
NON-COOPERATIVITY IN BAYESIAN SOCIAL LEARNING

Stan Palasek
Princeton University

July 2014

ABSTRACT

We describe a Bayesian model for social learning of a random variable in which agents might observe each other over a directed network. The outcomes produced are compared to those from a model in which observations occur randomly over a complete graph. In both cases we observe a nontrivial level of observation which maximizes learning, though individuals have strong incentive to defect from the societal optimum. The implications of such competition over information commons are discussed.

I. INTRODUCTION

Trends and fads, though at times ill-informed, are ubiquitous features of human societies. The mathematical examination of the origins of these trends as cascades of questionable wisdom was perhaps pioneered in 1992 by Banerjee [1] and independently Bikhchandani *et al.* [2] Although it is rational for agents to observe the actions of others and to update their own beliefs accordingly, Banerjee notes the dilemma that “the very act of trying to use the information contained in the decisions made by others makes each person’s decision less responsive to her own information and hence less informative to others.” Vives similarly showed that when agents’ observations are governed by a benevolent organizer, social learning is more effective than when each individual is acting strictly in his or her own interest. The observed inefficiency is a manifestation of an “underinvestment in production of public information,” which limits exploration of unpopular decisions and propagates herding. [3]

These results may go a long way in explaining the emergence of societal trends which may seem irrational to those with access to a different stream of information. Indeed, theories concerning social learning have found a number of applications including finance, [4, 5] policy, [6] and the adoption of new ideas and technologies in the developing world. [7, 8, 9]

II. SOCIAL LEARNING MODEL

We consider a system similar to that in Salge & Polani [10]. Here, there are n slots, of which one hides a reward which is being sought by n agents. At each time step, with probability $p \ll 1$ the reward is randomly relocated. A random agent a is then selected uniformly and independently who observes the slot s that it believes most likely to contain the reward. All m agents perform a naïve Bayes update: the a th agent’s based on the result of its observation of s , and the other agents’ based on the fact that a elected to observe s . This

update takes into account the possibility that the target has moved, removing the risk that agents get “stuck” having unsuccessfully checked all n loci.

As discussed in [10], an agent’s actions are chosen in consideration of its previous observations and therefore contain potentially useful information. For instance, in a simple model with no Bayesian updating and $p = 0$, it can be shown that the conditional probability of the reward being at s given that an agent chooses to observe s is twice the conditional probability given that an agent chooses to observe a slot other than s . We might expect that introducing social updating only injects more useful information into agents’ choices.

Described so far has been a model for social learning where the underlying social network is the complete graph of order m . More generally, we can consider the same process but with agents constrained to perform Bayesian updates only for those with whom they are connected on a graph G . Note that at times we will allow G to be directed; ie. for some (perhaps artificial) reason, an agent i employs information from j but not *vice versa*.

We make several key changes to Salge & Polani’s model which we believe make this formulation more realistic. First, here we do *not* reset the beliefs of an agent once it discovers the location of the reward. We admit that occasionally injecting maximum entropy makes sense when $p = 0$ since the game would otherwise be “solved” and quickly reach a steady state. However, the fact that the location of the reward changes simultaneously for the entire ensemble suggests that all the agents are within a single environment and it would therefore be unnatural for some to abruptly forget information or be replaced.

Second, here we use a social network

to limit information consumption (as in [11, 12, 13]) rather than throwing out observations with a uniform probability $1 - p_o$ (reminiscent of [14]). Although the latter model is appealing in the sense that it provides a linear continuum between a non-social situation ($p_o = 0$, or all isolated nodes) and complete observability ($p_o = 1$, or G is complete), it is far more natural for agents to form social networks and learn from those with whom they are connected. Indeed, one might expect the uniformly random information selection to be analogous to learning on an Erdős-Rényi random graph since they must both exhibit low diameter, clustering, and a binomial degree distribution. The authors suggested the random ignorance model as something that social agents might form strategically rather than something that is inherent within their decision-making processes, the plausibility of which we will investigate later.

Finally, here we order the agent observations uniformly and independently rather than in a predetermined sequence. This change was necessitated by the present use of the network model in lieu of random observation. The latter formulation is symmetric because the choice of information disposal is i.i.d. The arbitrary designation of an order is therefore of little concern. In the current model, however, the underlying graph is generally asymmetric so an arbitrary ordering of the agents may introduce harmful bias. Random ordering avoids this complication.

III. GRAPH CONNECTIVITY RESULTS

A major result of [10] is that maximal learning occurs at some intermediate $0 < p_o < 1$. We will illustrate an analogous fact for our

network-based model with the modifications detailed above. In particular, we execute the described process for graphs chosen uniformly at random from the set of graphs with n vertices and i edges, for each $i = 1, 2, 3, \dots, \binom{n}{2}$.

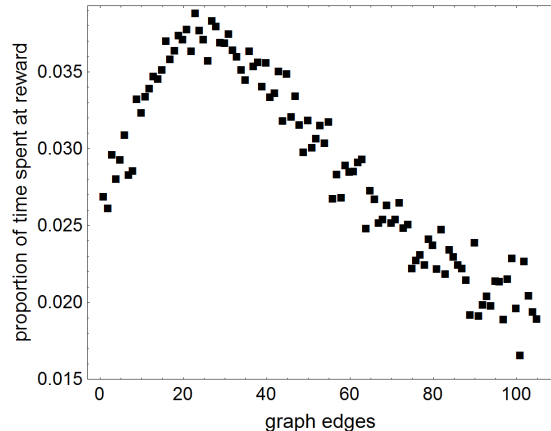


Figure 1: Simulation results for $n = 15$, $p = 0.01$, and a Bayes' factor of 2 between positive and negative observations. Depicted are the results of 50 trials each a mean of 1000 iterations at each edge count from 1 to 100.

The results are shown in Figure 1. We see, as in [10], that there indeed exists an optimal quantity of information to be collected by social agents, as learning is maximized when the graph has approximately 25 edges, or when the mean degree is 3.3. For very disconnected networks, agents have little access to the information contained in others' actions which, as we argued, has a Bayes' factor of at least two. Conversely, when the network has most of its possible edges present, we observe that herding behavior to incorrect slots is prevalent. This is illustrated in Figure 2 which depicts the simulation over a complete network during which the agents spend most of their time at only a handful of spots which rarely contain the reward.

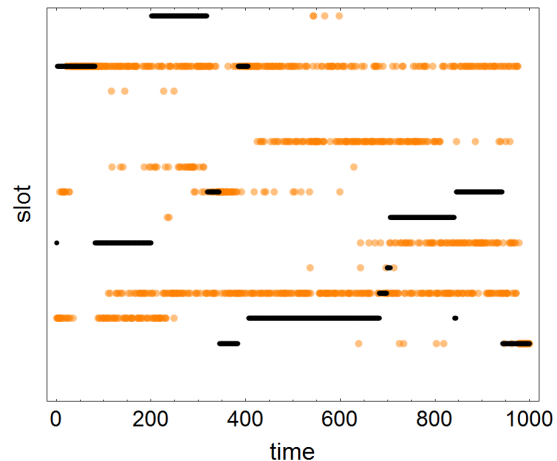


Figure 2: The location of the treasure (black) and the slot chosen at a given time (orange) for the same parameter set as Figure 1. Representative data for a society over a complete graph are shown, and misinformed herding behavior is evident.

On these grounds alone, [10] suggests that rational agents might voluntarily ignore information to achieve $p_o < 1$ or, in this setting, a non-complete graph. In the next sections we will evaluate whether agents truly have any incentive to do so, or whether defecting from an optimal configuration would suit one's individual interests at the expense of those of the group.

IV. STABILITY OVER COMPLETE G

The notion that agents freely decide whose information to take into account and to what extent suggests that by default they have at their disposal more information than they would optimally use. In this section we will therefore take the complete graph as a starting point of our investigation of stability and consider deviations thence. Should we see that from here no one has incentive to discard information, then we must conclude that the complete graph constitutes a stable equilibrium. However, the prevalence of harmful herding as demonstrated in the last section

would lead one to expect that agents are eager to defect from complete networks so as to minimize their susceptibility to cascades of false information. We test this by modifying a complete graph such that a test node, originally of degree $m - 1$, instead has degree i for $i = 0, 1, 2, \dots, m - 1$. Since the original graph was complete and therefore symmetric, there is a unique unlabeled graph for each i . This will be an instance, as we hinted at earlier, in which the graph is directed (except when $i = 14$) since an agent’s decision to disregard data from its neighbor has no bearing on its neighbor’s reciprocal observation.

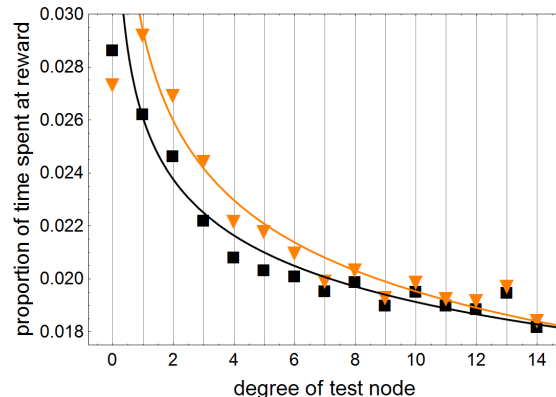


Figure 3: Simulation results for every possible single-agent deviation from a complete graph. Depicted are time series for both the defecting agent (orange) and the remaining $m - 1$ collectively (black). The parameter set is the same as that from Figure 1. Both series for $t \geq 1$ follow power laws with $R^2 = 0.999$ with forms $P = 0.0293 \times d^{-0.177}$ (individual) and $P = 0.0261 \times d^{-0.134}$ (society). We do not show data for the individual non-defecting agents for they vary in proportion with the aggregate data.

The result of the simulation is shown in Figure 3. As expected, single agents are generally rewarded for severing ties from a complete network. The only robust exception is in the transition from maintaining one connection to becoming isolated, in which the defecting node takes a slight loss

in performance by disconnecting itself completely. Furthermore, we see that the improvement in the defector’s performance is mirrored in the welfare gain experienced by the rest of society. We must then ask, given that deviations from the complete network are incentivised, to what density the network must be cleared for no defection to be motivated, should such a stable network exist at all.

V. STABILITY OVER SPARSE G

Next, rather than beginning with a complete graph, we will instead consider the the highest performance graph with edge count near the extremum of Figure 1, chosen out of the 100,000 random 15-order graphs examined. This graph, call it G_0 , is depicted in Figure 4 and it may be near the equilibrium, should one exist.

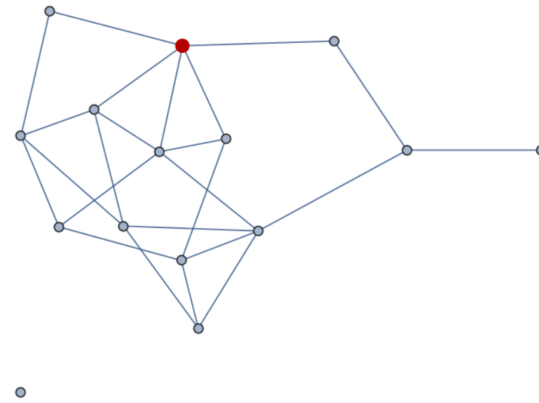


Figure 4: G_0 , the highest-performing graph with edge count near 25 among the 100,000 sampled graphs depicted in Figure 1. The “test vertex” is highlighted in red.

For the purpose of illustration we will arbitrarily designate the node highlighted in red as the test vertex which will defect from its current situation in G_0 during our experiments. In particular, we will repeat the simulation from the previous section upon this graph G_0 by varying the order

of the text vertex. This procedure differs in that now inasmuch as the starting graph is asymmetric, there is *not* a unique graph with the required properties. We will therefore select a uniformly random degree inclusively between 0 and m and then choose the network randomly from the class of graphs in which the test vertex has the selected degree and connections between non-test vertices are unchanged.

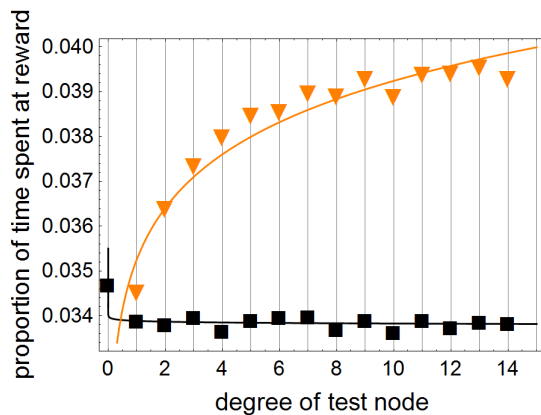


Figure 5: Learning performance over optimal graph G_0 with the test model’s connections systematically modified. 50,000 trials were executed with the same parameter set used in Figure 1. Depicted are both the performance of the test node (orange) and the rest of the network (black). Both time series are fit to power laws with $R^2 = 0.9999$ with forms $P = 0.0353 \times d^{0.0466}$ (individual) and $P = 0.0339 \times d^{-0.00762}$ (society).

Figure 5 depicts the results, which show that the test node can expect a robust increase in its average performance by observing more agents. Meanwhile, as long as the test agent is connected to *someone*, its behavior ceases to have a measurable effect on the success of the others. Because the test agent, already occupying one of the better-connected nodes in the network, is motivated to expand his observation of the other agents, we conclude that this network does not present a stable environment when individuals are free to adjust their informa-

tion consumption. Figure 6 confirms that this observation is a general one and that all nodes—not just the test node of Figure 4—are motivated to defect to complete observation.

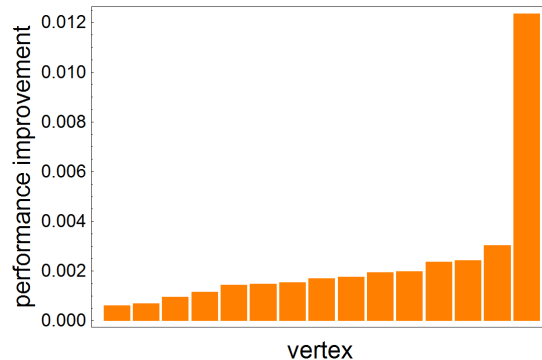


Figure 6: Mean performance improvements for each node of G_0 upon independently defecting to complete observation. The same parameter set from Figure 1 is employed with 1000 trials for each node. The outlying datum is that of the isolated node seen in Figure 4.

VI. STABILITY WITH RANDOM OBSERVATION

One might come to believe that the instability we have identified when individuals are free to choose which information to consider is a feature of our graph model and is not present in the random observation model of [10]. Here we will show that the phenomenon is present regardless.

Now instead of each agent forming its beliefs according to its connections in a directed graph, it will use information from a given agent at a given time with some probability p_o . As noted before, varying p_o between 0 and 1 interpolates between the null and complete graphs. In order to understand the incentives of agents to deviate from the p_o that is ideal for society as a whole, we first set all agents’ observation probabilities to $p_0 = 0.196$, the optimal probability identified among 10,000

trials with the same parameter set that we have used throughout. We then allow the observation probability of a single test node to vary.¹ The results are depicted in Figure 7. We conclude that the test agent, and therefore every agent by symmetry of the model, has incentive to increase its observation probability far past the optimum for society. Therefore, despite $p_o = 0.196$ being ideal for society as a whole, it will not be adhered to by rational agents who can freely choose their observation rates.

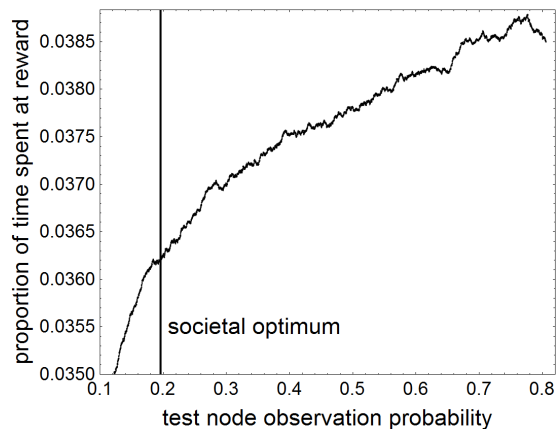


Figure 7: Learning performance for test node as it defects from the societal ideal $p_o = 0.196$. 10,000 trials were executed, each time selecting a random observation probability for the test node. Data are subjected to exponential moving averaging with a smoothing factor of 0.0005.

VII. GAME THEORY

Up to this point we have been employing language as though we are examining a competitive game without being explicit about its components. It will become apparent as we do so how naturally our formulation can be viewed as a normal-form

game.

Let the game have m players corresponding to the society’s agents. A pure strategy profile is any directed graph G over m labeled nodes in which the i th agent has control over the i th row of G ’s adjacency matrix. The payoffs are defined as the average performance for each agent when the described simulation is carried out over G .² We assume that the agents’ information on the simulation’s rules, including the value of p , is complete.³ In terms of this game, we can summarize the results of the previous two sections by noting that neither the complete graph nor G_0 is a Nash equilibrium since someone, and in fact everyone, gains a profit on average by defecting. Furthermore, the complete graph is Pareto inefficient by Figure 3. We might wish to find a stable outcome, ie. a Nash equilibrium, but as we are considering only pure strategies, Nash’s theorem does not guarantee their existence. [18] To find equilibria exhaustively is intractable for the number of unlabeled undirected graphs when $m = 15$ is on the order of 10^{19} . [19] When directed edges are introduced the space becomes larger still by orders of magnitude.

We may still, however, computationally identify an equilibrium within the observation probability model. Here, although the strategy space is continuous, it is one-dimensional and thus easy to search. We proceed as follows. Let p_t be the observation probability used by the test individual, and p_s the probability for the rest of society. Let the expected performance of the test individual be given by $U(p_s, p_t)$. In equilibrium, no individual has incentive to

¹The choice of the test node does not matter because here there is no underlying graph to introduce asymmetry.

²It is reasonable to expect that agents act upon the expectation of their utility. [15] For a treatment that does not neglect the stochastic nature of learning performance, one might consider games with “fuzzy” payoffs; see, for instance, [16]

³One need not make this assumption. See Dekel *et al.* [17] who examine learning without it.

deviate from p_s . Therefore, we seek a solution $(p_s, p_t) \in [0, 1]^2$ to

$$p_s = \arg \max_{p_t} U(p_s, p_t). \quad (1)$$

Clearly when $p_o = 0$, there is no widespread risk of herding so the test agent’s best-response is a p_t near 1. Furthermore, as we showed in Section IV, when $p_o = 1$ the best-response is a near-zero level of observation. These facts combined with the continuity of the right-hand side of (1) and the intermediate value theorem would imply that there exists a solution to the equation. However, due to the prevalence of phase behavior as the degree of connectivity of societies varies, [20] we cannot be certain as to how the function behaves in practice. With this in mind we evaluate the function *in silico*. The left- and right-hand sides of (1) are shown respectively in orange and black in Figure 8.

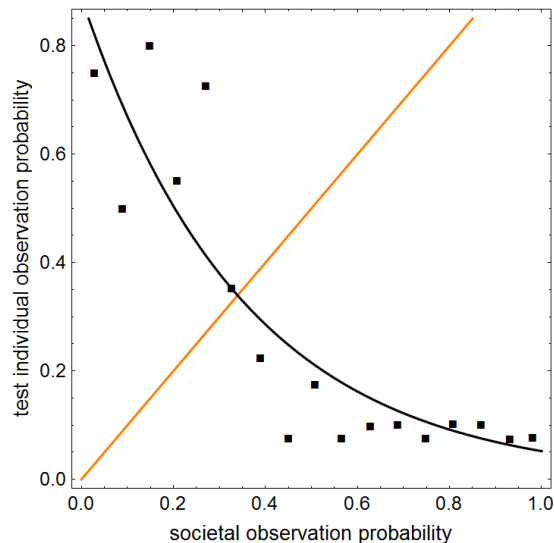


Figure 8: The test node’s best-response observation probability is depicted at various societal observation probabilities. The data represent 25,000 random samples of values for p_s and p_t . The best response function is fitted to an exponential with $p_t^* = 0.89 \times \exp(-2.84 \times p_s)$. (1) is satisfied when the best-response series (black) coincides with the locus $p_s = p_t$ (orange).

VIII. DISCUSSION

We, as in [10], determine that maximal social learning occurs when not all information is taken into account by all agents. The authors go on to detail the reasonable corollary that rational agents will voluntarily discard much of their information—between 70 and 80% according to both our analyses. Doing so gives individuals’ own observations greater weight in their Bayesian beliefs, thus globally limiting the impetus for herding which, as seen in Figure 2, is often to the wrong location. Our objection to this reasoning is as follows. A single agent, by defecting and letting the extent of its observation be universal (or nearly so), has little effect on the prevalence of herding, as it is necessarily a population-wide phenomenon. However, its defection yields in its favor a significant improvement in its ability to take advantage of emerging trends. Since the propensity for false information cascades has hardly increased, the defecting agent enjoys greater performance. Such an argument applies to any and therefore every member of the society. Consequently, we cannot expect rational individuals that have the ability to increase the extent of their observation to adhere to the welfare-maximizing level of information consumption.

This dilemma is highly reminiscent of Hardin’s tragedy of the commons. [21] There is a common resource in our system: information uncontaminated by negative information cascades. Acting in its own self-interest, each agent is strongly motivated to maximally consume this information (see Figures 5, 6, and 7). However, as the common pool of information becomes over-consumed, small random flocks to potentially incorrect loci are accentuated and herding behavior emerges; thus the resource is of far lesser use.

As Hardin famously pointed out, there is no technical solution to such problems of overconsumption. Fortunately, nature seems to impose a limit herself, for human societies are typically arranged in non-complete social networks. This raises the question of whether the properties exhibited by graphs of actual human communities are particularly conducive to such learning. One might also wonder, if such learning processes played any role in the adoption of technologies by early man which would have in turn correlated with survival probabilities, why defectors did not invade populations of mutual cooperators. For this we refer to the extensive literature on the evolution of cooperation in which these same questions are addressed in a general setting. [22, 23, 24]

REFERENCES

- [1] Banerjee, A. V. (1992). A Simple Model of Herd Behavior. *The Quarterly Journal of Economics* 107(3), 797-817.
- [2] Bikhchandani, S., Hirshleifer, D., & Welch, I. (1992). A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades. *Journal of Political Economy* 100(5), 992-1026.
- [3] Vives, X. (1995). Learning from Others: A Welfare Analysis. *Games and Economic Behavior* 20, 177-200.
- [4] Welch, I. (1992). Sequential Sales, Learning, and Cascades. *The Journal of Finance* 47(2), 695-732.
- [5] Chamley, C. & Gale, D. (1994). Information Revelation and Strategic Delay in a Model of Investment. *Econometrica* 62(5), 1065-1085.
- [6] Dobbin, F., Simmons, B., & Garrett, G. (2007). The Global Diffusion of Public Policies: Social Construction, Coercion, Competition, or Learning? *Annual Review of Sociology* 33, 449-472.
- [7] Conley, T. & Udry, C. (2001). Social Learning through Networks: The Adoption of New Agricultural Technologies in Ghana. *American Journal of Agricultural Economics* 83(3), 668-673.
- [8] Munshi, K. (2004). Social learning in a heterogeneous population: technology diffusion in the Indian Green Revolution. *Journal of Development Economics* 73, 185-213.
- [9] Banerjee, A., Chandrasekhar, A. G., Duflo, E., & Jackson, M. O. (2013). The Diffusion of Microfinance. *Science* 341, 1236-1248.
- [10] Salge, C. & Polani, D. (2014). Don't Believe Everything You Hear; Preserving Relevant Information by Discarding Social Information. *arXiv:1406.1034*.
- [11] Bala, V & Goyal, S. (1998). Learning from Neighbors. *The Review of Economic Studies* 65(3), 595-621.
- [12] Gale, D. & Kariv, S. (2003). Bayesian Learning in Social Networks. *Games and Economic Behavior* 45(2), 329-346.
- [13] Acemoglu, D. & Dahleh, M. A. (2011). Bayesian Learning in Social Networks. *Review of Economic Studies* 78, 1201-1236.
- [14] Smith, L. & Sørensen, P. (2013). Rational Social Learning by Random

- Sampling. (*unpublished working paper*).
- [15] Bernoulli, D. (reprinted 1967). *Specimen theoriae novae de mensura sortis*. Gregg Press.
- [16] Clemente, M., Fernández, & Puerto, J. (2011). Pareto-optimal security strategies in matrix games with fuzzy payoffs. *Fuzzy Sets and Systems* 176(1), 36-45.
- [17] Dekel, E., Fudenberg, D., & Levine, D. K. (2004). Learning to play Bayesian games. *Games and Economic Behavior* 46, 282-303.
- [18] Nash, J. (1951). Non-Cooperative Games. *The Annals of Mathematics* 54(2), 286-295.
- [19] OEIS Foundation Inc. (2003). The On-Line Encyclopedia of Integer Sequences, A000088.
- [20] Erdős, P. & Rényi, A. (1960). On the Evolution of Random Graphs." *Publications of the Mathematical Institute of the Hungarian Academy of Sciences* 5, 17-61.
- [21] Hardin, G. (1968). The Tragedy of the Commons. *Science* 162(3859), 1243-1248.
- [22] Axelrod, R. (1984). *The Evolution of Cooperation*. Basic Books: New York, NY.
- [23] Nowak, M. A., Sasaki, A., Taylor, C., & Fudenberg, D. (2004). Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428, 646-650.
- [24] Ohtsuki, H., Hauert, C., Lieberman, E., & Nowak, M. A. (2006). A simple rule for the evolution of cooperation on graphs and social networks. *Nature* 441, 502-505.