On-line Appendix for "Social Networks and Collective Action": Detailed Model Specification and Summary of Network Regularities

David A Siegel[1]

Each of the seven networks described in this paper is created during the initialization of the model. For networks with a stochastic component to their creation, every realization of the model thus uses a different realization of a particular network type. Networks are created according to the following rules:

- **Fully-Connected**: every possible connection between pairs of individuals is made.

- **Random**: each pair of individuals is randomly checked once to see if a connection between them will be made. The chance that each connection will be made is given by the parameter *Random Density*.

- **Ring**: individuals are arrayed in a ring, and connected to a number of other individuals to both sides of them equal to the parameter *Connection Radius*. A *Connection Radius* of 5 thus indicates a connectivity of 10.

- **Small-World**: First, a ring substrate is created according to the rules for creating a Ring network with parameter *Connection Radius*. Then, each network tie has some chance of being severed and reconnected randomly to any other node in the network according to the parameter *Rewire Probability*. Varying this probability takes the network from a ring (Rewire Probability 0) to purely random (Rewire Probability 1), while roughly maintaining the same number of ties per person. As described elsewhere (Strogatz 2001), even a slight increase in the rewire probability allows for much faster spreading of things like information or disease relative to a Ring network. The numerical values on the x-axes of Small-World plots correspond to values of the *Rewire Probability*; each line corresponds to a different value of the *Connection Radius*.

- **Village**: First, the population is split up into an array of equally-sized subsets called villages or cliques that are of size *Village Size*. If the population does not divide evenly in this way, then any left over individuals are placed into a last, smaller village. Second, every possible connection within each village is made. Finally, each individual has some probability, called *Far Probability*, of being connected to each other individual outside

---
[1]Assistant Professor, Department of Political Science, Florida State University, Tallahassee, FL 32306-2230; e-mail: dsiegel@fsu.edu; phone: 850-645-0083; fax: 850-644-1367; web: http://mailer.fsu.edu/~dsiegel/.

of her village. These probabilities are checked twice, so the true probability of any individual's being connected to a particular person outside her own village is equal to twice *Far Probability*. Note that, unlike in the Small-World network, individuals within a Village network may have different degrees of connectivity, though the average connectivity is the same for all.

- **Opinion-Leader**: First, each individual is assigned a number of ties according to the distribution $p(k) \propto k^{-\gamma}$, where k is the number of ties a particular individual has. The parameter γ thus determines the characteristics of the network, with smaller values corresponding to greater overall connectivity due to the presence of a greater number of elites. The x-axes of Opinion-Leader plots correspond to γ. Then, connections are made between individuals at random, beginning with those assigned the most number of ties in the first step, until either all connections assigned in step one are accounted for, or no additional connections can be made without adding to the assigned number of ties of some person. This method generally reproduces the distribution of ties fairly accurately, as measured by the difference between realized and assigned ties, summed across all individuals, but it is less accurate for very small values of γ where there are many elites, each assigned many connections. Thus, small values of γ do not display exactly the distribution of ties given above. Finally, the parameter *B Correlation* dictates how motivations are correlated with network position. These networks are also known as Scale-Free networks in the literature, and they have recently come into popularity due to studies of internet susceptibility and the like (Albert et al. 2000).

- **Hierarchical**: First, I create the skeleton of the hierarchy according to the parameter *Expansion Rate*. One individual is placed at the top, and each individual in the network is connected to a number of individuals below him equal to *Expansion Rate*, continuing until no more individuals are left in the population. Thus, while each level of the hierarchy before the last one contains a number of individuals equal to a power of *Expansion Rate*, the last level may have fewer than this if the total population does not divide appropriately. Once this skeleton is created, each potential tie between individuals within the same level has a probability equal to *Level Connection* of being made. Unlike as in the Village network, these probabilities cannot uniformly be converted to an average number of connections between levels due to the differing sizes of levels; thus the x-axes of Hierarchical Plots directly display *Level Connection*. Finally, the parameter *B Correlation* dictates how motivations are correlated with network position.

Each realization of the model carries out the following sequence of events: 1) The network is initialized in the manner described in the preceding paragraph, and net internal motivations are distributed. Uncorrelated motivations are distributed randomly to individuals throughout the network. To initialize correlated motivations, a thousand values are first drawn from the appropriate normal distribution, and then ordered according to the type of correlation desired. For Hierarchical networks, high (or low) motivations are assigned to those at the top first, decreasing (increasing) as one moves down the network. For Opinion-Leader networks, individuals are ordered by connectivity, and then matched with motivations beginning at the top (or bottom) of the ordered list of motivations and moving down (or up). 2) Individuals update external motivations according to the equation $c_i = -(1 - lpr_i)$ and participate in that period if $b_i + c_i > 0$. 3) Participation rates are (re)calculated, both population-wide and locally for each individual. The latter corresponds to what each individual observes within her local social network. 4) If no individual has changed her participation status for fifty consecutive periods, the realization of the model ends and final data for that realization are recorded. After 1000 independent realizations for a given model parameterization are completed, means and standard deviations of equilibrium participation levels over these realizations are calculated. All simulation data were obtained via a JAVA program coded by the author. A fraction of the simulation data taken appear in the figures; additional data can be obtained from the author upon request.

I also explored a variant model in which the participation decision in 2) was replaced with a stochastic version: $b_i + c_{i,t} > \varepsilon_{i,t}$ with $\varepsilon_{i,t}$ a random variable. This changed little of the analysis, save that it encouraged higher levels of participation than seen in the paper. Misperception (or probabilistic behavior) creates moments of high participation that feed on themselves to become self-sustaining.

The following is a summary of the model's parameters (default values used in figures in parentheses if applicable), the effect increasing them has upon steady-state (equilibrium) participation rates, and the general impact of network structure on participation rates.

N (1000): Non-monotonic: more likely to decrease participation if participation is already unlikely, more likely to increase participation if it is likely.

$b_{mean}(0.6)$: Positive

$b_{stdev}(0.23, 0.25, 0.3)$ (weak, intermediate, strong): Generally Positive, though negative for very high b_{mean}

Network Type: By network (network parameters in parentheses)

- **Random** (*Random Density*): Efficient at behavioral spread; fails to achieve high levels of participation outside of the strong motivation class. Increasing *Random Density*

increases participation in the strong class and has a non-monotonic effect otherwise.

- **Ring** (*Connection Radius*): Inefficient at behavioral spread, but with high connection radii can achieve high levels of participation thanks to its encouraging of small enclaves. Increasing *Connection Radius* increases participation in the strong class and has a non-monotonic effect otherwise.

- **Small-World** (*Connection Radius, Rewire Probability*): Combines aspects of Random and Ring networks to achieve higher levels of participation more efficiently than either. Increasing *Rewire Probability* increases participation in the strong class and has a non-monotonic effect otherwise.

- **Village** (*Village Size, Far Probability*): Similar to the Small-World network, but with less efficient behavioral spread and slightly lower maximal participation levels. Connections between Villages are extremely important. Increasing *Village Size* and *Far Probability* increases participation in the strong class and has a non-monotonic effect otherwise.

- **Opinion-Leader** (γ, *B Correlation*): Relies on centrally-located opinion leaders to spread behavior; generally achieves low levels of participation, but can attain near-total participation when internal motivations are positively correlated. Increasing γ can have varied effects, depending on correlation of motivations.

- **Hierarchy** (*Expansion Rate, Level Connection, B Correlation*): Very dependent on type of motivation correlation. Uncorrelated it behaves much like a Village, but with lower levels of participation. Negatively correlated it yields little participation, unless the bottom level is large and inter-level connectivity is substantial. Positively correlated it yields high participation, unless inter-level connectivity is substantial. Increasing *Expansion Rate* and *Level Connection* can have varied effects, depending on motivation correlation.

Further Justification of Method of Model Analysis

The strong dependence of equilibrium participation both on network structure and on the distribution of internal motivations suggests that nonlinearities and even non-monotonicities will be pervasive in the model, and this is borne out in the results. At least two useful approaches to deriving comparative statics in the presence of such complexity are available. Firstly, one can sample randomly across the parameter space and then utilize either regression analysis or explicit computational testing of directional hypotheses at each sampled point

(e.g. Kim and Bearman 1997, Smirnov and Fowler 2007). This sampling technique can be very powerful when a model is not analytically solvable. Secondly, one can attempt to limit the dimensionality of the parameter space to the point where a direct sweeping of the parameter space—essentially evaluating the dependent variable(s) at a number of points fine enough to illustrate nonlinearities—becomes feasible. This too can be a powerful technique, but requires theoretical justification.

I adopt the second approach for two reasons. First, the utility of the sampling approach decreases when one expects important non-monotonicities and conditional dependencies. In such cases one would need to sample sufficiently in the regions of nonlinearity, a difficult task when one does not know *a priori* where these might be. The uniform random samples typically used in such analyses are frequently ill-suited to such cases. Second, specification of regressions and/or directional hypotheses is difficult given a lack of foreknowledge regarding non-monotonicities and conditional dependencies.

Of course, without prior theoretical direction as to the expected nature of the non-monotonicity, we cannot overcome these problems via the second approach, and random sampling techniques may remain our best choice. In the case of this model, however, we do have prior beliefs over the nature of at least a subset of the non-monotonicities and dependencies present in the model. As detailed in Granovetter and Soong (1986) (see also Yin (1998) and Rolfe (2005)), we can expect in general up to three equilibria in an infinite-population threshold model when all individuals update behavior based on all other individuals, absent reverse bandwagon effects. These correspond to regions of low, intermediate, and high participation, with the exact levels of participation being functions of the distribution of thresholds in play. I take advantage of this theoretical finding to break up the space spanned by the parameters (b_{mean}, b_{stdev}), and N into three regions, each corresponding to a different class of distribution of net internal motivations. In the first region, when everyone is connected to everyone else, average levels of participation are low due to a rarity of cascades. In the second, average levels of participation are intermediate, and cascades occur approximately half of the time. Finally, in the third, average participation is nearly complete due to the ease of achieving cascades. I call these three classes of distribution of net internal motivations "motivation classes" for short, and denote them instrumentally as, respectively, the "weak motivation class," the "intermediate motivation class," and the "strong motivation class."

This vastly reduces the parameter space under consideration, allowing analysis to focus tightly on differences in the equilibrium level of participation achieved (the dependent variable) across networks. If a particular network parameterization yields high levels of participation within the weak motivation class, for example, one can attribute the difference

5

solely to the structure of the network. For single-parameter networks like the Opinion-Leader network, I am thus able to derive detailed comparative statics via simulation analysis for each of the three classes. For two- or three-parameter networks, I first map out the behavior of a single parameter, and then use the information so derived to split the parameter space into additional regions, within each of which I again carefully trace the impact of the remaining parameters. Though this can be quite time consuming, even with increases in computational power, the payoff is the ability to describe non-monotonicities and conditional dependencies that would otherwise be difficult to discern. For more on different computational methodologies, see Miller and Page (2007), de Marchi (2005), and Laver and Sergenti (2007). While space constraints limit the number of such results I can report, a fuller accounting may be obtained from the author's website.

References

Granovetter, Mark S. and Roland Soong. 1986. "Threshold Models of Interpersonal Effects in Consumer Demand." *Journal of Economic Behavior & Organization* 7(1):83–99.

Kim, Hyojoung J. and Peter S. Bearman. 1997. "The structure and dynamics of movement participation." *American Sociological Review* 62(1):70–93.

Laver, Michael and Ernest Sergenti. 2007. "Do ideologically intransigent parties affect the policy positions of other parties? And rigorously characterizing output from computational models of party competition." vol. Paper presented at the Annual Meetings of the American Political Science Association, Chicago, Annual Meetings of the American Political Science Association.

de Marchi, Scott. 2005. *Computational and Mathematical Modeling in the Social Sciences.* Cambridge University Press.

Miller, John H. and Scott E. Page. 2007. *Complex adaptive systems: an introduction to computational models of social life.* Princeton, New Jersey: Princeton University Press.

Rolfe, Meredith. 2005. "Conditional choice: social networks and decision-making." Presented at the 2005 Annual Meeting of the Midwest Political Science Association.

Smirnov, Oleg and James H. Fowler. 2007. "Policy-Motivated Parties in Dynamic Political Competition." *Journal of Theoretical Politics* 19(1):9–31.

Yin, Chien-Chung. 1998. "Equilibria of collective action in different distributions of protest thresholds." *Public Choice* 97(4):535–567.