | L$ model (the uniform distribution of random graphs with a given number of nodes, conditional on the presence of L edges.) The second model posits that an edge between pairs of nodes occurs independently and with fixed probability p , sometimes referred to as the Bernoulli graph distribution (Frank, 1981; Frank & Nowicki, 1993). The local specification of a Bernoulli graph is clear because the probability of a tie between two nodes is independent of anything else in the graph. If friendship networks could be reasonably described as Bernoulli graphs (which they cannot), we could claim that any pair of humans has a given propensity to friendship, irrespective of whatever other friendships may occur. A local specification permits a ready translation into a process or behavior at the level of the actors, the nodes in the network.

³ Watts has more recently developed new locally specified models to describe the results from the original small world experiment of Milgram (1967), recognizing that his cyclic substrate model does not allow for strategic search by actors for effective network partners (Watts, Dodds & Newman, 2002). Indeed, without allowing for a search capacity, there is doubt that the original model can account for actors' *use* of short paths, given that they only have local information available to them (Kleinberg, 2000).

Local specifications generally assume a level of *homogeneity*, that there are some effects observed locally that are reproduced across the entire network (for instance, a fixed p for a Bernoulli graph). A homogeneity constraint assumes a system-wide property, but in almost all social networks, the only intentionality in the system is at the level of the actor, that is, locally (Robins & Pattison, 2001). Similar arguments provide the theoretical underpinning for Snijders' actor-oriented models (Snijders, 1997, 2001; Van De Bunt, Van Duijn & Snijders, 1999). As Robins and Pattison (2001) argued, the systemic property inherent in a homogeneity constraint may reflect shared norms or behaviors across actors, norms or behaviors that could be construed as inherent in the particular social relation for this group of people. Modeling based on local effects presupposes some such level of sharedness of local behaviors across the system (otherwise modeling is impossible.) Randomness in such models arises from non-shared local behaviors.

The extent of homogeneity is usually an unheralded issue. Many social network modeling assumes at least homogeneity of nodes⁴ (i.e. the identity of nodes is irrelevant to the model, so that the model does not presuppose that one node is more likely than another to have, for instance, a given degree.). Node homogeneity is also a feature of the Watts (1999a) cyclic substrate model, although not of the preferential attachment model, where the nodes are ordered across time.

A purely local specification is not always appropriate. Clearly there may be exogenous effects that operate at a global level (Robins & Pattison, 2001). The CEO of a company may impose an organizational restructuring; television does shape the transmission of information; a computer company with a large market share may bundle certain website addresses as pre-set features into its internet vehicles. These exogenous actions are not local processes. Even so, they may influence but not determine the underlying network effects. The restructuring may not work, and workers may revert to some of their old alliances; the television news may be doubted and different information transmitted locally; the computer user may choose not to utilize some features of the package. So even when exogenous effects are apparent, some form of local specification may still be required.

In this article, we use Markov random graph distributions (Frank & Strauss, 1986) – described in detail below – as a basis for locally specified models. We have three principal aims: (a) to establish methods to assess and compare global structures for *distributions* of small, possibly non-connected, graphs; (b) to use simulated Markov graph distributions to investigate various types of global structures that may arise, using different parameter values that can be related to a few very simple postulated local processes; and (c) to illustrate how the scaling of parameters affects the transition from randomness to more regular structure. We shall give examples of what we call *small world distributions* of graphs with a given number of nodes, but we shall also show how to simulate distributions of *long path worlds*, of relatively dense graphs with no clustering, and of “degenerate”, highly structured worlds, such as the *caveman graph* of Watts (1999a).

Our third aim of parameter scaling can be distinguished from the *scale-up problem*, which concerns the change in global outcomes as the number of nodes in the network grows. Although the scale-up problem has a long tradition in random graph theory (Bollobas, 1985), we concentrate here on the scaling-up of parameter values in

⁴ Except in cases where some form of blockmodeling might be the objective of the analysis (e.g. Nowicki, & Snijders, 2001), or where the modeling explicitly includes observed or latent actor attributes (e.g., Hoff, Raftery & Handcock, 2002; Lazega & van Duijn, 1997; Robins, Elliott & Pattison, 2001).

networks with a constant number of nodes⁵. The parameter scale-up issue has been ignored in the existing graph literature but we demonstrate that it is inherently connected with the transition from randomness to regularity in graph structures (Watts, 1999a). We make some comments on the impact of the size of the node set in our conclusions, as a matter for ongoing work.

Our focus is the nexus between local specifications and global properties. Small world features are just one instance but as they have featured so prominently in recent work, and have been so helpful in permitting new ways of thinking about local-global connections, it is useful to introduce our approach specifically through a small-world lens. In what follows, then, we begin with a discussion of certain aspects of small world problems. We then introduce our simulation strategy, including a description of exponential random graph (p^*) models, of which Markov random graphs are a subclass. We give examples of the various global structures that emerge with different parameter values. We go on to discuss the effect of increasing parameter values, with examples of the resulting regularized structures. We conclude with a discussion of further work to elaborate our general approach.

2. THE SMALL WORLD

Stanley Milgram (1967) famously concluded that the median number of intermediaries required for one person in the U.S. to contact another person was five. Despite some important work in the intervening years (e.g., Kochen, 1989; Pool & Kochen, 1978), small world ideas rather languished until given new currency in the 1990s (Watts & Strogatz, 1998; Watts, 1999a, 1999b). A social network of acquaintances can be represented as a graph, that is, a structure comprising *nodes*, and *edges* that connect pairs of nodes⁶. Watts (1999a) specified the properties of small world networks in graph-theoretic terms (defined in detail in the immediate paragraph below): a small world graph has low *density* and is highly *clustered* but has short *characteristic path lengths*. Watts (1999b) noted some additional relevant properties about the human social world: the number of nodes is very large, and the network is highly decentralized in that there is no dominant node to which other nodes are directly connected (or *adjacent*). As he noted, the juxtaposition of all of these characteristics seems counterintuitive, yet they are features of many empirical networks, including social networks (see also Albert & Barabási, 2002).

2.1 Terminology

Before discussing Watts' approach to constructing a probabilistic model for small world graphs, we first review some graph-theoretic and social network terminology:

- a *graph* $G = (N, E)$ comprises a set N of nodes, and a set $E \subset N \times N$ of edges, each of which connects a pair of distinct nodes in N (we say that the edge $e = ij$ connects nodes i and j);
- the *order* of a graph is the number n of nodes in N ;
- the *size* of the graph is the number L of edges in E ;

⁵ As noted earlier, the vast bulk of empirical network analysis focuses on networks with a fixed number of nodes.

⁶ In this article, we focus on non-directed networks.

- the *density* of the graph is the number of observed edges as a proportion of the total possible number of edges, namely, $2L/n(n-1)$;
- the *degree* of a node is the number of edges incident to it, so that if there are L edges in a graph of n nodes, the average degree per node k equals $2L/n$, and the density equals $k/(n-1)$;
- the *degree distribution* for a graph is the vector $(d_0, d_1, \dots, d_{n-1})$ of degree frequencies, where d_k is the number of nodes having degree k ;
- $G' = (N', E')$ is a *subgraph* of G if $N' \subset N$ and $E' \subset E$;
- a q -*star* is a subgraph of $(q+1)$ nodes in which one central node is connected by an edge to exactly q other nodes;
- a *triangle* is a subgraph comprising a set of three nodes (a *triad*), each pair of which is connected by an edge;
- a *path* of length m is a sequence $(e_0, e_1, \dots, e_{m-1})$ of edges such that $e_h = i_h j_h$, $i_{h+1} = j_h$ and all nodes i_h and j_h are distinct;
- a *geodesic* between two nodes is the shortest path between them, taken to have infinite length if there is no path between the two nodes;
- a *connected graph* has paths between all pairs of nodes (i.e. no geodesics of infinite length);
- the *local clustering coefficient* C_i of a node i is the proportion of pairs of nodes to whom node i is connected that are connected to each other.

In general, clustering in a graph refers to a propensity for a pairs of nodes to be connected to each other if they are connected to a common third node⁷. The *characteristic path length* (P) of a graph is a measure of central tendency (across the graph) of the length of geodesics. Watts (1999a) defined clustering and characteristic path length in the following terms:

- the clustering coefficient C is the average value across nodes of the local clustering coefficient C_i ;
- the characteristic path length P is the mean geodesic length.

Clearly the small world phenomenon relates to characteristic path length. Some major results relating to path length and connectedness are well-known in graph theory (Bollobas, 1985). For U | L and Bernoulli graph distributions, the Erdős-Renyi theorem (Erdős & Renyi, 1959) establishes that almost all simple random graphs with more than $n \ln(n)/2$ edges are connected. Other important results used by Watts (1999a) include that, for large n and k , the characteristic path length, P , for a simple random graph is of the order of $\ln(n)/\ln(k)$, and C is of the order of k/n . So for low density graphs, the tendency for clustering in a random graph is small. Moreover, P for random graphs is typically short. So when a graph is an agglomeration of independent, randomly added edges, we expect short characteristic path length but little tendency for clustering. But certain highly structured graphs, such as Watts's (1999a) "connected caveman graph" (a graph of small clustered components connected in one large cycle), have high levels of clustering with long characteristic path length.

2.2 Simulating random graph distributions

⁷ There is an alternative network usage of the term *clustering* to refer to the clustering of ties around certain nodes through the formation of stars, with transitivity referring to triangulation (e.g. Frank & Strauss, 1986). We use the term *clustering* as relating to triangulation in accord with the usage adopted by the recent small world literature.

Watts's (1999a) argument used connected graphs (i.e. with finite geodesics). In particular, he focused on graphs with a cyclic substrate, a starting structure for his simulations constituted by a cycle of edges through all nodes. A cyclic starting point ensures that the network is connected, so one can work with more analytically tractable mean path lengths rather than (say) median path lengths. As discussed below, connectedness may or may not be a problematic assumption depending on the type of network under consideration.

Under the algorithm, edges are added to the graph allowing for a tendency towards clustering but with a variable degree of randomness (i.e. the algorithm permits edges to be added at random, as well as according to the clustering criterion.) By stopping the algorithm at a specified density, Watts observed some expected results. With no randomness, graphs were highly clustered, but path lengths tended to be long, as is the case of the connected caveman graph. With too much randomness, on the other hand, the random component of the algorithm overwhelmed the clustering aspect, and low clustering resulted, as is the case for graphs with edges added randomly. But Watts also found that a small amount of randomness in the model resulted in a class of graphs with relatively high clustering and short characteristic path length, that is, small world graphs. Watts developed an explanation of this shortening of paths through the establishment of "short-cuts" across the large cycle. Although we do not use a cyclic substrate, the notion of short-cuts across a cycle is an important element in what we present below.

Conceptually, our approach described below draws heavily on this work, but the implementation differs in several important respects. We do not concentrate on connected graphs, or graphs that originate from various substrates. Rather than using a central tendency of characteristic path length, we investigate the *distribution* of geodesics. We simulate distributions of random graphs that have certain global properties emergent from local dependencies among edges. To simulate, we use the Metropolis algorithm with several advantages⁸. The first is that it permits us to simulate stochastic models for social networks that are derivable from first principles using the Hammersley-Clifford theorem (Besag, 1974) and so possess a principled statistical basis. Different models embody different assumptions about the relevant local social processes that might generate the network. Accordingly, we are able to interpret any conclusions specifically in terms of local social processes that may or may not generate global properties. The second advantage is that, in principle, the algorithm ensures convergence to a given distribution irrespective of the starting point of the simulation. As a result, we need not be concerned about a starting substrate, nor do we need to rely on an arbitrary stopping rule. And we are able to derive statistics from a large sample of graphs from a given distribution, giving generality to our conclusions. A third advantage is that model parameters can also, in principle, be estimated from observed network data (e.g. Snijders, 2002), allowing useful comparisons between models under consideration and empirical observations.

Before we describe the details of our approach, we discuss briefly some aspects of clustering, and of characteristic path-length and connectedness.

⁸ Several algorithms could have been used; see Snijders (2002) for a review of some other possibilities.

2.3 Clustering

In network-theoretic terms, the idea of clustering is closely related to that of *structural balance* (Cartwright & Harary, 1956), whereby individuals who are friends (for instance) are consistent in their friendships (and non-friendships) with other people. As Holland and Leinhardt (1970, 1971) recognized, structural balance essentially involves the notion of *transitivity* in graphs, which in non-directed graphs is represented as the presence of triangles. If there is a triangle on a triad of nodes, the triad is transitive. It will be apparent that a triangle comprises three 2-stars. If a 2-star (which also involves a triad of nodes) is not part of a triangle, then that triad is *intransitive*.

Structural balance is a local process. We do not need to inspect the whole graph to determine whether a particular triad of nodes is balanced or not. The proportion of transitive to intransitive triads in the graph is an aggregate measure of the tendency towards transitivity and structural balance, and hence of clustering. Because any triangle contains three 2-stars, in what follows we use the proportion $3T / S_2$ as the clustering coefficient C , where T is the number of triangles and S_2 the number of two-stars (Newman, Strogatz & Watts, 2001) in the graph as a whole⁹. (Note that $C = 1$ if all triangles are transitive and the graph can be partitioned into completely connected components.) In other words, the clustering coefficient assesses the *sharedness* of balance-type processes across the entire network. As will be seen below, the statistics T and S_2 are also important in our models for Markov random graphs, so this proportion is also naturally related to our method of graph simulation.

2.4 Characteristic path length and connectedness

Watts (1999a) used the mean geodesic across the graph as the measure of characteristic path length. The problem with utilizing a mean, however, is that it can only be computed for connected graphs. Can we expect the social networks we study to be connected? Erdos and Renyi (1959) showed that almost all random graphs with number of nodes n will be connected when the average degree is greater than $\ln(n)$. On a global scale, a network of a billion nodes would require average degree of around 20 for the Erdos and Renyi theorem to apply, assuming of course that the graph was indeed Bernoulli.

But ultimately it is unhelpful to do this simple calculation without taking into account substantive considerations, in particular, the nature of the network under consideration. The point is that individuals are embedded in many different social networks, each of which arises from a social relationship of a particular type. The form and properties of the underlying social relationship patterns a particular network. For instance, we assume that average degree in a network of sexual relations is lower than the average degree in a network of friends. If we ask whether sexual networks are small world (and hence, for instance, possibly vulnerable to rapid HIV transmission), we cannot presume that the networks are connected. Even within the category of “friendship”, where an average degree of 20 may be plausible, there are those who are close and trusted (strong ties), and those who are not (weak ties). Granovetter (1973) argued that strong tie clusters are bridged principally by weak ties. Accordingly, we do

⁹ This definition differs slightly from that used by Watts (1999a), the average of local clustering. The two measures are in fact highly correlated.

not necessarily expect that strong tie networks will be connected, even for very large graphs.¹⁰ And if we are dealing with small networks, connectedness clearly may not be observed.

So our preference is to summarize the path length distribution of a graph in a way that does not depend on it being connected. We therefore characterize the distribution of geodesic lengths in terms of certain order statistics, some of which may be infinite if the graph is not connected. Where we require a single measure we use the median. Moreover, as explained below, we examine geodesic distributions across a distribution of graphs, so, for instance, a distribution of Bernoulli graphs will have a particular profile of geodesic percentiles. We can then compare the profiles of distributions of graphs with different properties¹¹.

3. EXPONENTIAL RANDOM GRAPH (p^*) MODELS AND THEIR SIMULATION

We simulate exponential random graph, or p^* , models for social networks (Frank & Strauss, 1986; Pattison & Wasserman, 1999; Robins, Pattison & Wasserman, 1999; Wasserman & Pattison, 1996; Pattison & Robins, 2002). First, we summarize the formulation of the models, and then we comment on the application of the Metropolis algorithm.

3.1 Exponential random graph (p^*) models

To begin, we need some standard terminology and notation. Let $N = \{1, 2, \dots, n\}$ be the set of network nodes and let the two-way $n \times n$ (binary) array \mathbf{x} denote an observed network on N (that is, $x_{ij} = 1$ if there is an observed edge from node i to node j , and $x_{ij} = 0$ otherwise, with x_{ii} a structural zero). Also let \mathbf{X} denote a *random* graph or network on N , with each *possible edge*, or *tie*, regarded as a random variable X_{ij} . For the purposes of this article, we regard possible ties as *nondirected* (so that $X_{ij} = X_{ji}$).

In order to construct general models for $\Pr(\mathbf{X} = \mathbf{x})$, we need to recognize that there may be dependencies among the network variables in \mathbf{X} . An analogous substantive statement is that particular social ties may tend to be observed in the presence of other ties. For instance, a tendency towards clustering and balance in a social system is, as noted above, a tendency towards transitivity in the network. That is, the presence of a 2-star tends to be associated with the presence of a triangle, or, in other words, the presence of an edge linking i and j is likely to be affected by the presence of edges between i and k , and k and j . If this effect is indeed present in the network, then there is interdependency among the variables X_{ij} , X_{ik} and X_{kj} .

¹⁰ Consideration of a network of only strong ties may be relevant, depending on the type of question being addressed: for instance, it may be that certain types of information (e.g. with sensitive or risky content) are transmitted only through trusted partners, and hence this information may not be spread through small-world processes, even though for other types of less delicate information, small-world processes may apply.

¹¹ These steps imply that we are engaged in empirical examination of simulation results, and not seeking closed-form analytical results, for which a median (or any percentile measure) is not usually amenable. As a reviewer has pointed out, there are of course alternatives to using a median geodesic: for instance, the inverse of the sum of inverse geodesics. This would be a measure of connectivity for which infinite paths contribute 0. See also Newman (2003).

Frank and Strauss (1986) recognized that some fundamental theorems for interdependent observations developed in spatial statistics could be applied to assumed dependencies among network ties. Application of these results yields a general expression for $\Pr(\mathbf{X} = \mathbf{x})$ from a specification of which pairs of possible ties are conditionally dependent, given the values of all other ties. This arises from the Hammersley-Clifford theorem (Besag, 1974) which yields an expression for $\Pr(\mathbf{X} = \mathbf{x})$ in terms of parameters and sub-structures corresponding to sets of variables that are mutually dependent¹²; that is

$$\Pr(\mathbf{X} = \mathbf{x}) = (1/\kappa) \exp\{\sum_A \lambda_A z_A(\mathbf{x})\} \quad (1)$$

where:

- a) the summation is over all subsets A of the set of network variables in \mathbf{X} ;
- b) $z_A(\mathbf{x}) = \prod_{(i,j) \in A} x_{ij}$ is the *network statistic* corresponding to subset A ;
- c) $\kappa = \sum_{\mathbf{X}} \exp\{\sum_A \lambda_A z_A\}$ is a normalizing quantity; and
- d) the parameter $\lambda_A = 0$ for all \mathbf{x} unless all the variables in A are mutually (conditional) dependent.

Equation (1) is the general form; for the purposes of this article the specific version we use is presented below as equation (2), which we will interpret in fuller detail. But before doing so, there are some features of the general form (1) that are worth noting. Each of the subsets A of network variables may be construed as a *local social neighborhood* (Pattison & Robins, 2002), that is, a site of mutual contingency among network ties, with such contingencies expressing a social process. For instance, in the example of clustering above, the subset of possible ties $A = \{ij, ik, jk\}$ constitutes a local social neighbourhood and the associated network variables are mutually conditionally dependent. The binary statistic $z_A(\mathbf{x})$ is computed from \mathbf{x} and takes the value 1 if all the possible ties in the subset A are present in \mathbf{x} . If all the possible ties in A are observed, there is a triangle (a subgraph configuration) observed in \mathbf{x} . If the parameter λ_A is large and positive, the probability of observing the network \mathbf{x} is enhanced if that triangle (the configuration corresponding to A) is observed. If we take A to be not just a particular triangle, but a generic triangular effect (i.e. we suppose that the parameter λ_A is the same for all possible triangles in the network¹³), then the statistic $z_A(\mathbf{x})$ becomes a count of triangles in the network, and the parameter λ_A represents the tendency for transitivity (i.e., clustering) in the network.

More generally, the models express the importance of various types of network configurations (e.g. triangles), with each network configuration relating to a parameter λ and with the associated z statistic being a count of that configuration in the network. The expression in (1) then represents a probability distribution of graphs with the properties implied by the configurations (e.g. models with a triangle parameter express tendencies to clustering.)

¹² More formally, dependencies among possible network ties may be represented by a *dependence graph* \mathbf{D} (Robins and Pattison, in press). In a dependence graph, each network variable is represented as a node, with an edge between two variables if they are dependent, conditional on all other network variables. A clique of a dependence graph \mathbf{D} is either a single node (network variable) or a subset of nodes that are all connected with each other by edges (i.e. a subset of variables that are all conditionally dependent on each other.) There is one and only one parameter for each subset of variables that are mutually conditionally dependent.

¹³ Technically this step imposes *homogeneity* across isomorphic network configurations (Wasserman & Pattison, 1996). It corresponds to an assumption of sharedness of the behavior across the network.

The simplest models arise if we assume no dependencies among network variables at all (i.e. social ties are independent of each other.) Allowing that the probability of any tie is constant across all possible ties, we then have the class of Bernoulli graphs, and the model is:

$$\Pr(\mathbf{X} = \mathbf{x}) = (1/c) \exp\{\theta L(\mathbf{x})\}$$

where $L(\mathbf{x})$ is the number of ties in the network \mathbf{x} and θ is a parameter relating to the number of edges (sometimes called the *edge* or *density* parameter.) Here the relevant subgraph configuration is a single edge, and the model is expressed in terms of such configurations, i.e., the number of edges.

For social networks, the assumption of independent ties is generally implausible. Frank and Strauss (1986) introduced *Markov dependencies*, whereby two possible edges are assumed to be conditionally dependent if they share a node. For nondirected networks, the resulting model has parameters relating to a small number of configurations: stars of various types, and triangles. For the purposes of this article, we shall concentrate on the four Markov configurations of Figure 1: single edges, 2-stars, 3-stars and triangles. The model then becomes:

$$\Pr(\mathbf{X} = \mathbf{x}) = (1/c) \exp\{\theta L(\mathbf{x}) + \sigma_2 S_2(\mathbf{x}) + \sigma_3 S_3(\mathbf{x}) + \tau T(\mathbf{x})\} \quad (2)$$

where σ_k and $S_k(\mathbf{x})$ refer to the parameter associated with k -star effects and the number of k -stars in \mathbf{x} , respectively; and τ and $T(\mathbf{x})$ refer to the parameter for triangles and the number of triangles, respectively. The interpretation of the triangle parameter has been discussed above in terms of clustering and balance; a k -star parameter relates to the propensity for k -stars to be present in the network, that is, a tendency for individuals to have connections with multiple network partners¹⁴. Note that the Bernoulli graph distribution is a special case of (2) with only the edge parameter non-zero. For a Markov dependency assumption, the use of 2-stars and triangles has become standard. The rationale for the inclusion of the 3-star parameters in the models is given below¹⁵.

(Figure 1 about here)

The model in (2) then represents a distribution of graphs based on edge, star and triangle parameters. The parameters relate to local subgraph configurations. By setting the parameters to particular values, and then simulating this distribution, we can investigate certain global network properties across the distribution of these graphs.

With the model in equation (2) as the basis for our simulations, the next few subsections cover a number of technical details, before we turn to our results. In section 3.2, we outline the simulation approach; in section 3.3, we present our framework for comparing graphs and graph distributions in order to determine important features in global structures; in section 3.4, we discuss the appearance of “frozen” patterns when parameter values are large, of which we provide examples in the results; in section 3.5, we provide further detail on model interpretation, and the rationale for the inclusion of 3-stars; and in section 3.6, we briefly discuss our choices of parameter values.

¹⁴ It is well-known that the parameters in exponential random graph models are not independent. Such interdependence is to be expected because the models treat 2-stars as 2-way interactions among single edge variables, and triangles as 3-way interactions. Of course 2- and 3-way interactions in statistical methods are not independent of each other, nor from main effects. The complexity of the interactivity, however, does mean that small changes in parameter values may or may not have large effects. This is the expected behavior of complex systems, and in part this article is intended to explore such behavior.

¹⁵ Model specification appropriate for observed networks is a vexed question. This is not an issue that immediately concerns us here, but our current recommendation, based on the simulation results described below, is that at least 3-stars should be included in models fitted to empirical data.

3.2 The simulation strategy

We simulate graph distributions based on (2) using the Metropolis algorithm (Gilks, Richardson, & Spiegelhalter, 1996). This technique was first utilized for social networks by Strauss (1986). In summary, the algorithm operates as follows. We decide on the model we wish to simulate by fixing the values of the parameters, and we start with a randomly chosen initial graph with a fixed number of nodes. At each iteration, we propose a new graph as a candidate for the next step in a Markov chain. The candidate graph \mathbf{x}' is determined from the current graph \mathbf{x} by selecting at random a possible edge from i to j and changing x_{ij} to $1 - x_{ij}$. We accept the candidate graph as the current graph whenever the candidate graph has an increased probability of being observed according to (2). When the probability is not increased, we accept the candidate graph with probability

$$r = \exp\{\theta(L(\mathbf{x})-L(\mathbf{x}'))+\sigma_2(S_2(\mathbf{x})-S_2(\mathbf{x}'))+\sigma_3(S_3(\mathbf{x})-S_3(\mathbf{x}'))+\tau(T(\mathbf{x})-T(\mathbf{x}'))\} \quad (3)$$

The algorithm establishes a Markov chain on the state space of all graphs of order n and converges to a distribution of the form of (2). There may be a “burn-in” period as the chain moves away from the starting graph towards the distribution.

It should be emphasized that the simulation does not converge to a particular graph. What is produced is a distribution of graphs. Graphs in this distribution will be more probable, and so more likely to appear in our sample, if they express the properties implied by (2) once parameter values have been set. For instance, if we set τ as positive, then there is a tendency for graphs in the distribution to have more triangles than would be the case if the parameter value were 0 or negative. In this way, we can simulate graphs with greater clustering. Of course, for triangles to occur at all there have to be some 2-stars present, so that below we typically set the 2-star parameter as positive to ensure this. As discussed below, this can be interpreted as a tendency for actors to have multiple network partners. Of course, there needs to be a ceiling on this tendency, else there is no reason why people should not be partners with everyone else, and in realistic social networks, having too many network partners is costly in terms of time and effort. Accordingly it is useful to set the 3-star parameter as negative. We discuss the relevance of the 3-star parameter more fully below. The edge parameter, on the other hand, simply represents a baseline tendency for edges to be present.

We have simulated graphs of various sizes from 30 to 500 nodes. Most of the results we present here are based on graphs of 100 nodes, in part because larger graphs cannot readily be depicted to illustrate various points. Generally, we seek to simulate distributions of graphs with low density, principally because high density graphs have high clustering and short paths anyway, so that the small world nature of such graphs is hardly in question. And, once again, it is easier to illustrate certain points with depictions of low density graphs, as the clarity of an illustration is sometimes lost when there are many edges present. We achieve lower density graphs by adjusting the edge parameter θ in (2) as required.¹⁶

¹⁶ We note that despite the usefulness of graphical depictions, we have chosen examples rather carefully to make the representations clear. Of course, empirical networks may not result in such clarity when depicted, so in general the analysis is more important than the graphical representation.

In our simulations, we use 500,000 iterations. Burn-in is checked by observing the stabilization of the statistics across iterations. As burn-in typically occurs relatively early, we generally remove the first 50,000 iterations to ensure we are sampling from the distribution. We then sample every 100th graph, giving a typical sample size of 4500 from each distribution.

For each sampled graph, apart from the model statistics (number of edges, 2- and 3-stars, and triangles), we calculate the following aggregate measures: degree distribution; geodesic distribution in percentiles (although for simplicity we shall use quartiles here); and clustering coefficient. (We also collect, for depiction, a number of graphs as typical representatives of the distribution.¹⁷) So for our graph distribution, we have an accompanying distribution of clustering coefficients. The situation is somewhat more complicated for degree distribution, for each sampled graph from the graph distribution has its own distribution of node degrees. For instance, in our sample of graphs the frequency of nodes with degree 1 varies across a certain range. Below we use a series of boxplots to represent this “distribution within a distribution”.

Although it can be proven that in principle the Metropolis algorithm yields convergence to the desired probability distribution, in practice convergence may take an exceptionally large number of steps. For example if x is a high probability graph, and its neighboring graphs x' much less probable, then the algorithm may retain x for a very large number of steps. When this happens, we refer to x as a *frozen* structure, and to the Markov chain as showing poor mixing (Gilks et al, 1996). These circumstances are discussed further in section 3.4.

3.3 Comparing graphs and graph distributions

Our measures of global structure can be used to describe any of our sampled graphs, but it is difficult to interpret them in the absence of a basis for comparison. For instance: on what basis do we judge a median geodesic length as “short”? In the most general sense, there are a number of possible comparisons that could be made: Pattison, Wasserman, Robins & Kanfer (2000) illustrated how to draw judgments about global structure through comparing a graph of interest against a range of different graph distributions of increasing complexity. Here, however, we make comparisons against a relevant Bernoulli graph distribution, because as explained above we expect graphs in that distribution generally to have short average geodesics and low clustering.

To consider the structure of an individual graph, our approach is as follows: we simulate a comparative Bernoulli distribution of graphs. We then examine the graph’s clustering coefficient C against the distribution of clustering coefficients from the Bernoulli sample. If C is extreme in that distribution – that is, if it is greater than the 95% percentile – we say that the graph is highly clustered. Similar, we say that the graph’s median geodesic ($G50$) is *short* if it is not extreme compared to the distribution of median geodesics from the Bernoulli sample – that is, if it is less than the 95% percentile. We can define short first quartile geodesics ($G25$) and short third quartile geodesics ($G75$), similarly. We then define a small world graph as having a short $G50$ and high clustering¹⁸.

¹⁷ We depict graphs using *Pajek* (Batagelj & Mrvar, 2002).

¹⁸ One could, of course, adopt more or less stringent versions of these criteria. For instance, Watts (1999a) investigated caveman graphs with clustering coefficients close to 1. We show such examples below. For the most part, however, we do not insist on clustering coefficients close to 1 but adopt the traditional statistical approach of investigating extremes in a comparison distribution, in this case the

We need to determine the appropriate comparative Bernoulli distribution. We simulate the Markov graph distribution and calculate the mean number of edges from the sample. We then simulate a Bernoulli distribution with the same expected number of edges. We can then assess each graph in the Markov sample against the Bernoulli sample as in the previous paragraph, and determine, for instance, the percentage of small world graphs (noting that in the Bernoulli sample no more than 5% of graphs can be small world by definition). This enables us to assess whether the local parameterization tends to result in small world graphs or some other structure. At the same time, a comparison of degree distributions across the two graph distributions is helpful in understanding the graph structure, as will be seen below.

3.4 Frozen patterns

As noted above, for some choices of parameter values the simulation may reach a particular high probability graph that remains as the current graph for a large number of steps in the simulation. Such behavior is common as the size of the parameters increases. Within the context of general pattern theory, Grenander (1993) described such behaviour as *freezing* and posed what he termed the *first limit problem*: how does the probability distribution behave as the interactions among variables are made stronger? In the context of model (2), the question becomes one of characterizing the random graph distribution as the parameters become larger in absolute value. As Grenander demonstrated, scaling the parameters by an increasingly large constant positive value $1/\gamma$ inevitably leads to freezing. In particular, as the parameters become larger (small values of γ), the probability density becomes concentrated uniformly on a subset of structures having minimum energy¹⁹: such structures are often very regular in form and are referred to as *frozen patterns*. Conversely, as γ increases in magnitude, the parameters of the model become very small and the graph distribution approximates a Bernoulli distribution. In the context of statistical mechanics, the parameter γ can be regarded as temperature – hence the description of patterns associated with small values of γ as *frozen*. Frozen patterns often occur in “degenerate” regions of the parameter space, as discussed in the next subsection 3.5.

The presence of freezing may be examined in several ways: for instance, by observing low variance in the degrees of individual nodes across the iterations, or by tracking the number of times a candidate graph is accepted by the simulation algorithm, since freezing is characterized by a very small acceptance rate.

distribution of Bernoulli graphs. Of course, in a long enough simulation a Bernoulli graph distribution will produce a number of highly clustered graphs, so adopting the extreme 5% criterion is not unreasonable.

¹⁹ Define the *energy* of a graph \mathbf{x} as $H(\mathbf{x}) = -\sum_A \lambda_A z_A(\mathbf{x})$ and the *minimum energy* m as the minimum value of $H(\mathbf{x})$ over all graphs in the state space. If K denotes the set of graphs with minimum energy m and $1/\gamma$ ($\gamma > 0$) is a scaling constant for parameters, then a scaling of the parameters of model (1) yields the model given by $\Pr(\mathbf{X} = \mathbf{x}) = (1/\kappa) \exp\{\sum_{A \in \mathcal{M}} (\lambda_A/\gamma) z_A(\mathbf{x})\}$. This model can be re-written in the form:

$$\Pr(\mathbf{X} = \mathbf{x}) = \exp\{-(H(\mathbf{x})-m)/\gamma\} / [|K| + \sum_{\mathbf{x}' \notin K} \{-(H(\mathbf{x}')-m)/\gamma\}],$$

from which it is clear that as γ becomes smaller (and the model parameters become larger), the probability of graphs in the minimum energy set K approaches $1/|K|$ and the probability of all other graphs approaches 0.

3.5 Model interpretation and the importance of the 3-star parameter

We have discussed above a substantive interpretation of the triangle parameter as expressing the outcomes of a local social process relating to structural balance²⁰. The edge parameter represents a baseline tendency to form social ties. The star parameters are somewhat more complex. The two-star parameter may be interpreted as a tendency to have multiple network partners, as indeed may the three-star parameter. What we present here are models with positive two-star and negative three-star parameters. Our substantive interpretation of this parameter pattern is that actors tend to want multiple network partners (positive two-star) but experience a cost in having *too many* (negative three-star). It should be noted that higher order stars contain lower order stars, so that a node with degree four, for instance, is at the center of 6 two-stars and 4 three-stars. Accordingly, there comes a point in the degree distribution, determined by the relative values of the two- and three-star parameters, at which a tendency for additional partners moves from positive to negative.

Models with 3-star parameters are helpful in order to construct models that yield even rough approximations to observed networks. If the highest order star parameter in the model is sufficiently positive, the complete graph will be a minimum energy graph, and the simulation will move towards the complete graph and stay there with very high probability. Such models have been termed “degenerate” by Handcock (2003)²¹.

Likewise, if the higher order star is sufficiently negative, then the null graph will be a minimum energy network, and the simulation will gravitate towards the null graph. For example, if the only star parameter in the model is for a two-star, then the null graph will be a minimum energy graph for sufficiently negative values of σ_2 and the complete graph will be a minimum energy network for sufficiently positive values of σ_2 . Indeed, the range of values of σ_2 for which probability is quite widely dispersed across graphs in the state space may be quite small. For example, in Figure 2, we show graph statistics for graphs in the distribution for the model with a fixed edge parameter θ and varying 2-star parameter σ_2 . It is clear from these figures that while the effect of σ_2 is dependent on the value of θ , there is a relatively low range of parameter values for which the corresponding graph distributions have average densities in other than the very low or very high range. Thus, models with just an edge and a 2-star parameter may be too simple to capture some important characteristics of observed networks (such as some simple aspects of connectedness). Of course, the behavior of the models changes with the addition of a triangle parameter, but many of the same features persist.

Accordingly, we have included a non-zero 3-star parameter in all of the models that we simulate below. (It is worth noting that stars are related to the degree

²⁰ We want to be clear that we are not interpreting the Metropolis algorithm as instantiating such processes – rather that algorithm is used to simulate graph distributions with certain properties. These properties can be interpreted as the outcomes of (possibly unobserved) social processes. An exponential random graph distribution may be construed as the stationary distribution of structures emergent from more overtly process-based models (Snijders, 2001, 2002).

²¹ For this class of models, the issue of model degeneracy for certain parameter values was first discussed by Strauss (1986). By model degeneracy, Strauss (1986) meant situations where $H(x)$ tended in probability to a minimum as the number of nodes became large. In fact, this notion of degeneracy relates to what Grenander (1993) terms the *second limit problem*: how does the probability distribution behave as the order of the graph (and hence the number of random edges in the model) tends to infinity? In terms of Grenander’s first limit problem (subsection 3.4), with fixed numbers of nodes, Handcock (2002, 2003) has extensively investigated degenerate parameter regions for 2-star models for very small networks.

distribution, so that to model up to 3-stars is in effect to model the first three moments of the degree distribution.)

(Figure 2 about here.)

3.6 Choosing parameter values

In this article, we have chosen various parameter values based on our exploration of the parameter space, simply to illustrate various points. In the main, we seek to demonstrate that it is possible to achieve properties akin to the small-world with a few locally specified parameters that can be interpreted in terms of local behaviors. More generally, we wish to illustrate certain behaviors of the model under different parameter conditions. Our results are not intended as claims to universality, but rather to draw attention to possibilities, so in what follows we have pragmatically chosen particular parameter values for these purposes. The following then is not intended as an exhaustive investigation of the parameter space, as has been done by Handcock (2002, 2003) for 2-star models in very small networks. Ultimately, of course, any investigation of whether particular parameter values are more likely than others has to rest on empirical investigations of observed networks in the context of careful model specification. In related work not reported here (Robins, 2003), our approach has been to estimate parameters from empirical network data, using newly available Monte Carlo maximum likelihood procedures, and then to simulate from these estimates using the procedures in this paper to investigate the global properties implied by the empirical estimates.

4. SIMULATION RESULTS

4.1 A distribution of small world networks

With our model parameterization, it is in fact not difficult to produce small world graphs through a judicious choice of parameter values. So we begin by presenting a distribution in which a substantial majority of graphs satisfy our small world criteria, and then by adjusting the parameter values we change the features of the distribution in ways that are revealing. For these simulations, we are aiming to produce graphs on 100 nodes that have around 100 edges, so density is quite low, approximately 2%.

The Markov random graph distribution we present is based on the following vector of parameter values $(-4.0, 0.1, -0.05, 1.0)$, where the values are for the edge, two-star, three-star and triangle parameters, respectively. The mean number of edges for graphs in this sample was 119.5; the sample from the comparable Bernoulli distribution had a mean number of edges of 119.9. Table 1 presents the basic graph statistics for the Markov random graph sample, with those for the Bernoulli sample. We see that the mean of the basic statistics are similar, except for triangles and hence for clustering.

Of course, we could perform the usual null hypothesis statistical tests to determine whether there are significant differences between the means in Table 1, but at conventional alpha levels for two independent samples of 4500 each, we have rather excessive power, with a capability of detecting even slight differences. It is more appropriate in these circumstances to use a measure of effect size. There are various possible measures of effect size that could be used. We present here the differences between means scaled in terms of the standard deviation from the Bernoulli graph distribution, $(\mu_1 - \mu_2)/\sigma$. For the means in Table 1, the difference in means does not exceed 0.07 of a standard deviation, except for the triangles, where the difference is 2.3

standard deviations, and for the clustering coefficient, where the difference in means is 2.4 standard deviations. We infer that there is a substantial effect for clustering in the Markov distribution.

Table 1 about here

In the Markov random graph distribution, 83% of graphs have short G25 (first quartile of geodesic distribution), 82% have short G50 (the 95% cut-off in the Bernoulli graph distribution is a median path length of 7), but only 48% have short G75; 66% of these graphs have high clustering and in this sample 54% are small world according to our criteria. What seems to be happening here is that the increased clustering is not greatly affecting median path length (G50) but is at the cost of increasing the “long” geodesics (i.e. above the third quartile G75). This is illustrated in Figure 3, which compares median geodesics and third quartile geodesics for both distributions. The patterns of median geodesics are not greatly different, but for the longer geodesics there are substantially more infinite G75 in the Markov graph distribution. And the comparison of degree distributions indicates that there are not great differences between the two samples (Figure 4).

Figures 3 and 4 about here

Finally Figure 5 presents a representative graph from the Markov graph distribution. This is a small world graph with 120 edges, a median path length of 5, and a clustering coefficient of 0.07. To the eye, the graph does not appear greatly different from a Bernoulli graph (not presented here): a couple of important features are a tendency for some long cycles and plenty of nodes with degree of four, five or more (in line with Figure 4). It is worth noting that many of the cycles intersect one another. In terms of Watts’ (1999a) notion of short-cuts, two cycles that *jointly* intersect create shortcuts *for each other*. This seems a highly efficient way to create short paths. It is also worth noting that a node that is at the intersection of two cycles must be at least a four-star.

Figure 5 about here

4.2 Long path networks

We now present results for a simulation with parameters $(-1.2, 0.05, -1.0, 1.0)$. Compared to the previous example, the most important change is to adjust the three-star parameter to be strongly negative, at the same time as to increase the edge parameter to achieve a comparable density of around 2%. We produce a Markov graph sample with a mean number of edges of 123.6 (SD=4.1). Not surprisingly, the graphs in this sample have fewer stars, especially higher order stars, compared with the Bernoulli: for the Markov graph sample, the mean number of three stars is 69.7 (SD = 10.0) compared to 252.3 (SD=74.6) for the Bernoulli sample (difference in means is 2.4 Bernoulli standard deviations). But despite the positive triangle parameter, the mean clustering coefficient for the Markov graph sample of 0.030 (SD = 0.02) is not substantially different from the mean clustering coefficient of the Bernoulli sample (0.023, SD=0.02, difference in means is 0.4 of a standard deviation). According to our criteria, less than 15% of the Markov graphs are highly clustered. The difference in the number of stars is clearly illustrated by the truncated degree distribution of the Markov graph sample (see Figure 6). In the sample, all have long first quartile geodesics (G25) and 99.8% have long median geodesics (G50), although interestingly, few of the Markov graphs have infinite G75, whereas a substantial proportion of the Bernoulli graphs do. Apparently, as the short and median geodesics tend to be longer in the Markov graph sample, more nodes

become connected, so that there are fewer infinite geodesics. This is borne out by Figure 6, where the number of isolated nodes (D0) is lower in the Markov graph sample.

Figure 6 about here

In Figure 7, we present a graph from the Markov graph sample that illustrates the effect of the longer G25 and G50. This graph has 122 edges and is typical of the distribution with low clustering coefficient and long G25 and G50. The rather tight “circularity” of the graph in Figure 5 still seems to be apparent, but closer inspection reveals that paths tend to be more implicated in larger cycles. There are fewer isolated nodes but, importantly, few stars of higher order than three. In the graph in Figure 7 there are only four 4-stars, compared to 194 4-stars in the graph in Figure 5. As a result, there are many fewer mutually intersecting cycles.

Figure 7 about here

To illustrate the point further, we present in Figure 8 a graph from a distribution generated from parameter values $(-2.2, 0.05, -2.0, 1.0)$. For this model, we have made the negative three-star effect even stronger, and adjusted the edge parameter. Without presenting all the details, some 80% of these graphs have long G50, and 9% have high clustering, with only 1% small world. It is clear that the graph with 94 edges in Figure 8 has long paths and zero clustering. We compare it with an example of a Bernoulli graph with 96 edges in Figure 9. At this lower density, we see that the Bernoulli graph has itself a number of long paths, but the higher order stars continue to create several larger intersecting cycles.

Figures 8 and 9 about here

4.3 A local process to produce a global small world

Of course, we are not claiming that these types of simulations represent the only processes that can give rise to a small world. What we do wish to illustrate is how locally specified effects may produce particular global structures. Substantively, we argue that a network is likely to be small world if the following conditions apply:

- individuals seek more than one network partner,
- but the costs of maintaining many partners are high, so there is a tendency against a multitude of partners;
- there is some tendency for network partners to agree about other possible partners (i.e. structural balance and clustering);
- but this tendency is neither too strong (else the network becomes too clique-like with insufficient links between cliques for smaller geodesics – which we illustrate below), nor too weak (else there is little clustering in the network).

The last condition might be interpreted in terms of Granovetter’s (1973) strong and weak ties argument. If all we have are strong ties, which tend to cluster, then our “reach” across the network is limited.

In an organizational context, DeCanio, Dibble and Amir-Atefi (2000) argued that an efficient communication structure in a network can emerge through mutually intersecting cycles, resulting from a tendency for multiple partners combined with a cost against having too many partners. They performed a series of simulations based on cost and reward functions that in effect implemented such rules. This result does not seem to be widely known among small world researchers. Of course, a short-cut as defined by Watts (1999a) is a global feature (because one can never be sure that a path is indeed a short cut unless one knows about the structure of the rest of the graph). But the basis of the simulations by DeCanio et al (2000), as embodied in their cost and reward functions, is locally specified. The fact that these simulations produce short-cuts through mutually

intersecting cycles, in conjunction with our results reported here, provides one possible local description of how short paths may emerge.

4.4 A denser non-clustered “world”

We now give an example of a denser network simulation, where what is of interest is not so much geodesic length (because as noted above path length tends to be short in denser graphs anyway). Rather, in this example we wish to make some points about the degree distribution and also to illustrate a tendency against clustering. We simulate for a strong propensity for multiple partners, while retaining some cost against too many partners, but with a large negative clustering effect, using the parameter values $(-3.2, 1, -0.3, -4.0)$. In this case, we have not fully compensated with a large negative edge parameter to create very low density graphs. The strong tendency for two-stars results in a sample with a mean number of edges of 371.5 (SD=5.0), implying a mean density of 7.5%, three to four times greater than in the previous samples²². The first point to note is that the negative triangle parameter successfully removes most triangles from the Markov graphs, with the mean number of triangles being 1.0 (SD=1.0) compared to 67.9 (SD=12.5) in the comparative Bernoulli sample. The two degree distributions are presented in the top two panels of Figure 10. It is immediately apparent that the Markov sample has a degree distribution that is truncated at both ends compared to the Bernoulli sample.

Figure 10 about here

Of course, this result in part arises from the restriction on the number of nodes. Substantively, these parameter values can be interpreted to describe persons who wish to have many network partners but have only 99 others from whom to choose. Hence we see very few low degree nodes. (In that sense, the effect of the actual parameter value is dependent on the number of nodes, an issue we discuss further below.)

It is interesting to compare this result with the degree distribution for a simulation *without* the negative triangle (negative τ) effect, with parameters $(-3.2, 1, -0.3, 0.0)$. The sample for this distribution has a mean number of edges of 383.5 (SD=5.0), some two standard deviations above that of the negative τ distribution, but has a mean number of triangles of 52.7 (SD=8.3), over six standard deviations above the negative τ distribution. This is not surprising. What is interesting, however, is to consider the degree distribution, which is presented in the bottom panel of Figure 10. While there are some differences between the two Markov degree distributions, they nevertheless seem quite similar in comparison to the Bernoulli distribution. An important point to make here, then, is that knowledge of the degree distribution does not necessarily give knowledge about clustering.

The absence of triangles is not just a curiosity but has substantive interpretation. There are real social networks where we see fewer triangles than in random networks, for instance, networks of heterosexual partners. What we have in the sample of negative τ networks are graphs with hardly any triangles (three-cycles) but many four-cycles²³. We might expect to see such patterns in promiscuous heterosexual networks, with triangles perhaps more prominent in homosexual (and maybe bisexual) networks. The balance between three- and four-cycles in a sexual network may then relate to

²² At this density distinctive features in the structure are not readily discernible to the eye, so we do not present examples of actual graphs.

²³ That is, cyclic paths that involve four nodes.

levels of multiple partnerships among and across actors of different sexual orientation. Such outcomes have clear implications for the understanding of HIV networks.

Sexual networks are a simple example of the possible relevance of four-cycles as opposed to three-cycles. But they are not the only example. In organizational contexts, for instance, four-cycles have important theoretical implications in terms of generalized exchange and the building of trust (Lazega & Pattison, 1999).

5. RANDOMNESS AND STRUCTURE

We introduced frozen patterns earlier as minimum energy graphs and we observed that there is a tendency for the simulation algorithm to become “trapped” by such graphs. Although this poor mixing on the part of the Metropolis algorithm is deleterious in terms of producing an adequate simulation of the graph distribution, such regions of the parameter space are often interesting in their own right, because they represent regular structures. When such “frozen” patterns represent regular and stable structural forms, a description of the forms that they take, and the class of models from which they are derived, are of theoretical interest in their own right. As we noted earlier, we can consider γ as a “temperature” factor, so that as the temperature increases, the model can move from a frozen to a less regular structure.

5.1 From caveman graphs to random graphs

Watts (1999a) introduced *caveman graphs* as part of his initial argument about the importance of paths. These graphs comprise subsets of nodes (the *caves*), which are fully connected within subsets but have no connections between subsets (i.e. several complete components in the graph, with no connections between components.) A *connected caveman graph* has a cyclic pattern connecting the caves. These two graph types are clearly highly structured, and the chance of observing them in a Bernoulli distribution of graphs is miniscule.

We present here a series of simulations with a fixed set of values that, once scaled by a temperature factor γ , become the parameter values. For ease of visual presentation of graphs, we fix the number of nodes to 30 in this illustration. We select a vector $\lambda = (-3.2, 1, -0.3, 3.0)$; these values are based on the previous example, except that here we need a large and positive triangle value (necessary to create the “caves”)²⁴. In Figure 11, we compare two graphs from simulations for $\gamma=1$ and $\gamma=6$. The graph for $\gamma=1$ in the top panel of Figure 11 is a highly regular caveman structure, comprising two fully connected components of 15 nodes each. The second graph, however, is a much less regular structure, as we explain below.

Figure 11 about here

When $\gamma = 1$, every sampled graph from our simulation has 210 edges and clustering coefficient of 1. Across the sample (after burn-in), the degree of each vertex never changes (which means that no changes to the graph are accepted for every 100 iterations). Clearly this simulation has “frozen”, mirrored by the perfect transitive structure presented in Figure 11. We have checked this simulation with a number of

²⁴ The size of the triangle parameter, relative to the other parameter values, determines the number and size of the caves. As shown in Figure 11, the value of 3 chosen here produces two caves when $\gamma=1$, whereas if we had chosen a value of 2, the result would be three caves.

random starts to ensure we have not found a local energy minimum. There are variations depending on random starts, but in all cases, two large “caves” resulted although there might be variation in the size of the caves with the number of nodes varying between 13 and 17, and possibly some isolates appearing.²⁵ As the temperature increases, some small variation begins to appear. For instance, by the time $\gamma=4$, a mean of 2.5 changes to the graph are accepted for every 100 iterations, so the structure is not completely frozen. In the $\gamma=4$ sample the mean number of edges is 205 (SD=2), and the mean clustering coefficient is 0.97 (SD=0.01). The median G50 is infinite. Yet, at some point between $\gamma=4$ and $\gamma=6$ we see a phase transition, as the structure “melts”. (At $\gamma=5$, for instance, much longer runs are required to burn in to a highly regular two cave structure, indicating closeness to a transition.)

At $\gamma=6$, however, the distribution shows no signs of being frozen at all. The graph in the lower panel of Figure 11 has 180 edges and a clustering coefficient of 0.44. Across the sample, the mean number of edges is 173.0 (SD=5.7), and the mean clustering coefficient is 0.427 (SD=0.02). In this simulation, an average of 65 changes to the graph are accepted for every 100 iterations. And the median G50 is now 2. Yet this is not quite a totally random graph distribution. A comparable Bernoulli graph sample indeed has a median G50 of 2, but has a mean clustering coefficient of 0.397, over one standard deviation below that of the $\gamma=6$ sample. So a large proportion (94%) of graphs in the Markov sample are more highly clustered than the Bernoulli graphs, and have similar median path lengths, that is, they are small world graphs by our criteria. But by the time we increase the temperature to $\gamma=10$, there is little to differentiate the sample from a comparable Bernoulli sample²⁶. For instance, the mean clustering coefficient is 0.42 (SD=0.02) compared to 0.41 (0.03) for the comparable Bernoulli sample, with median G50 remaining at 2 for both samples.

This result concurs with those of Watts (1999a), in that at a certain temperature a small addition to “randomness” (in our case, a small increase in temperature) pushes a structured graph distribution with long median geodesics into a less regular distribution with frequent small world graphs. The ratios of the parameter values are of course retained as temperature changes, but these ratios determine the frozen structure. So the particular ratios chosen here lead to a clustered caveman world for low temperatures, whereas for our first small world example above, decreasing the temperature results in an empty graph, principally because in that case we chose parameters to produce a low density graph.

Figure 12 about here

We have experimented with various parameter ratios, and with judicious choices we can produce various structural forms at low temperatures. In Figure 12 we show some interesting examples that involve negative triangle parameters, producing cyclic structures of various types. In the top panel of the figure we have a structure of cycles produced from the parameter vector $(-5.0, 5.0, -1.0, -6.0)$. The first point to note is that this structure has four structural equivalence classes and could be simplified into a perfectly fitting block model (White, Boorman & Breiger, 1976). There are no within-block ties, and each node is tied to every other node in two other blocks. This structure can be construed as a “frozen” generalized exchange structure, highly reminiscent of Bearman’s (1997) cycle of marriage exchange among residents of Groote Eylandt

²⁵ We have similarly checked other simulations described in this section with multiple starts to reach consistent conclusions.

²⁶ Although the Bernoulli sample does seem to have somewhat wider dispersion for the statistics.

(except that the marriage cycle had eight equivalence classes.) By adjusting the parameters slightly, with a parameter vector $(-4.2, 1.0, -0.05, -6.0)$, we produce the graph in the bottom panel of Figure 12, a complete bipartite graph.

6. CONCLUSIONS

In this article, we have (i) shown how to simulate a distribution of Markov random graphs based on assumptions about simple local social processes; (ii) developed a method to examine the resulting global structures by comparison with an appropriate Bernoulli distribution of graphs; (iii) provided examples of various stochastic global “worlds” that may result, including small worlds, long path worlds and dense non-clustered worlds with many four-cycles; (iv) suggested one locally-specified social process that may result in small-world global properties; (v) shown how parameter scaling relates to the movement from structure to randomness, with a phase transition occurring at a certain scaling (“temperature”) so that a highly regular structure “melts” into a less regular form; and (vi) provided examples of “frozen” structures, including highly clustered “caveman” graphs, bipartite structures, and global cyclic structures involving structurally equivalent groups.

There are some important points to make that bear on future work. First of all, our methods are deliberately based on small-scale fixed-node networks, which constitute a large proportion of the empirical work involving social network analysis. Using these methods, an empirically observed network can be compared to an appropriate Bernoulli graph distribution to investigate levels of clustering, geodesic length and degree distribution, permitting an examination of global structure. The methods could be further built on by the comparison of an observed network against a range of graph distributions with given properties expressing certain structural hypotheses, as in Pattison et al (2000). Such empirical examination need not be confined to unimodal networks: we are currently studying global structures of interlocking directorships through similar approaches to bipartite graphs (Robins & Alexander, in press). On the other hand, to consider local structure, one might fit an exponential random graph (p^*) model to the network, using recently developed Monte Carlo maximum likelihood techniques (Handcock, 2002, 2003; Snijders, 2002)²⁷. Even so, model specification remains an important issue, especially with the need to avoid degenerate regions of the parameter space.

But these methods open additional possibilities for model examination: by fitting models and then producing a distribution of graphs as described above, a researcher can investigate whether the model is successful in reproducing aspects of the observed global structure. If the global structure is not successfully reproduced, then it may be that model specification should be reconsidered. In our introduction, we argued that observation, modeling and simulation should not be seen as separate, but rather as a package of complementary tools. To examine simulated global structure from a model fitted to an observed network may be a demanding test of model specification²⁸, but it clearly illustrates the value of combining these approaches.

Even though our focus has been on fixed node networks, the effect of given parameter values for networks with different numbers of nodes is an important issue.

²⁷ Approximate pseudo-likelihood techniques (Strauss & Ikeda, 1990), previously a standard approach, should now be relied on only when maximum likelihood is not feasible.

²⁸ Although not an impossible one – see Robins (2003).

This is a matter for ongoing work. There are conceptual problems here, however, as the number of nodes becomes increasingly large. The Markov random graph dependence assumption imposes dependency among all possible ties from one actor to all other actors. This seems to us, ultimately, to be untenable on substantive grounds: individuals do not “know” all other actors in a large social network, and may have no plausible opportunity to meet some “distant” actors. In what sense can such highly unlikely ties be said to influence possible ties to those who are much more socially “proximate”? It is for such reasons that Pattison and Robins (2002) introduced the notion of overlapping *social settings* as “social locales” within which dependency amongst ties can come into being. In this case, examination of node scale-up issues may not require extension to graphs of indeterminate size.

There is much to be done on parameter scale-up and the transition from less to more regular structures. The more regular structures are those on which much of social network analysis has focused. We have presented some simple examples here but there are more complex possibilities: in a multiple network framework, Pattison (2002) showed how “freezing” a stochastic model can result in classic balance and strong-weak tie descriptions.

We began by referring to the global structure of the Medici network as the outcome of local social processes. Clearly, the Medici were particular in that network, and Padgett and Ansell make clear why they were historically special. In the end, our models may not be able to rest on node homogeneity. The qualities of the actors are likely to count. It is in the interplay of actor attributes and network ties that more realistic local social processes will be described. Thus, an important next step is to extend the approach described here to social selection models (Robins, Elliott and Pattison, 2001), to social influence models (Robins, Pattison & Elliott, 2001) and to dynamic conjunctions of the two. It remains to be seen exactly what the implications of these extensions will be for global structure, but it is likely that such a move will have the added benefit of modeling the greater dispersion of degrees typically seen in empirical networks compared to homogeneous random graphs (Albert & Barabási, 2002; Pattison & Wasserman, 2002).

References

- Albert, R., & Barabási, A.-L. (2002). Statistical mechanics of complex networks. *Review of Modern Physics*, 74, 47-97.
- Amaral, L.A.N., Scala, A., Barthélémy, M., & Stanley, H.E. (2000). Classes of small-world networks. *Proceedings of the National Academy of Sciences of the United States of America*, 97, 11149-11152.
- Barabási, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286, 509-512.
- Barrat, A., & Weigt, M. (2000). On the properties of small-world network models. *European Physical Journal B*, 13, 547-560.
- Batagelj, V., & Mrvar, A. (2002). *Pajek 0.79*. University of Ljubljana.
- Bearman, P. (1997). Generalized exchange. *American Journal of Sociology*, 102, 1383-1415.
- Berthélémy, M., & Amaral, L.A.N. (1999). Small-world networks: Evidence for a crossover picture. *Physical Review Letters*, 82, 3180-3183.
- Besag, J. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B*, 36, 96-127.
- Bohland, J.W., & Minai, A.A. (2001). Efficient associative memory using small-world architecture. *Neurocomputing*, 38, 489-496.
- Bollobas, B. (1985). *Random graphs*. London: Academic Press.
- Cartwright, D., & Harary, F. (1956). Structural balance: A generalization of Heider's theory. *Psychological Review*, 63, 277-292.
- Comellas, F., Ozon, J., & Peters, J.G. (2000). Deterministic small-world communication networks. *Information Processing Letters*, 76, 83-90.
- DeCanio, S.L., Dibble, C., & Amir-Atefi, K. (2000). The importance of organizational structure for the adoption of innovations. *Management Science*, 46, 1285-1299.
- Erdős, P., & Renyi, A. (1959). On random graphs. I. *Publicationes Mathematicae (Debrecen)*, 6, 290-297.
- Frank, O. (1981). A survey of statistical methods for graph analysis. Pp. 110-155 in S. Leinhardt (Ed.), *Sociological methodology 1981*. San Francisco: Jossey-Bass.
- Frank, O., & Nowicki, K. (1993). Exploratory statistical analysis of networks. In J. Gimbel, J.W. Kennedy & L.V. Quintas (Eds.), *Quo Vadis, Graph Theory? Annals of Discrete Mathematics*, 55, 349-366.
- Frank, O., & Strauss, D. (1986). Markov graphs. *Journal of the American Statistical Association*, 81, 832-842.
- Faust, K., & Skvoretz, J. (2002). Comparing networks across space and time, size and species. *Sociological Methodology*, 32, 267-299.
- Gilbert, E.N. (1959). Random graphs. *Annals of Mathematical Statistics*, 30, 1141-1144.
- Gilks, W., Richardson, S., & Spiegelhalter, D. (1996). *Markov Chain Monte Carlo in Practice*. London: Chapman Hall.
- Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78, 1360-1380.
- Grenander, U. (1993). *General Pattern Theory*. Oxford: Oxford University Press.
- Handcock, M. S. (2002). *Assessing degeneracy in statistical models for social networks*. Paper presented at Workshop on Dynamic Social Network Analysis, Washington, Nov 7-9, 2002.
- Handcock, M.S. (2003). Statistical models for social networks: Inference and degeneracy. In R. Breiger, K. Carley, & P. Pattison (Eds.), *Dynamic social network modeling and analysis* (pp. 229-240).

- Hoff, P., Raftery, A.E., & Handcock, M.S. (2002). Latent space approaches to social network analysis. *Journal of the American Statistical Association*, 97, 1090-1098.
- Holland, P. W., & Leinhardt, S. (1970). A method for detecting structure in sociometric data. *American Journal of Sociology*, 70, 492-513.
- Holland, P. W., & Leinhardt, S. (1971). Transitivity in structural models of small groups. *Comparative Group Studies*, 2, 107-124.
- Kirillova, O.V. (2001). Communication networks with an emergent dynamical structure. *Physical Review Letters*, 87,
- Kleinberg, J.M. (2000). Navigation in a small world. *Nature*, 406, 845.
- Kochen, M. (1989) *The small world*. Norwood N.J.: Ablex.
- Lazega, E., & Pattison, P. E. (1999). Social capital, multiplex generalized exchange and cooperation in organizations: A case study. *Social Networks*, 21, 67-90.
- Lazega, E., & van Duijn, M. (1997). Position in formal structure, personal characteristics and choices of advisors in a law firm: A logistic regression model for dyadic network data. *Social Networks*, 19, 375-397.
- Mathias, N., & Gopal, V. (2001). Small worlds: how and why. *Physical Review E*, 63,
- Milgram, S. (1967). The small world problem. *Psychology Today*, 2, 60-67.
- Newman, M.E.J. (2000). Models of the small-world: a review. *Journal of Statistical Physics*, 101, 819-841.
- Newman, M. (2003). The structure and function of complex networks. *SIAM Review*, 45, 167-256
- Newman, M.E.J., & Watts, D.J. (1999). Scaling and percolation in the small-world network model. *Physical Review E*, 60, 7332-7342.
- Newman, M.E.J., Strogatz, S.H., & Watts, D.J. (2001). Random graphs with arbitrary degree distributions and their applications. *Physical Review E*, 64, 026118.
- Nowicki, K., and Snijders, T.A.B. (2001). Estimation and prediction for stochastic blockstructures. *Journal of the American Statistical Association*, 96, 1077-1087.
- Padgett, J.F., & Ansell, C.K. (1993). Robust Action and the Rise of the Medici, 1400-1434. *American Journal of Sociology*, 98, 1259-1319.
- Pandit, S.A., & Amritkar, R.E. (2001). Random spread on the family of small-world networks. *Physical Review E*, 63, 041104.
- Pastor-Satorras, R., & Vespignani, A. (2001). Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63, 066117.
- Pattison, P. E. (2002). *Social networks, social space, social structure*. Keynote address, Sunbelt XXII International Social Network Conference, New Orleans, February 2002.
- Pattison, P., & Robins, G.L. (2002). Neighborhood based models for social networks. *Sociological Methodology*, 32, 301-337.
- Pattison, P. E., & Wasserman, S. (1999). Logit models and logistic regressions for social networks, II. Multivariate relations. *British Journal of Mathematical and Statistical Psychology*, 52, 169-194.
- Pattison, P. E., & Wasserman, S. (2002). Multivariate random graph distributions: applications to social network analysis. In Hagberg, J. (Ed.), *Festschrift for Ove Frank* (Pp. 74-100).
- Pattison, P. E., Wasserman, S., Robins, G.L., & Kanfer, A. M. (2000). Statistical evaluation of algebraic constraints for social networks. *Journal of Mathematical Psychology*, 44, 536-568.
- Pool, I., & Kochen, M. (1978). Contacts and influence. *Social Networks*, 1, 1-48.

- Robins, G.L. (2003). *The small worlds of small social networks*. Paper presented at the American Association for the Advancement of Science Annual Meeting, Denver CO, 13-18 February.
- Robins, G.L., & Alexander, M. (in press). Small worlds in the social networks of interlocking directorships: Structure and distance in bipartite graphs. *Journal of Computational and Mathematical Organization Theory*.
- Robins, G.L., & Pattison, P.E. (2001). Random graph models for temporal processes in social networks. *Journal of Mathematical Sociology*, 25, 5-41.
- Robins, G. L., & Pattison, P. E. (in press). Interdependencies and social processes: Dependence graphs and generalized dependence structures. In P. Carrington, Scott, J., & S. Wasserman (Eds.), *Models and methods in social network analysis*. NY: Cambridge University Press (forthcoming).
- Robins, G. L., Elliott, P., & Pattison, P. E. (2001). Network models for social selection processes. *Social Networks*, 23, 1-30.
- Robins, G.L., Pattison, P., & Elliott, P. (2001). Network models for social influence processes. *Psychometrika*, 66, 161-190.
- Robins, G. L., Pattison, P. E., & Wasserman, S. (1999). Logit models and logistic regressions for social networks, III. Valued relations. *Psychometrika*, 64, 371-394.
- Snijders, T.A.B. (1997). Stochastic actor-oriented models for network change. In P. Doreian & F.N. Stokman (Eds.), *Evolution of Social Networks* (pp.185-208). Amsterdam: Gordon & Breach.
- Snijders, T.A.B. (2001). The statistical evaluation of social network dynamics. M.E. Sobel and M.P. Becker (eds.), *Sociological Methodology 2001* (pp.361-395). Boston and London: Basil Blackwell.
- Snijders, T.A.B. (2002). Markov chain Monte Carlo estimation of exponential random graph models. *Journal of Social Structure*, 3, 2.
- Strauss, D. (1986). On a general class of models for interaction. *SIAM Review*, 28, 513-527.
- Strauss, D., and Ikeda, M. (1990). Pseudolikelihood estimation for social networks. *Journal of the American Statistical Association*, 85: 204-212.
- Strogatz, S.H. (2001). Exploring complex networks. *Nature*, 410, 268-276.
- Van De Bunt, G.G., Van Duijn, M.A.J., and Snijders, T.A.B. (1999). Friendship networks through time: An actor-oriented dynamic statistical network model. *Computational and Mathematical Organization Theory*, 5, 167-192.
- Wasserman, S., & Pattison, P. E. (1996). Logit models and logistic regressions for social networks, I. An introduction to Markov random graphs and p^* . *Psychometrika*, 60, 401-425.
- Watts, D.J. (1999a). *Small worlds: The dynamics of networks between order and randomness*. Princeton, NJ: Princeton University Press.
- Watts, D.J. (1999b). Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, 105, 493-527.
- Watts, D.J., Dodds, P.S., & Newman, M.E.J. (2002). Identity and search in social networks. *Science*, 296, 1302-1305.
- Watts, D.J., & Strogatz, S. H. (1998). Collective dynamics of 'small world' networks. *Nature*, 393, 440-442.
- White, H.C., Boorman, S.A., & Breiger, R.L. (1976). Social structure from multiple networks: I. Blockmodels of roles and positions. *American Journal of Sociology*, 87, 517-547.

Yang, X.S. (2001). Small-world networks in geophysics. *Geophysical Research Letters*, 28, 2549-2552.

Table 1
Graph statistics from Markov random graph and Bernoulli samples

	<u>Markov random graph sample</u>	<u>Bernoulli sample</u>
	Mean	Mean
	(Standard deviation in	(Standard deviation in
	brackets)	brackets)
Edges	119.5 (12.3)	119.9 (12.7)
Two stars	289 (60)	285 (62)
Three stars	226 (74)	223 (80)
Triangles	6.2 (3.1)	2.3 (1.7)
Cluster coefficient	0.06 (0.03)	0.02 (0.02)

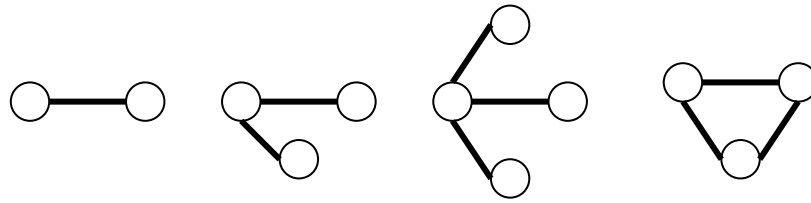


Figure 1
Configurations represented by parameters in the model:
Single edges; 2-stars; 3-stars; triangles

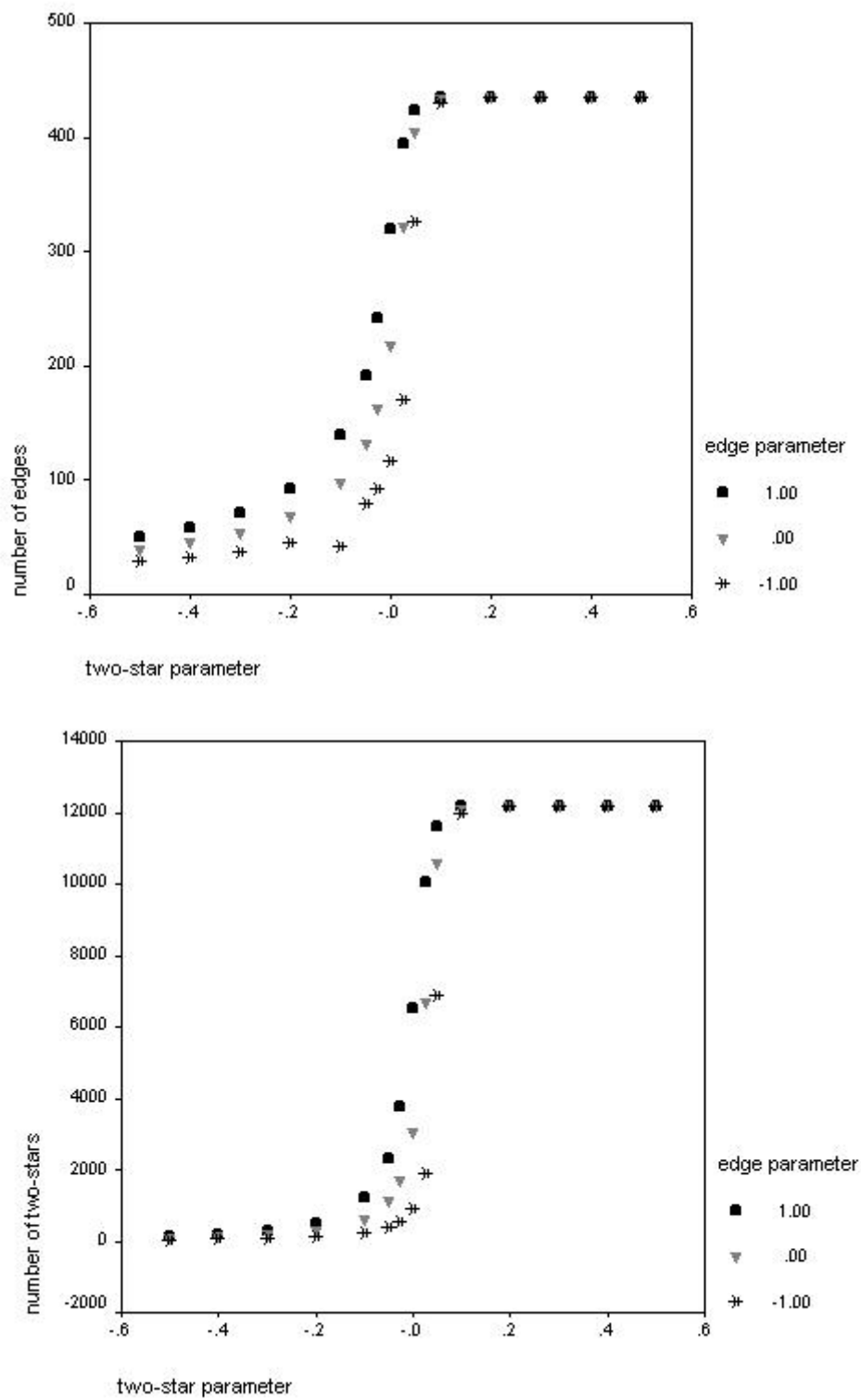
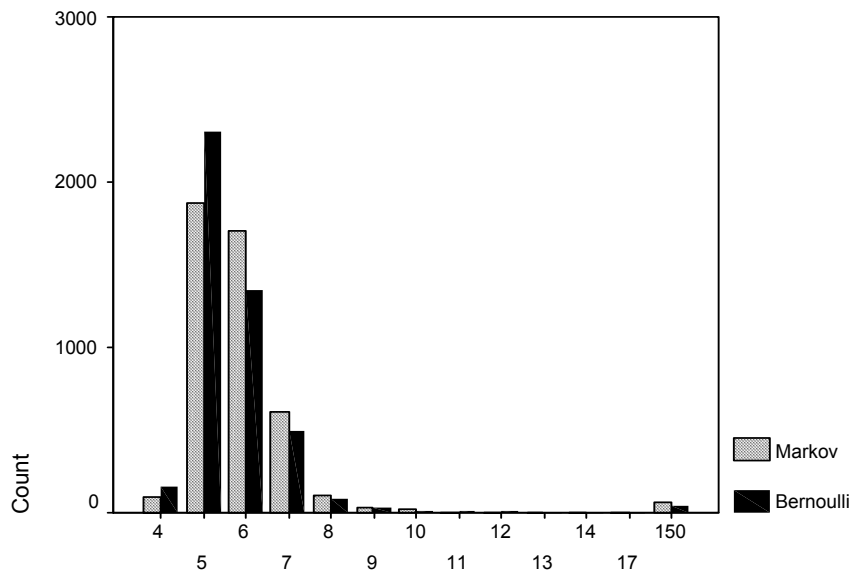
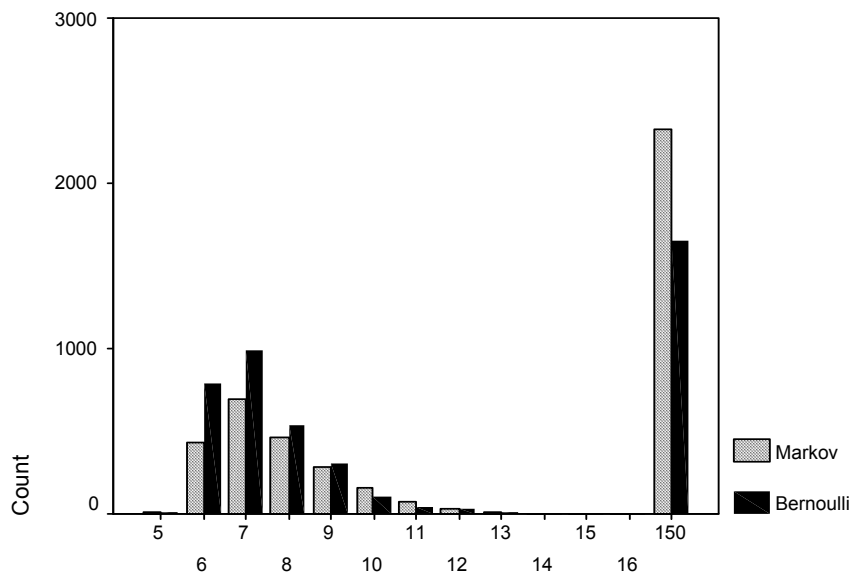


Figure 2
Mean numbers of edges and 2-stars in simulated graph distributions on 30 nodes for edge/2-star models with various parameter values.



G50



G75

Figure 3
 Frequencies of median geodesic (G50) and third quartile geodesic (G75):
 Markov graph sample and comparison Bernoulli graph sample

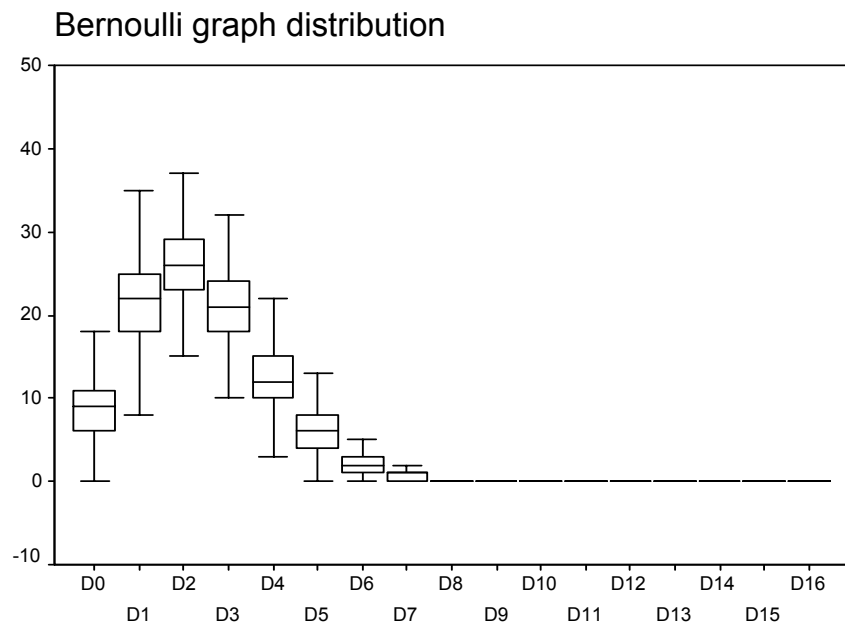
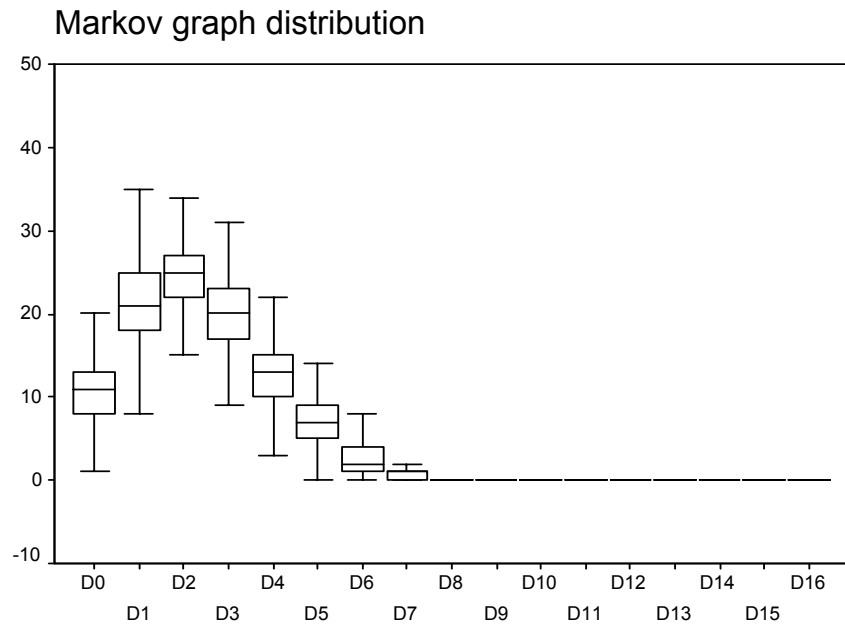


Figure 4
 Small worlds: Degree Distribution for Markov graph and Bernoulli samples
 (NB: D_n refers to the number of nodes of degree n . Boxplots indicate the range of D_n across the sample; outliers are excluded.)

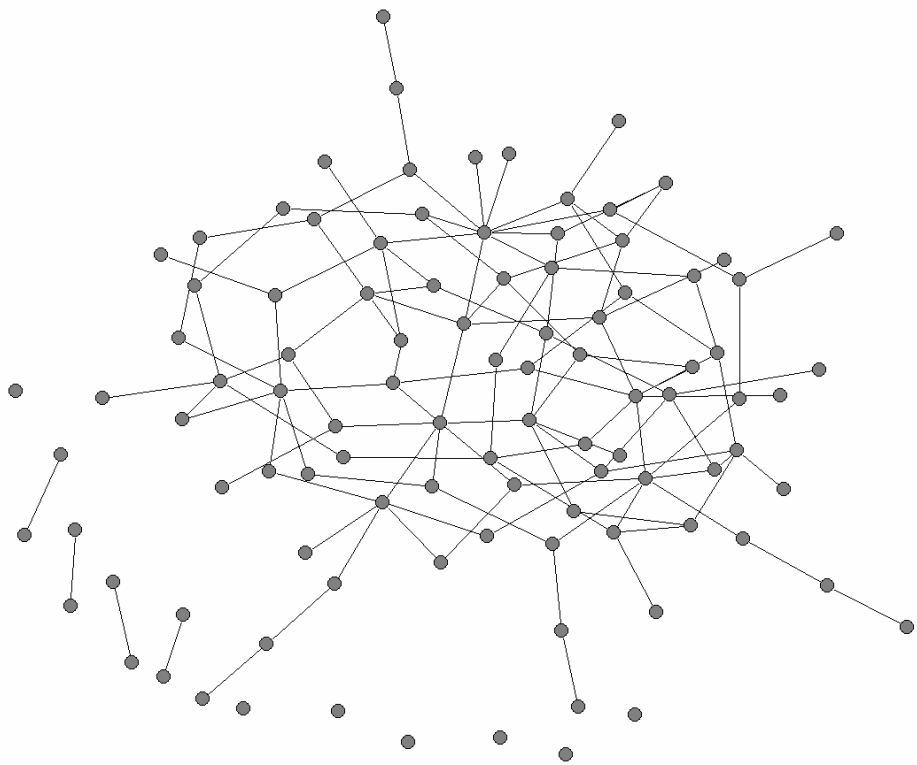


Figure 5
A small world graph

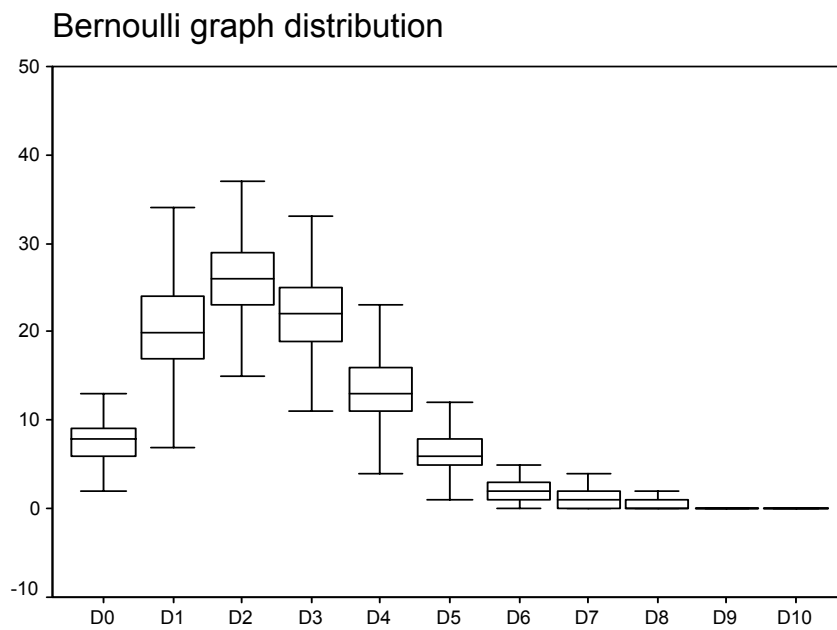
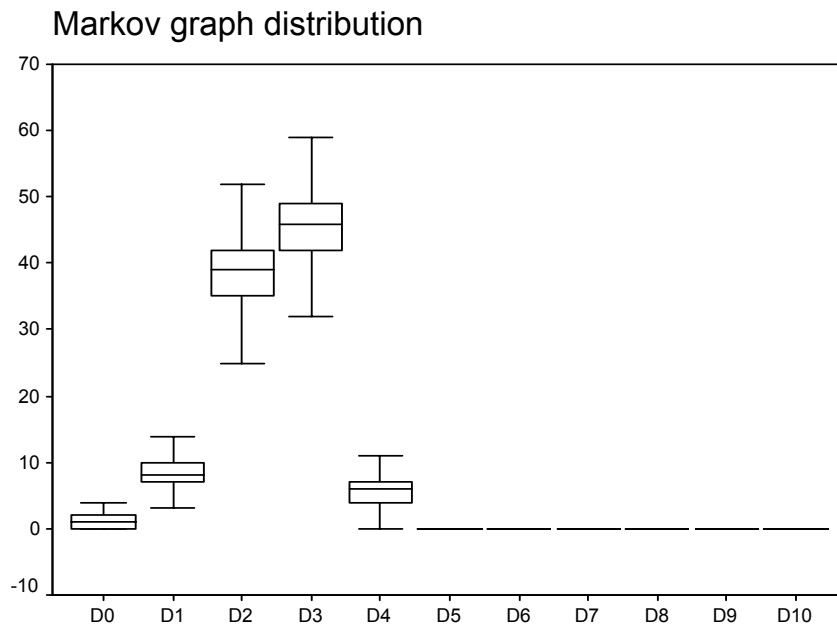


Figure 6
 Long paths: Degree Distribution for Markov graph and Bernoulli samples

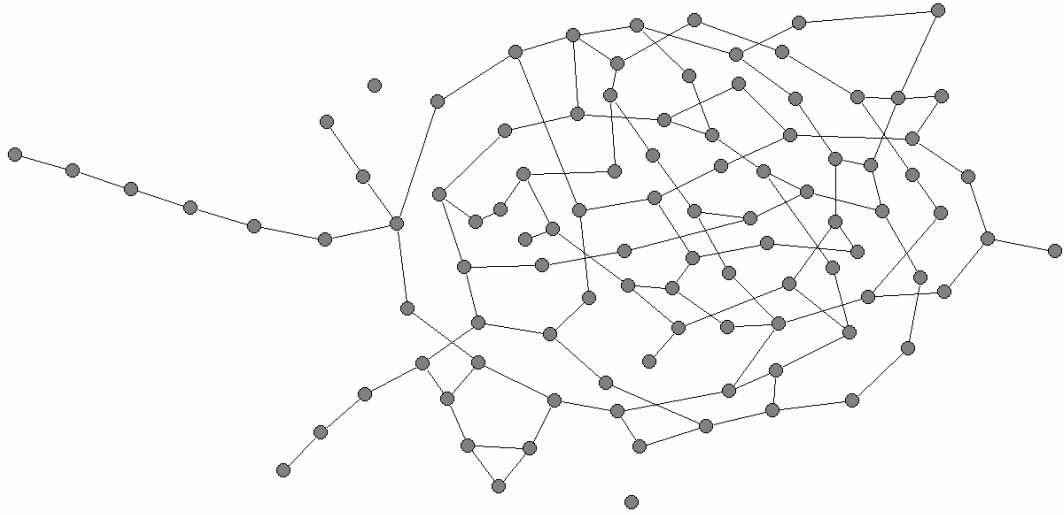


Figure 7
A graph with long median paths

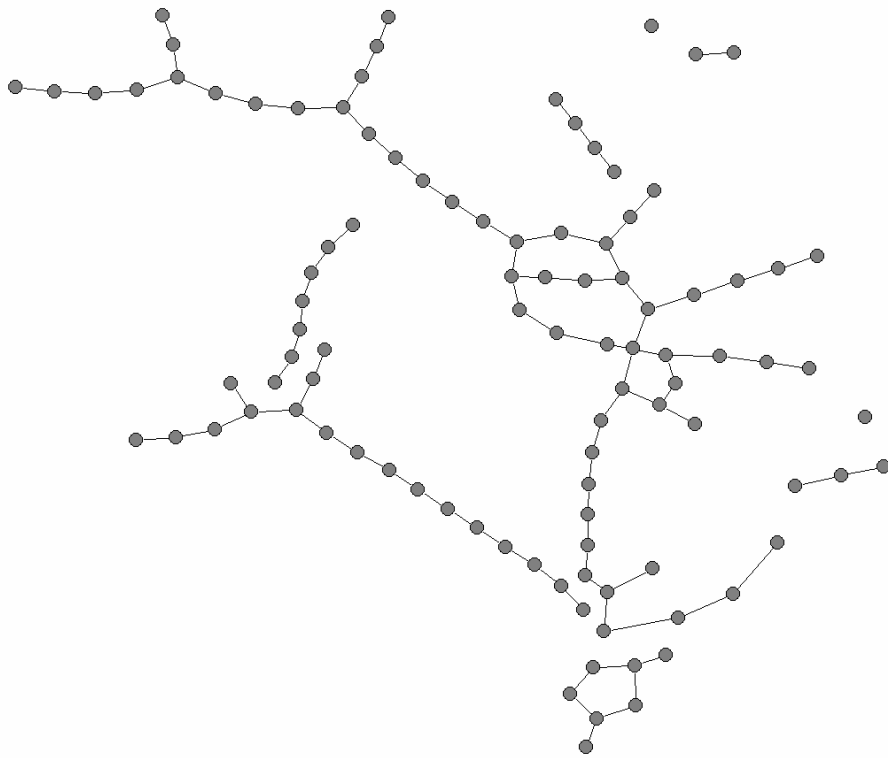


Figure 8
A long path graph with low clustering

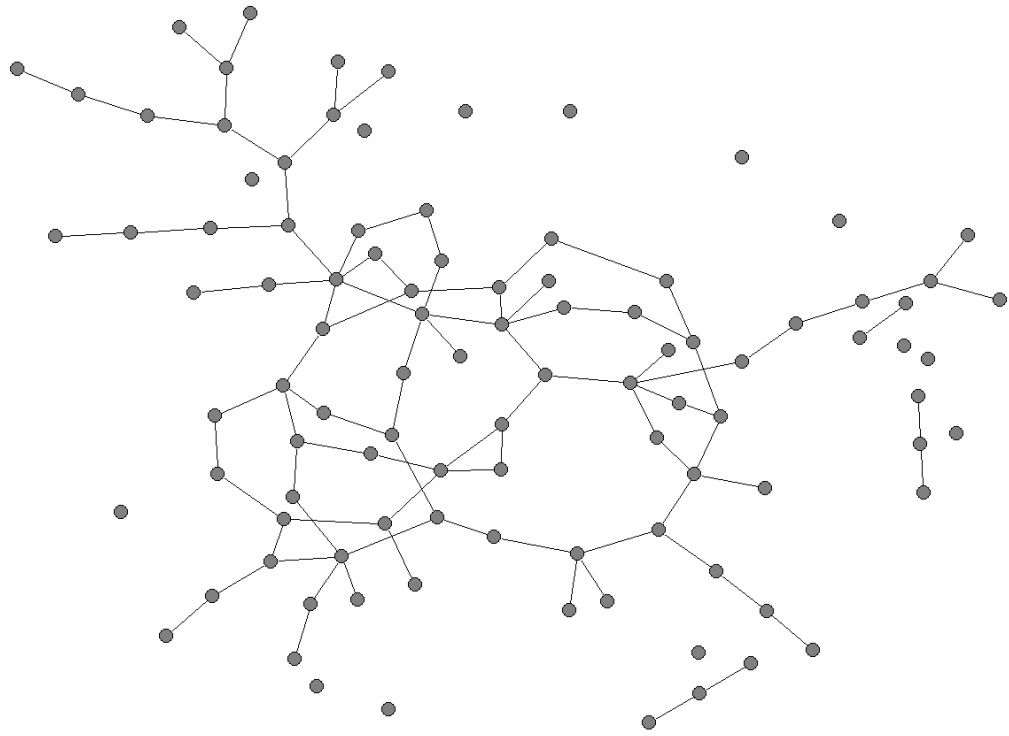
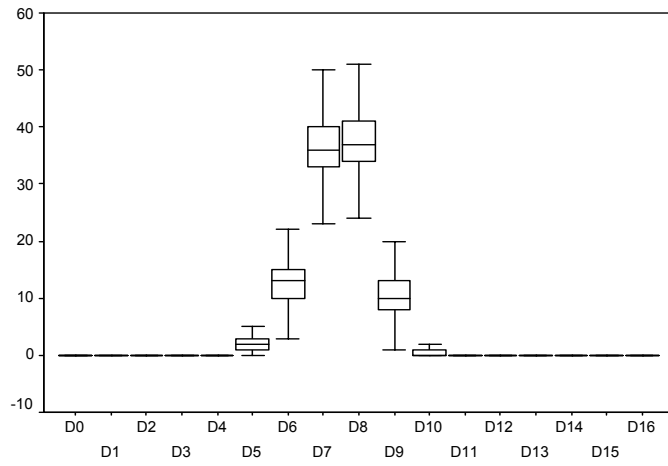
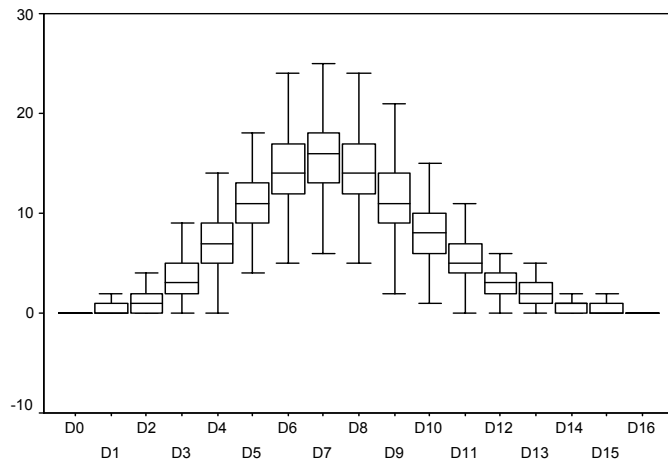


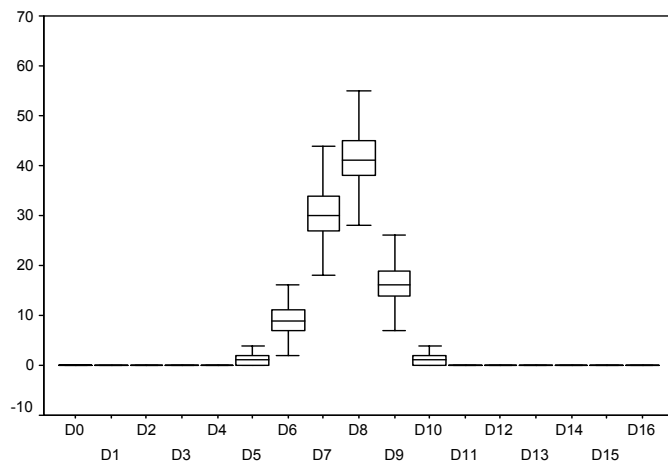
Figure 9
A Bernoulli graph



a. Negative triangle parameter



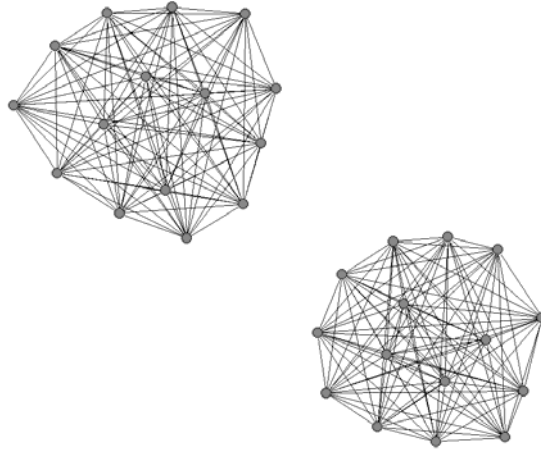
b. Bernoulli sample



c. Zero triangle parameter

Figure 10
Degree distributions for some denser graphs

$\gamma=1$



$\gamma=6$

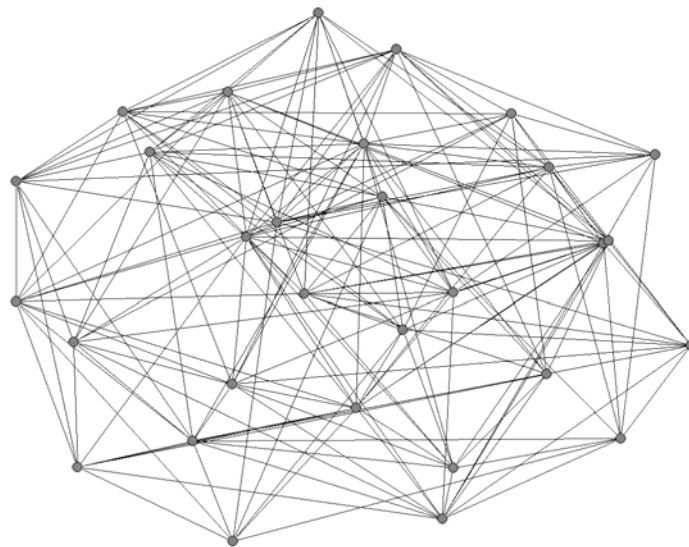


Figure 11

Effects of parameter scaling for two temperatures

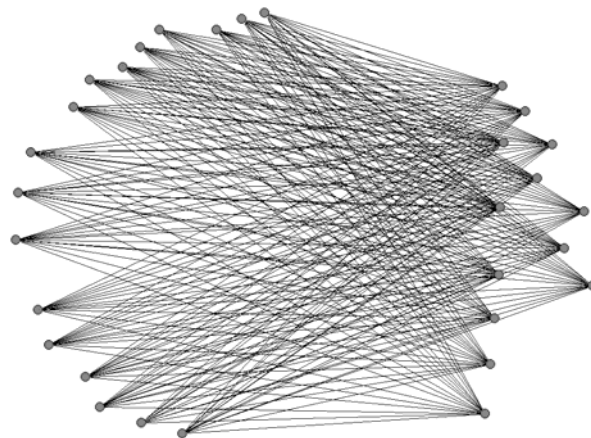
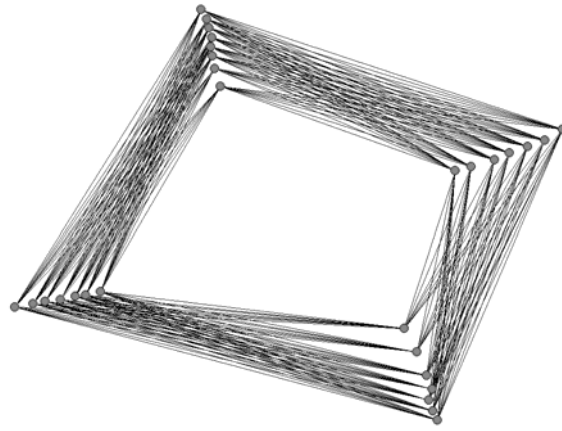


Figure 12
Some “frozen” graph structures:
Top panel: Cyclic structures with 4 structural equivalence classes
Bottom Panel: Complete bipartite graph