A Simple but More Realistic Agent-based Model of a Social Network

Lynne Hamill and Nigel Gilbert

Centre for Research in Social Simulation
University of Surrey, Guildford, Surrey, GU2 7XH, United Kingdom
lynne.hamill@surrey.ac.uk, n.gilbert@surrey.ac.uk

Abstract. None of the standard network models fit well with sociological theory. This paper presents a simple agent-based model of social networks that have fat-tailed distributions of connectivity, that are assortative by degree of connectivity, that are highly clustered and that can be used to create a large variety of social worlds.

Keywords: social networks, personal networks, agent-based models

1 Introduction

For many social simulation models, an underlying network model is required. There are currently four basic types of models of networks: regular lattice, small-world, scale-free and random. However, while these models accurately reflect some real networks, they do not seem to be very good models of *social* networks.

Models created by random linking have been analysed since the mid-twentieth century, starting with Erdos and Renyi [1:12]. Yet social networks are not random: for example, similar people link with others who are similar [2], although Aiello et al's recent analysis of phone call data [3] suggested that, at the very large scale, random patterns may appear.

In 1999 Barabasi & Albert [4] proposed a scale-free network model created by preferential attachment, in which new nodes link to those that already have many links. But this, too, does not in general apply to social networks, the only exception found in the literature being sexual partners in Sweden [5]. People do not necessarily know who has many links and even if they did would not necessarily want to link to them, or the 'target' may not want to reciprocate. The failure of Milgram's and subsequent 'small world' experiments suggest that people lack a global view because the majority were unable to find paths to the targets ([6], [7]). However, an important characteristic of this scale-free model is that the cumulative degree of connectivity follows power laws, often called "fat-tailed" distributions because there are more nodes with high connectivity than is found in the random model. This "fat-tail" accords with the social world, where evidence suggests that a few individuals are very well-connected. For example, Fischer [8: 38-9] found that while the average size of personal networks was 18, the number varied from 2 to 67. (By personal network, we mean egocentric network in contrast to a social network, which is the aggregation of personal networks, the whole set of social relationships.) Recently it has been suggested that another key feature of social networks that distinguishes them from other networks is assortivity of the degree of connectivity i.e. those with many links are linked to others with many links ([1: 555], [9]). Yet the scale-free model generates a hub-and-spoke pattern which is not assortative. Furthermore, clustering is not high although it is a noted feature of real social networks. For example, Wellman's work suggested it averaged 33% among close associates, often kin, with a fifth having a density exceeding 50% [10: 80-82].

At the other extreme is a regular lattice, a grid, often found in cellular automata models. These are characterised by high clustering. In 1998 Watts & Strogatz [11] discovered that a few random re-wirings of a regular lattice produced a model with high clustering and short paths which they labeled a 'small-world'. In effect, the small world model inherits its clustering from the regular lattice and its short paths from a random model [12: 105] However, it is not clear how this 'rewiring' would be caused in social networks. Watts [13: 86] suggested mobile phones create a small world because they enable people to contact someone "chosen at random from the entire network". But the prime use of mobile phones is to increase connectivity with those we already know (e.g.[14]). Newman, Barabasi and Watts [1: 292] argued that: "the small-world model is not in general expected to be a very good model of real networks, including social networks" and Crossley [15] concurred. In particular, the small-world model does not produce nodes with high degrees of connectivity or assortativity.

Pujol et al [16] concluded that the small world and scale free models are based on "unrealistic" sociological assumptions. However, they based their critique on social exchange theory which implies that people weigh the costs and benefits of social relationships. This is highly contentious among sociologists (see e.g.[17]). A model that does not rely on such strong sociological assumptions is needed.

To sum up: none of the standard network models seem to be appropriate for social networks because these tend to contain a few very well-connected people as found in the scale-free attachment model but not the small world, and the high clustering found in the small world model but not the scale-free model. Neither model is assortative. What is needed, it appears, is something between the two and which is assortative.

Furthermore, as Gilbert [18] noted, social simulation models have assumed that the maintenance of social networks are costless, which of course in reality they are not. As has been observed, there are cut-offs in real networks for this very reason [11, 19, 20]. Thus any model should limit the size of personal networks because of the costs to individuals of maintaining them. But the model should also permit the size of personal networks to vary, unlike, for example, [21].

In addition, a model of a social network should:

- create relationships between those who are physically proximate and have similar characteristics (homophily)
- create relationships that are reciprocal: if A knows B, B knows A
- create some very well connected individuals to provide short cuts
- permit modelling of ties of different strengths.

This paper presents an agent-based social network model, with weak but sociologically realistic assumptions that meet these criteria. The main inspiration has come from Watts et al [21] in which "the probability of acquaintance between individuals" "decreases with decreasing similarity of the groups to which they belong". In their model, by tuning a single parameter, they could create a "completely homophilous world or isolated cliques" or at the other extreme "a uniform random graph in which the notion of individual similarity or dissimilarity has become irrelevant". Newman et al [1:292] suggested that this model is "possibly moderately realistic…based on a hierarchical division into groups". The idea of grouping is not new. Pool & Kochen [22] used "stratum". They also used the idea of social space, as in effect did Wasserman & Faust [23:385-7] who used multidimensional scaling to map people's relative positions so that those "that are more similar to each other are closer in the space". More recently, Edmonds [24] argued that it is important to bring together physical and social spaces and the only way to do that is by using agent-based models. Models similar to that proposed below have been reported in the physics literature e.g. [20] and [25].

Section 2 describes the basic structure of the model and Section 3 extends it. Section 4 concludes. The models were implemented using NetLogo version 4.0.2 [26] and can be found at: www.hamill.co.uk/misc/essa08.zip.

2 Basic Structure of the Model

The setting for the model is what could be called a social map. While a geographical map shows how places are distributed and linked, the social map does the same for people. Thus two individuals will be located close to each other on this map if they are close socially: the closer the agent, the stronger the tie. Social distance can be defined as the acceptable degree of interpersonal closeness [27: 191] and assessed according to numerous characteristics, including geographical distance. At one extreme, if homophily is ignored, the social map collapses into a geographical map with distance measured in miles or travel time.

The proposed model is based on the concepts of social circles, an idea dating back to at least Simmel [28] in the early twentieth century. The term circle was then used as metaphor. Yet a circle has a very useful property in this context: the formal definition of a circle is "the set of points equidistant from a given point", the centre [29: 246]. The circumference of a circle will contain all those points within a distance set by a radius and creates a cut-off, limiting the size of personal networks. For a given distribution of agents across the map, a small radius – which will henceforth be called the 'social reach' – can create a disconnected, geschellshaft-type society; a large social reach, a connected, gemeinshaft-type society. Alternatively, if the social reach is very small, it can be said to replicate McPherson et al's 'confidants' [30]: if larger, it becomes a model for larger networks as described by, for example, [8].

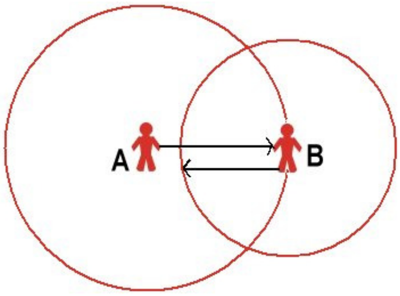
The use of two dimensions of course imposes limits on the structure of the network. This is best illustrated by considering an example of four agents **A**, **B**, **C** and **D**. If **A**, **B** and **C** are linked and **A**, **B** and **D** are also linked, then the distance between **C** and **D** is fixed. For relatively abstract modelling, this constraint is, we suggest, acceptable because social networks are created by kinship and homophily and because the number of links are limited by the social reach:

- kinship: for example, **A** and **B** could be the parents of **C** and **D**.
- homophily implies that friends are likely to know one another.
- social reach: **C** and **D** may be too far apart to be linked.

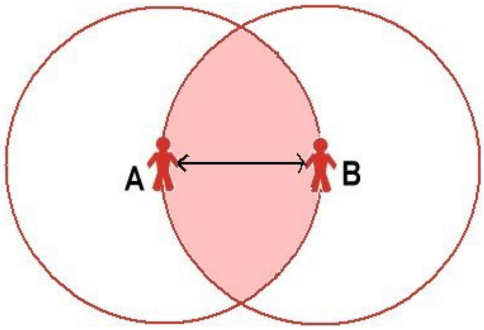
Furthermore, Heider's theory of balance [27] and the simulation results of Wang & Thorngate [31] support the idea of agents creating groups of this kind. However, if **C** and **D** both know a fifth agent, **E**, who does not know **A** and **B**, it may not be possible to show both links on a two dimensional map although it would be using three dimensions. But to provide insights, models must be simple. This model is intended to reproduce certain key features of social networks, and to do that simplifications have to be accepted.

Agents are only permitted to link with agents who can reciprocate; in other words, alters whose reach includes ego. If **A** were to have a bigger reach than **B** then **B** could be in **A**'s circle but not vice-versa, implying that **A** 'knows' **B** but **B** does not 'know' **A** as illustrated in the left hand panel of Figure 1. Now there may be all sorts of asymmetries in the relationship between **A** and **B** and in their communication pattern, but they must in some sense both 'know' each other. This definition thus excludes, for example, 'knowing' a celebrity seen on TV where there is no reciprocal contact. The simplest way to achieve this is for all agents to have the same reach, as shown in right hand panel of Figure 1. But this is not essential, as will be explained later. However, we start by exploring the properties of the simplest model, where all agents have the same social reach.

Ceteris paribus, the size of personal networks will vary with the reach: the larger the reach, the larger the size of the personal network. To look at personal networks larger than 'intimates', a large number of agents are required. The simulations presented in this paper use 1,000 agents, meaning that there is a total of almost half a million possible undirected links ($1000 \times 999 / 2$). These agents are randomly distributed across a non-bounded grid of just under 100,000 cells. All reported results are based on 30 runs.



(a) No reciprocity: different social reaches: **A** knows **B** but **B** does not know **A**.



(b) Reciprocity with the same social reach.

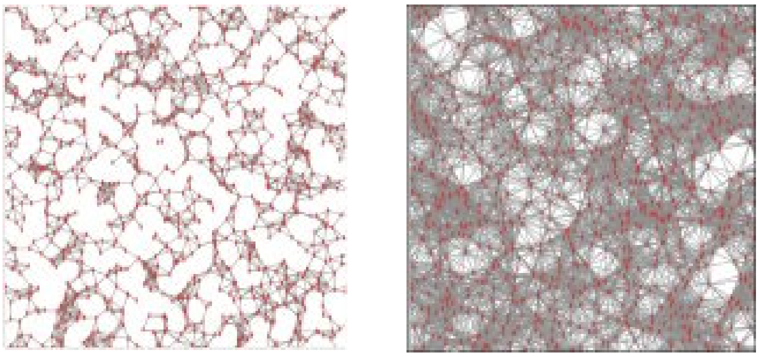
Figure 1. Reciprocity and social reach

The minimum number of steps, the path length, is determined by the size of the 'world' and the social reach. For example, a world of about 100,000 cells is created by a wrapping grid of 315 by 315 cells. An agent sitting at the centre of this grid will be at least 157 units from the edge (314/2). But the diagonal provides the furthest distance and by Pythagoras's theorem, this diagonal will be 222 units. So if the social reach were set at 40, it would take a minimum of six steps to reach the farthest point, consistent with the famous six degrees ([6], [13]). However, this optimum may not be attainable, depending on how agents are distributed and there is no guarantee that agents could find it.

Clustering is determined by the overlap of circles. If two individuals are located very close to each other on the map, their circles will almost coincide and they will know most of the same people. At the other extreme, if an individual is located on the circumference of another's circle, the overlap will cover 39 percent of the area of each circle [29: 250]: this is shown by the shaded area in right hand panel of Figure 1.

Although the personal networks of all the agents have the same social reach, the numbers in each personal network will vary due to the randomness.

- Setting the social reach at 15 produces personal networks ranging from zero to 20 with an average of 7. With this small reach, many agents have few, or even, no links. In total there are some 3½ thousand undirected links giving a whole network density of 0.7 percent. This is illustrated in the left hand panel of Figure 2: the (red) dots indicate agents and the (grey) lines, the links between them.
- Setting the social reach at 30 produces personal networks ranging in size from 11 to 52 with an average of 28. Now there are some 15 thousand undirected links giving a whole network density of about 3 percent. This is illustrated in the right-hand panel of Figure 2.



Social reach = 15 Social reach = 30

Figure 2. How networks vary with the size of the social reach. (Red nodes, grey links.)

Hermann et al [25] suggested that in such a spatial model, as the number of nodes increases and the reach reduces, the connectivity distribution tends towards a Poisson. Figure 3 shows how the connectivity of the nodes changes as the social reach is increased. For a social reach of up to about 30, the connectivity of nodes follows a Poisson distribution (the mean is the same as the variance) but beyond that, the mean tends to exceed the variance. The Poisson distribution implies that the network is random [1: 233], which is to be expected as the agents are distributed randomly across the social map.

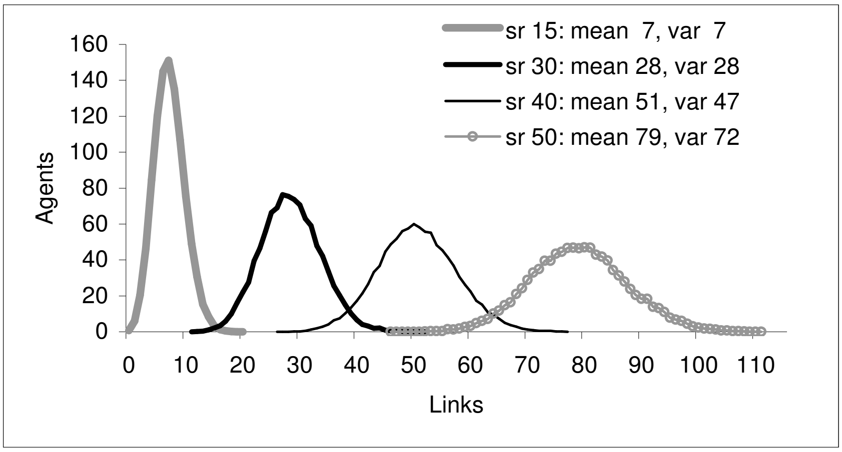


Figure 3. Degree of connectivity by social reach (sr).

Intuition suggests that this model should produce assortative networks because those in densely populated regions will tend to have many links, as will those to whom they are linked (and Hermann et al [25] agree). This proves to be the case. The relationship between an agent's degree of connectivity and the average for those to which it is linked is positively correlated as indicated by the Pearson correlation coefficients (following [20]). For example, for a social reach of 30, the correlation coefficient averages 0.83 (sd 0.03). (A typical example is shown in Figure 4.) For the lower reach of 15, it is 0.78 (sd 0.03) and for the higher reach of 50, 0.84 (sd 0.05).

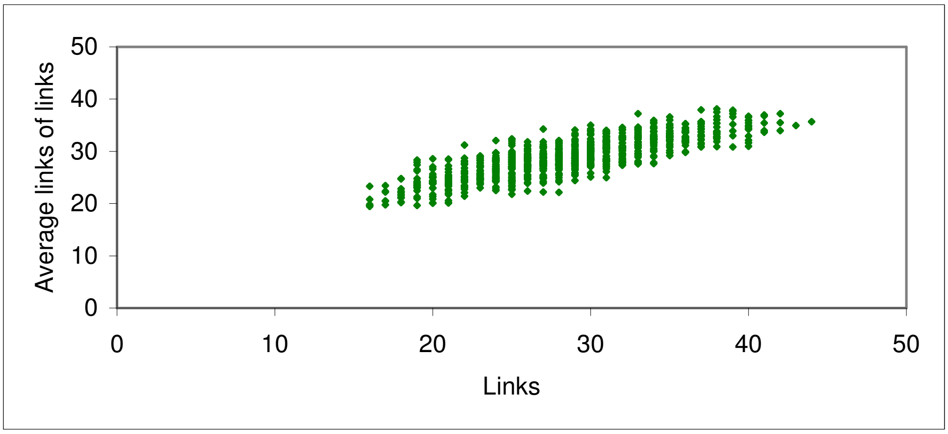


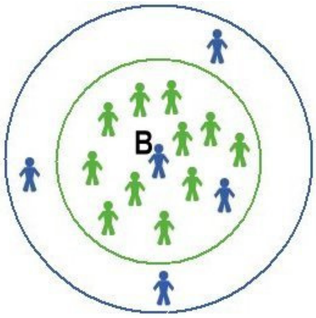
Figure 4. Assortativity: typical example of correlation between degrees of connectivity: social reach of 30.

3 Extending the Model

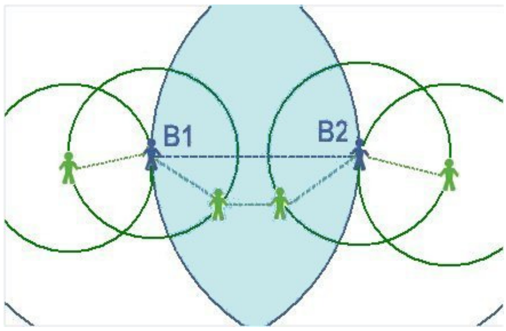
The simple one circle model is inflexible, the only parameters being population density and the size of the social reach, and while assortative it does not produce a fat-tailed distribution of connectivity and the resulting short cuts. Also all agents will know at least 39 percent of their neighbours' neighbours. These issues can be addressed by splitting the population in two and giving one group – let's call them Blues – a larger social reach than the other – let's call them Greens – but only permitting links between those who can reciprocate. Thus Green agents link only to other agents – Greens and Blues – within their small reach. But Blues with a large reach not only link to the Greens within their smaller reach but also Blues within the larger reach (see left hand panel of Figure 5). There are therefore two more parameters to adjust: the percentage of Blues with the larger social reach and the size of that reach.

For Blues the sharpness of the discontinuity created by the cut off is reduced, blurring the edge of their personal networks, and also reducing the clustering. For example, a Blue may share no Greens with a neighbouring Blue. For the Greens, a Blue in their personal network will provide a short cut to agents beyond their reach. In this way, a hierarchy is created. These features are illustrated in the right hand panel of Figure 5.

The two-circle model in effect adds together two Poisson distributions and as a result produces a distribution with larger variance, a fatter tail. Of course, if the percentage of Blues is small or if there is little difference between the two social reaches, the results from the two circle model will tend towards that of the one circle model.



B, a Blue, links with everyone in the smaller circle plus other (darker) Blues within the larger circle.



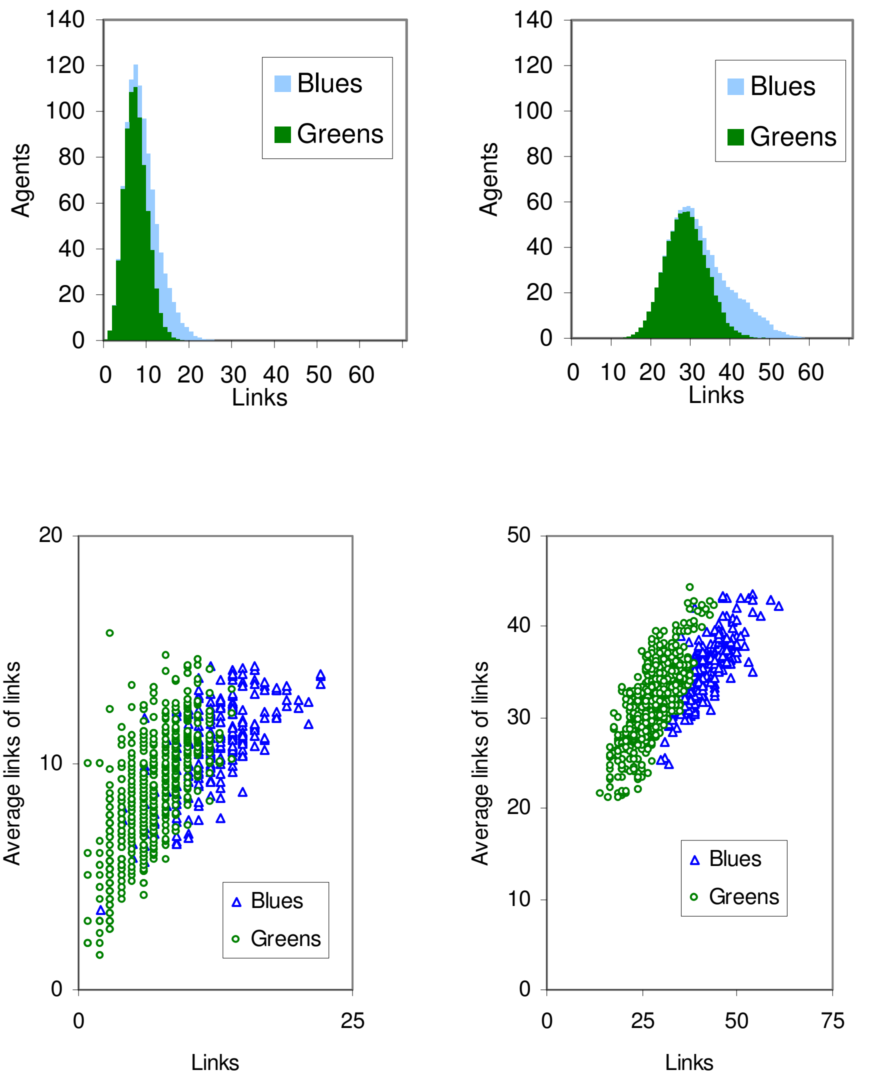
Links between Blues **B1** and **B2** creates short-cuts and, for Blues, reduces clustering. Shaded area indicates overlap between the Blues' circles.

Figure 5. Two circle model

Figure 6 shows results for a pair of two-circle models with 25 percent Blues. In the first case (illustrated in the left column of Figure 6) the well-connected Blues have a social reach of 30 while that of the Greens is only 15; in the second (illustrated in the right column), the Greens have a social reach of 30 while that of the Blues is 50. Three results emerge:

- The size of personal networks of the better connected Blues is constrained by the relatively few Blues. In both cases the average personal network of the Greens is the same as if all agents had their social reach, but that of the Blues is much lower than would be expected if all agents had the their larger reach. For example, in the first case, the Greens with a social reach of 15 have an average network of 7, which is the average if all agents have a reach of 15 (see Figure 3). In contrast although the Blues have a reach of 30 their personal networks average only 12, far fewer than the average of 28 that is found when all agents have a reach of 30.
- The better-connected Blues add a "fat tail" to the distribution of connectivity: in both cases, the variance is significantly greater than the mean and the distributions spread more widely than a Poisson, although for the Greens and Blues separately, the distribution of connectivity is approximately Poisson. In both cases about half the links involve at least one Blue even though only a quarter of the agents are Blues.
- Overall the assortativity is slightly weaker than in the one circle case. The correlation coefficients are still high (see bottom row) but are lower than in the single circle case because although the Blues are well-connected to other Blues, more than half of their links are to the less well-connected Greens. (Typical examples are illustrated in the middle row of Figure 6.)

Greens' reach = 15: Blues' reach = 30: **Greens' reach = 30: Blues' reach = 50:**



Greens' reach = 15: Blues' reach = 30:

	Blues	Greens	All
Average circle sizes			
	12	7	8
Connectivity correlation			
	0.69	0.69	0.71
% links	B-B	B-G	G-G
	21	32	47

Greens' reach = 30: Blues' reach = 50:

	Blues	Greens	All
Average circle sizes			
	41	28	32
Connectivity correlation			
	0.78	0.79	0.75
% links	B-B	B-G	G-G
	16	34	51

Figure 6. Examples of two circle models: Blues 25 percent.

Adding a third circle increases the flexibility of the model still further. To illustrate this, we offer an example that demonstrates how it is possible to create two very different types of networks. In both cases, the reaches are set at 30, 40 and 50 but in the 'elitist' case agents are distributed between the three groups 75/20/5 per cent while in the 'democratic' case they are split more evenly at 40/30/30 per cent. As before, agents can only link to those who are able to reciprocate. The results are shown in Figure 7. In both cases the distribution of connectivity is much wider than a Poisson distribution, notably so for the democratic case. The whole network densities are around 3 percent. However, in the elitist case, 6 out of 10 links are within group compared to only 4 out of 10 in the democratic example. The personal network size of the better connected groups is constrained by them being minorities. The least well-connected, with a social reach of 30, have average personal networks of the same size as if all agents had a reach of 30, i.e. about 28 as shown in Figure 3. But the best connected groups have an average of 44 even though a social reach of 50 for all would produce an average of 79; and the middle group, with a reach of 40, has an average of 40 instead of 51.

Whether or not this flexibility is required and whether the additional complication is justified compared to the two-circle model will depend on the questions to be addressed by the modelling. For instance, one might adopt a three circle model if there were three distinct groups involved in the process being modeled, e.g. those who are globally mobile, nationally mobile or only regionally mobile.

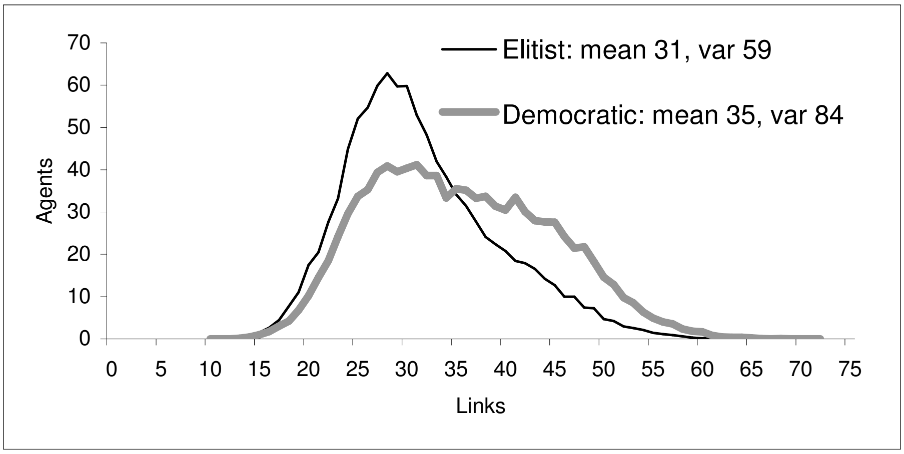


Figure 7. Example of degrees of connectivity in a three-circle model.

This paper has not addressed the dynamics but we suggest just two processes be used to maintain the basic structure while allowing change at individual level: one to reflect demographic changes (ageing and death) and the other, geographical and social movement.

4 Conclusion

We have presented a simple agent-based model of a social network that meets all the criteria set out in Section 1:

- the fat-tail of the degree distribution (compared to a Poisson distribution) shows that the model includes some very well-connected agents, creating short-cuts
- the model is assortative: well-connected agents tend to be connected to other well-connected agents
- the overlap between the social circles ensures clustering
- the random distribution across the social map, together with varying the size of the social reaches, ensures that the size of personal networks varies between individuals
- the size of personal networks is limited by the cut off imposed by the social reach that defines the circles
- drawing circles on the social map creates relationships between those who are physically proximate and with similar characteristics
- relationships are reciprocal
- the use of circles also potentially permits the modelling of ties of different strengths.

The two-circle model offers considerable advantages over the one-circle model. Three (or more) circles can be used but this complication may not be necessary. The model can be used to look at networks of close associates or wide groups, to model gemeinshaft or geschellshaft.

Acknowledgements

This work was supported by Microsoft Research through its European PhD Scholarship Programme.

References

[1] Newman, M., Barabasi, A-L. & Watts, D.J. The Structure and Dynamics of Networks. Princeton University Press. (2006)

[2] McPherson, M., Smith-Lovin, L. & Cook, J.M Birds of a Feather: Homophily in Social Networks. Annual Review of Sociology. 27. 415-444 (2001)

[3] Aiello,W., Chung F. & Lu, L. A random graph for massive graphs. In [1: 259-258] (2000/2006)

[4] Barabási A-L. & Albert, R. Emergence of Scaling in Random Networks. Science, New Series, Vol. 286, No. 5439., pp. 509-512 (1999) (In [1] 349-352)

[5] Liljeros, F., Edling, C. Amaral, L., Stanley, H.E., & Aberg, Y. The web of human sexual contacts. Nature. Vol 411, 21 June. 907-8. In [1: 227-8] (2001/2006)

[6] Travers, J. & Milgram, S. An Experimental Study of the Small World Problem. Sociometry. Vol (32) 4. pp 425-443(1969)

[7] Dodds, P.S., Muhammad, R. & Watts, D.J. An Experimental Study of Search in Global Social Networks. Science, 8 August pp 827-829. (2003)
[8] Fischer, C.S. To Dwell Among Friends. University of Chicago Press. Chicago. (1982)
[9] Newman, M. & Park, J. Why social networks are different from other types of networks. Physical Review E. Vol 68, 036122 (2003)
[10] Scott, J. Social Network Analysis: A Handbook. Sage. London (1991)
[11] Watts, D. J. &. Strogatz, S. H. Collective dynamics of 'small-world' networks. Nature. Vol 393 4 June (1998)
[12] Dorogovtsev, S. & Mendes, J. Evolution of Networks: From biological nets to the Internet and WWW. OUP. (2003)
[13] Watts, D.J. Six Degrees: The Science of a Connected Age. Vintage. London (2004)
[14] Vincent, J. Emotion and Mobile Phones. In Nyeri, K. (ed) Mobile Democracy, Essays on Society, Self and Politic*s*. Passengen-Verlag. Vienna. 215-224 (2003)
[15] Crossley, N. Small-world networks, complex systems and sociology. Sociology. Vol 42, No. 2, pp 261-277 (2008)
[16] Pujol, J. M., Flache, A., Delgado, J. & Sanguesa, R. How Can Social Networks Ever Become Complex? Modelling the Emergence of Complex Networks from Local Social Exchanges. J. of Artificial Societies and Social Simulation 8(4)12 <http://jasss.soc.surrey.ac.uk/8/4/12.html>. (2005)
[17] Harper, R. Are mobiles good for society? In Nyiri, K, (ed), Mobile Democracy: Essays on Society, Self and Politics, Passagen Verlag, Vienna, 185-214 (2003)
[18] Gilbert, N. Putting the Social into Social Simulation. Keynote address to the First World Social Simulation Conference, Kyoto. (2006)
[19] Amaral, L. A. N., Scala, A. Barthelemy M., & Stanley H. E. Classes of small-world networks. PNAS. Vol. 97. No. 21 11149–11152 (2000)
[20] Barthélemy, M. Crossover from scale-free to spatial networks Europhys Lett, 63 915-912, (2003)
[21] Watts, D.J., Dodds, P.S., & Newman, M. Identity and search in social networks. Science. Vol 296. 17 May.1302-1305 (2002)
[22] Pool, I.S. & Kochen, M. Contacts and Influence. Social Networks. 1. 5-51 (1978/9). Also in [1:83-129].
[23] Wasserman, S. & Faust, K. Social Network Analysis. Cambridge University Press. (1994)
[24] Edmonds, B. How are physical and social spaces related? In Billari, F.C., Fent, T.; Prskawetz, A. & Scheffran, J. (eds.) Agent-Based Computational Modelling Springer. (Downloaded on 10 March 08 from cfpm.org/cpmrep127.html) (2006)
[25] Herrmann, C. Bathélemy, M & Provero, P. Connectivity distribution of spatial networks. Phys. Rev. E 68 026128 (2003)
[26] Wilensky, U. NetLogo. http://ccl.northwestern.edu/netlogo/. Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL. (1999)
[27] Heider, F. The Psychology of Interpersonal Relations. Wiley. New York. (1958)
[28] Simmel, G. The Number of Members as Determining the Sociological Form of the Group.I. American J. of Sociology. Vol 8, No.1. 1-46. (1902)
[29] Weisstein, E. CRC Concise Encyclopedia of Mathematics. CRC Press. Boca Raton.(1998)
[30] McPherson, M., Smith-Lovin, L. & Brashears, M. Social Isolation in America: Changes in Core Discussion Networks over Two Decades. American Sociological Review. Vol 71 353-375 (2006)
[31] Wang, Z. & Thorngate, W. Sentiment and social mitosis: implications of Heider's Balance Theory. J of Artificial Societies & Social Simulation vol. 6, no. 3 <http://jasss.soc.surrey.ac.uk/6/3/2.html> (2003)