Social Networks and the News: An Agent-Based Model of a Local Media Market

Oleg Smirnov

oleg.smirnov@stonybrook.edu

Allan Ingraham

aingraham@empiris.com

We propose an agent-based computational model of a local media region that consists of two components: local media voices reporting local news and a social network of households receiving the news. To our knowledge, this is the first attempt to model news dissemination at the individual level in a media market of up to 600,000 households, which approximately corresponds to a media region roughly the size of the New Orleans "designated market area" (DMA). The model can be applied to address a variety of research questions. Here we report how basic characteristics of the DMA in question and the social network affect local news dissemination among households.

Oleg Smirnov (corresponding author), Department of Political Science, Stony Brook University, Stony Brook, NY 11794. Phone: 786-566-1185. Allan Ingraham, Managing Partner, Empiris, LLC, Washington D.C. We thank Enrique Armijo, Jonathan Blake, Ellen Goodman, and Matthew Spitzer for helpful comments.

I. Introduction

A problem central to the field of mass communication is the understanding and measurement of the dissemination of news within a society (e.g., Ostgaard 1965; Price and Zaller 1993; Price and Czilli 1996; Rogers and Seidel 2002; Monge and Contractor 2003; Watts and Dodds 2007). Since one's ability to receive timely and accurate information contributes to one's welfare, the effect of the amount of news coverage in a given area on news awareness in the household is a subject worthy of study. For example, a lack of access to local news regarding the Chernobyl disaster of 1986 and, as a result, a lack of information sharing with regard to that event caused thousands of people to be exposed to radiation with grave health consequences (Baverstock and Williams 2006). Although Chernobyl is an extreme example, it shows the importance of understanding the amount of news coverage necessary to promote news exposure.

Despite the importance of understanding the relationship between local media coverage and the diffusion of local news within a social network, this relationship, to our knowledge, has yet to be analyzed in a formal (mathematical) model. This paper represents an initial attempt to fill this void in the literature. In addition to the contribution this model and the associated analysis below offers, the model could be adjusted to study other relevant research questions such as electoral campaigns and candidate issue positioning, advertising efficiency, the effects of media consolidation or closure, or the effect of a new media form's emergence on information dissemination. To further this research, we publicly share the code of our program in an appendix to this paper.

The paper is organized as follows. In section II, we provide the reader with necessary background knowledge relating to media study, and we review the relevant academic literature.

Section III summarizes the model both informally and formally. Section IV provides the results of the analysis, and section V concludes.

II. Relevant Background and Literature Review

A formal model of media coverage and news diffusion would be based on explicit assumptions regarding the agents in the model, the agents' environment, information in the environment, and interaction between agents. The fully specified model would then be used to analyze, for example, the speed and breadth of information transmission within a given area. Although analytic computation of the results would be preferred, computational analysis would be used in the event that direct solutions to the model are not tractable.

A model of a local media region with local media voices, news stories, and viewing households represents a major modeling challenge. To appreciate fully this challenge, consider that a realistic model will consist of the following elements:

1) different local media consisting of television, radio, daily and weekly newspapers, and the Internet;
2) household specific preferences for different types of media or for specific media voices within a type of media (for example, a house might prefer a TV to radio and the Internet, and may prefer station "A" to station B);
3) different scopes of new coverage across media voices—that is, some events may be covered by several media voices, some by only a few voices, and others may not receive local news coverage;
4) news events may be transmitted within the network by individuals that share information with each other;

5) certain individuals sharing news with one another may have frequent interactions with others or infrequent interactions with others, and the probability of sharing news with a person with whom one interacts may vary across the population.

As the outline above illustrates, modeling the diffusion of news within a society is a complicated endeavor. We therefore propose a two-level agent-based computational model of a local media region. We model a *local* media region consisting of local news distributed through local media voices (first level) to a local audience (second level). This model substantively corresponds to the “designated market area” (DMA), a term introduced by Nielsen Media Research that represents a set of geographic boundaries used in analyzing media subscription and viewership.[1]

The model's purpose is to analyze the diffusion of local news into a social network consisting of households and those households' resulting *news exposure*, defined as agents’ “knowledge of local news.” News exposure represents the proportion of all local news on a given day to which households have access. For example, if there are ten local newsworthy events and the average household has access to news coverage of seven of the ten events, then the media

---

[1] Neilsen Media Research has defined 210 different DMAs in the United States. The most populous is the New York City DMA with approximately 7.5 million TV households. Glendive, MT is the lowest population-ranked DMA, with approximately 4,000 television households). Running the model on a standard workstation—namely, a machine with Intel Core2Duo processor, 2GB or RAM, and Windows Vista 32 bit operating system—we are able to analyze DMAs with approximately 600 thousand or fewer television households. This amount encompasses about 75 percent of the DMAs in the United States (DMA 50-210). If run on a more powerful machine, the model could be used to study even the largest local media markets.

market news exposure is 0.7 (or 70%). Furthermore, we can study news exposure to a single story from one or more media voices (e.g., Chernobyl nuclear disaster of 1986 (Baverstock and Williams 2006), or September 11th terrorist attack (Rogers and Seidel 2002)). Before we turn to detailed description of the model, we discuss the place of our work in a larger context of the mass communication and social science literature.

Although we primarily focus on news diffusion within the social network of households, the news transition among media voices themselves is an important and complex research question (Leskovec et al. 2009). The model allows one to specify local media voices as a set of independent agencies with idiosyncratic news programs. It also allows one to model local media voices as a social network of media agencies that exchange information and with some probability report similar news stories. In either case, some of the local events may receive coverage from many local media voices while other events may be covered only by a single media voice. Some local newsworthy events may be not covered at all. For example, if there are 10 local events and only one local media voice with a capacity to cover 8 stories, then 2 events will not be reported at all. In the modeling section below, we describe the difference between a “potential news story” and a “reported news story.”

Empirical examination of news exposure is challenging. For example, Price and Zaller (1993) explore the measurement of individuals' use and absorption of the news by examining audience recollection of prominent news stories. They find that self-reported measures of news exposure did not significantly predict the likelihood that individuals would ultimately retain the news they claimed to have viewed. The authors, therefore, conclude that it may be more prudent to focus on the general political knowledge of subjects as opposed to self-reported measures of news exposure when analyzing problems relating to the news.

In a related paper, Price and Czilli (1996) hypothesize that (1) the most important factor determining whether someone recognizes or recalls a news story is how much coverage it receives; (2) more attentive audience members will both recognize more stories and recall stories better; and (3) personality-focused or domestic stories are more important to most people than other types of news. Audience survey data come from the 1989 pilot study of the National Election Studies, and news categorization data come from Vanderbilt University's *Television News Index and Abstracts*. The most significant, positive determinants of story recognition and recall were minutes of coverage per week, personality- or domestically focused nature of the story, and audience members' background knowledge of the news.

In this paper, we avoid the issue of news absorption and analyze whether a household has access to a news story. For our purposes, whether individuals actually internalize the news is a separate research question. We do assume, however, that some households are capable—with a small probability—of transmitting a news story to their social neighbors.

In another paper, Rogers and Seidel (2002) study the effect that the news of the 9/11 terrorist attacks had on Albuquerque residents and find that the news traveled quickly, and the cumulative diffusion of the story followed the usual S-shaped logistics curve such that the rate at which the news spread peaked between half an hour and two hours after the events occurred. By noon almost everyone, whether at work or at home, in the sample had heard of the attacks. In our terms, the single story news exposure variable was approaching one. We believe that the nearly universal knowledge of the news story was a function of both media coverage of the event as well as information exchange within the social network of households.

In our model, we assume that all households in the network have the same level of influence when it comes to information transmission—an assumption that could, in theory, be problematic. Watts and Dodds (2007) study the effect of agents in the network that have a disproportionately high level of influence on information passed within the network. The authors refer to such agent as "influentials." By modeling and simulating information diffusion, the authors find that there are circumstances under which influentials would become the primary drivers of information within the network. These circumstances, however, are rare. Relying on this research, we therefore find that endowing some households with greater influence than others would not substantially enhance the model.

Watts and Dodds (2007) also found that the primary driver behind information dissemination was the existence of easily influenced agents that transfer information to one another. This result highlights the importance of one’s understanding of the effects of the parameters of the social network on the behavior of that network. That is, there are assumptions that will produce effects from influentials that are, at times, stunning. One must understand, however, whether those effects are indeed realistic based on the parameters that one has chosen in constructing the social network. A solid understanding of the model’s construct, its parameterization, and the information dissemination within it must be present to evaluate correctly the existence of influentials and their effect on the network. In constructing our Monte Carlo simulation, which we discuss in detail below, we treat this result in Watts and Dodds (2007) seriously. Specifically, the simulations are performed over a wide range of network parameters, which reduces the likelihood that the general results are produced by a specific parameter value or set of parameter values in particular groups of model runs.

As a final note, it bears mention that the overall literature on information diffusion and social networking is vast (see, for example, Wheeler and Mitchelson 1989; Borgatti and Cross 2003; MacFarlane 2003; Watts 2003). The field of communications research has drawn on a formidable arsenal of tools and data to study various network phenomena (Wasserman and Faust 1994; Newman, Barabási, and Watts 2006), such as the spread of contagious diseases (Ball, Mollison, and Scalia-Tomba 1997), the proliferation of rumors, fashions, and opinions (Watts 2002), and, most recently, the advancement of phenomena such as obesity, happiness, or smoking (Christakis and Fowler 2009). In contrast to the social networking model presented in this paper, much of the research in social networking is geared toward estimation of parameters that characterize structural features of a real world network such as reciprocity, balance, and density (Monge and Contractor 2003).[2] Unlike this research, we rely on *exogenous* network parameters to create an artificial social network that we use to analyze the flow of information from the sources of that information. For this reason, the correspondence of the social network to the real world network of households, news, and media will depend, in part, on whether the choice of underlying network parameters is informed by real world characteristics.

III. The Agent-Based Model

First, we provide an informal summary of the model. We then discuss our motivation for using an agent-based model and the advantages of this approach. Finally, we present the formal model.

A. Non-technical Overview of the Model

[2] A recent advancement in this area of research on social networks is, for example, the exponential random graph models—commonly referred to as "p*" in the literature. This method and others are powerful tools to analyze social network data and estimate the parameters that characterize the underlying network (Robins, 2009; Robins, Pattison, Kalish, and Lusher, 2006).

We use a computational agent-based model (Epstein 2006) to study local news coverage and the resulting information diffusion in a DMA. Like an analytical model, an agent-based model is built on formal assumptions about agents (players in games) and how they interact. As in analytical models, the assumptions are clearly defined and the results are quantitative and easy to replicate (Gilbert and Troitzsch 1999). Computational models generate data to show the relationships between variables of interest. An agent-based model simplifies analysis of equilibrium paths, recognition of emergent patterns of interaction, and rapid generation of a model (such as the one in this paper) where an element of interaction is particularly complicated (reference withheld). Simply put, agent-based computational modeling sheds light not only on the outcome of a process but also on the dynamics of the process itself without sacrificing the rigor of formal modeling (Nelson and Winter 2002).

The model is a multi-level agent-based computational model of a local media market. We focus on the basic characteristics of a generic local media market that may affect the levels of news exposure among households. Unlike media markets in the real world, however, our market is completely transparent. For instance, we know each household's exact media voice preferences for news consumption and the news stories to which it is exposed. In the model, we control the parameters that define both the market and the society. We can alter population size, change the structure of the social network, vary the number of television stations, or increase the online news media audience size. We then can observe how each change affects the aggregate population's news exposure, its exposure to a specific set of news stories, or its exposure to a single news story.

The agent-based model consists of two main components: households that form the social network and media voices that provide news coverage to the households. A household represents a specific “agent” within the network, and a household is connected to other neighboring households. Information is initially added to the network from one or more external sources—namely, media voices. With some probability, neighboring households exchange information over time, and therefore contribute to the propagation of news. We do not, however, model cognitive psychology of agents. That is, we study news exposure rather than news awareness (Price and Zaller 1993; Price & Czilli 1996). Given the object-oriented nature of the computer program, adding an psychological component to the model that could account for individual beliefs, memory, or desire to seek and retain the news would be feasible.

We model the news as a vector of data that represents the total number of *potential* news stories in a given market within a given day. Although it is an abstract concept, one can nevertheless estimate its size empirically. The size of the news vector is estimated from the demographic characteristics of the society, population size, and violent crime statistics (which we explain in greater detail below). In the model specification used here, news coverage by any particular media voice is random and independent of the new coverage selected by all other media voices. However, no one voice in a particular market is assumed to cover all the news in that market.[3]

With the model in place, we then use it to determine how the properties of the social network and the characteristics of the media market affect news exposure for the average household. Variables of specific interest include the structure of the social network, population

[3] Under an alternative model, media voices would form their own social network (e.g., Leskovec et al. 2009). In this case, news coverage by a media voice would affected by the information received from other media voices. This effect is not modeled in this paper, and is instead reserved for future work.

size, total number and number of each type of media voices, amounts of news production,[4] the percentage of households within the market that obtain news from each type of media voice, and other variables. The use of computer simulation allows one to examine the same media market numerous times while selectively changing these and other market attributes.

To parameterize the model, we rely on empirical evidence from academic research, official government sources, and various media surveys. In this fashion, we are able to create a computational model that approximates a generic real-world DMA. We are then able to compute repeatedly the diffusion of news within the social network within a given period of time while holding parameters constant or by changing only a set of the model’s parameters. We are therefore able to track how the addition of, say, 10 percent to the size of the local population affects the total population’s ability to receive the news quickly, or how the addition of a particular media voice affects news exposure.[5]

---

[4] It is our understanding that "news hole" is the term that industry uses to refer to the amount of news produced by a media voice. For example, the "news hole" for a program broadcast on television or radio corresponds to the amount of airtime for the program. The news hole for a newspaper then refers to the number of printed pages of news content.

[5] The model is programmed using the *R programming language* (R Development Core Team 2008) and the igraph library (Csárdi and Nepusz 2006). Both R and *igraph* can be downloaded free of charge at <http://www.r-project.org/>, and our results can be replicated using the computer code and data available at <author’s website>.

B. Why Use an Agent-Based Social Network Model?

The modeling of information diffusion in a media market is complex. There are thousands of connected households and dozens of media voices that differ in a multitude of ways. Analytical modeling of such a complex system requires simplifying assumptions to achieve closed form solutions (Taber and Timpone 1996). For example, economists study advertising and social welfare in a media market by assuming that there are two firms competing for viewers and each firm carries one program with advertising (Gabszewicz et al. 2000; Gal-Or and Dukes 2003; Anderson and Coate 2005). It is also assumed that the audience consists of a single representative viewer, who is choosing which program is closer to his or her own preferences.[6] Because making the model more realistic by relaxing certain assumptions would render it analytically intractable, and because the simplified model is unrealistic when compared to the complexity of a representative DMA, an alternative solution is needed. Hence, the agent-based social network model is valuable in tackling this problem, as the degree of realism in our model makes computational methods the only feasible option.

C. The Computational Model

The population of the media region consists of N households. The structure of the population within the media market is the "social network." Each household in the network has a number of connections—formally called “edges”—to other households in the network. The social network is created computationally using the Watts-Strogatz small-world network model (Watts and Strogatz 1998). Characteristics of the small-world network closely correspond to the characteristics of social networks in the real world (Newman, Barabási, and Watts 2006). Most households in the small-world network model are not connected to each other directly. However,

[6] The model is based on Hotelling's spatial model (Hotelling 1929).

they are linked together through a small number of intermediate connections. The small-world network model is one example of the well-known phenomenon of the “six degrees of separation” (Watts 2003), in which any two individuals are separated by just six neighboring steps in the network.

Each household receives local news from a number of media voices. News media is comprised of television, radio, print, and the Internet. Within each type of news media, there are a certain number of voices, such as specific television stations, particular newspapers, and local news websites. *M* denotes the total number of media voices. The probability that any given household receives news from any given media voice is an exogenous parameter based upon empirical evidence such as Nielsen Media Research (www.nielsenmedia.com) survey data (see details below).

The population matrix $\underset{N \times M}{\mathbf{P}}$ describes each household’s access to each media voice. As an *N* by *M* matrix, each element of the matrix describes the transmission of news from a particular media voice to a particular household. If a household *i* receives local news from a media voice *j*, then $\mathbf{P}[i, j] = 1$; otherwise, $\mathbf{P}[i, j] = 0$.

The *daily vector of news*, $\underset{V \times 1}{\mathbf{D}}$, represents news coverage of all daily events in the designated market area (DMA). Each element of $\mathbf{D}$ corresponds to a single news story. The total number of events—the length of the vector of news—is equal to *V*. Strictly speaking, the length of the vector of news is an arbitrary number since the exact length is impossible to measure. Moreover, significant day-to-day variation in the number of daily events in a DMA is likely to occur. However, one can estimate $\mathbf{D}$ empirically as a function of the DMA population size, average household size, violent crime statistics, and data on the percentage of news devoted to the coverage of violent crimes (see the Results section below).

Coverage of the news by a particular media voice is chosen randomly, and no single media voice covers all local news events. The media coverage matrix $\underset{M \times V}{\mathbf{C}}$ describes *each* media voice's coverage of *all* elements of the daily vector of news. If media voice *m* produces news coverage of an event *v*, then $\mathbf{C}[m,v]=1$; otherwise, $\mathbf{C}[m,v]=0$. Notice that in the simulation, we examine how news coverage in time period 1 is then diffused over time within the social network of households.

Together, the population matrix, $\underset{N \times M}{\mathbf{P}}$, and the media coverage matrix, $\underset{M \times V}{\mathbf{C}}$, allow one to examine household exposure to the coverage of all daily events. The *N* by *V* matrix $\underset{N \times V}{\mathbf{W}}$, describes each household's access to the news coverage of any particular event on a given day. Thus, each row in the matrix represents a household whereas each column represents exposure to a particular news story. The matrix $\underset{N \times V}{\mathbf{W}}$ can be found by multiplying the matrices $\underset{N \times M}{\mathbf{P}}$ and $\underset{M \times V}{\mathbf{C}}$ and then determining whether each element of the resulting *N* by *V* matrix is nonzero. Specifically, let $\underset{N\times V}{\mathbf{Z}} = \underset{N\times M}{\mathbf{P}} \times \underset{M\times V}{\mathbf{C}}$. Then define the news matrix as follows:

(1) $$\mathbf{W}[i,v]=\begin{cases}1 \text{ if } \mathbf{Z}[i,v]\geq 1\\ 0 \text{ otherwise}\end{cases}$$

In words, if household *i* is exposed to news coverage of event *v at least once* by any combination of media voices, then $\mathbf{W}[i,v]=1$; otherwise, $\mathbf{W}[i,v]=0$. This expression means that the variable of interest is news exposure *per se* and not how many times an individual was exposed to a news story.[7]

[7] An alternative convention—see, for example, Price and Czilli (1996)—would make news absorption contingent on multiple instances of news exposure. Such an assumption would be

The primary dependent variable is the average news exposure of a household, which given by the elements of $\mathbf{W}$. For example, in a media market with 4 households, 3 media voices, and 5 daily events, the population matrix, media coverage matrix, and news exposure matrix might be as follows:

$$\underset{N\times M}{\mathbf{P}}\quad \underset{M\times V}{\mathbf{C}}\quad \underset{\mathrm{N}\times V}{\mathbf{Z}}\quad \underset{N\times V}{\mathbf{W}}$$

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 2 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \tag{2}$$

In this case, the average household news exposure is the *sum* of all elements of the $\mathbf{W}$ matrix divided by the total number of elements: $(2+4+4+1)/20 = 0.55$. Here, news exposure is 0.55, which means that an average household is exposed to coverage of 55 percent of *all local events*.

To simulate information diffusion within the social network, we assume that households can share news with each other. We examine all edges (connections) in the social network for each time period. Each household has a randomly drawn probability of news transmission to a "neighboring" household. The term "neighboring" is used in a social sense—individuals need not live next door to one another to be socially connected. The probability of news transmission decreases with time, meaning that older news is less likely to be transmitted. The news that households share with each other is chosen randomly with the obvious constraint that households cannot share news to which they are unexposed. Also notice that news transmission may be redundant if the recipient of the story was already exposed to it.

---

more appropriate in a model of news *awareness*, which would account for individual cognitive processes.

IV. Results

To analyze computational results from the model, we examine the dynamics of individual simulation runs (sample runs) as well as the aggregate data from multiple runs (Monte Carlo simulations). The advantage of analyzing individual sample runs is that it allows one to inspect the inner workings of the model and gain insight into the model's dynamics (Gilbert and Troitzsch, 1999). Here we start from a baseline model and use that model to generate comparative static graphs.

The advantage of the Monte Carlo analysis is that it allows one to examine data derived from a large parameter space and account for possible interaction effects among the explanatory variables (Mooney 1997; Robert and Casella 2004). The Monte Carlo study also allows one to examine the robustness of results produced using the individual sample runs. Below we discuss results first from various sample runs and then from Monte Carlo simulations.

A. Sample Runs

In this section, we examine data from individual sample runs and use those data to perform comparative statics graphically. That is, we plot the behavior of dependent variables in response to changes in the independent variables of interest, holding all other variables constant. A number of parameters will be held constant throughout the sample runs. The illustrative simulations persist for 14 time periods with 10 percent news salience discounting; in other words, the probability that a household shares old news with a neighboring household is 10 percent lower today than it was yesterday. The 10 percent discounting leads to non-linear asymptotic loss of salience, since only 90 percent of the prior period's relevance remains today.

Unless otherwise specified, we examine a local news media market with 5 TV stations, 10 radio stations, 3 Internet sites, and 2 daily newspapers. We further assume—using Nielsen

Media Research data—that approximately 60 percent of the population uses television for local news, 30 percent relies on radio, 50 percent uses newspapers, and 30 percent uses Internet sites. Since Nielsen Media Research data may not be entirely reliable, in Monte Carlo simulation below we examine a much larger parameter set in order to account for local media markets where these data may differ substantially from those reported by Nielsen Media Research. For example, and as we explain in the Monte Carlo section below, the newspaper reliance parameter varies from just 29% in some simulations all the way to 68% in others.

We also assume (unless otherwise noted) that each media voice can produce news coverage of 8 local events (in addition to the coverage of local events from the previous days), which roughly corresponds to a third of all local events in a generic DMA with 150,000 households (see Results). The size of the news coverage is another exogenous parameter of the model that we vary randomly in Monte Carlo simulation below.

First, we examine how the social structure and the population size affect the average household news exposure. We compare two population sizes, 10,000 households and 100,000 households for three types of social networks: (1) lattice with an average of about 4 neighbors, (2) small-world network with an average of about 4 neighbors, and (3) small-world network with an average of about 8 neighbors. In the lattice network, individuals are located on a two-dimensional grid and have connections only with the four immediate neighbors (located to the north, south, east, and west from the individual). Individuals on the edge of the lattice and in the corners may have a smaller number of neighbors. The small-world network, on the other hand, can be viewed as a lattice with distant connections (in fact, one way to create a small-world network is to start with a lattice and then “rewire” some of the connections to distant individuals). A higher rewiring probability implies a more inter-connected society, in which space does not inhibit inter-personal communication. In this respect, the small-world network structure potentially allows a greater number of social connections.

Sample simulation runs yield a number of non-trivial results (see Figure 1).

[FIGURE 1 IS ABOUT HERE]

Population size alone does not seem to affect the average household news exposure, and this is true for all three types of social structures that we study. Keeping all other variables constant, including the size of the news vector and the number of media voices, the populations of 10,000 households and 100,000 households have nearly identical percentages of news exposure. Interestingly, news exposure for the lattice (approximately 0.62) and small-world with an average of 4 neighbors (approximately 0.64) are very similar. The small difference may be attributed to the fact that the small-world network with an average of 4 neighbors has a marginally greater number of total connections than the lattice since the agents on the edges of the lattice have less than 4 neighbors. By contrast, news exposure in a small-world with an average of 8 neighbors appears to be much higher (0.77 for 10,000 households and 0.78 for 100,000 households). Thus, the primary variable that affects the news exposure is the average number of connections, not the population size or the social network structure. The latter result may come as a surprise for the researchers studying communication networks. Recall, however, that a media market is different from a standard social network. Unlike rumors, ideas, diseases, and other transmitted elements characterized by a single source or small number of sources, the news stories are transmitted to a large number of households at once from a large number of sources (media voices).

It is unrealistic, however, to assume that the size of the news vector is the same for small and large populations. Recall that the daily vector of news $\underset{V \times 1}{\mathbf{D}}$ represents all daily events in the designated market area. We can estimate the approximate size of the news vector using a number of empirical findings. The State of the News Media 2006 study conducted by the Project for Excellence in Journalism (http://www.journalism.org) estimates that 30.7 percent of total broadcasting is devoted to crime (most of which is violent crime); for local television the number is 41.7 percent. The percentage is smaller for print and online media, 23.4 percent. We make a conservative assumption that the percentage of news stories devoted to crime should not exceed 20 percent. That is, we may be *overestimating* the total number of local events that the journalists should cover. However, this should not affect our substantive results based upon the comparative statics. We further assume that the percentage of local events which are *not* reported in the media is comparable for crimes and non-crimes. Thus, knowing the actual number of violent crimes in the market region we can estimate the approximate total number of local events (potential news stories) assuming that the violent crime stories constitute 20% of all the news.

To model a specific DMA, we would need to use the total number of violent crimes in the area, the actual proportion of news coverage in the local media devoted to crimes, and the average household size (which may differ as well, for example, from 2.08 in D.C. to 3.01 in Utah). For a generic simulation run and an abstract DMA we can assume that the number of violent crimes in the area is equal to the U.S. average of 473.5 violent crimes per 100,000 individuals per year (FBI: http://www.fbi.gov/ucr/ucr.htm) and the U.S. average household size of 2.6 individuals per household (U.S. Census Bureau: http://www.census.gov). The size of the vector of news, therefore, can be found according to the following formula:

$V = \left(\left(N * 2.6 * \left(473.5/100000\right)\right)/365\right)/20\%$, where N is the total number of households in the market region[8].

Making the size of the news vector a function of population size causes one to re-evaluate the precise relationship between the population size and news exposure. Figure 2 shows the dynamics of news exposure for the markets with 50,000, 150,000, and 250,000 households, while keeping all other variables constant.

[FIGURE 2 IS ABOUT HERE]

Clearly, greater population size has a negative effect on the news exposure. Although, the speed of news dissemination is the same in small and large markets (as we show above), there are more local events—and, therefore, news stories—to report in larger markets. Our baseline set of media voices results in ~0.95 news exposure in a DMA with 50,000 households. The same media voices can only achieve ~0.61 news exposure in a DMA with 150,000 households, and ~0.41 in a DMA with 250,000 households. Essentially, we confirm the intuitive idea that there is a greater natural supply of news in larger markets; consequently, there is a greater demand for news production.

---

[8] In the formula, $\left(473.5/100000\right)$ represents the number of violent crimes per person per year. $N * 2.6$ is the total number of people in the region. Thus, $\left(N * 2.6 * \left(473.5/100000\right)\right)$ is the total number of violent crimes in the region per year, which we then divide by 365 to get the average number of violent crimes per day. Since violent crimes constitute approximate 20% of the total number of potential news, the total number of potential news is 5 times larger (hence, division by 20%).

One obvious variable that should affect the amount of news production is the number of news stories produced by a media voice. Empirical evidence indeed suggests that media voices in larger markets produce more news. The State of the News Media 2006 reports that the average amount of time spent on local TV news is just over 3 hours per day in small markets (DMA 150-210) and well over 4.5 hours per day in large markets (DMA 1-25). To examine the effect of the size of news production, market size, and news exposure, we conduct a simulation of the following markets: markets with 100,000 and 400,000 households, and markets in which TV stations cover 8 and 16 local events per day (along with the news coverage of past events and other programming). Figure 3 shows that markets characterized by more productive media voices have higher levels of news exposure for both small and large markets.

[FIGURE 3 IS ABOUT HERE]

The figure further suggests that increasing news exposure by means of greater news production is effective but more challenging in larger markets. Doubling TV news production has a greater impact in a market with 100,000 households than in the market with 400,000 households, despite the fact the smaller market has a smaller news vector.

Another variable that may increase news exposure is the total number of media voices in the market. To determine whether a greater number of media voices alone increases news exposure, we keep the news production and audience size constant (each media voice covers 8 events; the total audience size for TV, radio, newspapers, and Internet sites is, respectively, 60 percent, 30 percent, 50 percent, and 30 percent). We examine how a different number of TV stations (2, 5, and 8) affect news exposure in two kinds of markets, with 50,000 households and 400,000 households. Interestingly, increasing the number of TV media voices in the small market does not seem to increase news exposure (Figure 4a).

[FIGURE 4 IS ABOUT HERE]

In fact, having two TV stations with 30 percent of the audience in a small market results in greater news exposure than having five TV stations with 12 percent of the audience. A closer look at the media coverage matrix $\underset{M \times V}{\mathbf{C}}$ for the small market reveals that two TV stations are sufficient to cover most elements of the news vector. In a larger market, however, two TV stations are not enough to cover most elements of the news vector (Figure 4b). Additionally, a greater number of TV stations in a larger market have a lower probability of overlapping news coverage.

Thus far the dependent variable has been the average household news exposure, which measures the aggregate public knowledge of all local events. Many scholars, policy-makers, and entrepreneurs may be also interested in public knowledge of a specific set of news, or even a single news story. Advertisers, for example, may be interested in how a single favorable news story produced by a single media voice is disseminated in the population. Similarly, politicians may be interested in greater news coverage of some issues (over which they have an advantage) and not others (over which they are likely to lose). Our computational model makes it very easy to examine news exposure to a selected set of stories. Consider, for example, our baseline market configuration: 150,000 households, 5 TV stations (60 percent total audience), 10 radio stations (30 percent), 3 Internet sites (30 percent), and 2 daily newspapers (50 percent), and uniform coverage of 8 local events per media voice. In this market, we can examine how a unique news story produced by a single media voice gets disseminated in the population. An explicit comparison of various media types for this market shows that a single newspaper story will eventually reach 59 percent of the population, 42 percent for a TV story, 29 percent for an Internet story, and just over 9 percent for a unique radio story (Figure 5).

[FIGURE 5 IS ABOUT HERE]

What happens if Internet stories are more likely to be shared within the social network (presumably, by means of email)? This and other similar questions can be easily addressed within the present computational framework. For example, if we assume that Internet news stories are twice more likely to be shared, a single Internet story will reach close to 48 percent of the population—more than that of a TV story despite the fact that TV stations have a larger total audience.

B. Monte Carlo Simulation Runs

To examine how the average population news exposure is affected by a variety of factors, we conduct Monte Carlo simulations (Mooney 1997; Robert and Casella 2004). We run the computational model 10,000 times. For each run, the simulation parameters (independent variables) are drawn randomly from an exogenously specified probability distribution. Thus, we examine a different hypothetical market with each Monte Carlo simulation run. The goal of the simulations is to generate data with enough variance to detect statistically significant relationships between the news exposure and characteristics of a social network and a media market. The sheer number of simulations allows us to examine a large parameter space.

A number of variables are kept constant throughout the simulation runs. Specifically, each simulation lasts 10 time periods. We also assume that the probability that an individual will transmit older news is discounted by 5 percent. The initial ($t = 1$) probability of news transmission for any given household is between zero and one, uniformly distributed.

The network size—that is, the number of households—varies uniformly between 50,000 and 600,000. In this respect, the smallest possible market that we model is DMA 201 (St. Joseph, MO), whereas the largest possible market is DMA 53 (New Orleans, LA) (source: http://www.tvb.org). The households form a social network created on the basis of the Watts-Strogatz small world model (Watts and Strogatz 1998). The number of outgoing connections that a household has is a random integer between 1 and 4 (thus, the total number of connections per household is, on average, between 2 and 8). The rewiring probability, which is a key element of the Watts-Strogatz model, varies between 0 and 0.5. This probability is a likelihood that household i loses a connection with one of the neighbors (e.g., $i-1$ or $i+1$) and forms a new connection with a random household from the population (e.g., $i+532$).

The total number of media of voices covering local news in the media market is the sum of four types of media voices: television, radio, daily newspapers, and the Internet. The total number of TV and radio stations varies uniformly between 2 and 10 (for each type), and the total number of daily newspapers and local news web-sites varies between 1 and 5. Clearly, each market region is unique with its own composition of media voices. Our computer model allows examination of a DMA with any variety of media voice counts. The numbers of TV and radio stations, newspapers, and Internet sites are parameters in the model that can be changed by the user given the relevant empirical evidence. The generic market region that we model is fairly representative of a DMA with 100,000 to 400,000 households (DMAs ranked between 150-210). Data for TV and radio can be obtained from the U.S. Federal Communications Commission's website: http://www.fcc.gov/mb/video/, http://www.fcc.gov/mb/audio/. For the number of local newspapers we examined a sample of small DMAs and found two newspapers to be the most common case. For example, a typical DMA 192 (Charlottesville VA) has two locals newspapers reporting local news: The Daily Progress and The Cavalier Daily.

To simulate the audience size we rely on the Nielsen study mentioned above, which reports that approximately 60 percent of individuals learn about the local events from television, 30 percent from radio, 50 percent from daily newspapers, and 30 percent from the Internet. In the simulation, the total audience of a media type is drawn from a normal distribution with the mean of the distribution corresponding to the above percentages; the standard deviation was assumed to be 10 percent of the value of the mean. The audience of a single media voice is equal to the total audience of the corresponding media type divided by the number of media voices of that type in the market (for example, if the TV total audience is 50 percent and there are 5 TV station, it is assumed that each station gets 10 percent of the total audience).

Each media voice can cover a certain number of elements of the daily news vector, which corresponds to the number of unique daily news stories covering local events (events from the previous days would count as additional elements of the news vector, but we model a news vector for a single day). The selection made by each media voice of news items to cover is random and independent from the selections made by other media voices. For each simulation run, the maximum number of stories that a media voice can cover varies uniformly between 1 and 10, subject to constraints given by the size of the vector of news (for example, if the total number of daily events is 9, the maximum number of stories can only be 8). We assume that no media voice can cover all the elements of the news vector.

In each run, after *each time period*, we record the average household news exposure, defined above as a proportion of stories in the vector of news to which a household is exposed. The Monte Carlo simulation parameters also can be inferred from the descriptive statistics of the generated data (Table 1). The table also contains information about the dependent variable, average household news exposure, after each of the ten time periods.

[TABLE 1 IS ABOUT HERE]

The average amount of news exposure across all simulation runs is approximately 24 percent after the first time period and 39 percent after ten time periods. Figure 6 shows the frequency distribution of the news exposure after ten periods for all 10,000 simulation runs.

[FIGURE 6 IS ABOUT HERE]

As evident from the figure, we achieve our primary goal of generating enough variance in the dependent variable to allow us to conduct more detailed statistical tests and examine how simulation parameters affect it. Of interest is the difference between news exposure at $t = 10$ and news exposure at $t = 1$; the value of the variable corresponds to news exposure generated by information diffusion within the social network, or the news that people learn from each other. The average news exposure gain from inter-household news transmissions within the network was 15 percent with a standard deviation of 8 percent. The number does not reflect the *total* number of news transmissions, but rather the total number of *effective* transmissions, or the news that people first learned from each other and not from the mass media. The total number of news transmissions within the network is much larger but most of the transmissions are redundant as households share the news that they already knew from the media.

We use regression analysis to determine the factors that affect news exposure in the media market (Table 2).

[TABLE 2 IS ABOUT HERE]

The three statistical specifications only differ by the dependent variable, which is the *news exposure* at $t = 1$ in Model #1 ($R^2 = 0.77$), the *news exposure* at $t = 10$ in Model #2 ($R^2 = 0.86$), and the *difference* between the news exposure levels at $t = 10$ and $t = 1$ in Model #3 ($R^2 = 0.89$). Varying parameters of the computational model such as the population size and the number of media voices, we could examine the rates of news exposure for different kinds of social networks and media markets.

From the regression analysis, it is evident that population size has a negative effect on news exposure. All other variables being equal, the average household news exposure (at $t = 10$) decreases by 11.3 percent for each additional 100,000 households in the media market (and 8.2 percent at $t = 1$). In other words, the average household in a larger market is exposed to a smaller *percentage* of news coverage of local events. Notice that this does not have to be true in *absolute* terms. Consider an example of two households, one in a small market and one in a large market. The household in a large market may be exposed to 10 news stories while the household in a small market may be exposed to 8 stories. If the total number of local events is 20 in the large market and 10 in the small market, the resulting levels of news exposure are 50 percent and 80 percent respectively. Although the household in the large market is exposed to 2 more stories than the household in the small market, the latter is 30 percent more knowledgeable about the local events. The Model #3 also shows that the extra amount of news exposure generated by information diffusion among households is 3.1 percent smaller for each additional 100,000 households. Not only it is more difficult for media voices to cover a greater percentage of local events in a larger market; the news transmission within the social network is also slower.

Among the characteristics of the social network, both the number of social connections and the rewiring probability have a positive effect on news exposure. Each additional [outgoing] connection per household increases the average household news exposure by 6.5 percent (at $t = 10$). Similarly, a greater likelihood of having distant social neighbors increases news exposure. If an average household is 10 percent more likely to have such neighbors, the average household news exposure increases by 0.15 percent. Both results suggest that in a modern society characterized by a greater number of long-distance information diffusion, such as over the phone or the Internet, the population is exposed to a greater percentage of news coverage of local events.

Among the media characteristics, we observe that the popularity—that is, the audience size—of all types of media increases news exposure. The same is true for the quantity of news reported for all types of media. For each additional news story reported by an average television station, the average household news exposure increases by 0.9 percent. Similarly, an average radio station reporting an additional news story increases news exposure by 0.5 percent. The number is the same for each additional news story per web-site (0.5 percent). Finally, an additional story in an average newspaper increases news exposure by 0.7 percent.

Although an increase in the number of media voices in a local market increases the average household's news exposure (at both $t = 1$ and $t = 10$), the share of news exposure generated by the information diffusion within the social network actually decreases. A simple explanation of this result is the fact that a greater number of households receive the news from the media, as opposed to social neighbors. In this case, many news transmissions among the households become ineffective, or redundant (that is, the household learns from another household about a news item to which it has already been exposed by a media voice).

The number of media voices has a statistically significant effect on news exposure at t = 10, but no effect on news exposure at t = 1. At t = 10, with all other variables being equal, each additional television station increases news exposure by 0.2 percent, each additional Internet site increases news exposure by 0.3 percent, and each additional newspaper increases news exposure by 0.4 percent. The number of radio stations does not have a statistically significant effect on news exposure. The results may appear counter-intuitive. One would think that a greater number of media voices would dramatically increase news exposure. This is likely to be true if we make the audience size and the number of media voices positively correlated. In the computational model, however, the size of the audience and the number of media voices are independent random variables. For instance, a media market with 3 television stations is equally likely to have 60 percent of the total TV audience as the market with 6 television stations.

V. Concluding Remarks

This paper presents an agent-based computational model of a media market. The model consists of two major levels: media voices reporting news and the households that form a social network. Similarly, there are two kinds of information diffusion in the model: (1) vertical, from the media voices to the households, and (2) horizontal, news transmission among the households themselves. The model is based upon clearly and formally stated assumptions, and the computer code is publicly available. Not only can one easily replicate our results and analysis, but one can also work with our code and modify it to examine various other aspects of mass communication in a media market.

To our knowledge, this is the first such model that explicitly simulates the diffusion of news from dozens of media voices and among thousands of households. The largest market size that we are able to simulate consists of 600,000 interconnected households, which approximately corresponds to DMA 53 (New Orleans, LA). However, we expect that our substantive results – specifically, the causal relationship among variables – will hold for even larger markets as well.

In our view, such a computational approach is a valuable tool in communications research as it allows the researchers to examine a multitude of various market strategies and public policies without actual implementation, which is likely to be costly and may be prone to error.

Although we find computer simulation of media markets very promising, it is important to recognize the standard limitations of computational models as such recognition should only improve the quality and analysis of these models. Two such limitations are the underlying assumptions of the model and an incomplete coverage of the parameter space. To deal with the first limitation, we base our model on empirical evidence (such as population statistics and media research surveys) and prior research (such as the social network analysis and agent-based modeling). Because our assumptions are likely imperfect, we do not focus on the descriptive statistics and the exact levels of news exposure. Instead, we emphasize the regression analysis, which allows one to study the *relationships* among the variables of interest. To deal with the second limitation—the incomplete coverage of the parameter space—we rely on computational power. Although the total number of parameter configurations, or hypothetical markets, is infinite, we managed to achieve a substantial coverage of the parameter space by conducting 10,000 simulation runs. These simulations allowed us to generate enough variance in the dependent variable to conduct meaningful statistical tests.

References

Anderson, Simon P., and Stephen Coate. 2005. "Market Provision of Broadcasting: A Welfare Analysis." *Review of Economic Studies* 72(4):947-972.

Ball, F., D. Mollison, and G. Scalia-Tomba. 1997. Epidemics with two levels of mixing. *Annals of Applied Probability* 7:46-89.

Baverstock, Keith, and Dillwyn Williams. 2006. "The Chernobyl Accident 20 Years on: An Assessment of the Health Consequences and the International Response." *Environmental Health Perspectives* 114(9):1312-1317.

Borgatti, Stephen P., and Rob Cross. 2003. "A Relational View of Information Seeking and Learning in Social Networks." *Management Science* 49(4): 432-445.

Christakis, Nicholas A., and James H. Fowler. 2009. *Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives*. Little, Brown & Company.

Csárdi, Gábor, and Tamás Nepusz. 2006. "The igraph software package for complex network research." *InterJournal Complex Systems*, 1695.

Epstein, Joshua M. 2006. *Generative Social Science: Studies in Agent-Based Computational Modeling*. Princeton, NJ: Princeton University Press.

Gabszewicz, D., J. Laussel, and N. Sonnac. 2000. "TV-Broadcasting Competition and Advertising." Discussion Paper 00/6, CORE, Universite Catholique de Louvain.

Gal-Or, E. and A. Dukes. 2003. "Minimum Differentiation in Commercial Media Markets," *Journal of Economics and Management Strategy* 12(3):291-325.

Gilbert, Nigel, and Klaus Troitzsch. 1999. *Simulation for the Social Scientist*. Philadelphia, PA: Open University Press.

Hotelling, H. 1929. "Stability in Competition," *Economic Journal* 39(153):41-57.

Leskovec, Jure, Lars Backstrom, and Jon Kleinberg. 2009. “Meme-tracking and the Dynamics of the News Cycle.” *Proceedings of the 15th ACM SIGKDD international conference on Knowledge Discovery and Data Mining*.

MacFarlane, Alistair G. J. 2003. "Information, Knowledge and the Future of Machines." *Philosophical Transactions: Mathematical, Physical and Engineering Sciences* 361(1809): 1581-1616.

Monge, Peter R., and Noshir S. Contractor. 2003. *Theories of Communication Networks*. New York, NY: Oxford University Press.

Mooney, Christopher Z. 1997. *Monte Carlo Simulation*. Thousand Oaks, CA: Sage University Press.

Nelson, Richard R., and Sidney G. Winter. 2002. “Evolutionary Theorizing in Economics.” *Journal of Economic Perspectives* 16(2):23–46.

Newman, Mark, Albert-László Barabási, and Duncan J. Watts (eds.). 2006. *The Structure and Dynamics of Networks*. Princeton, NJ: Princeton University Press.

Ostgaard, Einar. 1965. "Factors Influencing the Flow of News." *Journal of Peace Research* 2(1):39-63.

R Development Core Team. 2008. *R: A Language and Environment for Statistical Computing*. URL: http://www.R-project.org.

Robert, Christian P., and George Casella. 2004. *Monte Carlo Statistical Methods (2nd ed.).* New York, NY: Springer.

Taber, Charles S., and Richard J. Timpone. 1996. *Computational Modeling*. Thousand Oaks, CA: Sage University Press.

Wasserman, Stanley, and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. New York, NY: Cambridge University Press.

Watts, Duncan J. 2002. “A simple model of global cascades on random networks.” *Proceedings of National Academy of Sciences USA* 99:5766-5771.

Watts, Duncan J. 2003. *Six Degrees: The Science of a Connected Age*. New York, NY: Norton.

Watts, Duncan J. and Steven H. Strogatz. 1998. "Collective dynamics of 'small-world' networks." *Nature* 393:409-410.

Wheeler, James O., and Ronald L. Mitchelson. 1989. "Information Flows among Major Metropolitan Areas in the United States." *Annals of the Association of American Geographers* 79(4): 523-543.

Figure 1. News Exposure and Network Structure (Exogenously Fixed News Vector).

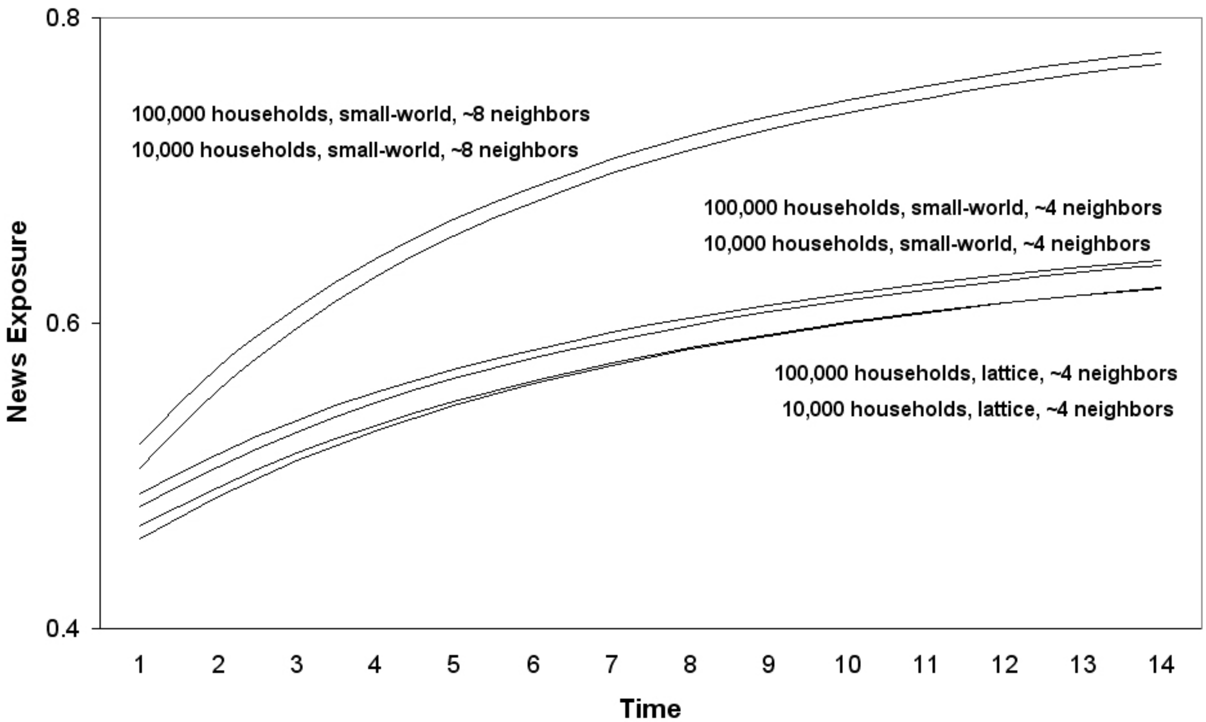

Figure 2. News Exposure and Population Size (Endogenous News Vector).

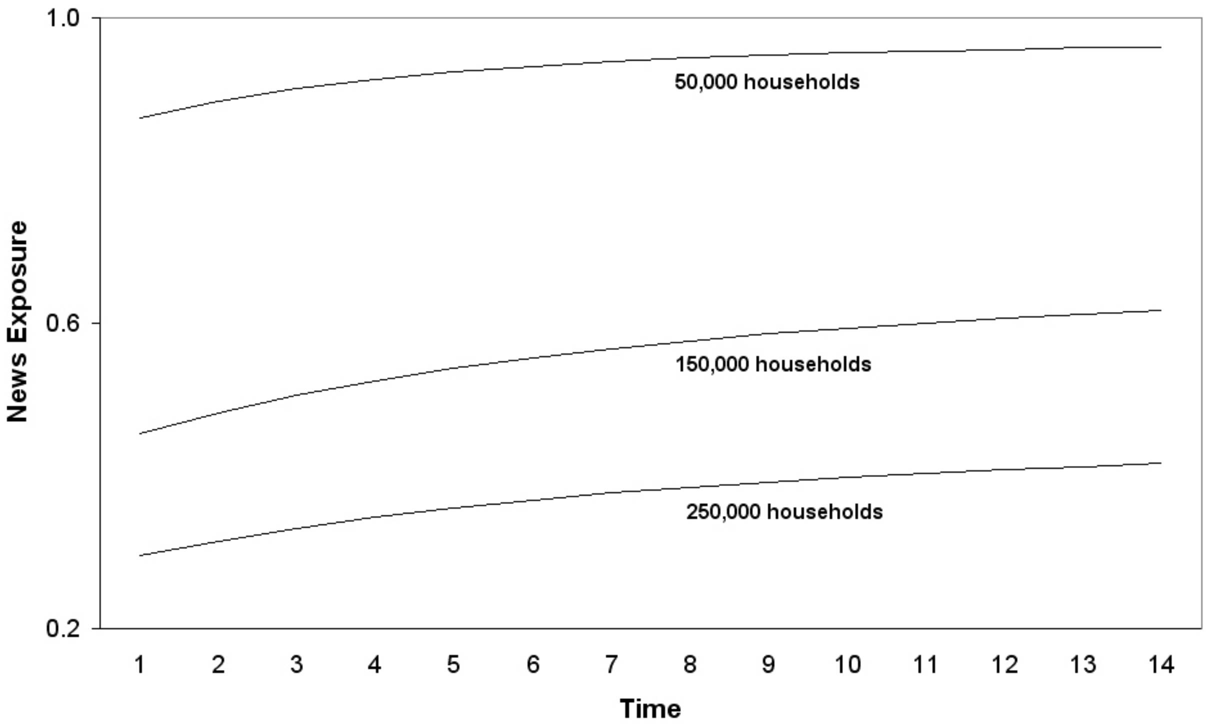

Figure 3. News Exposure and News Production.

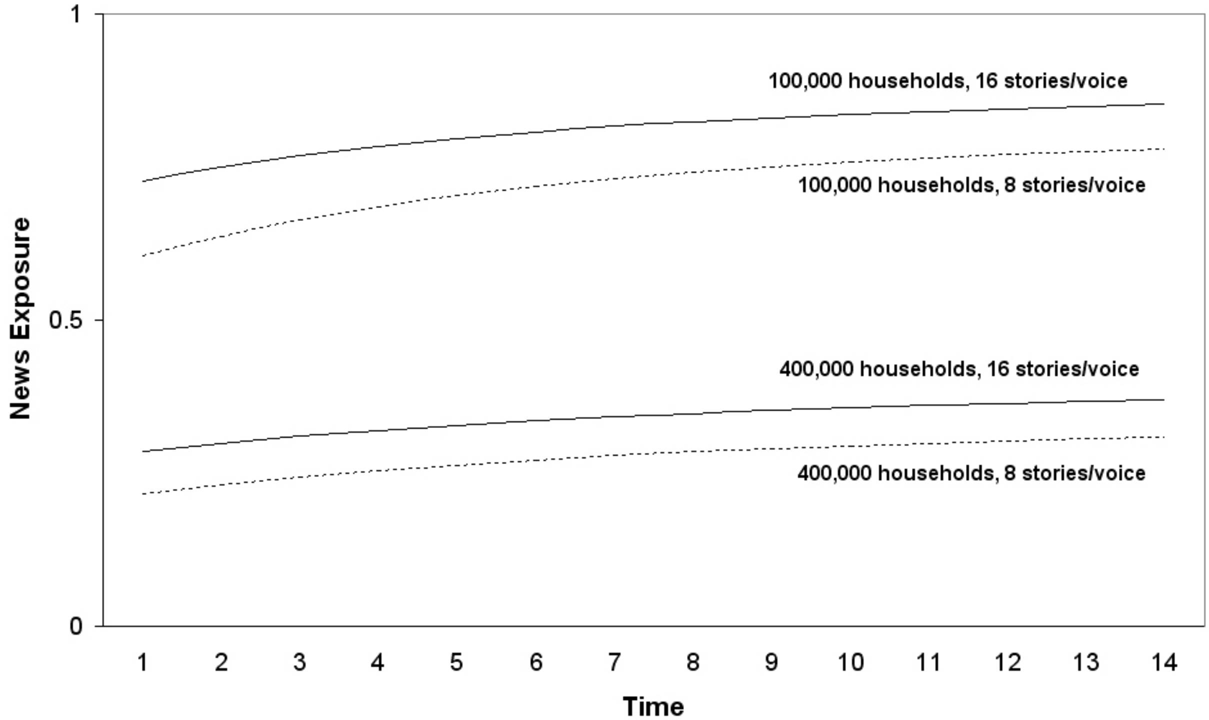

Figure 4a. Number of Voices in a Small Market (50,000 households).

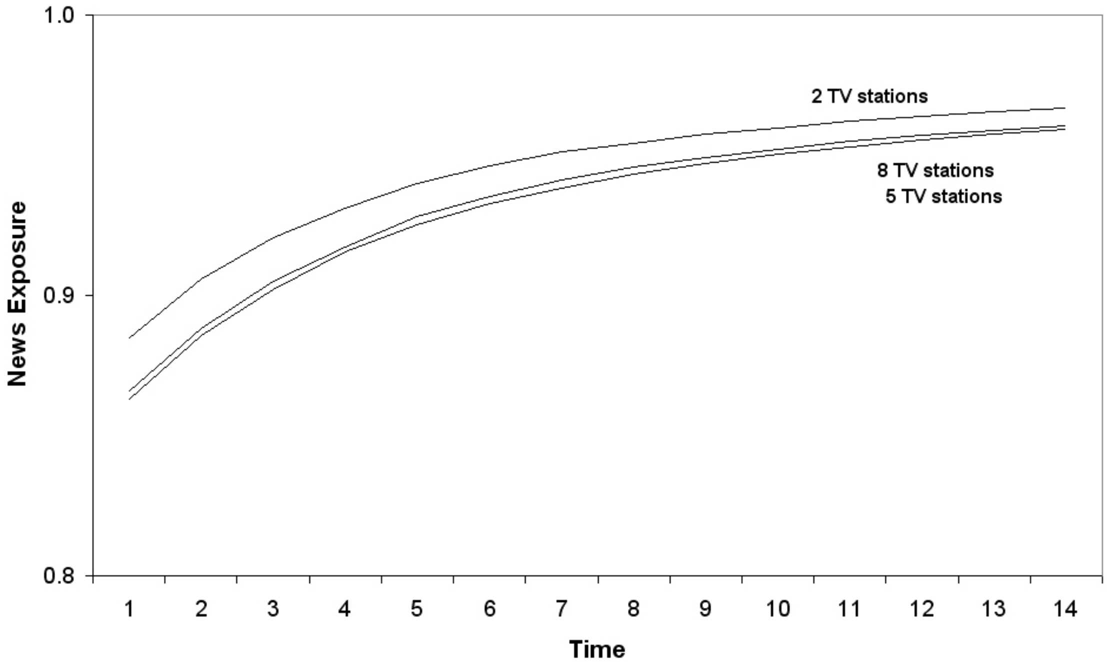

Figure 4b. Number of Voices in a Larger Market.

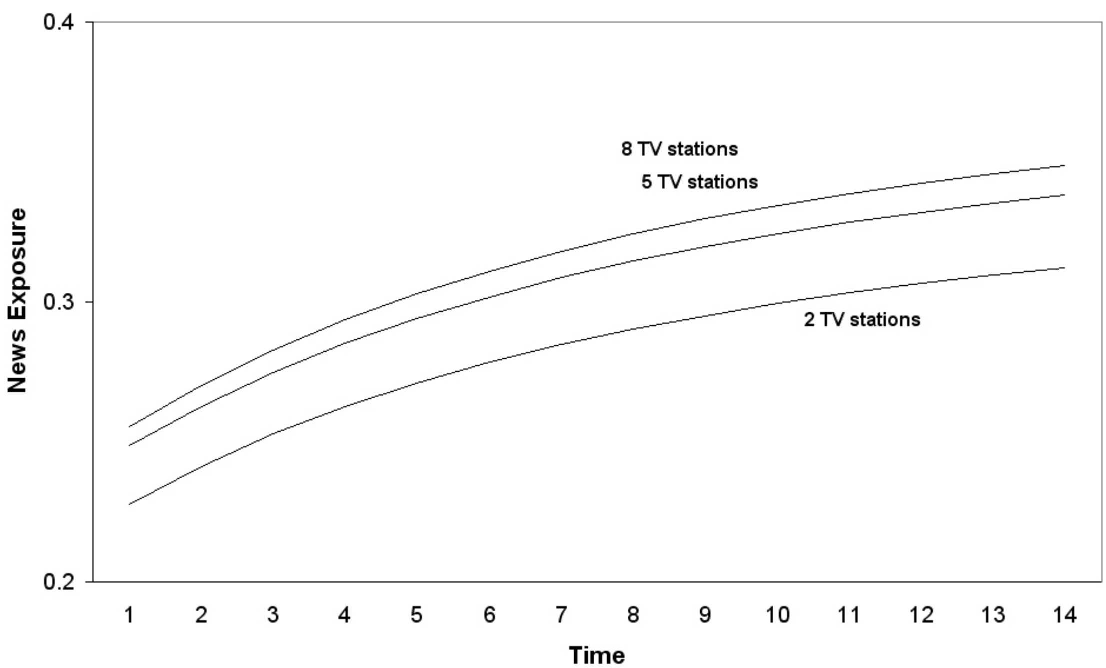

Figure 5. Single Story Exposure by Media Type.

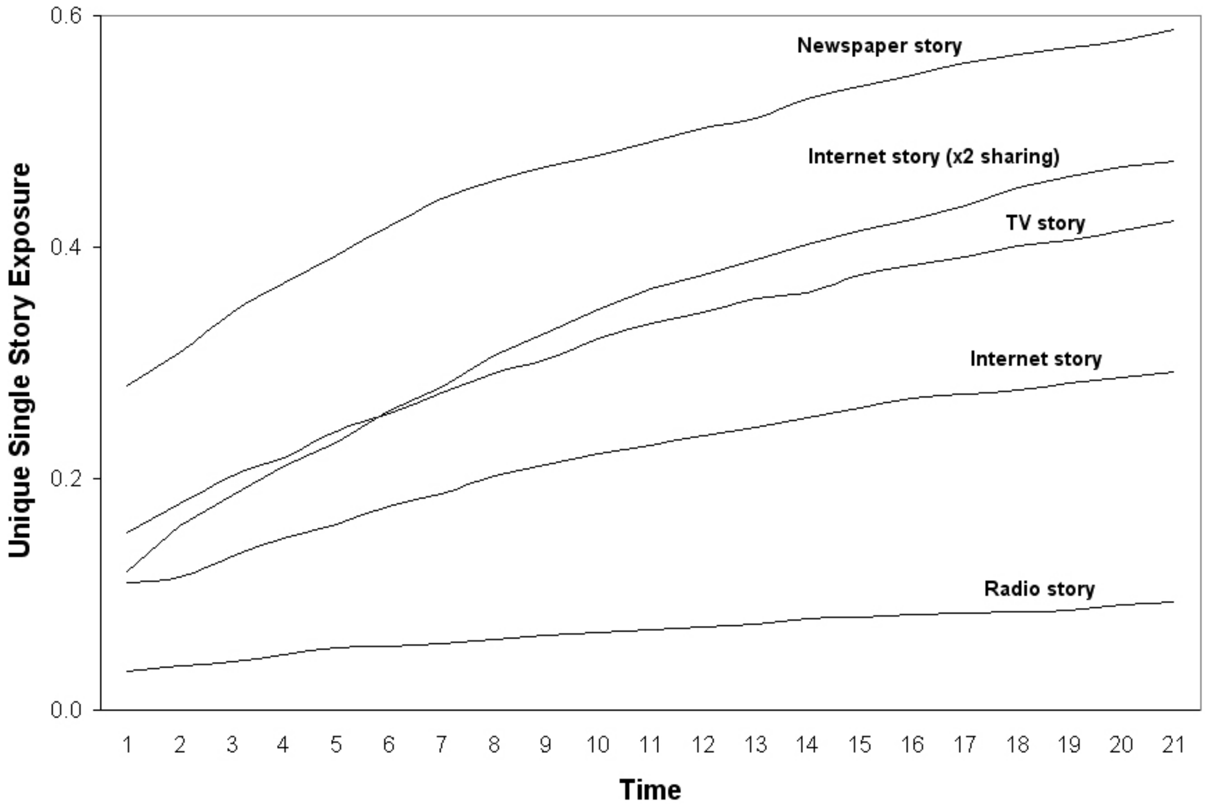

Figure 6. Frequency Distribution of the Dependent Variable at t=10.

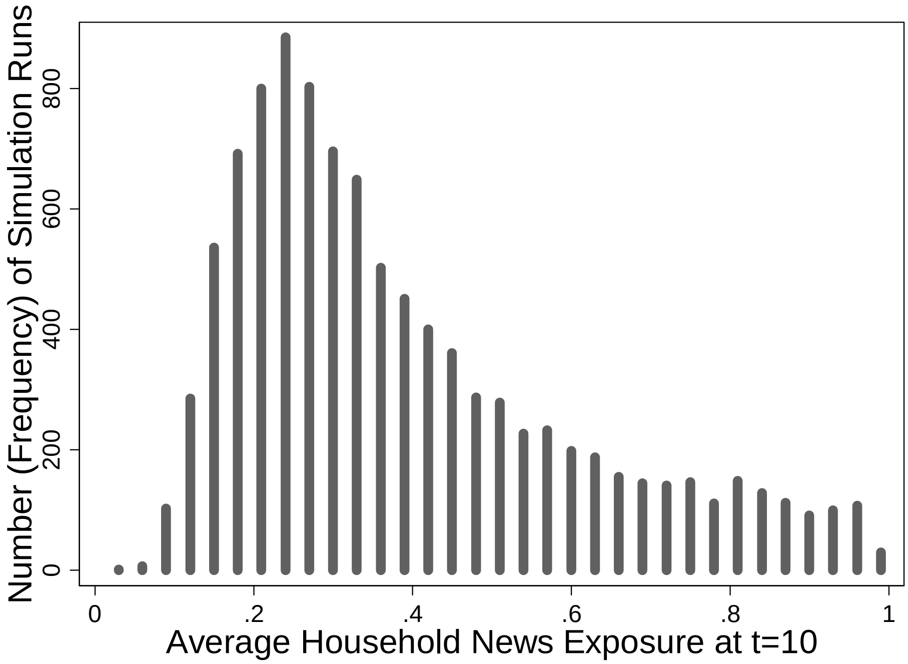

Table 1. Monte Carlo simulations (n=10,000). Descriptive Statistics.

	Mean	Std. Dev.	Min	Max
Social Network Variables				
Households (in thousands)	326.95	157.89	50	600
News vector length	54.64	26.62	8	101
Rewiring probability	0.25	0.15	0.00	0.50
Outgoing connections / household	2.50	1.12	1	4
Number of Media Voices				
Number of TV stations	6.01	2.56	2	10
Number of radio stations	6.01	2.57	2	10
Number of Internet sites	3.00	1.41	1	5
Number of newspapers	2.99	1.41	1	5
Media Audience Size				
Total TV audience	0.60	0.06	0.37	0.85
Total radio audience	0.30	0.03	0.18	0.42
Total Internet audience	0.30	0.03	0.18	0.41
Total newspaper audience	0.50	0.05	0.29	0.68
Local News Coverage				
TV daily local news	5.51	2.86	1	10
Radio daily local news	5.50	2.89	1	10
Internet daily local news	5.47	2.87	1	10
Newspaper daily local news	5.45	2.86	1	10
Average Household News Exposure				
News exposure, t1	0.24	0.16	0.02	0.89
News exposure, t2	0.27	0.17	0.02	0.92
News exposure, t3	0.29	0.18	0.03	0.95
News exposure, t4	0.31	0.19	0.03	0.96
News exposure, t5	0.33	0.19	0.03	0.97
News exposure, t6	0.34	0.20	0.04	0.98
News exposure, t7	0.35	0.20	0.04	0.98
News exposure, t8	0.38	0.21	0.04	0.99
News exposure, t9	0.38	0.21	0.04	0.99
News exposure, t10	0.39	0.21	0.04	0.99
News exposure, t10 – t1	0.15	0.08	0.02	0.45

Table 2. Regression Analysis of the Average Household News Exposure.

Model #	(1)	(2)	(3)
DV: News exposure (NE)	*NE(t=1)*	*NE(t=10)*	*NE(t=10) – NE(t=1)*
Households (in 100,000s)	-0.082 (173.42)**	-0.113 (221.90)**	-0.031 (181.72)**
Rewiring probability	0.003 (0.57)	0.015 (2.66)**	0.012 (6.37)**
Outgoing connections / household	0.013 (19.97)**	0.065 (90.31)**	0.052 (214.22)**
Number of TV stations	0.000 (1.55)	0.002 (6.89)**	0.002 (16.24)**
Number of radio stations	-0.001 (1.85)	-0.000 (0.19)	0.000 (4.55)**
Number of Internet sites	0.001 (1.45)	0.003 (4.97)**	0.002 (10.81)**
Number of newspapers	0.000 (0.37)	0.004 (7.05)**	0.004 (20.02)**
Total TV audience	0.092 (7.36)**	0.060 (4.48)**	-0.032 (7.04)**
Total radio audience	0.081 (3.26)**	0.071 (2.62)**	-0.011 (1.20)
Total Internet audience	0.143 (5.67)**	0.117 (4.31)**	-0.026 (2.86)**
Total newspaper audience	0.106 (7.06)**	0.061 (3.75)**	-0.045 (8.37)**
TV daily local news	0.010 (37.57)**	0.009 (32.73)**	-0.001 (6.42)**
Radio daily local news	0.005 (18.39)**	0.005 (17.61)**	0.000 (1.58)
Internet daily local news	0.005 (18.78)**	0.005 (16.38)**	-0.000 (3.14)**
Newspaper daily local news	0.008 (32.27)**	0.007 (24.97)**	-0.001 (14.92)**
Constant	0.143 (8.99)**	0.293 (17.08)**	0.150 (26.08)**
Observations	10000	10000	10000
R-squared	0.77	0.86	0.89

Absolute value of t statistics in parentheses
* significant at 5%; ** significant at 1%

Online Appendix

```
#R code to accompany Social Networks and the News: An Agent-Based Model of a Local Media Market
#author information
#igraph library must be installed
#edit parameters, copy, and paste into R console
  library(igraph)
#index variables
  m=t=j=x=z=0
# salience/memory discounting
  df=0.95;
#single run duration in time periods
  T=21
#single run data matrix
  data=array(0,dim=c(T,5))
#50k to 150k households corresponds to DMAs 150-200
  N2=200;
#the size of the daily vector of news (what there is to cover)
  vN=as.integer((((N2*2.6)*(473.5/100))/365)/0.2)
# number of households to model
# for faster approximate results multiply N2 by 10 or 100
# otherwise, multiply by 1000 (slow for large nei and vN)
  N=N2*10;
#rewiring prob
  rp=0.25;
#number of neighbors (even number only)
  nei=8
#min and max(pr(transmission))
  ptL=0.0;ptU=1.0;
#media voices total and the initial distribution
  tTV=5; tRA=10; tIN=3; tNE=2;
  tM = tTV + tRA + tIN + tNE;
#total audiences for media types
  aTV=0.6; aRA=0.3; aNE=0.5; aIN=0.3;
#all media voices from the same category have the same audience size
#MV can be also manually reflecting exact audience sizes for each media voice
  MV=array(c(rep(aTV/tTV,tTV),rep(aRA/tRA,tRA),rep(aNE/tNE,tNE),rep(aIN/tIN,tIN)),dim=c(tM))
#media voice coverage/capacity (local events to cover)
  MVCtv=MVCra=MVCin=MVCne=8;
#create media voices and corresponding news vectors for each media voice
  MM=array(0,dim=c(tM,vN))
  z1=1;z2=tTV; MM[z1:z2,1:vN]=rbinom((z2-z1+1)*vN,1,(MVCtv/vN));
  z1=z2+1;z2=z1+tRA-1; MM[z1:z2,1:vN]=rbinom((z2-z1+1)*vN,1,(MVCra/vN));
  z1=z2+1;z2=z1+tNE-1; MM[z1:z2,1:vN]=rbinom((z2-z1+1)*vN,1,(MVCne/vN));
  z1=z2+1;z2=z1+tIN-1; MM[z1:z2,1:vN]=rbinom((z2-z1+1)*vN,1,(MVCin/vN));
#special cases, e.g., dissemination of a story from a single media voice
#below: media voice 1 is only producer of story 1
#MM[,1]=0;MM[1,1]=1;
#MM[,2]=0;MM[6,2]=1;
#MM[,3]=0;MM[16,3]=1;
#MM[,4]=0;MM[19,4]=1;
#create social network, small world network, lattice
  g=simplify(watts.strogatz.game(1,N,nei/2,rp))
#g=graph.lattice(length=as.integer(N^0.5), dim=2)
#edgelist population matrix
  ed=get.edgelist(g)
#assign initial transmission probability
  PT=runif(N,ptL,ptU)
#media coverage population matrix
  pop=array(0,dim=c(N,tM))
#initial distribution assignment
```

```
  for (m in 1:tM) {pop[1:N,m]=rbinom(N,1,MV[m])}
#news awareness population matrix and initial awareness
  news=array(1,dim=c(N,vN))
  news=pop%*%MM;news[news>1]=1
#total number of edges
  TE=length(ed)/2
#start single run lasting T periods
  for (t in 1:T) {
#applying salience discounting to PT
  PT=PT*df;
#examine each edge starts below
  for (j in 1:TE) {
#is there a news story transmission? if 1, then yes
  x1=rbinom(1,1,PT[ed[j,1]+1])
  x2=rbinom(1,1,PT[ed[j,2]+1])
#z is 1 if individual has access to at least one story
  z1=1;z1[sum(news[(ed[j,1]+1),])==0]=0
  z2=1;z2[sum(news[(ed[j,2]+1),])==0]=0
  if (x1==1 & z1==1) {
#find a story which an individual knows
  story=sample(which(news[(ed[j,1]+1),]==1),1)
#transmit story... no effect if the target already knows the story
  news[(ed[j,2]+1),story]=pmax((news[(ed[j,2]+1),story]),(news[(ed[j,1]+1),story]))
  }
#repeat in the opposite direction
  if (x2==1 & z2==1) {
  story=sample(which(news[(ed[j,2]+1),]==1),1)
  news[(ed[j,1]+1),story]=pmax((news[(ed[j,1]+1),story]),(news[(ed[j,2]+1),story]))
  }
  }
#record news exposure for time t
  data[t,1]=sum(news[1:N,1:vN])/(N*vN)
#single news story dissemination
#data[t,2]=sum(news[1:N,1])/N
#data[t,3]=sum(news[1:N,2])/N
#data[t,4]=sum(news[1:N,3])/N
#data[t,5]=sum(news[1:N,4])/N
  }
plot(data[,1], ylim=c(0.0,1),type="l",xlab="time periods",ylab="news exposure")
#lines(data[,2], ylim=c(0.0,1),col="red",type="l",xlab="time periods",ylab="single story exposure")
#lines(data[,3], ylim=c(0.0,1),col="blue",type="l",xlab="time periods",ylab="single story exposure")
#lines(data[,4], ylim=c(0.0,1),col="green",type="l",xlab="time periods",ylab="single story exposure")
#lines(data[,5], ylim=c(0.0,1),col="orange",type="l",xlab="time periods",ylab="single story exposure")
#save data
  write.csv(data, file="ABMMM.csv");
#done;
```