Trust Degradation in a Simple Model of Prisoner's Dilemma among Reciprocating Agents on a Network

Zackary Dunivin[*] and Qi Hao[†]
[*]Indiana University
[†]Michigan State University

Introduction

Trust has been shown to be a very important part of social life. Simulations have been used to explore the dynamics of trust diffusion and it effects in economic contexts (e.g. Morvan & Sené, 2006), organizations (e,g, Williams, 2005) or complex socio-economic networks (e.g. Hauke, Pyka, Borschbach, & Heider, 2010). It was demonstrated that the trust network had great influence on the network's robustness against coalitions and trojan attacks and the organization's transparency.

This study is going to ask the question of: is it better that trust stays between dyads or is it better that trust diffuses through the network? It might be interesting to explore the difference between the two situations in terms of how much trust can exist in the system and how much good information will be exchanged due to trust over time.

Methods

An agent-based model is used to simulate the behaviors of this system. The parameters that are being modeled (in the global environment) include:

Number of agents. 100

Number of generations. 200

Networks. Two kinds of networks are used as the underlying network on which the simulated diffusion happens. One is the regular random network with uniform degree distribution, with everybody having a degree of 10. The other network is a random scale free network, with power-law degree distribution and average degree of 10.

Lag. A certain number of generations is assumed to pass before the receivers of information are able to figure out the quality of the information and make evaluations about the sender of that information. This is the lag, L.

Information Quality. The quality of the information is modeled as a binary variable. Each piece of information is either of good quality or bad quality.

Fidelity. Each agent is assigned a fidelity level. The fidelity is a probability that each agent is able to send the information of intended quality. For example, an agent with fidelity level of 50% can only send good information with 50% chance when the agent intends to send a good information to a certain receiver. Fidelity of agent i is represented as an operator $p_F(x)$, which generates the value x with probability p_F. In this study, it is assumed that all agents have the same fidelity level, so that p_F does not have a sub-index i.

Trust. In the model, trust of agent i towards agent j at generation t, $T_{ij}(t)$, is ensured by the latest revealed information quality of the information sent from agent j to agent i, $Q_{ji}(t-L)$ with L being the number of generations it takes for the information quality to be revealed.

$T_{ij}(t) = Q_{ji}(t-L)$, or in matrix form $T(t) = Q^T(t-L)$

For example, if the information was of good quality, the receiver trusts the sender for the current generation, $T_{ij}(t) = 1$. But if the receiver does not trust the sender due to bad revealed information from the sender, the receiver will lose trust for the current generation, $T_{ij}(t) = -1$.

In this simulation, it is assumed that people all start trusting until they find other people sending bad information. As a result, for all the generation numbers smaller than the lag, **T(1), T(2), T(3)… T(L) = T(0),** with**T(0)** being the matrix representing the underlying network.

With certain combinations of the above settings, two different microlevel mechanisms are being simulated and compared, one without diffusion of trust and one with diffusion of trust.

Reciprocation Only Model. If an agent trusts another agent, $T_{ij}(t) = 1$, the information sending behavior will be good-willed in the next generation, sending good quality information, $Q_{ij}(t+1) = 1$, with a probability of p_F. But if $T_{ij}(t) = -1$, agent reciprocate with bad information in the next generation, $Q_{ij}(t+1) = -1$, with a probability of p_F.

$Q_{ij}(t+1) = p_F(T_{ij}(t))$.

*Reciprocation Diffusion Model.*After the information sending and receiving of each generation, agents will hold the information pieces they just received and sort them into two groups according to their trust towards the senders of these information pieces at this time **T(t)**. So if an agent trusts the sender at the time of the reception of the information, $T_{ij}(t) = 1$, the agent will sort this piece of information as good. And if $T_{ij}(t) = 1$, bad. For each of its trusted neighbors, the agent will select and relay one piece of what it believes is high quality information (information received at t-1 from trusted neighbors). Likewise, the agent selects from its store of bad information to relay to untrusted neighbors. Note that this behavior is based on knowledge about the quality of information and neighbors L generations ago. Also note that, due to this reason, the agents could mistakenly classify information and send bad information to good people. As a result of this, bad reputation could diffusion.

Results

We are interested in the degradation (or maintenance) of trust in our model. We take the total percentage of high quality information (1 in our binary system) as a proxy for trust. High quality information is typically exchanged when there is trust, and only outside of trusted relationships as a result of error in transmission or by mistake due to a lag in identification of information quality.

Each plot in Figure 1-3 shows the behavior of our model as we vary the fidelity of information transmission in our model. Figures 1-3 show compare the effects of transmission error in three different network scenarios. Figure 1 is run on a regular random graph with the non-diffusion model. As transmission error increases, the rate at which the model converges to its equilibrium increases. Figure 2 is also run on a regular random graph, but using the diffusion model. While the slopes appear very similar between the two models, the runs are considerably noisier. Figure 3 runs the diffusion model on a scale free network (Barabasi-Albert graph). Here the slopes seem slightly less steep than either model run on the regular graph, and noise increases considerably.

Figure 1. Amount of good info in population over time w/o diffusion on regular random NW

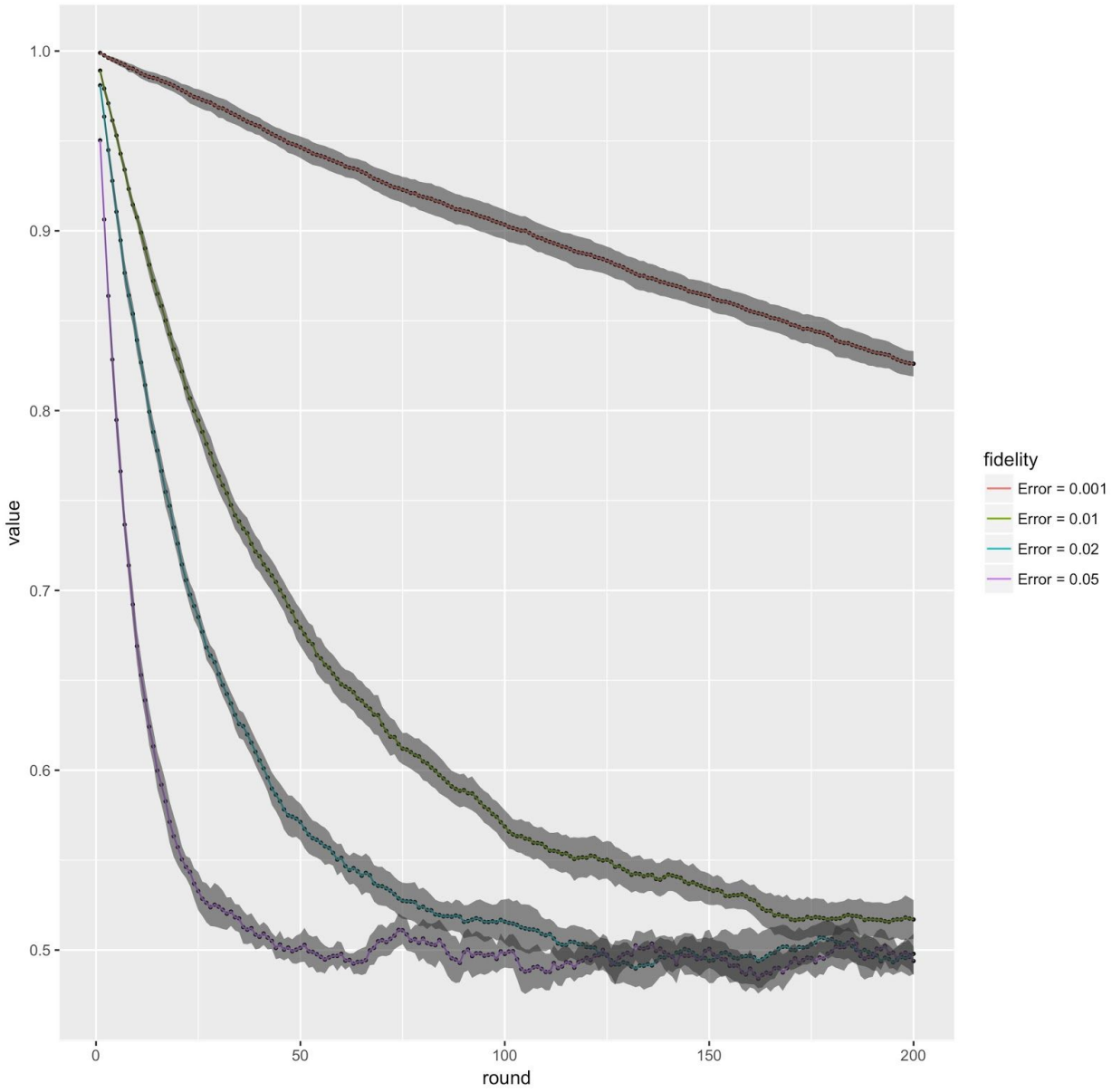


Figure 2. Amount of good info in population over time on regular random network

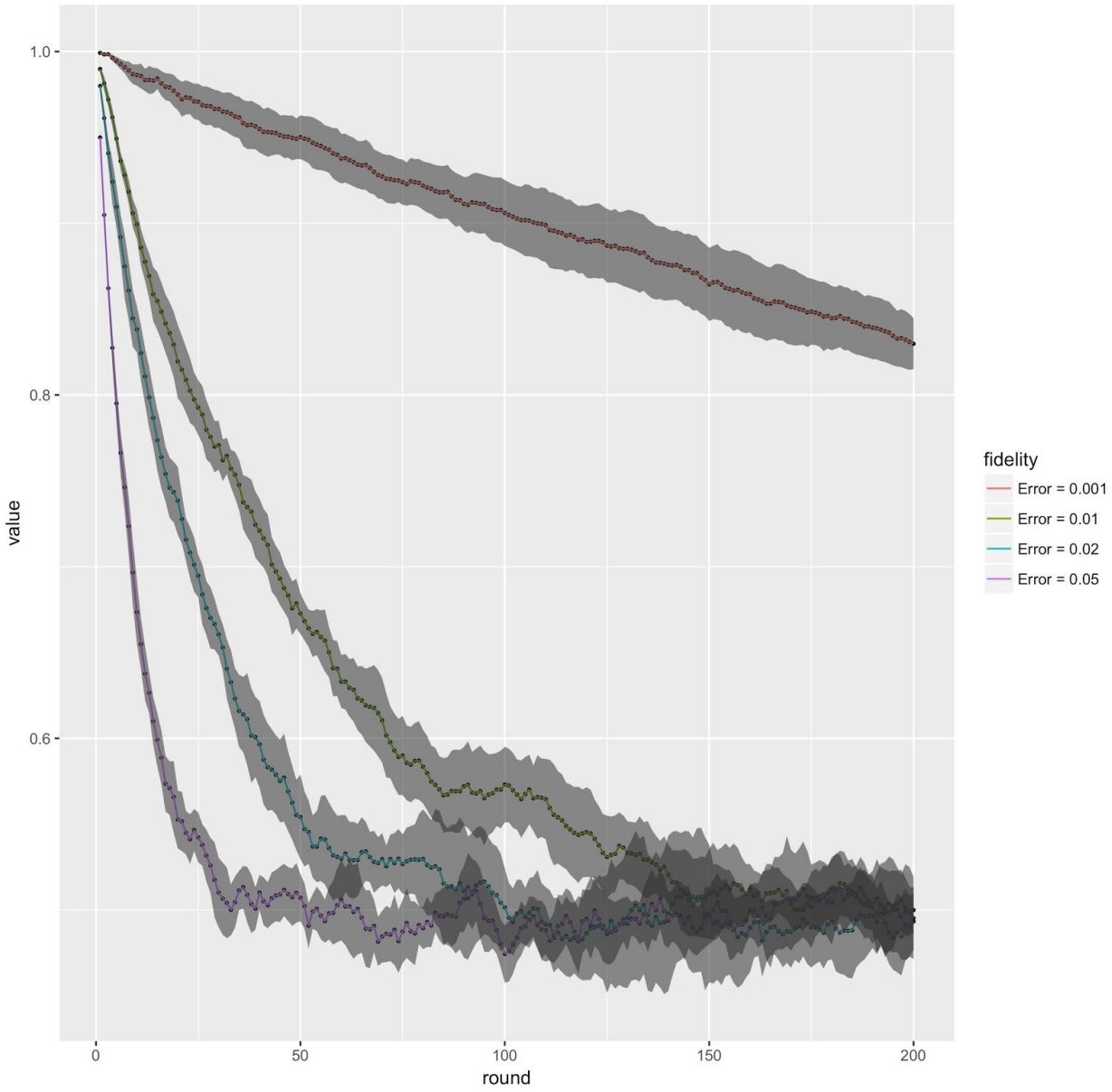
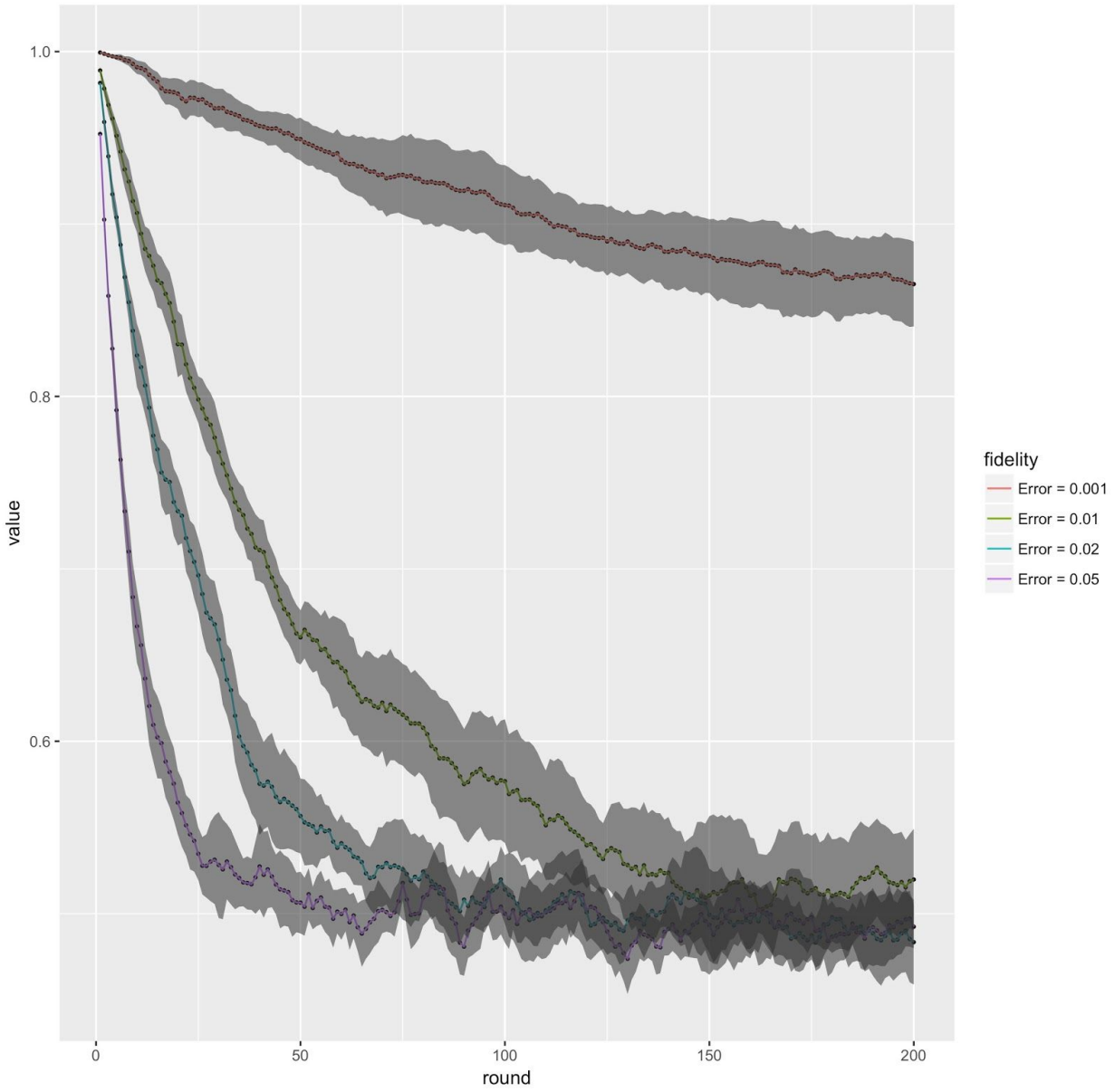


Figure 3. Amount of good info in population over time on scale free network



Figures 4-6 share parameters with Figures 1-3, but instead hold transmission error rate at 0.001 and vary the number of timesteps it takes to verify the quality of information (lag). In the non-diffusion model (Figure 4), increase the lag decreases the slope of the decline in information quality. This pattern is not observed in the diffusion models on either the regular random or scale-free networks (Figures 5 and 6 respectively). Rather, the slope is not affected by increasing lag. Diffusion appears to cancel out the effect of lag observed in the non-diffusion model.

Figure 4. Amount of good info in population over time w/o diffusion by lag level

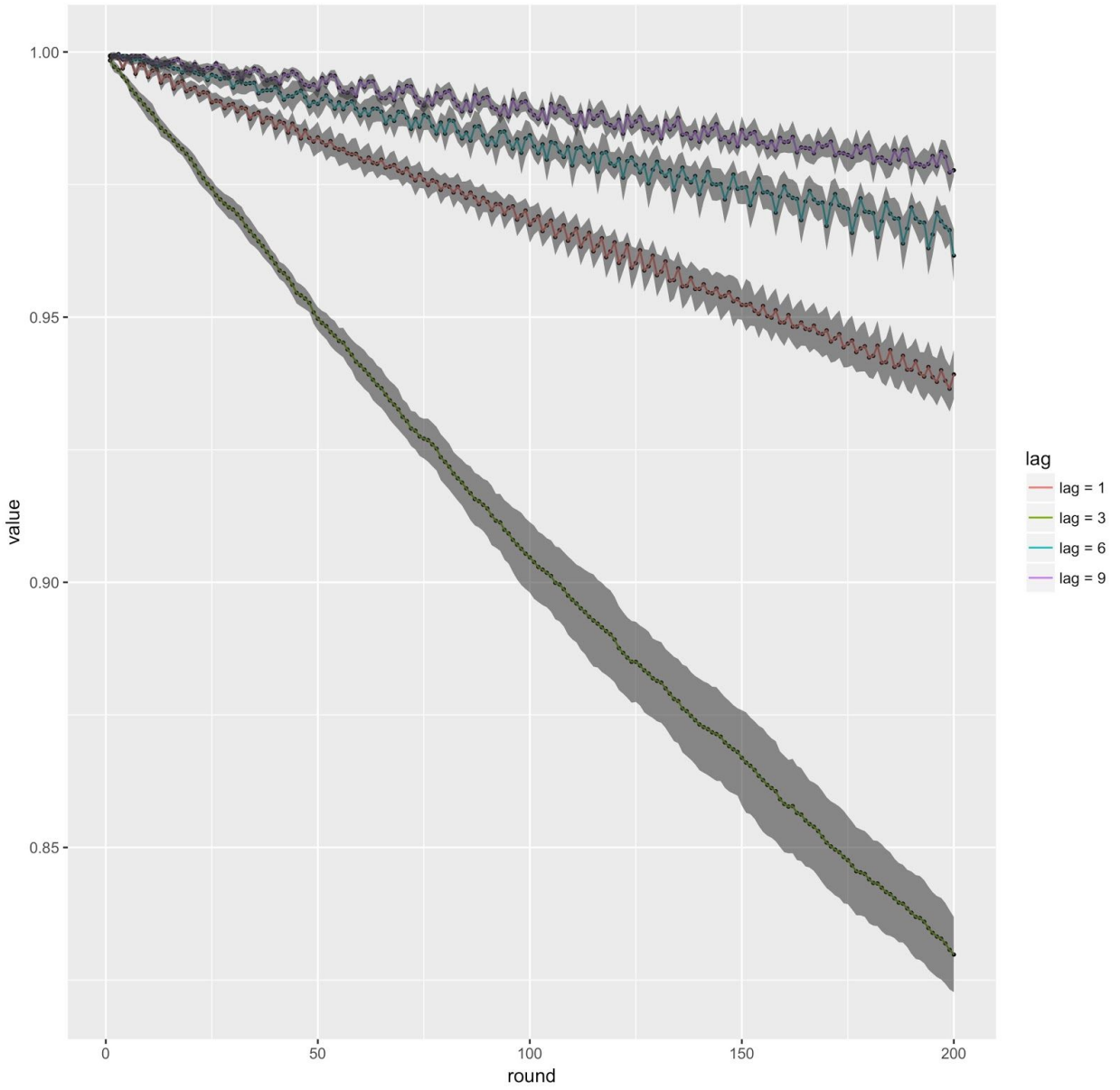


Figure 5. Amount of good info in population over time w/ diffusion by lag level

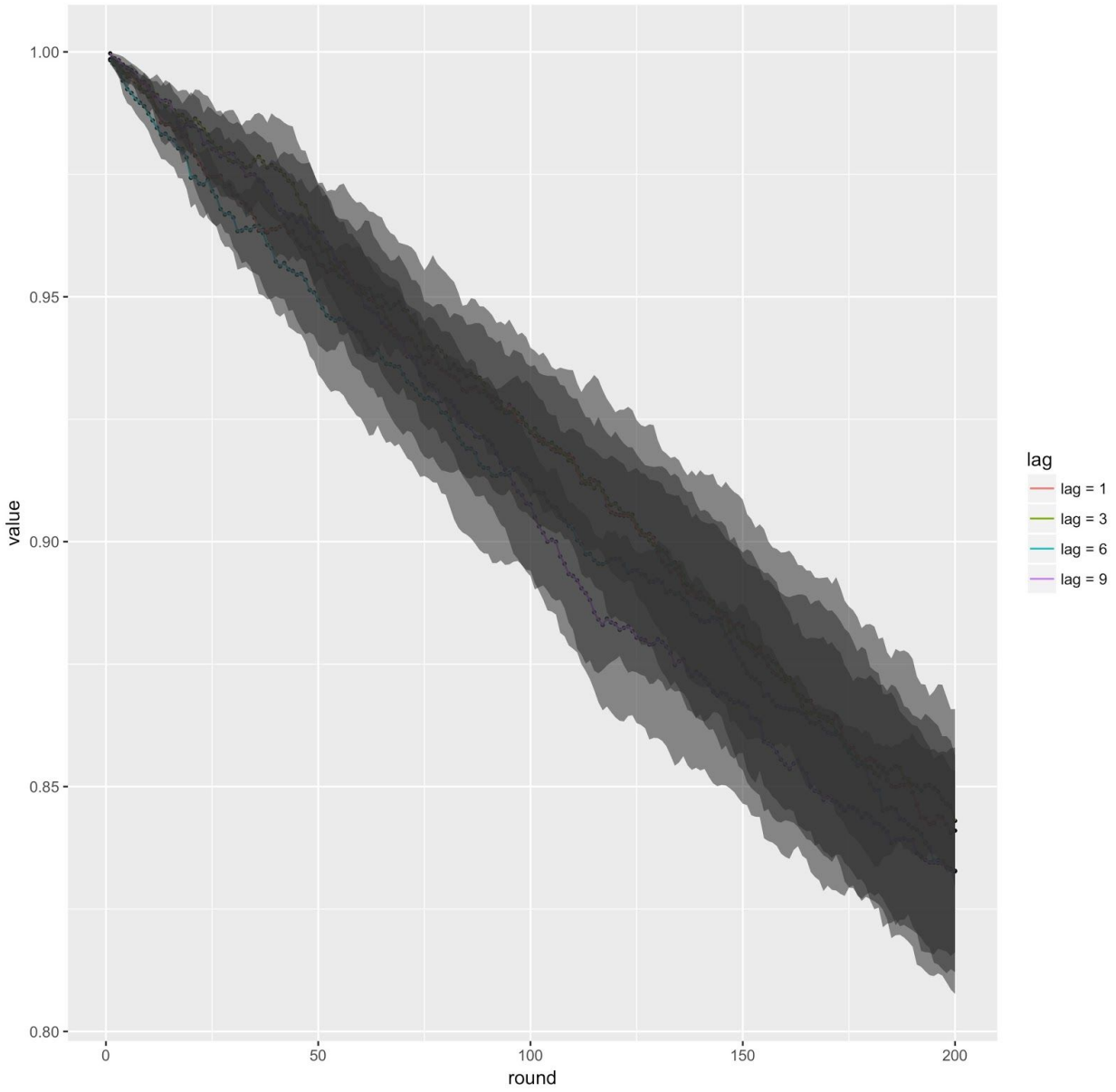
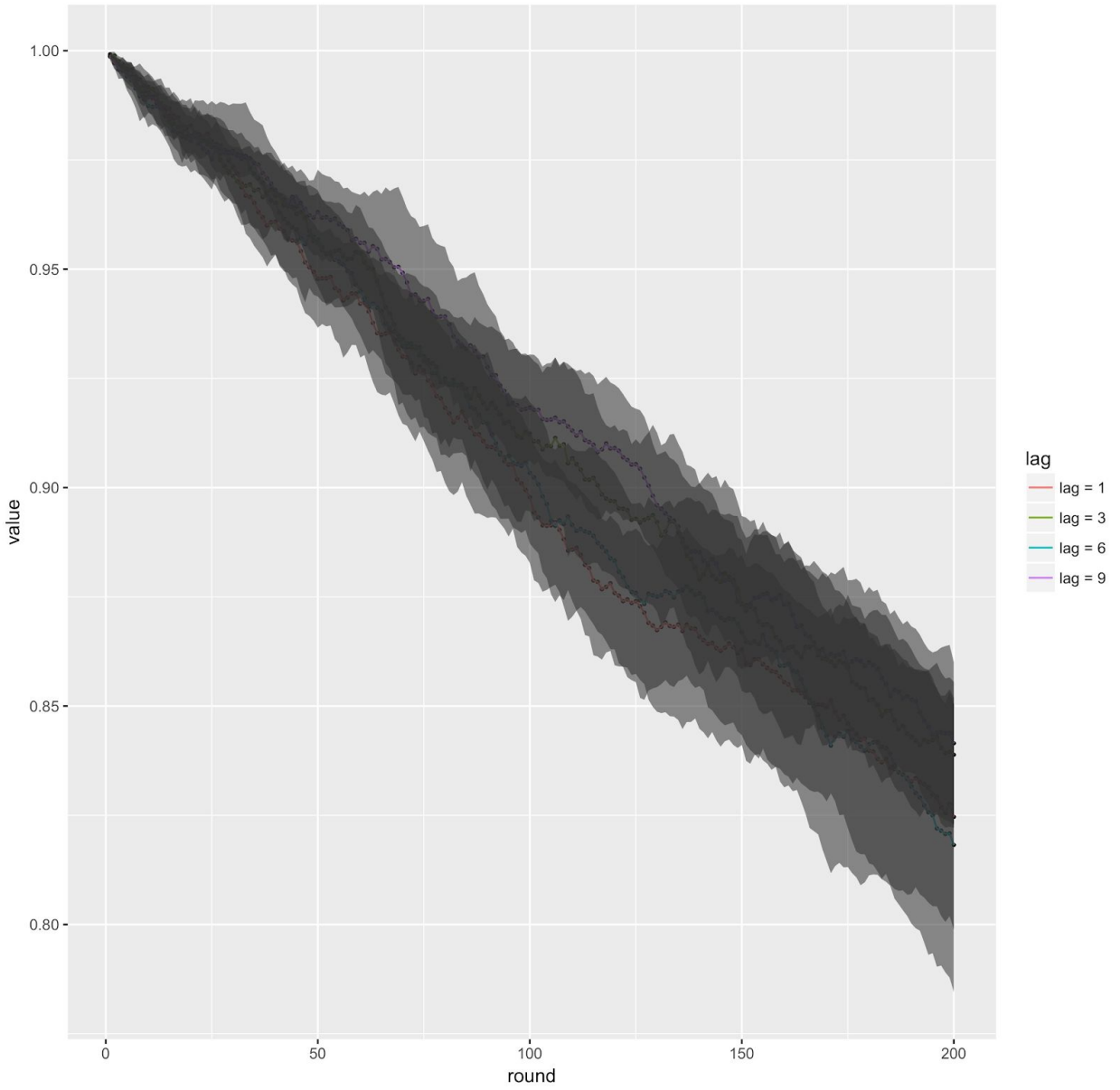


Figure 6. Amount of good info in population over time w/ diffusion on scale free NW



Discussion

Our base model is a prisoner's dilemma among reciprocating agents. When decisions are based on the exchange immediately prior, a single mistransmission will trigger a stable defecting relationship (until it is switched by an accidental high quality transmission). Lag introduces robustness into the tit-for-tat exchanges. Instead of requiring 1 mistransmission of low quality information to trigger a stable defecting relationship, a lag of 1 timestep demands 2 consecutive mistransmissions by one of the agents. A lag of 2 timesteps requires 3 consecutive mistransmissions, etc.

Introducing lag into the non-diffusion model decreases the slope of the curve of good information in the system, prolonging the time it takes to reach equilibrium. In the diffusion model, the slope does not change as the lag varies. While the time to lock in defection does increase with the lag, so does the accidental spread of low quality information. What is surprising is that these two effects offset each other

equally. Further probing of the model, either through deduction or exploration of the data, is required to account for the equal effect of delayed mutual defection and increased accidental defection due to contagion.

References

Williams, C. C. (2005). Trust diffusion: The effect of interpersonal trust on structure, function, and organizational transparency. *Business & Society*, *44*(3), 357-368.

Morvan, M., & Sené, S. (2006, July). A distributed trust diffusion protocol for ad hoc networks. In *Wireless and Mobile Communications, 2006. ICWMC'06. International Conference on* (pp. 87-87). IEEE.

Hauke, S., Pyka, M., Borschbach, M., & Heider, D. (2010). Reputation-based trust diffusion in complex socio-economic networks. In *Information Retrieval and Mining in Distributed Environments* (pp. 21-40). Springer, Berlin, Heidelberg.