The Tipping Point

An Agent Based Model of Information Diffusion

And Social Change

John Doty

Abstract

This agent-based model seeks to explore the relationship between the structure of social networks, the properties of the agents of which they consist and the affect these dynamics have on the way information spreads and social change occurs. It has two main inspirations. The first is the book *The Tipping Point*, by Malcolm Gladwell, in which he explores the dynamics of epidemic social change. The second is the development of web 2.0 and evidence indicating that knowledge of the underlying structure of social networks on the Internet and the properties of the agents of which they consist can be leveraged to build useful and profitable applications.

Table of Contents

I. A Brief History of the Project

This idea for this model was birthed spring quarter of 2009 in Professor McKelvey and Professor Bragin's course on Complexity Theory. I was concurrently enrolled in an accompanying Seminar on Agent Based Modeling. Having recently read Malcolm Gladwell's *The Tipping Point,* a book about the dynamics of epidemic social change I noticed that he had utilized many concepts originating from the Complex Adaptive Systems literature. I mentioned my interest in the possibility of producing and agent based model according to Gladwell's theory to Gregg Rugolo, a graduate student in the Information Science program at UCLA was also enrolled in these two classes. He was also fascinated by Gladwell's approach to studying the tipping point phenomenon and suggested we make it the basis of our final project.

After co-writing the final research paper for the Complexity Theory course on the formalization of Gladwell's theory for the basis of an agent based model with Gregg and another student, Sachini Ranasinghe, Gregg and I embarked, under the guidance of Professor Robin Liggett of the school of Design and Architecture, to produce a agent based model based upon our work. In the course of this independent research project Gregg produced an in depth report on our project and the collaborative tools we used for its creation. My focus was on the software design and production aspect of the project. This resulted in the first iteration of the random-network generating algorithm used by this simulation.

Gregg's expert knowledge of collaborative tools and incredible level of research were essential to this project. Professor Liggett's and Professor Bragin's continuing consultation and support have also been a fundamental guiding force in its creation.

II. Motivation and Theoretical Background

A. Malcolm Gladwell's *The Tipping Point*

In his book *The Tipping Point*, Malcolm Gladwell (2000) sets out to answer two questions, "Why is it that some ideas or behaviors or products start epidemics and others don't? And what can we do to deliberately start and control positive epidemics of our own?" (TP 14) Gladwell identifies three properties involved in the process of epidemic social change. The first of these is the existence nonlinear causal relationships. "We need to prepare ourselves for the possibility that sometimes big changes follow from small events, and that sometimes these changes can happen very quickly." (Gladwell. 2000 p, 11) The second is the idea of contagion; because of the high degree of interconnectedness of the agents that make up a social network, change can spread through them much like a virus. A third property is the propensity for change to occur rapidly rather than as a gradual change. This is related to the notion of self-organized criticality. Bak (1996) says of this phenomenon that, "…complex behavior in nature reflects the tendency of large systems to evolve into a poised "critical" state way out of balance… most of the changes take place through catastrophic events rather than following a smooth path."(p. 1)

Gladwell goes on to identify three sources of change within a social network. Gladwell labels sources the law of the few, the stickiness factor and the power of context.

The stickiness factor is property of the message itself. It is a measurement of how effective the content and presentation of the message are in perpetuating its spread. This is more closely related to individual's psychology than the structure of the social system itself. The power of context references the powerful effect small changes in the physical environment in which the social network is embedded can have on particular messages spreading throughout the network. Gladwell's quintessential example of this is known as the "broken windows" theory of crime. The central tenet of this theory is that crime inevitably follows disorder. The simple mending of windows and painting over of graffiti can be powerful anti-crime actions. (Gladwell, 2000, p. 141)

The focus of this agent-based model is on what Gladwell calls the law of the few. This is the theory that there are a special group of individuals that have a disproportionately powerful role in the creation of social epidemics. Gladwell identifies these special few as connectors, salesmen and mavens. All people are endowed with varying degrees of persuasiveness, curiosity, and social connectedness; Gladwell argues that it is the few outliers that are the driving force of social epidemics.

Connectors are individuals that have a very large and diverse social network consisting of many different social "neighborhoods." They function as the vessels that allow a message to spread to different social clusters rather than stagnating in a limited portion of the social network. Salesman is the term Gladwell uses for people who have a knack for persuasion. They are the people whose friends emulate their style, are convinced by them to vote Obama, and try out a never before visited restaurant. Salesmen have the ability to instigate deviance from the status quo and normal routine. The last agent Gladwell discusses is the maven. Mavens are the early adopters, those who waited in line at six in the morning to get that first iPhone, have extensive knowledge of the newest consumer products, and the newest Burning Man clone. They are the first to introduce new sources of social change into the social network.

The most appealing aspect of Gladwell's theory is its ambition. He seeks to demonstrate that the same agent properties and system dynamics are the source of all varieties of rapid, epidemic social change. This supplies a strong incentive to test the validity of Gladwell's theory. The first step towards determining the veracity of Gladwell's theory is to probe its internal consistency. This means determining whether or not the following conditional relationship holds: if agents have varying degrees of the three properties of mavenhood, salesmanship, and connectedness and these properties are distributed in such a way that a small number of agents excel far beyond the average in one or more of these traits, then if these outliers are indoctrinated by a new message this message will rapidly spread throughout the social network.

Because of the nonlinear causal relationships present in social networks and the heterogeneous agents that are their constituent parts, modeling the tipping point phenomenon using classical approaches from mathematics and statistics is insufficient. John Holland states that, "Nonlinearities mean that our most useful tools for generalizing observations into theory –trend analysis, determination of equilibria, sample means, and so on –are badly blunted." (Hidden Order pg. 5) Agent based models are an ideal tool for studying the emergent behavior of systems that that are made up of interconnected heterogeneous actors that have nonlinear causal relationships to one another. (Gilbert, 2008)

B. Facebook: A Case Study and Proposal

In this section I give a general overview of the current marketing strategy available to organizations using Facebook as an advertising platform and argue that it unduly focuses on individuals as isolated units rather than as agents embedded in a complex adaptive system. This conceptual shift, coupled with an agent based modeling approach to study information diffusion in Facebook's social network could be used to produce a new marketing tool focused on the creation of tipping points for advertisers' products.

Facebook is the largest social networking site in the world. It boasts over four hundred million active users. The average user is linked to 130 "friends", as well as 60 pages, groups and events.(Facebook) This is an incredibly dense and interconnected social network made up of people of all ages, nationalities and backgrounds who are constantly interacting with one another, generating content and clicking on advertisements.

Facebook is a free web application. Like Google, it generates revenue by selling ad-space on its site. What makes Facebook unique is the detailed knowledge it has about its individual users, knowledge that it is able to leverage to offer unheard of micro-targeting for advertisers.

Placing an advertisement on Facebook is a simple and automated process. After creating an advertisement according to Facebook's template, an advertiser is given a list of Facebook user characteristics that allow the advertiser to specify its target audience. These characteristics are incredibly far ranging and include information that many Facebook users would be unwilling to share on a survey or market research interview.

Some examples of the user information one is able to specify includes basic geographic information, demographic information including their relationship status, sexual orientation, religion, industries or specific companies in which the user has worked, whether or not any of the user's friends have visited a Facebook page, group or event administered by the advertiser's organization and whether or not it is currently the user's birthday. Of particular value is the ability to target people based upon their activities and interests. Facebook states that this information "…is based on information users list in their Facebook profiles like favorite movies and music, groups and Pages they have connected to and other information they have shared on the site." (Facebook)

Hy Mariampolski (2001) has described qualitative research as, "a family of approaches methods and techniques for understanding and thoroughly documenting attitudes and behavior… Qualitative approaches rely upon personal expressions and behavioral observations either in laboratories or in the natural environment and have a strong preference for unstructured or semi-structured questions --with a notable improvisational and interactive bent." (p. 7-8). In this sense, Facebook has created a qualitative market researcher's information utopia. They are granted unfettered access, not directly, but through the results of Facebook's data-mining and processing, to the details of the day to day lives of a dedicated user base whose behavior is generally unperturbed by the knowledge that their data is being studied. Facebook is the unobserved or perhaps more accurately ignored observer.

This extremely personalized form of advertising can have a high conversion rate over the set of individuals exposed to a particular advertisement. For instance one of my friends was targeted as someone who might be interested in an extreme race involving obstacles including leaping over flames and crawling under barbed wire. It is likely that she was targeted for this advertisement because of because of her age and sports driven interests, both of which are posted on Facebook. She actually signed up and had a great time. Within two days of the event she had commemorated it by posting pictures from the race to her Facebook page.

But is there a way to ensure our message reaches people that we are unable to directly target due to budgetary constraints, ignorance of an interested group or other factors? How can we go beyond a targeted audience and help our message spread virally? The qualitative approach to marketing focuses on the properties that a person as an isolated individual rather than the properties they have in virtue of their relationship to other agents in a social network. Gladwell's theory supports the argument that equally, or perhaps even more, importantly than the micro-targeting of individuals based upon the properties that they have in isolation is understanding the dynamics of the complex adaptive system in which they are embedded, the properties they have that are relevant to their role within this system, and how this knowledge can be leveraged to predict and cause the widespread dissemination of a message beyond a highly focused audience.

In their paper *Theories of Complexity*, Chu et al. describe this shifting of relevant agent properties using the concept of contextuality. If I am interested in gaining a deep understanding of a person's behavior, interests and general lifestyle there is a broad but specific set of properties she has that are relevant to gaining this understanding. If I am interested in the role that she plays within a larger social network I will study a different set of properties and ask a different set of questions. To whom is she connected? How does she interact with them? What are the products of these interactions?

The ability to acquire a detailed topographical knowledge of social networks as they exist in the physical world can be a daunting task. Fairly involved experiments need to be conducted in order to gain even a limited knowledge of their structure. A good example of this is the now famous small-world problem, better known as six degrees of separation proposed by Stanley Milgram(1967). The experiment he conducted was addressing the following question, "Starting with any two people in the world, what is the probability that they will know each other."(p. 62) In order to answer this question Milgram constructed the following experiment. He chose a person at random who lived in the United States. In one case a stockbroker from Boston. He then chose a distant town in Nebraska. He gave several volunteers from this town a package along with the name of the target person and as well as a few very general details about him. Their task was to try to deliver the package to him. If they did not know this stockbroker directly they were not to try to reach him directly but rather to pass on the package, regarding the target rules of the experiment and a roster archiving who the package had passed through on to someone who they thought would be more likely to know their target. On average, a chain of five people sufficed to connect the stockbroker to one of the initial volunteers in Nebraska.

In graph theory this is known as the characteristic path length. It is a global property of a graph that indicates the average distance between any two vertices within it. (Watts & Strogatz, 1998, p. 440) Studying the structure of virtual social networks in order to determine properties such as these is very different than studying their reality embedded counterparts. Facebook would not need to conduct an analogous experiment to Professor Milgram's; it could simply query their servers using well-known shortest path finding algorithms. But this does not mean they are able to predict how trends will spread through the network or what kind of marketing techniques will be most effective.

If you are one of the over one million Facebook developers and entrepreneurs (Facebook), the thought of producing a product that goes viral is mouthwatering. Take for instance FarmVille, a Facebook application that is a relatively simple social game in which you create and tend to a virtual farm. Since its creation in 2007, it has spread through Facebook garnering just shy of 75,500,000 users. That is approximately a quarter of all Facebook users and one percent of the world's population. It is inconceivable for a startup company to be able to afford to advertise to even close to the number of individuals that constitute FarmVille's user base.

Facebook has access to in-depth knowledge of the structure of their social network, and the vast amount of data about and produced by the agents that comprise it. An agent based network model similar to the one presented in this paper could be constructed utilizing this knowledge in order to predict how effective various marketing tactics would be at instigating tipping point phenomenon.

As mentioned previously such an approach would entail a focus on a different set of user properties than is currently available for Facebook advertisers. The set of properties of interest in this case would revolve around inter-agent interactions. The propensity for a user to adopt new Facebook applications and tools could mark them as a maven and early adopter. A user whose generated content produces a disproportionately large volume of responses from others could indicate the Facebook equivalent of a salesman. A high degree of connectivity to disparate friend groups could indicate a connector.

Although the set of possible actions a Facebook user can take in interacting with other Facebook users is large, it is finite and well defined. This fact demonstrates that it is possible to construct an artificial Facebook composed of agents modeled after Facebook users. This ABM could then be used to test en vitro the dynamics of information diffusion within Facebook. The results of these experiments could then be used to increase the precision of Facebook advertising focusing on the creation of the epidemic diffusion of information through the social network. The lever for the tipping point would be placed in the hands of Facebook's advertisers.

III. The Simulation

The logic for the tipping point simulation was coded entirely in JAVA. There were a variety of reasons for this. Agent-based models naturally lend themselves to an object oriented programming design paradigm. Agents consist of a set of properties and a set of actions that they can take in the world. This maps well to the state variables and member functions of a class. Java is machine agnostic which makes any java based ABM easily portable to different machines. When Gregg Rugolo and I originally set out to construct this ABM we wanted to build a visual representation of the simulation and JAVA has a wide variety of graphics libraries.

The simulation is made up of four main modules. These modules are a random connected graph generator, a random number generator that generates numbers according to various probability distributions, an agent module, and an overarching simulation module. First I will discuss the parameters and options governing each run of the simulation. Then I will give a conceptual overview of the four modules that constitute the simulation. This overview will entail a discussion of the algorithms necessary to their implementation as well as their role in the control flow of a run of the simulation.

Random Graph Generator:

While planning our Tipping Point research project Gregg Rugolo and I decided that the most effective way to model our agent interactions would be to embed them in a social network represented as a connected graph. This raised the interesting problem of how to efficiently generate connected graphs given a set degree sequence. I was unsatisfied with the existing tools I found and searched the graph theory and computer science literature until I came upon the paper, "Efficient and Simple Generation of Random Simple Connected Graphs with Prescribed Degree Sequence" (Viger, Latapy). This paper offers a study of an optimized version of random connected graph generation that would be particularly useful for our simulation, which would require many such network generations. The random graph generator in this project is essentially my JAVA implementation of their paper. Although this paper is not directly related to agent based modeling theory, because of the importance it plays in the Tipping Point simulation and the fact that it was the most work intensive portion of the project, I feel that it is relevant to present a conceptual overview of its functioning. The algorithm proceeds in three main steps:

Step 1: Realize the Sequence

This step involves utilizing the Havel-Hakimi algorithm in order to realize the degree sequence for the graph. The Havel-Hakimi algorithm takes as input a degree sequence and, if possible, realizes it by connecting each vertex to its designated number of neighbors. To do this, the algorithm iteratively selects the vertex that has the most unassigned vertices left and creates an edge between it and another node. It has been demonstrated that this algorithm is guaranteed to realize the sequence if it is possible.

Step 2: Connect the Disparate Components

Although the Havel-Hakimi algorithm realizes the sequence, it is very unlikely that in doing so it created one single connected graph. In order to connect all of the disconnected components I use a breadth first search algorithm that returns all of the vertices that are within a component as well as any redundant edges if the component is not a tree and has cycles.

Once this has been completed all of the components that are trees are connected to components containing cycles by performing an edge swap. This edge swap takes any two connected vertices from the tree and two vertices whose edge is redundant (e.g. its removal will not disconnect the component) from the cyclical component and switches the edges in order to connect the two components. For example if A-B represents the edge in the tree and C-D represents the edge in the cyclic component the swap would be the removal of these edges and the creation of the edges A-C and B-D. This maintains the degree of each vertex while connecting the components. A similar process is then performed in order to connect the disconnected cyclic components into one connected graph.

Step 3: Random Edge Swaps

So far we have only succeeded in creating a very specially structured graph in which a collection of densely interconnected components are tenuously connected by just two edges. In order to have an equal probability of creating any possible graph having this degree sequence, it is necessary to perform enough random edge swaps. Enough in this case was the square of the number of edges. This was a slightly arbitrary choice, but Viger and Latapy's paper hint that it is a reasonable one.

For any non-trivially sized graph, a naive implementation of this part of the algorithm will take a prohibitively long time. In a naive-implementation, two edges will be selected at random, then a depth first search of the entire graph will be conducted in order to ensure that this swap has not disconnected any components. If it had done so, the swap would be reversed. I had initially coded this implementation, and it took roughly 1.5 hours to perform the random edge swaps for a graph of ten thousand nodes. This would make performing many runs of the Tipping Point simulation very difficult

The optimized version of this step involve a heuristic technique that Viger and Latapy call K-Isolation testing. This is based upon two facts. Firstly, the chance of the graph being disconnected by any single edge swap is very low. Secondly, it is much more likely for a very small component of the graph to be disconnected than a larger one by an edge swap. This second fact suggests that if instead of performing a full depth first search of the graph after each edge swap, it is only necessary to check that the components attached to the new edges are sufficiently large for it to be incredibly unlikely that a disconnection has occurred.

After each edge swap, a breadth first search of the graph is conducted from one of the vertices in each of the two new edges. After seeing that each component contains at least K vertices the search terminates. K is assigned a value of one on the first iteration of this portion of the algorithm. After conducting the necessary number of swaps, the full graph is checked for connectedness. If it has been disconnected, the graph is reverted back to its previous state before the swaps, and the swaps are performed again with the value of K having been incremented by one. In my implementation of this technique it takes an around 5 seconds to perform the random edge swaps for a graph with ten thousand nodes, a much more reasonable amount of time for simulation purposes.

Random Number Generator:

In order to test how different probability distributions of traits (connectedness, salesmanship, mavenhood) would affect the rate of the dissemination of information in our model, it is necessary to have a random number generator that can output numbers based upon a variety of different probability distributions. The probability distribution generator class in the Tipping Point simulation is effectively a wrapper class for functions from the Colt library, an open source library for high performance scientific and technical computing developed by CERN. The probability distribution generator is able to generate random numbers based upon the power, uniform and normal distributions.

Agent Class:

The agents in our simulation are relatively simple. They are each coupled with a node in the graph that indicates their level of connectedness and the other agents to whom they are connected. They are each assigned a degree of mavenhood and salesmanship. Their heterogeneity stems from having differing levels of these traits.

The only action that the agents can take is to interact with each other. These interactions are trivial if they are both already carriers of the message or are neither carriers. In either such case a message will neither spontaneously emerge nor be destroyed.

However, if one of the agents is a carrier and the other is not, they will attempt to convince the other one to come over to their point of view. The likelihood of the other agent changing stances is determined by the salesmanship of the persuader and the mavenhood of the agent being persuaded.

Simulation Class:

The simulation class takes as input the parameters necessary for generating the social network and the agents. It then conducts a run of the simulation. This entails iterating through every agent and having the agent interact with one of its neighbors chosen at random. The number of cycles through the population that occurs is a parameter that is designated by the user.

V. Experimentation:

A. Hypothesis and Experiment Design

Although I believe the tipping point agent based model is general enough to test a wide range of hypothesis, for the purposes of this paper I will seek to test two that I find particularly interesting and relevant.

1) I would like to test the validity of Gladwell's assertion that the outliers he discusses within his book, the connectors, mavens and salespeople are deeply important to social change taking place. Implicit within Gladwell's discussion is that these traits are distributed relative to a fat tail distribution where a small number of the agents are producing the greatest affect.

In order to test this hypothesis I will generate a social network and distribute properties according to a power law distribution. I will then conduct a series of runs altering the way in which a message is initially introduced into the network and see what affect this has in the eventual outcome.

I will use two different ways to initialize the first carriers of the message. The first way is to introduce it to a group of agents selected at random. This would be analogous to the way unselective or less selective mass marketing techniques and mediums such as billboards and T.V commercials attempt to spread their message. The second way would be to introduce the message to an equivalent number of agents but to ensure that these agents are outliers in at least one of the three category traits identified by Gladwell. An example of this might be similar to companies paying Facebook to place targeted advertisements focusing on users having specific traits such as account activity and number of friends. Although this might cost more money initially, the dynamics of the social network in which the agent they are attempting to persuade might be such that their message will have a disproportionally powerful impact relative to money spent. If Gladwell's belief is correct, the message will propagate through the social network more rapidly and successfully given this second form of initialization.

B. Results

I conducted ten runs of the simulation that used a power law distribution to generate the character traits and initialized the connectors, mavens and salesmen. This resulted in seven runs in which the message failed to disseminate through the population and three in which it remained relatively constant until the maximum number of cycles parameter was reached.

I conducted ten runs of the simulation using a normal distribution with an equivalent average degree sequence for each agent property and initialized the connectors, mavens and salesmen. This resulted in two successful runs in which 90% message saturation was reached, one neutral run and seen failed runs.

There was an incredibly large array of graph shapes and there were no clear general trends that occurred for one type of run or the other. The tentative conclusion that the results of this experiment suggests is that the role the outlying agents play is not as essential to the diffusion of information as Gladwell has postulated. However, do to the unpredictable results of runs using various parameter combinations I believe to draw stronger conclusions about the validity or invalidity of Gladwell's claim it will be necessary to implement an automated parameter searching algorithm to locate parameter combinations of interest as discussed in the Future Work section.

VI. Future Work:

A. Active Nonlinear Tests

The parameter combination space for the tipping point agent based model is extensive. For any particular run of the simulation there are eighteen to twenty different parameters that can be altered (depending on the type of probability distributions selected). To explore the entire parameter space would be computationally infeasible because of the computational complexity of the algorithms involved and thus the amount of time required for each run. This makes it difficult to discover parameter combinations that lead to the production of interesting behavior in the model.

Something I would like to experiment with is known as active nonlinear tests of complex simulation models (ANTs). John Miller (1998) describes ANTs as the, "use [of] a nonlinear optimization algorithm to search across a set of reasonable model perturbations with the objective of maximizing the deviation between the original model's prediction and that obtained from the perturbed model (p. 820) This system could be just as easily used to discover the parameter combinations that produce results most consistent with a given theory or to discover some other behavior that can be formalized in a fitness function.

An example of this framework applied to the Tipping Point ABM would be use a genetic algorithm to search over a bounded parameter set and to define the fitness of the individuals in the GA population to be directly proportional to the speed with which a population becomes saturated with a message.

B. Creation of a General Network ABM Framework

Although we are able to produce dramatically different networks by generating them using degree sequences given by different probability distributions, this model is still fairly limited in the kinds of network structures it is able to produce. An important component of social networks (and many other phenomenon that are isomorphic to a graph) is what is known as clustering.

A simple extension of the graph generation algorithm could be used to generate graphs with this kind of structure. An algorithm that would perform this function would initially use the graph generation class to generate a collection of connected graphs according to the number of clusters desired in the final network. Random edge random swaps would then be performed to connect all of these graphs. The more swaps performed the less radical the clustering will be. The amount of clustering could be measured an analysis of the clustering coefficient. Watts and Strogatz (1998) define the clustering coefficient as a local property that, “the cliquishness of a typical neighborhood.” (p. 440)

Without have an instantiation of a graph produced using such an algorithm and thus being able to study its properties it is difficult to say exactly what kind of graph structures could be produced. Given that the generation process loosely interconnects many densely connected graphs, it seems reasonable to assume that a short characteristic path length as well as tightly clustered neighborhoods will be maintained. These are properties common to small-world graphs. Watts and Strogatz (1998) have shown that many disparate phenomenon including, “The neural network of the worm Caenorhabditis elegans, the power grid of the western United States, and the collaboration graph of film actors,” are small-world networks. (p. 440) Because of the wide array of phenomenon that can be modeled using small-world networks, a programmatic extension allowing for their generation would be a valuable tool.

VI. Bibliography

P. Erdos and T. Gallai. "Graphs with prescribed degree degree of vertices." Mat. Lapok, 11:264-274, 1960.

Facebook Press Room 2010. Facebook. 19 May 2010.
http://www.facebook.com/press/info.php?statistics

Facebook Advertising. 2010. Facebook 19 May 2010.
http://www.facebook.com/advertising/

Stanley Milgram, "The Small World Problem," Psychology Today (1967) vol. I, p. 60-67.

Gregg Rugolo, "The Development of a "Tipping Point" Agent Based Model:
A Collaborative Research and Programming Project by John Doty & Gregg Rugolo" 2009

F. Viger and M Latapy. "Efficient and simple generation of random simple connected graphs with prescribed degree sequence." Computing and Combinatorics (2005) p.440-449

FarmVille Facebook Group. 2010. Zynga. 19 May 2010.

Chu et al. "Theories of Complexity" Complexity (2003) vol. 8 p. 19-30

Watts & Strogatz. "Collective Dynamics of 'Small-World' Dynamics" Nature (1998) vol. 393

Holland J. "Hidden Order" (1995)

Gladwell, M. "The Tipping Point" (2000)

Bak. "How Nature Works". (1996)

Gilbert. "Agent Based Models" (2008)