PROPAGATION, CASCADES, AND AGREEMENT DYNAMICS IN COMPLEX COMMUNICATION AND SOCIAL NETWORKS

By

Qiming Lu

A Thesis Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: PHYSICS

Approved by the
Examining Committee:

György Korniss, Thesis Adviser

Boleslaw K. Szymanski, Thesis Adviser

Saroj Nayak, Member

John Cummings, Member

Chjan C. Lim, Member

Rensselaer Polytechnic Institute
Troy, New York

November 2009
(For Graduation December 2009)

CONTENTS

LIST OF FIGURES

ACKNOWLEDGMENT

Foremost, I would like to express my sincere gratitude to my advisor Prof. György Korniss for the continuous encouragement, guidance and support of my Ph.D study and research. He offered me the possibility to enter and work in the fascinating field of statistical physics and complex systems. I am also heartily thankful to Prof. Boleslaw K. Szymanski. His perpetual energy and enthusiasm in research had motivated all his advisees, including me. Working with them has been a pleasure and an honor for me.

I would like to thank the rest of my thesis committee, Prof. Nayak, Prof. Lu, Prof. Cummings, and Prof. Lim for their encouragement and insightful comments. I thank Dr. Charles Reichhardt, Dr. Cynthia Reichhardt, and Dr. Christof Teuscher from Los Alamos National Laboratory for their mentorship when I was an intern in Center for Nonlinear Studies (Summer 2006) and Information Science Group (Summer 2008). The associated experiences broadened my perspective on many aspects. I would like to thank my fellow group members Hasan Guclu, Balazs Kozma, Lauren O'Malley, and Rui Huang. They and all my friends here at Rensselaer Polytechnic Institute made it a convivial place to work.

My deepest gratitude goes to my family for their unflagging love and support throughout my life; this dissertation is simply impossible without them. I am grateful to my dad for giving me the life I ever dreamed. I cannot express my gratitude for my mom in words, whose unconditional love has been my greatest strength. Finally and wholeheartedly, I would like to thank my wife Yu Jin for her constant love, encouragement, and support for all these years.

I thank Y.-Y. Ahn, E. Anshelevich, and A.-L. Barabási for enlightening discussions we have had. I also thank Prof. Barabási for sharing his mobile phone data set for my research. I acknowledge the financial support of the National Science Foundation Grant No. DMR-0426488, Office of Naval Research Grants Nos. N00014-06-1-0466 and N00014-09-1-0607, and Defense Threat Reduction Agency Award No. HDTRA1-09-1-0049. I also acknowledge the National Nuclear Security

Administration of the U.S. Department of Energy at Los Alamos National Laboratory Grant No. DE-AC52-06NA25396, and U.S. Department of Energy through the LANL/LDRD Program for the auspices of my internship in Los Alamos National Laboratory in the Summer of 2006 and 2008.

ABSTRACT

Many modern and important technological, social, information and infrastructure systems can be viewed as complex systems with a large number of interacting components. Models of complex networks and dynamical interactions, as well as their applications are of fundamental interests in many aspects. Here, several stylized models of multiplex propagation and opinion dynamics are investigated on complex and empirical social networks.

We first investigate cascade dynamics in threshold-controlled (multiplex) propagation on random geometric networks. We find that such local dynamics can serve as an efficient, robust, and reliable prototypical activation protocol in sensor networks in responding to various alarm scenarios. We also consider the same dynamics on a modified network by adding a few long-range communication links, resulting in a small-world network. We find that such construction can further enhance and optimize the speed of the network's response, while keeping energy consumption at a manageable level.

We also investigate a prototypical agent-based model, the Naming Game, on two-dimensional random geometric networks. The Naming Game [A. Baronchelli et al., J. Stat. Mech.: Theory Exp. (2006) P06014.] is a minimal model, employing local communications that captures the emergence of shared communication schemes (languages) in a population of autonomous semiotic agents. Implementing the Naming Games with local broadcasts on random geometric graphs, serves as a model for agreement dynamics in large-scale, autonomously operating wireless sensor networks. Further, it captures essential features of the scaling properties of the agreement process for spatially-embedded autonomous agents. Among the relevant observables capturing the temporal properties of the agreement process, we investigate the cluster-size distribution and the distribution of the agreement times, both exhibiting dynamic scaling. We also present results for the case when a small density of long-range communication links are added on top of the random geometric graph, resulting in a "small-world"-like network and yielding a significantly reduced

time to reach global agreement. We construct a finite-size scaling analysis for the agreement times in this case.

When applying the model of Naming Game on empirical social networks, this stylized agent-based model captures essential features of agreement dynamics in a network of autonomous agents, corresponding to the development of shared classification schemes in a network of artificial agents or opinion spreading and social dynamics in social networks. Our study focuses on the impact that communities in the underlying social graphs have on the outcome of the agreement process. We find that networks with strong community structure hinder the system from reaching global agreement; the evolution of the Naming Game in these networks maintains clusters of coexisting opinions indefinitely. Further, we investigate agent-based network strategies to facilitate convergence to global consensus.

CHAPTER 1
Introduction

A *network* is a collection of nodes, which we call *vertices*, and a collection of *edges* that connect pairs of vertices. Networks are also called *graphs* in mathematics and computer science. A graph maybe *undirected*, meaning that there is no distinction between the two vertices associated with each edge, or its edges may be *directed* from one vertex to another.

Systems that have the form of networks or graphs are abundant in the world. Examples range from the Internet (computers as vertices and physical network connections as edges), the World Wide Web (pages as vertices and hyper-links as edges), power grids (power plants and transmission lines), to social networks of acquaintance or other connections between individuals, networks of business relations between companies, even metabolic networks, neural networks, and chemical reaction networks, etc. Meanwhile, *network science*, the study of networks has emerged in diverse disciplines as a means of analyzing complex relational data. It is a new and emerging scientific discipline that examines the interconnections among diverse physical or engineered networks, information networks, biological networks, cognitive and semantic networks, and social networks. This field of science seeks to discover common principles, algorithms and tools that govern network behavior.

Leonhard Euler's solution to the famous Seven Bridges of Königsberg problem in 1736 is the earliest known paper in the theory of networks [115]. His mathematical description of vertices and edges was the foundation of graph theory, a branch of mathematics that studies the properties of pairwise relations in a network structure. In the 1930s, Jacob Moreno developed a sociogram representing the social structure of a group of elementary school students [114]. This network representation of social structure has found many applications and has grown into the field of social network analysis, the branch of sociology that deals with the quantitative evaluation of an individual's role in a group or community by analysis of the network of connections between them and others.

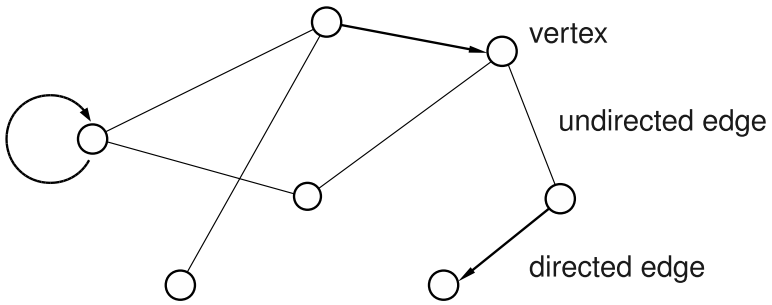


Figure 1.1: A small example network made up of seven vertices and eight edges (including five undirected edges and three directed edges)

Probabilistic theory in network science developed as an offshoot of graph theory with Paul Erdős and Alfréd Rényi's famous papers on random graphs [53, 23]. For social networks the exponential random graph model or p* graph is a notational framework used to represent the probability space of a tie occurring in a social network [128]. An alternate approach to network probability structures is the network probability matrix, which models the probability of edges occurring in a network, based on the historic presence or absence of the edge in a sample of networks [110].

More recently other network science efforts have focused on mathematically describing different network topologies. Duncan Watts reconciled empirical data on networks with mathematical representation, describing the small-world network [154]. Albert-László Barabási and Réka Albert developed the concept of scale-free networks [**?**, 11] which is a network topology that contains hub vertices with many connections, that grow in a way to maintain a constant ratio in the number of the connections versus all other nodes. Although many networks, such as the internet, appear to maintain this aspect, other networks have long tailed distributions of nodes that only approximate scale free ratios.

1.1 Networks in the Real World

Empirical network data are usually the starting point for network analysis. Recent work on mathematics of networks has been driven largely by observations of the properties of actual networks and attempts to model them. In the real world, networks may take various forms and many of them are more complex than a simple collection of vertices linked by edges. Based on originations and network properties, Mark Newman classifies real-world networks into four loose categories: social networks, information networks, technological networks and biological networks [115].

1.1.1 Social Networks

A social network is a social structure made of individuals or organizations. These individuals or organizations are tied or connected by one or more specific types of interdependency, such as friendship, kinship, financial exchange, dislike, sexual relationships, or relationships of beliefs, knowledge, or prestige [68]. Early works on the subject can be traced back in the 1920s and 30s by Jacob Moreno [114] on friendship patterns within small groups. Another important experiment is the famous "small-world" experiment of Milgram [112]. Instead of reconstructing the actual topology of acquaintance networks as people would usually do, the experiment probed the distribution of path lengths by asking participants to pass a letter through one of their acquaintances in an attempt to get to an assigned target individual. This experiment led to the popular concept of the "six degrees of separation", as the result indicates that the letters reached the target have passed on average through the hands of only about six people in doing so. In recent days, the contact or interaction patterns between individuals in social networks are not limited to traditional ways. For example, one could construct a network of email correspondences, instant messages, or telephone calls between individuals. And the number of vertices in such network may be significantly larger than before, which allows to have various statistical analysis on such networks.

1.1.2 Information Networks

Information networks, also sometimes called knowledge networks, are a category of network that stores the information at its vertices while the network structure reflects the relative structure of the information (hence the term "information networks"). The network of citations between academic papers, and the World Wide Web are two classic examples of the information network. In a citation network, the vertices are articles and a directed edge from article A to article B indicates that A cites B. While the World Wide Web is a network of Web pages containing information, linked together by hyperlinks from one page to another. Both networks reflect the infrastructure of how information is organized. Furthermore, the citation network is acyclic because papers can only cite other papers that have already been written, not those that have yet to be written [43, 115]. Therefore are no loops possible in a citation network. However, the World Wide Web dose not have such natural constraints to prevent the appearance of closed loops, thus it is cyclic. By studying the publication citation pattern quantitatively, Alfred Lotka made the important discovery in 1926 of the so-called Law of Scientific Productivity, which states that the distribution of the numbers of papers written by individual scientists follows a power-law [100]. That is, the number of scientists who have written k papers falls off as $k^{-\alpha}$ for some constant α. Price [43] further pointed out that both the in and out-degree distributions of the network follow power-laws. Interestingly, the World Wide Web also appears to have power-law in and out-degree distributions, as well as a variety of other interesting properties [5, 1].

1.1.3 Technological Networks

Technological networks are usually man-made networks designed typically for distribution of some commodity or resources, such as electricity or information [115]. Examples of technological networks include the electric power grid, which is a network of high-voltage transmission lines that transmit electricity across the country, the network of airline routes, the networks of roads, railways, and pedestrian traffic, sensor networks, etc. Another very widely studied technological network is the Internet. Note that here the Internet means the network of physical connections

between computers. While the World Wide Web indicates the web pages stored on computers and the hyperlinks between web pages.

1.1.4 Biological Networks

Many biological systems or processes can be represented by networks. E.g., the network of protein-protein interactions which involve both the direct-contact association of protein molecules and longer range interactions through the electrolyte, aqueous solution medium surrounding neighbor hydrated proteins [136], the protein phosphorylation network, the metabolic interaction network, or the genetic regulatory network [161]. However, there are many other studied networks can be also categorized into biological networks, such as the food web, in which the vertices represent species in an ecosystem with a directed edge from species A to species B indicating that A preys on B [37]. Neural networks are another class of biological networks of considerable importance [156, 56].

1.2 Models for Complex Networks

To analyze networks of various forms, it is then of fundamental interests to model networks mathematically. Perhaps the simplest useful model of a network is the random graph, first studied by Erdős and Rényi [53]. However, most of the interesting features of real-world networks are not like random graphs. And sometimes these features are crucial in modeling and exploiting the network structures.

1.2.1 Network Features

In this section we describe some features or statistic measurements that appear to be common to networks of many different types: The small-world effect; transitivity or clustering; degree distribution; and community structure.

1.2.1.1 The Small-world Effect

In Stanley Milgram's famous experiment of letter passing in 1960s, randomly selected individuals in the U.S. cities of Omaha, Nebraska and Wichita, Kansas were asked to deliver letters to someone in Boston, Massachusetts through a chain

of correspondence. These cities were selected because they represented a great distance in the United States, both socially and geographically. However, many letters are delivered in only a small number of steps – around six in the published cases. This result is one of the first direct demonstrations of the *small-world effect*, the fact that most pairs of vertices in most networks seem to be connected by a short path through the network. A modern version of "six degrees of separation" experiment was conducted by Duncan Watts et. al in 2003 [49], in which more than $60,000$ e-mail users attempted to reach one of 18 target persons in 13 countries by forwarding messages to acquaintances. Same result as in Milgram's experiment was repeated that in successful deliveries an email reaches the target in a median of five to seven steps, depending on the separation of source and target [49]. In 2007 Jure Leskovec and Eric Horvitz examined a data set of instant messages composed of 30 billion conversations among 240 million people. This is by far the largest data set examined for this purpose. They found the average path length among Microsoft Messenger users to be 6.6 [94]. Similar experiments also appear in Facebook, one of world's largest social networks. Over 100 groups under the name "six degrees of separation" experiment on proving the statement bearing in their titles. So far the largest group has $5,563,429$ members as of October 2009. The average separation among these members is 5.73, whereas the maximum degree of separation is 12 [69].

The most direct measurement of the small-world effect is the average shortest path length l in a network:

$$l = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i<j} d_{ij}, \tag{1.1}$$

where N is the number of vertices in the network, and d_{ij} is the shortest path length from vertex i to vertex j. In any regular lattices or spatial graphs of dimension d, the number of vertices within a distance r of a typical central vertex grows with r in a power of d, thus the value of l as well increases as a power-law of $1/d$,

$$l \sim N^{1/d}. \tag{1.2}$$

If for networks (the random graph for example) the number of vertices grows expo-

nentially with r, then l will increase as $\log N$. In recent years the term "small-world effect" more precisely means that the value of l scales logarithmically or slower with network size [115].

The small-world effect has obvious implications for the dynamics of processes taking place on networks. For example, the small-world effect greatly speeds up the global information spread across the network than on pure spatial graphs [101], or gives better synchronization in problems in distributed computing [79].

1.2.1.2 Network Transitivity

Another fundamental measure that has long received attention in both theoretical and empirical network research is the *network transitivity*, or the *clustering coefficient*. This measure assesses the degree to which nodes tend to cluster together. Evidence suggests that in most real-world networks, and in particular social networks, nodes tend to create tightly knit groups characterized by a relatively high density of ties [66, 154]. In the language of social networks, the friend of your friend is likely also to be your friend. In real-world networks, this likelihood tends to be greater than the average probability of a tie randomly established between two nodes [66, 154]. The clustering coefficient C is the quantified network transitivity [115]:

$$C = \frac{3 \times \text{number of triangles}}{\text{number of connected triples of vertices}}, \tag{1.3}$$

where a "connected triple" means a single vertex with edges running to an unordered pair of others. In effect, C measures the fraction of triples that have their third edge filled in to complete the triangle. Watts and Strogatz [154] proposed an alternative way to define the clustering coefficient through a local value

$$C_i = \frac{2E_i}{k_i(k_i - 1)}, \tag{1.4}$$

where k_i is the degree of vertex i, and E_i is the number of edges in the subgraph consisting of vertex i and all his neighbors. Then the clustering coefficient for the

whole network is the average of all local values,

$$C = \frac{1}{N} \sum_i C_i. \tag{1.5}$$

A purely random graph built according to Erdős and Rényi model exhibit both a small average shortest path length and a small clustering coefficient. Watts and Strogatz measured that in fact many real-world networks have a small average shortest path length, but also a clustering coefficient significantly higher than expected by random chance [154].

1.2.1.3 Degree Distribution

In a network, the degree of a vertex is the number of edges connected to that vertex. We define p_k to be the fraction of vertices in the network that have degree k. A histogram of p_k for any given network is thus the *degree distribution* for the network. The degree distribution is very important in studying both real networks, such as the Internet, social networks, and theoretical networks. The simplest network model, for example the random graph, in which each possible pair of n vertex is connected with independent probability p, has a binomial distribution of degrees, which, in the limit of large N, can be well approximated by a Poisson distribution with an exponentially decaying tail. Many networks in the real world, however, have degree distributions very different from this. E.g., some of them are highly right-skewed, meaning that a large majority of nodes have low degree but a small number, known as "hubs", have high degree. Some networks, notably the Internet, the world wide web, and some social networks are found to have degree distributions that approximately follow a power law. Such networks are called scale-free networks and have attracted particular attention for their structural and dynamical properties.

1.2.1.4 Community Structure

Community structure, i.e., groups of vertices that have a high density of edges within them, with a lower density of edges between groups, is assumed to be prevailed in most social networks and some other types of networks [137]. It is a matter

of common experience that people tend to divide into groups along lines of interest, occupation, age, and so forth. Thus it is quite common to observe community structures in real networks, particularly in social networks.

In Fig. 1.2 we show a visualization of the friendship network of high-school students. By implementing a suitable model for opinion dynamics (we use the Naming Game here [104]) on the friendship network, the network exhibits four different communities at some stage. These communities correspond to segregation along the two-schools involved in the particular network, high-school (HS) – middle-school (MS) pair, and along racial lines, whites students – black students in each. Checking the race and school-grade attribute of the node information in the raw data [113], we find most of the nodes in respective communities are coincide with their attributions of, as shown in Fig. 1.2, black HS (green), white HS (yellow), black MS (purple), and white MS (red) students.

1.2.2 Prototypical Models for Complex Networks

We now look at some mathematical models for networks of various types, mainly focus on following models: the random network by Erdős and Rényi [53]; spatially embedded random geometric graphs [40]; the "small-world model" of Watts and Strogatz [154]; and models of growing networks, in particular, "scale-free networks" of Price [44] and Barabási and Albert [11].

1.2.2.1 Random Graphs

The model of "random graph" or as sometimes called "Poisson random graph", was probably the first attempt to construct and model a certain type of large scale random network by Paul Erdős and Alfréd Rényi. They proposed an extremely simple model of a network: for a network with N vertices, connect each pair of vertices (or not) with probability p (or $1-p$). Purely random graphs, built according to the Erdős - Rényi model, exhibit a small average shortest path length (varying typically as the logarithm of the number of nodes $l \sim \log N$) along with a small clustering coefficient [4, 115].

Given the mathematic model, most properties of the random graph can be solved exactly. For example, the mean degree is $\bar{k} = (N-1)p$, and the degree k_i of

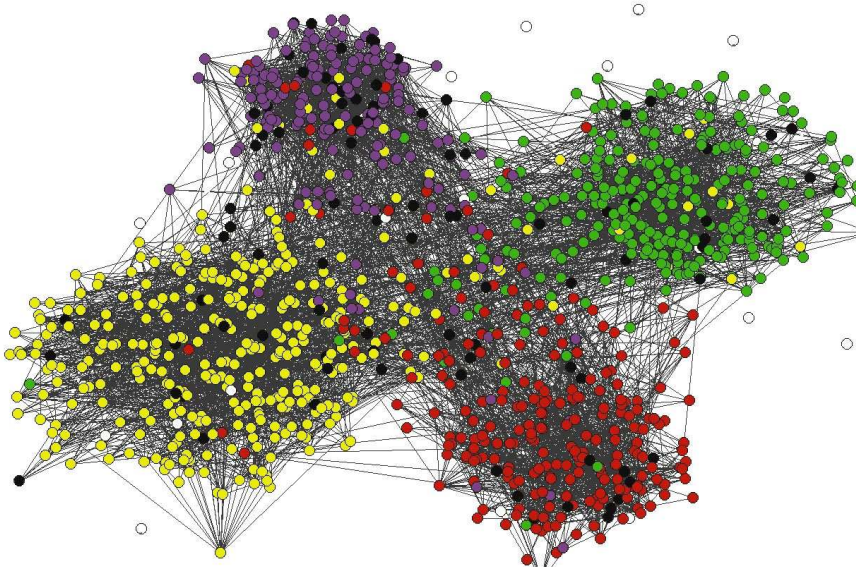


Figure 1.2: Community structure in a U.S. high-school friendship network. The network is constructed from in-school questionnaires by Ad-Health [113]. Vertices are colored according to races and grades of students. Yellows are white students in high-school, greens are black students in high-school, reds are white students in middle-school, and purples are black students in middle-school. The community separation is achieved with the naming game simulation.

a node i follows a binomial distribution with parameters $N-1$ and p:

$$P(k_i = k) = \begin{pmatrix} N-1 \\ k \end{pmatrix} p^k (1-p)^{N-1-k}. \tag{1.6}$$

This probability represents the number of ways in which k edges can be drawn from a certain node: the probability of k edges is p^k, the probability of the absence of additional edges is $(1-p)^{N-1-k}$, and there are C_{N-1}^k equivalent ways of selecting the k end points for these edges. Further more, in the limit of large network size, with a good approximation the binomial distribution can be replaced by a Poisson distribution [4],

$$P(k) \simeq e^{-pN}\frac{(pN)^k}{k!} = e^{-\bar{k}}\frac{\bar{k}^k}{k!}, \tag{1.7}$$

where the average degree $\bar{k} = (N-1)p \simeq Np$ for large N. The Poisson distribution decays rapidly for large values of k, the standard deviation of the distribution being $\sigma_k = \sqrt{k}$.

The network structure of random graphs is largely affected by the connection probability p. For low-p values, there only exists few edges and the graph is separated into small isolated component, i.e., subsets of vertices that are connected by paths through the network. For high-p values, these small components quickly merge into a giant connecting component whose size is comparable with the graph size. For random graphs, this is a typical phase transition from isolated components to a giant connecting component as the probability of connection p increases [53]. The criticality of p happens at

$$p_c \sim \frac{1}{N}. \tag{1.8}$$

1.2.2.2 Random Geometric Graphs

The model of random graphs can be extended in a variety of ways to make them more realistic. In many cases vertices of a network may have geographical information attached to them, such as randomly deployed sensor nodes, or geographical living spaces for individuals in a social network. Here we look at another network model called *random geometric graph* [125, 111, 40], also referred to as Poisson Boolean graphs, which carries both the random connectivity and spatial property of vertices. A random geometric graph is a network in which each vertex is assigned uniformly and independently at random coordinates in a geometric space of arbitrary dimensionality and two vertices are connected if and only if the distance between them is at most a threshold r. Random geometric graphs feature with a large average shortest path length (scales with the system size in a power-law of $l \sim N^{1/d}$, where d is the dimension of a random geometric graph), and a large clustering coefficient [125, 40].

The average degree α of a random geometric graph is the (average) number of

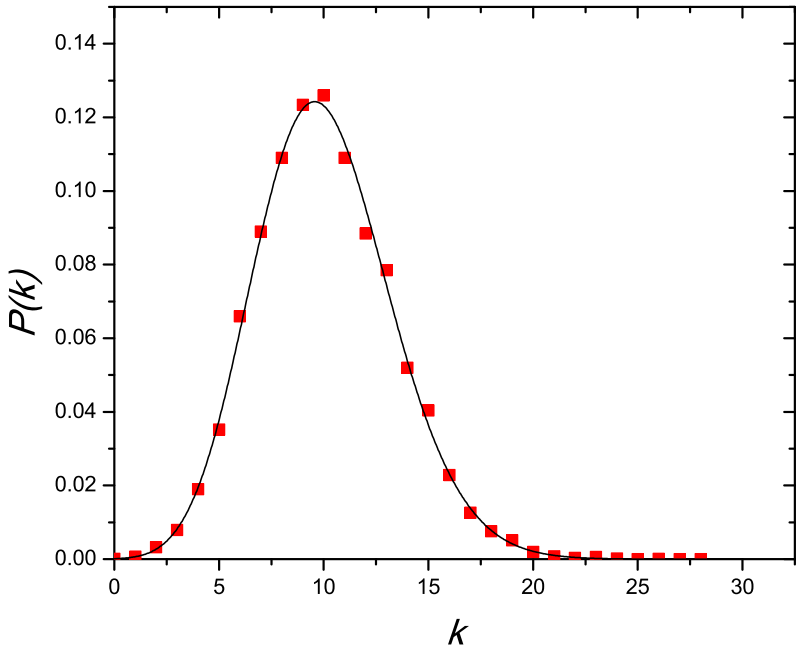


Figure 1.3: Degree distribution of a random geometric network. The network has $N = 10,000$ nodes with average degree $\bar{k} = 10$. The solid line is a Poisson distribution.

vertices inside a d dimensional sphere with radius r,

$$\alpha = \frac{NC_d r^d}{V} = \rho C_d r^d, \tag{1.9}$$

where $\rho = N/V$ is the number of vertices in a unit volume of the geometric space, and the constant proportionality is given by

$$C_d = \frac{\pi^{d/2}}{\Gamma(\frac{d}{2}+1)}. \tag{1.10}$$

In a $2d$ random geometric graph the average degree depends on r with a simple relationship

$$\alpha_{2d} = \rho\pi r^2. \tag{1.11}$$

Random geometric graphs also have the binomial degree distribution, which

approaches Poisson distribution for large N [Fig. 1.3] [40]:

$$p_k = \begin{pmatrix} N-1 \\ k \end{pmatrix} p^k (1-p)^{N-1-k} \simeq \frac{\alpha^k e^{-\alpha}}{k!}. \tag{1.12}$$

The phase transition from small separated clusters to a single giant component is also observed in random geometric graphs at percolation threshold [40]. In $2d$ random geometric graphs, the critical connectivity is about $\alpha_c \simeq 4.52$. Below α_c the network is mostly disconnected and average size of the components is considerably small comparing with the size of the network. Above the α_c the giant component emerges in the network covering majority of the nodes.

1.2.2.3 Small-world Networks

When measured by the average path length and transitivity, random graphs and random geometric graphs sit on the opposite side of each other: random graphs have small average path length as well as small transitivity; on the other side, random geometric graphs are usually able to achieve high transitivity accompanied with very large average path length. Here, another classic network model is the "small-world model" proposed by Watts and Strogatz in 1998 [154]. The model starts from a low-dimensional lattice – it thus conserves the spatial properties of vertices with high transitivity derived from lattices. And by proper manipulation (rewiring) of existing links the network is also able to achieve a fairly low path length.

Small-world models can be built on lattices of any dimension or topology, but the best studied case by far is one-dimensional one. If we take a one-dimensional lattice of L vertices with periodic boundary conditions, i.e., a ring, and connect each vertex to its nearest k neighbors, we get a system shown in Fig. 1.4(a), with Lk edges. The small-world model is then created by taking a small fraction of the edges in this graph and "rewiring" them. The rewiring procedure involves going through each edge in turn and, with probability p, moving one end of that edge to a new location chosen uniformly at random from the lattice, except that no double edges or self-edges are ever created. This process is illustrated in Fig. 1.4(b). Fig. 1.5 shows the

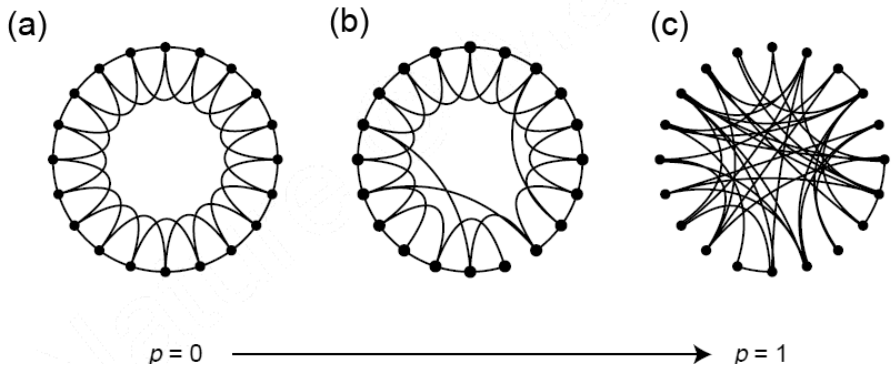


Figure 1.4: (a) A one-dimensional lattice with periodic boundary condition. (b) The small-world model is created by choosing at random a fraction p of the edges in the graph and rewire them to new ending vertices, also chosen uniformly at random. (c) A random graph is constructed with the rewiring probability $p = 1$. From Watts et al., 1998 [154].

degree distribution of a numerically constructed random small-world network with average degree $\bar{k} = 10$.

As we stated above, the rewiring process allows the small-world model to interpolate between a regular lattice and a random graph. When $p = 0$, we have a regular lattice. The clustering coefficient of this regular lattice is $C = (3k - 3)/(4k - 2)$. Such regular lattice does not show the small-world effect because the mean path length between vertices is around $L/4k$ for large L. When $p = 1$, every edge is rewired to a new random location and the graph is now a random graph, with typical path length on the order of $\log L/\log k$, and very low clustering coefficient $C \simeq 2k/L$ [115]. As Watts and Strogatz showed by numerical simulation, there exists a sizable region in between these two extremes for which the model has both low path lengths and high transitivity [153].

1.2.2.4 Models of Network Growth

All of the models discussed above reflect some properties observed in real-networks, such as degree distribution or transitivity, and try to incorporate these

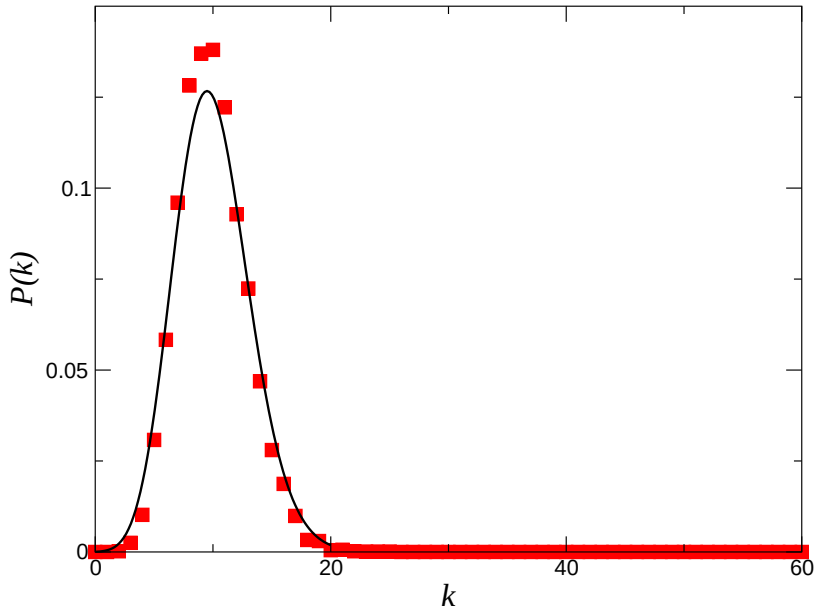


Figure 1.5: The degree distribution of a random small-world network. The network is constructed by adding $\frac{1}{2}pN$ random links with $p = 8$ on top of a $1d$ ring. The resulting network has average degree $\bar{k} = 10$. The solid line is a Poisson distribution. From G. Korniss [78]

properties when create new networks with the model. However they do not reveal how networks come to have these properties. Thus, it is worth to look at another category of network model whose primary goal is to explain network properties. In these models, the networks typically grow by the gradual addition of vertices and edges in some manner intended to reflect growth processes that might be taking place on the real networks, and it is these growth processes that lead to the characteristic structural features of the network. Here we introduce models of network growth aimed at explaining the origin of the highly skewed degree distributions, by Price [44], and Barabási and Albert [11].

In Price's earlier paper in 1965 [43], he studied the the network of citations between scientific papers and found that both in and out-degrees have power-law distribution. He later published another paper to propose a model for the explanation of the power-law degree distribution [44]. His work was intrigued by the idea of "the rich get richer", which was originally discovered in the problems of wealth dis-

tribution, and he later applied them to the growth of a network. This is an analogy of cumulative advantage in the context of network growth, that the rate at which a paper gets new citations should be proportional to the number that it already has.

Consider a directed graph of n vertices, such as a citation network. Let p_k be the fraction of vertices in the network with in-degree k (number of times a paper has been cited). At the time of creation, new vertices that are to be added to the network have a certain out-degree (citations to other papers). The out-degree may vary from one vertex to another, but the mean out-degree, which is denoted m, is a constant over time. When a new created vertex is added to the network, the probability that a new edge attaches to any of the existing vertices with degree k is thus:

$$\frac{kp_k}{\sum_k kp_k} = \frac{kp_k}{m}. \tag{1.13}$$

In the limit of large n, the degree distribution can be solved exactly [44]:

$$p_k \sim k^{-(2+1/m)}. \tag{1.14}$$

It has a power-law tail with exponent $\alpha = 2 + 1/m$, which typically gives exponents in the interval between 2 and 3.

The mechanism of cumulative advantage proposed by Price is now widely accepted as the probable explanation for the power-law degree distribution observed not only in citation networks but in a wide variety of other networks also, including the World Wide Web, collaboration networks, etc. The phenomena was later rediscovered by Barabási and Albert [11] in the context of complex networks, called *preferential attachment*, to explain the network properties of the World Wide Web. The network growth model of the Web proposed by Barabási and Albert has one important difference, that edges of the network are undirected, so there is no distinction between in- and out-degree. This model can also be solved exactly and the analytical solution gives a power-law degree distribution of $p \sim k^{-3}$, with only the single fixed exponent $\alpha = 3$ [Fig. 1.6], as well as a logarithmically increasing average shortest path length with the system size, $l \sim \log N$ [4].

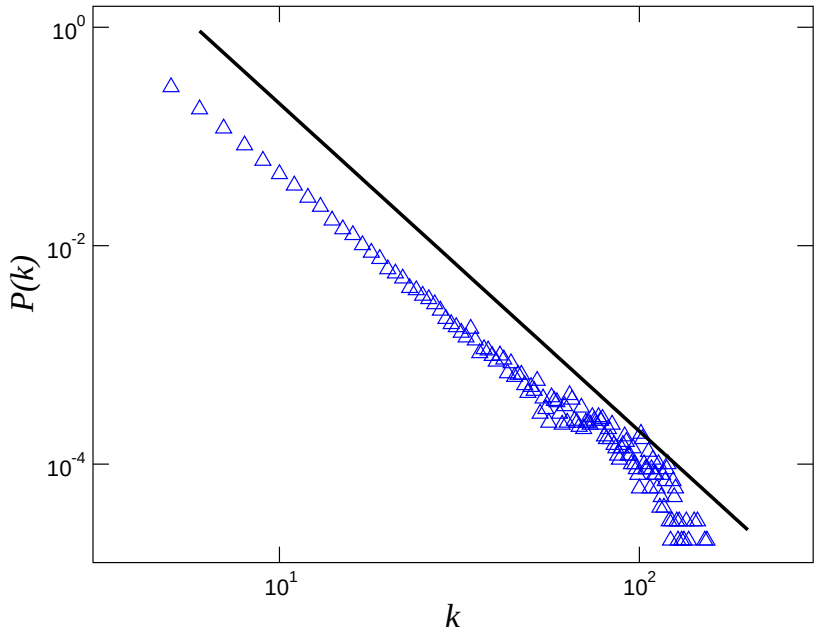


Figure 1.6: The degree distribution of the scale-free Barabási-Albert network with $m = 5$, yielding an average degree of $\bar{k} = 10$. The solid line shows the power-law behavior in log-log scale with the power $\alpha = -3$. From G. Korniss [78].

1.3 Network Dynamics

Many modern and important technological, information and infrastructure systems can be viewed as complex systems with a large number of interacting components, whose aggregate activity is usually non-linear and typically exhibits self-organization under selective pressures. In many of such systems, complex networks serve as the underlying topology to support the dynamics of interactions among system components (nodes) through network links. Examples include the package flow over the Internet, electric power grid, sensing networks, traffic flow on road network, and a large category of social dynamics and interactions. These systems and dynamics are usually autonomous, i.e., they lack a central regulator. Through local interactions they still exhibit collective behavior at some large scales, e.g., from disorder to order, achieving agreement or consensus without global regularities, etc. Here, we briefly review some well-known and interesting network dynamics models, such as information spreading models, random boolean networks, and opinion

dynamics in social interactions.

1.3.1 Simple and Multiplex Propagation Dynamics

A large number of system dynamics and social phenomena can be described by simple contagion processes. The most obvious examples include the epidemic spreading and information propagation. Even if we look into the deep reasonings of, e.g., how does a small initial shock cascade to affect a large system, and how do large, grassroots social movements, fads, riots start in the absence of centralized control or public communication, we are still able to find the simple contagion process, or so called *information cascades*, under the complex interactions. These processes share a common feature: for individuals in a population, either their status, or how they make decisions based on the status or the actions of other individuals rather than relying on their own information about the problem [155]. In case of epidemic spreading it is inevitable that certain disease may be transmitted through close contact regardless of the status of individuals. And in social systems, decision makers often pay attention to each other usually because they have limited information about the problem, or limited ability to process. For example, when deciding which restaurant to visit we usually rely on the recommendation of our friends if we have limited information to evaluate the alternatives.

For simplicity, we only consider binary decisions regarding the problem, which is suitable in many real cases such as infected or not, either decide to visit (the restaurant) or not, etc. If a single active node is sufficient to trigger the activation of its neighbors with however small probabilities, such as the spread of information or disease, it is called *simple propagation* [34]. In addition to simple propagation, *multiplex propagation*, in which node activation requires simultaneous exposure to multiple active neighbors, is also common in the real-world situation. One particular model for multiplex propagation is the *threshold model* for contagion [155], in which a node becomes active only if the fraction of its neighbors in the active state is equal to or larger than a pre-selected threshold φ, and it will remain active afterwards. This simple threshold-based binary-decision model has its application in various fields, such as decision making in social activities or detecting outliers in a sensor

network. We investigate more details about the model in chapter 2.

If a small perturbation can cascade to affect a large system, it is then of particular interests to study the robustness of technological networks against certain perturbations of failures, e.g., the spreading of a computer virus of the internet [38], failure propagation in power girds [131], or the perturbation of gene expression patterns in a cell due to mutations [126]. One possible way is to use random Boolean networks (RBNs) as a dynamics model to study the response of complex dynamical networks to external perturbations, also referred to as damage [129].

RBNs were originally introduced by Kauffman as simplified models of gene regulation networks [74, 75]. In its simplest form, RBN is a discrete dynamical system composed of N automata (nodes). And each automata can make decisions between two possible states (similar to binary decisions in threshold model) based on inputs of K randomly chosen other nodes according to associated (random) Boolean functions. Comparing with the simple or multiplex propagation, RBN model is able to achieve more complicated controls on decision making rules for individuals in the system. And this makes RBN a more suitable model in studying the robustness for technological networks with complex decision making rules, such as the Internet at the router level, and road traffic networks.

1.3.2 Social Dynamics

Social systems and related social interactions are another wide collection of topics about dynamics over complex systems. Common social dynamics include the dynamics of opinions (opinion formation and agreement dynamics), language dynamics (formation and evolution of a language, and the competition between different languages), cultural dissemination, etc. In most social systems the elementary components of the system are usually humans, which is probably the most complicated thing in the world, as it is nearly impossible to precisely predict the behavior of human individuals. So there usually has two folds of difficulties in modeling social dynamics, first is in the definition of relative simple while realistic microscopic models; the second is the problem of inferring the macroscopic phenomenology out of the microscopic dynamics of such models and obtaining useful results. In this

respect, the concept of *universality* brings a possible solution to this problem. In statistical physics, universality is the observation that some properties of large-scale phenomena are independent of the dynamical details of the system. Systems display universality in a scaling limit, when a large number of interacting parts come together. It is thus possible to model the social system by including only the simplest and most important properties of single individuals and looking for qualitative features exhibited by models. From early adoptions of the Ising model for ferromagnets [138], in which a collection of N spin particles are energetically pushed to be aligned with their nearest neighbors, to more general q-state Potts model [106, 138, 139], by far there are many stylized models for social dynamics, e.g., the voter model [96, 140, 71, 72], cultural dissemination model [10], and bounded confidence models [46] for opinion dynamics; the evolutionary language game [107, 120] and the Naming Game [142, 15, 14, 103] for language dynamics.

The voter model assigns to each agent a discrete variable, i.e. an opinion, that can assume two values. The dynamics then evolves following the rule that at each time step the opinion of a randomly chosen agent is made equal to the one of its neighbors (selected at random on its turn). The process ends up with a complete order on low dimensional lattices but different blocked configurations survive in more complex topologies [140, 72].

The Axelrod model [10], describes agents with a rich repertoire of opinions. More precisely, each individual is endowed with F cultural traits ($\sigma_{i,f}, f = 0, .., F-1$), each of which can assume one out of q values. The agents are embedded on a square lattice and at each time step two of them are randomly selected along with a cultural trait f. If $\sigma_{i,f} \neq \sigma_{j,f} (\forall f)$ nothing happens. Otherwise another cultural trait f' is randomly chosen and its value is set equal for the two agents, i.e. $\sigma_{j,f'} \rightarrow \sigma_{j,f'} = \sigma_{i,f'}$. Absorbing states are those in which, across each bond, all features are equal or different, and final configurations can be described in terms of distributions of different clusters of homogeneous agents. In this respect, it is worth mentioning that a non-equilibrium phase transition from a culturally polarized phase (all agents belong to the same cluster) and culturally fragmented one (finite size, $\mathcal{O}(1)$, clusters) has been shown to occur as q grows [30, 77].

The Hegselmann and Krause model [62] describes the opinion formation of an individual as an averaging process on the opinions of all its neighbors, provided they fall inside a given confidence bound ϵ. This is the fundamental parameter of the model, since three different final states can be reached depending on its value. In particular, starting from a uniform distributions of opinions, that here are real numbers between 0 and 1, the final state consists of a plurality of stable opinions (small ϵ), a consensus on a single opinion (large ϵ), or, interestingly, a polarization of the population on two or three different opinions for intermediate values of the confidence bound.

In the model proposed by Deffuant et al. [46], agents perform pairwise interaction. Opinions are real numbers, and when two agents i and j meet, they make their opinion O_i and O_j closer by a certain amount $\mu|O_i - O_j|$, with $\mu \leq 0.5$, provided that they find each other in their confidence bound, i.e. $|O_i - O_j| < \epsilon$. Also in this case the fundamental parameter is ϵ, and its magnitude determines the number of opinions that survive in the final state. The parameter μ, on the other hand, only influences the convergence time.

The Naming Game model was expressly conceived to explore the role of self-organization in the evolution of language [16, 41]. In the original paper [142], Luc Steels focused mainly on the formation of vocabularies, i.e. a set of mappings between words and meanings (for instance physical objects). In this context, each agent develops its own vocabulary in a random private fashion. But agents are forced to align their vocabularies in order to obtain the benefit of cooperating through communication. Thus, a globally shared vocabulary emerges, or should emerge, as a result of local adjustments of individual word-meaning association.

The work outlined in this thesis analyzes several stylized models of multiplex propagation and opinion dynamics on complex and empirical social networks.

We first investigate the threshold-based cascade dynamics on wireless sensor networks. The purpose is to effectively detect outliers in sensing networks while suppressing false alarms. It is also interesting to look at the impact of network topology including the small-world effect to the information cascading, energy con-

suming, and other network properties.

The underlying network structure for wireless sensor networks can be interpolated as random geometric graphs (RGGs), which are both random and spatial. Thus it is worth to look at some other complex dynamics on this particular network. We introduce the simplified agreement dynamics model on both pure and small-world embedded random geometric networks. Detailed scaling behavior of the Naming Game has been carefully examined.

The Naming Game (NG) was originally proposed as a social dynamics model to study the evolution of human languages. Later it is found that many other phenomena such as the collaborative tagging, autonomous development of shared lexicon, etc., can also be explained with the NG model. It is then of particular interest to look at the behavior of NG on empirical social networks. In the simulation, long lasting meta-stable state is found to indicate the existence of strong community structure in empirical social networks.

CHAPTER 2
Threshold Based Cascading in Random Geometric Graphs

Many of the modern and important technological, information and infrastructure systems can be viewed as complex networks with a large number of components [4, 50, 115]. The network consists of nodes (or agents) and (physical or logical) links connecting the nodes. These links facilitate some form of interaction or dynamics between the nodes. Spreading information fast across such networks with efficient and autonomous control is a challenging task. Wireless Sensor Networks (WSNs) provide an example where understanding dynamical processes *on* the network is crucial to develop efficient protocols for autonomous operation.

A sensor network is comprised of a large number of sensor nodes which monitor, sense, and collect data from a target domain and then process and transmit the information to the specific sites (e.g., headquarters, disaster control centers). There are many potential applications of sensor networks including military, environment and health areas (for a taxonomy of sensors networks see [63]). There are fundamental differences between a sensor network and other wireless ad-hoc networks. First, sensor nodes are often densely deployed (typically 20 sensor per cubic meter) [3] so that the underlying network has high redundancy for sensing and communications. Accordingly, the size of sensor networks may be several orders of magnitude larger than the other ad-hoc networks. Hence, scalability of sensor network operations is of utmost importance (see, for example [54] for scalable coordination challenges and solutions or [148] scalable, self-organizing designs for sensor networks and for limits on achievable capacity and delay in mobile wireless networks). Second, sensor nodes have limited battery power without recharging capabilities. Nodes running out of power may cause topology changes in sensor networks even without mobility (see for example [36] for scalable and fault-tolerant routing and [**?**] for communal routing in which some of the nodes take over routing for the sleeping neighbors). Third, new sensors with fresh batteries may be injected to a sensor network, already in use, to enhance and ensure its correct operation. Finally, the sensor nodes may be deployed

in adversarial environments such as battlefields, hostile territories or hazardous domains that make their management, control and security very difficult. Combined with diverse environments, ranging from deserts to rain forests, from urban areas to battlefields and habitats of protected species [2], these challenges make designing sensor networks that can operate reliably and autonomously (totally unattended) very difficult.

Our focus is on outliers detection in wireless sensor network but the challenge is the same as in the above mentioned work, to design scalable, energy efficient algorithms for communication and coordination. Outlier detection is an essential step which precedes most any analysis of data. It is used either with the intention of suppressing the outliers or amplifying them. The first usage (also known as data cleansing) is important when the analysis carried on the data is not robust. Examples for such applications are optimization tasks, including routing (where erroneous data may lead to infinite loops). The second usage is important when looking for rare patterns. This often happens in adversarial domains such as battlefield monitoring, controlling a boundary or a perimeter of protected objects or intrusion detection.

Outliers are caused not only by external factors, but also by imperfections in the acquisition of the data. They typify error prone systems, specifically those which ought to operate in harsh environmental conditions and make imperfect measurements of external phenomena. Another setting in which outliers may occur is whenever an adversary can control the measurement (but not the computation and communication) of a device. In this setting, outliers detection can either detect the manipulation of the data, or limit the extent to which the data is manipulated. Thus, in some settings, outliers detection limits the ability of an adversary to divert the result.

Several factors make WSNs especially prone to difficulties in outliers detection. First, WSNs collect their data from the real world using imperfect sensing devices. Next, they are battery operated and thus their performance tend to deteriorate as power is exhausted. Moreover, as sensor networks may include thousands of devices, the chance of error accumulates in them to high levels. Finally, sensors are especially exposed to manipulation by adversaries in their usage for security and

military purposes. Hence, it is clear that outlier detection should be an inseparable part of any data processing that takes place in sensor networks. In this part of work, we investigate a simple model of cleansing and amplifying outliers in wireless sensor networks. WSN environments pose several restrictions on outlier computation, such as: (i) it has to be done in-network because communication of raw data would deplete batteries, (ii) communication may often be asymmetric, (iii) the data is streaming, or at least dynamically updated, and (iv) both spatial and temporal locality of the data are important for the result – data points sampled by nearby sensors during a short period of time ought to be more similar than ones sampled by far off sensors over a large time interval. Hence, in this paper, we assume that the number of nodes reporting outliers is significant in making a decision to amplify or not the outlier discovery.

Sensor networks are both spatial and random. As a large number of sensor nodes are deployed, e.g., from vehicles or aircrafts, they are essentially scattered randomly across large spatially extended regions. In the corresponding abstract graph two nodes are connected if they mutually fall within each others transmission range, depending on the emitting power, the attenuation function and the required minimum signal to noise ratio. Random geometric graphs (also referred to as Poisson/Boolean spatial graphs), capturing the above scenario, are a common and well established starting point to study the structural properties of sensor network, directly related to coverage, connectivity, and interference. Further, most structural properties of these networks are discussed in the literature in the context of continuum percolation [111, 125, 40].

The common design challenge of these networks is to find the optimal connectivity for the nodes: If the connectivity of the nodes is too low, the coverage is poor and sporadic. If the node connectivity is too high, interference effects will dominate and results in degraded signal reception [61, 159, 88, 87]. From a topological viewpoint, these networks are, hence, designed to "live" somewhat above the percolation threshold. This can be achieved by adjusting the density of sensor nodes and controlling the emitting power of the nodes; various power-control schemes have been studied along these lines [61, 87]. We consider random geometric graphs above the

percolation threshold, as minimal models for the underlying network communication topology. The focus of this work is to study novel cascade-like local communication dynamics on these well studied graphs.

Here, we focus on the scenario where the agents (the individual sensors) are initially in an *inactive* mode, typically performing some periodic local measurements. However, if an alarm-triggering event is detected locally by a (few) agent(s), the network, as a whole should "wake-up" (all agents turning to an active state), to closely monitor the spatial and temporal behavior of the underlying phenomena that caused the alarm (e.g., spread of a fire or toxic chemicals). This process of turning agents from an inactive state to an active one, requires some kind of local rules between agents. There are three (somewhat conflicting) objectives for constructing an optimal protocol:

1. reliability, so local erroneous events or false-alarms are suppressed and do not result in a "global wake-up";

2. speed, so that sensors can monitor the underlying physical, chemical, etc. phenomena;

3. energy efficient, so main concern in sensor networks, namely energy limitation is addressed.

To this end, we will consider a simple threshold-based model (or multiplex propagation) [60, 155, 34] on the sensor network with the potential to efficiently facilitate the transition of the nodes from an inactive to an active state. First, we will consider the threshold-based cascade dynamics on random geometric networks. Then we will experiment with the "addition" of a few long-range communication links, representing multi-hop transmissions. In particular, we will investigate the benefits in shortening the global transition time versus the increase in communication (and therefore also energy) costs. Such networks, commonly referred to as small-world networks [154, 153], has long been known to speed up the spread of local information to global scales [4, 50, 115, 154, 153] and to facilitate autonomous synchronization in coupled multi-component systems [145, 12, 79, 83, 84].

2.1 Threshold-Based Propagation on Random Geometric Networks

2.1.1 Random Geometric Networks

As mentioned in the Introduction, we consider random geometric graphs [111, 125, 40] as the simplest topological structures capturing the essential features of ad hoc sensor networks. N nodes are distributed at uniformly random in an $L \times L$ spatial area. For simplicity we consider identical radio range R for all nodes. Two nodes are connected if they fall within each other's range. An important parameter in the resulting random geometric graph is the average degree or connectivity α (defined as the average number of neighbors per node $\overline{k}$), $\alpha \equiv \overline{k} = 2K/N$, where K is the total number of links and N is the number of nodes. In random geometrical networks, there is a critical value of the average degree, α_c, above which the largest connected component of the network becomes proportional to the total number of nodes (the emergence of the giant component) [111, 125, 40]. For a given density of nodes ρ, there is a direct correspondence between the degree of connectivity α and the radio range R of each node [111, 125, 40],

$$\alpha = \rho \pi R^2 \ . \tag{2.1}$$

In what follows, for convenience, we will use R instead of α, as the relevant parameter controlling the connectivity of the network.

2.1.2 Threshold-Controlled Propagation

The phenomenon of large cascades triggered by small initial shocks, originally motivated by propagation in social networks [60], can be described by a simple *threshold-based* model [60, 155, 34]. This model considers the dynamics *on* a network of interacting agents (wireless sensors in the present context), each of which must decide between two alternative actions and whose decisions depend explicitly on the actions of their neighbors according to a simple threshold rule. Unlike in simple diffusive propagation, such as the spread of a disease, where a single node is sufficient to "infect" (activate) its neighbors, in threshold-based (multiplex) propa-

gation node activation requires simultaneous exposure to multiple active neighbors. Here, we implemented these simple rules for random geometric networks, capturing the topological features of wireless sensor networks.

The detailed description of the threshold model is as follows. Each agent can be in one of two states: state 0 or state 1, corresponding to the agent being inactive or active, respectively. Upon observing the states of its k neighbors, an agent turns its state from 0 to 1 only if the fraction of its active neighbors is equal to or larger than a specific threshold φ. In this work we are interested in the temporal characteristics of global cascades (network "wake-ups"), hence once a node turns active, it remains active for the duration of the evolution of the system.

We consider a system of N agents located at the nodes of a random geometric network. Each agent is characterized by a fixed threshold $0 \leq \varphi \leq 1$. For simplicity all agent have the same threshold. Initially the agents are all off (in state 0). The network is perturbed at time $t = 0$ by a small fraction of nodes that are switched on (switched to state 1). The number of active nodes then evolves at successive time steps with all nodes updating their states simultaneously (synchronous updating) or in random, asynchronous order (asynchronous updating) according to the threshold rule above. Once a node has switched on, it remains on (active) for the duration of the experiment.

For sensor networks, we are interested in the behavior of the network under emergent situation which is represented by small perturbation in the initial condition. In our investigation, the focus is on:

1. the probability that a successful *global cascade* will be ignited by small fraction of initial seed(s);

2. time needed for a global cascade, that is how fast an initial shock will spread out to the entire network; and

3. the energy used for communication between agents in a successful global cascade.

The last quantity is an important factor in designing wireless sensor networks. Here the term *cascade* refers to an event of any size triggered by initial seed(s), whereas

global cascade is reserved for sufficiently large cascades (corresponding to a final fraction of active agents, larger than a cutoff fraction of large, but finite network).

2.1.3 Simulation and Analysis

We simulated systems consisting of $N = 10^4$ sensor nodes distributed randomly in a $L \times L = 10^3 \times 10^3$ (in arbitrary units) two-dimensional region with periodic boundary conditions. These nodes are employed (for simplicity) with identical communication range R to form a random geometric network. For a fixed density of sensor nodes, $\rho = N/L^2$, the relationship between the average degree α and the radio range R is given by Eq. (2.1).

Figure 2.1 displays the phase diagram for the cascade dynamics on the (φ, R) plane in terms of the probability of global cascades for two different initial seed size. Each point in the graph is obtained by averaging over 1000 simulations (including different network topologies and initial cascade seeds). For seed size one, only a single node is activated initially in the network. For seed size three, we randomly select three neighboring (connected) nodes as a seed and turn them active as the initial condition. In the phase diagram, by fixing the threshold φ and going along the line parallel to the R-axis, hence increasing the radius of communication R, the system exhibits two different phase transition as shown in Fig. 2.2. The first transition occurs at about $R_{c1} \simeq 12.5$ (corresponding to $\alpha \simeq 4.9$) where the probability of global cascades sharply rises from 0 to around 1 for both initial seed sizes. We refer to this phase transition as phase transition I. Further increasing the communication range R, the probability of a global cascade slowly drops to 0. This phase transition is referred to as phase transition II. Phase transition I is inherently related to the emergence of the giant component in a random geometrical network [111, 125, 40]. Below the critical value R_{c1}, the network is poorly connected, hence no cascade can spread to global scales. Above R_{c1} the giant component can support global cascades, depending on the threshold φ. This transition across R_{c1} is sharp, related to the scaling properties of the giant component of the random geometric network. For sufficiently small threshold values $\varphi < 1/\alpha$ the probability of global cascades is close to 1 in a well connected graph, yielding the upper boundary of the region

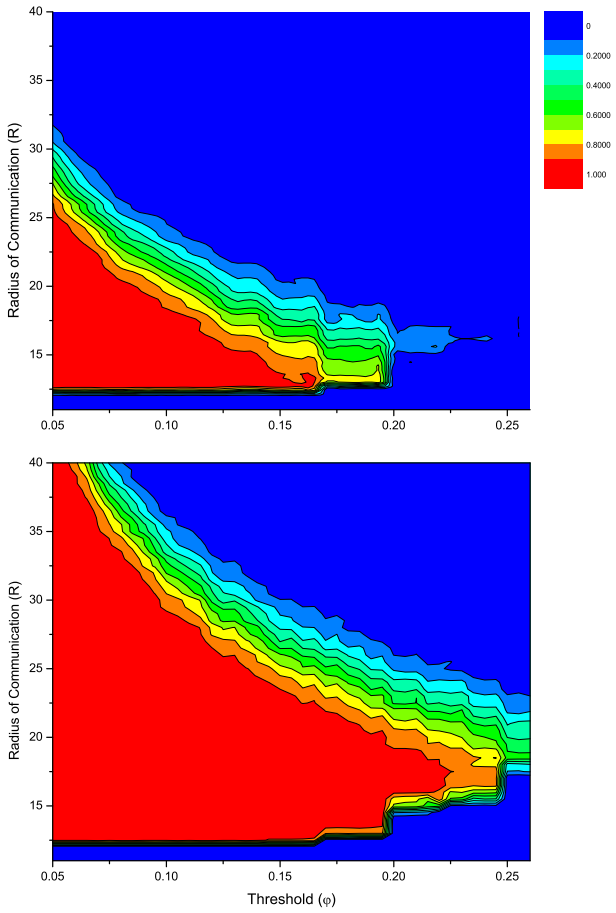


Figure 2.1: Phase diagram in the plane of threshold and radio range (φ, R). The cascade window is enclosed by two different types of boundaries. (a) the cascade window when the size of the initial seed is one. (b) the cascade window when the size of the initial seed is three. Both graphs are obtained at the system size $N = 10^4$ and averaged over 1,000 simulation runs.

where global cascades are possible, $R_c(\varphi)$. As the range R is increased, while the threshold φ is held fixed, nodes will have have so many neighbors that they cannot be activated by a single active agent. Since the relationship between the average degree and the range follows Eq. (2.1), the approximate location of the boundary associated with phase transition II scales as

$$R_c(\varphi) \sim 1/\sqrt{\varphi}\,. \tag{2.2}$$

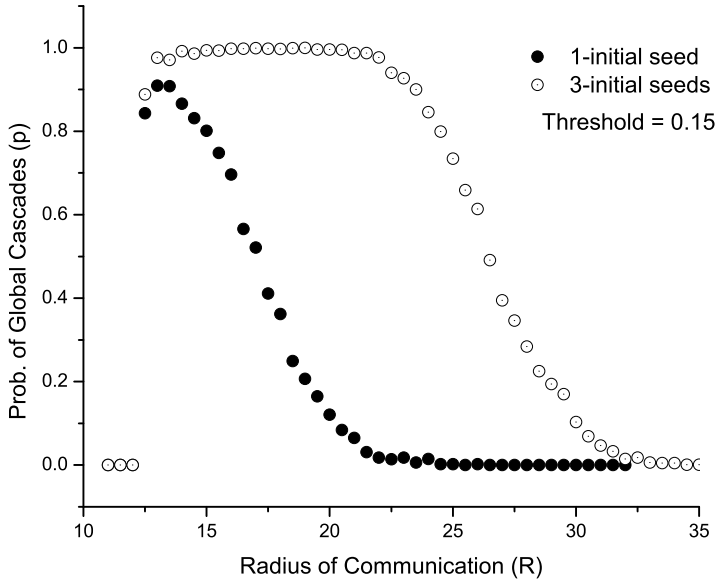


Figure 2.2: Cross section of cascade window from Fig. 2.1, at $\varphi = 0.15$, with one initial seed node (solid circles) and with three initial seed nodes (open circles). Choosing any value at which these two curves separate sufficiently far away from which other enables the network to suppress the global cascade triggered by poorly supported events. The radius R is connected to the average network degree α according to Eq. (1).

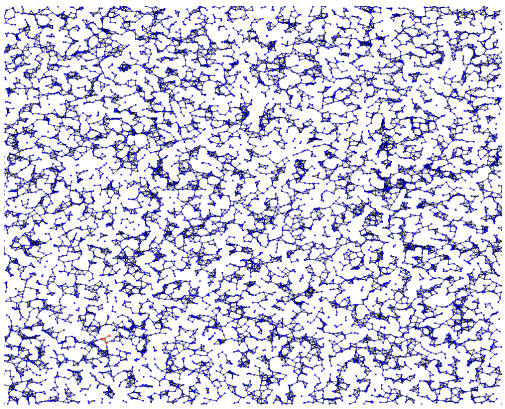
Figure 2.3 is the snapshots from a successful global cascade with non-periodical boundary condition. At time $t = 170$ the fraction of active nodes exceeds 0.85 and completes the global cascade.

In reality, it is possible that one of the sensors fails or turns active and sends out an alarm message by accident. One of our objectives is to construct local dynamics where global "wake-ups" due to erroneous events or miss-alarms are suppressed or possibly excluded. This should be accomplished in an autonomous fashion, i.e., without any external filtering or intervention. To this end, first, it is reasonable to assume that the probability that a specific node will send out an erroneous alarm is very small. Thus, if, e.g., three neighboring nodes become active simultaneously, it should be considered to be a real event. Comparing Fig. 2.1(a) (one initial seed node) and Fig. 2.1(b) (three initial seed nodes), it is clear that to prevent miss-

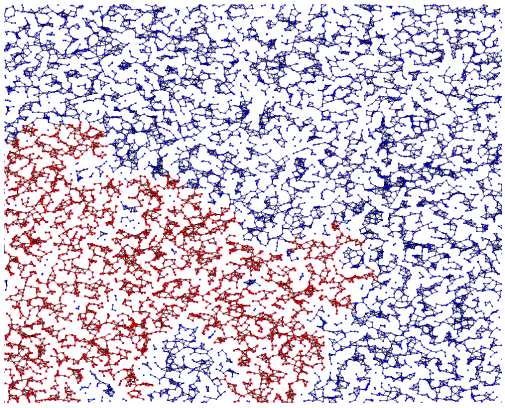
alarms by eliminating global cascades triggered by one initial seed, while allowing them when triggered by the real event defined as above, we should chose the region outside of the global cascades (blue/dark colored region) of Fig. 2.1(a) but inside the global cascade region (red/light colored region) of Fig. 2.1(b) by picking proper R and φ. Fig. 2.2 is the cross section of phase diagram at $\varphi = 0.15$. We can see that if the radius of communication range is within $20 < R < 25$, the probability of global cascades triggered by three initial seed nodes is still very close to 1 while the probability of global cascades triggered by a single initial seed essentially drops to 0. Hence, by choosing the proper values of R and φ, according to the phase diagrams, we can prevent miss-alarms and, at the same time, ensure that real events will trigger global cascades with near certainty.

2.2 Threshold-Model with Small-World Links

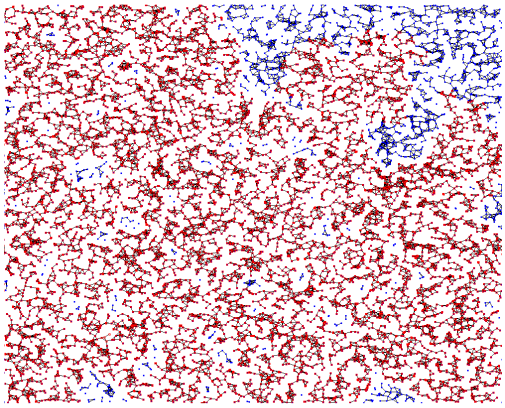
The small world phenomenon originates from the observation that individuals are often linked by a short chain of acquaintances. Milgram [112] conducted a series of mail delivery experiments and found that an average of "six degrees of separation" exists between senders and receivers. The small-world property (very short average path length between any pair of nodes) were also observed in the context of the Internet and the world wide web. Motivated by social networks [153], and to understand network structures that exhibit low degrees of separation, Watts and Strogatz [154] considered the re-wiring of some fraction of the links on a regular graph, and observed that by re-wiring just a small percentage of the links, the average path length was reduced drastically (approaching that of random graphs), while the clustering remains almost constant (similar to that of regular graphs). This class of graphs was termed small-world graphs to emphasize the importance of random links acting as shortcuts that reduce the average path length in the graph. In the following, we will continue to study random geometric graphs in the context of wireless sensor networks but with the goal of investigating the applicability of the small world concept to these networks. The topological properties (such as the shortest path and the clustering coefficient) of small-world-like sensor networks have been studied in [64]. Here, we focus on the effect of *adding* some random communication



(a)



(b)



(c)

Figure 2.3: Snapshots of the regular random geometric network evolving with time after the initial shock. Blue dots are in state 0, red ones are in active state. Black lines are local links. Snapshots are taken at time-step (a) t=0; (b) t=80; (c) t=170. The network implemented the synchronous updating with $N = 10^4$, $\varphi = 0.12$, and $R = 16.0$.

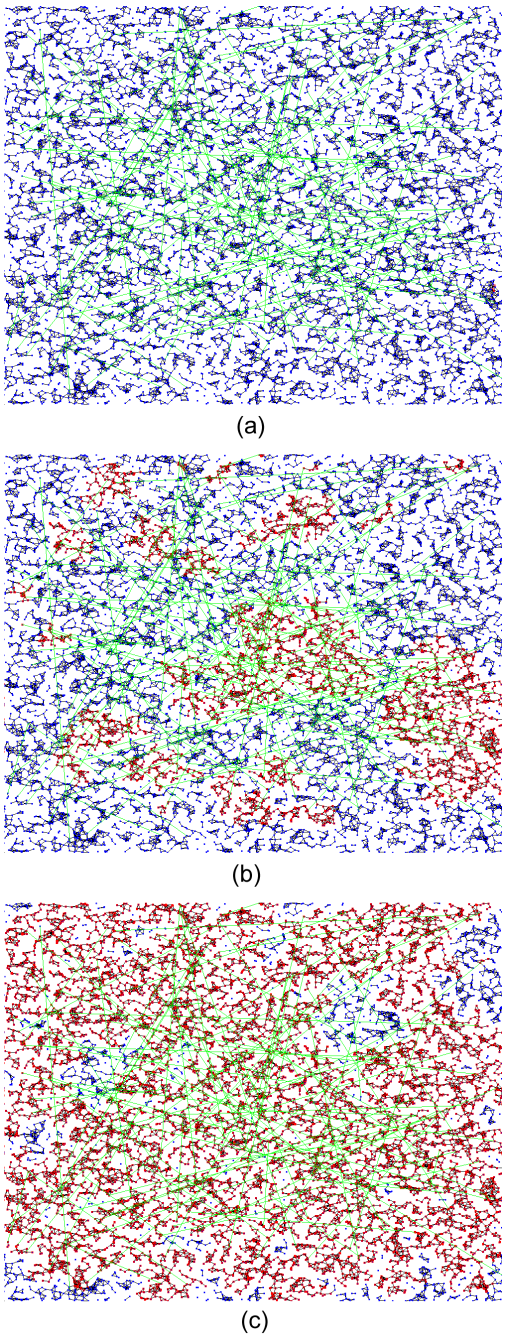


Figure 2.4: Snapshots of the small-world network evolving with time after the initial shock. Green lines are random long-ranged links. Snapshots are taken at time-step (a) t=0; (b) t=28; (c) t=45. Messages propagate much faster than in regular random geometric networks. The network implemented the synchronous updating with $N = 10^4$, $\varphi = 0.12$, and $R = 16.0$. The total number of long-ranged links is 200.

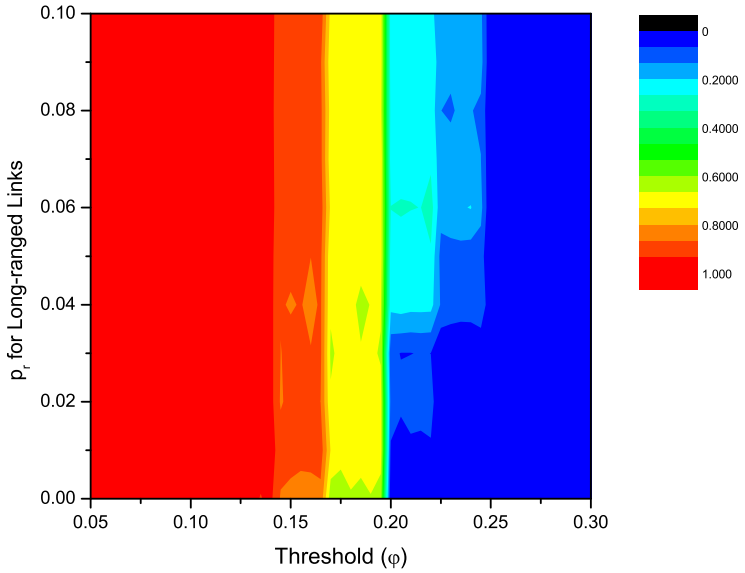


Figure 2.5: Phase diagram in the plane of threshold and the probability of long-ranged links (φ, p_r). The cascade window is enlarged as we increase the probability of long-ranged links (p_r). The graph is obtained at the system size $N = 10^4$ with $R = 14$ and averaged over $1,000$ simulations.

links between possibly distant nodes, on *the dynamics* on the network.

As previously, we randomly deploy $N = 10^4$ sensor nodes in a $L \times L = 10^3 \times 10^3$ area, set a fixed radio range R to connect them, and generate the corresponding random geometric network. In addition, we also add random "long-range" links to this backbone to construct a small-world-like network. Random (possibly long-range) connections are constructed by *adding* a fixed number of random links, so that the total number of random added edges is $p_r N$ with $p_r \ll 1$. Alternatively, one can construct statistically identical networks by adding a random link emanating from each node with probability p_r. The procedure has several different realizations, depending on how p_r (in the above probabilistic interpretation) varies with the underlying spatial distance between the randomly connected two nodes.

1. $p_r = constant$, in which case there are no restrictions on the length of random

long range links;

2. $p_r \propto 1/d^{\delta}$, where d is the distance between the two chosen nodes and δ is a parameter;

3. $p_r = constant$ for $d \leq d_c$ and $p_r = 0$ for $d > d_c$, where d is the distance between the two picked nodes and d_c is a parameter to which we refer as *cutoff distance*.

The dynamics on the network is defined by the same threshold model that was discussed earlier. Here we show results for the case when the size of the initial seed is set to one. Thus, the dynamics is controlled by threshold φ, the radio range R, and the probability of a long-range links p_r.

The addition of small-world links is expected to speed up propagation (reduce time for global cascades to complete) in the region where global cascades are possible [34]. Focusing on the three quantities outlined earlier, the probability of global cascades (yielding the phase diagram), the average global cascade times, and energy costs, compared to the original random geometrical network are discussed below. The measured observables are averaged over 1,000 simulations. The time needed for global cascades is averaged over all successful global cascades with the same model parameters.

Snapshots of a successful global cascade in the small-world network is shown in Fig. 2.4. Comparing the snapshots from regular random geometric networks, nodes in small-world networks can be ignited by his neighbors in distance via long-ranged links added to the network, thus expedite the speed of message propagation.

Next, we define the energy consumption during a successful global cascade. The total energy consumption in a wireless sensor network is, in most part, due to communication between nodes, computing, and storage (neglecting some smaller miscellaneous costs). The communication is the main part of energy consumption, and is directly related the network structure and dynamics, while energy used for computing can be regarded as a constant, so it can be easily evaluated. Consequently, in the following, we only consider the energy consumption in communications between sensor nodes.

There are two kinds of communication in small-world graphs. First is the local

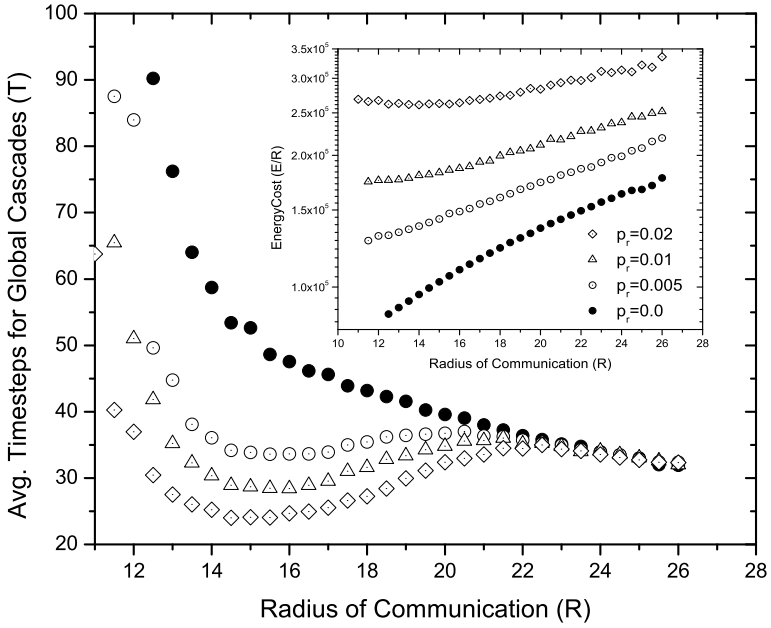


Figure 2.6: Effect of long-ranged links on the dynamics of random geometric graph under threshold rule. The introduction of long-ranged links decreased the time needed for cascades but increased the energy cost. Symbols mean: $p_r = 0$ (solid circles); $p_r = 0.05$ (open circles); $p_r = 0.01$ (open triangles); $p_r = 0.02$ (open diamonds). The graph is averaged over $1,000$ simulations with synchronous updating and the threshold fixed at $\varphi = 0.12$.

communication. According to the nature of sensor nodes, the wireless broadcast is the most commonly used method for local communication. Thus, the energy cost for local broadcast is proportional to the square of radio range R, $E_l = cR^2$, where c is a coefficient that we scaled to 1. It should also be noted that a message send by a specific node can be received by all his local neighbors, regardless of the number of them, by a single broadcast. The second type of communication is the long-ranged one. The energy cost for long-ranged communication depends on how this communication is implemented, including solutions such as a directional antenna, multi-hop transmission, global flooding, etc. In this paper, we consider

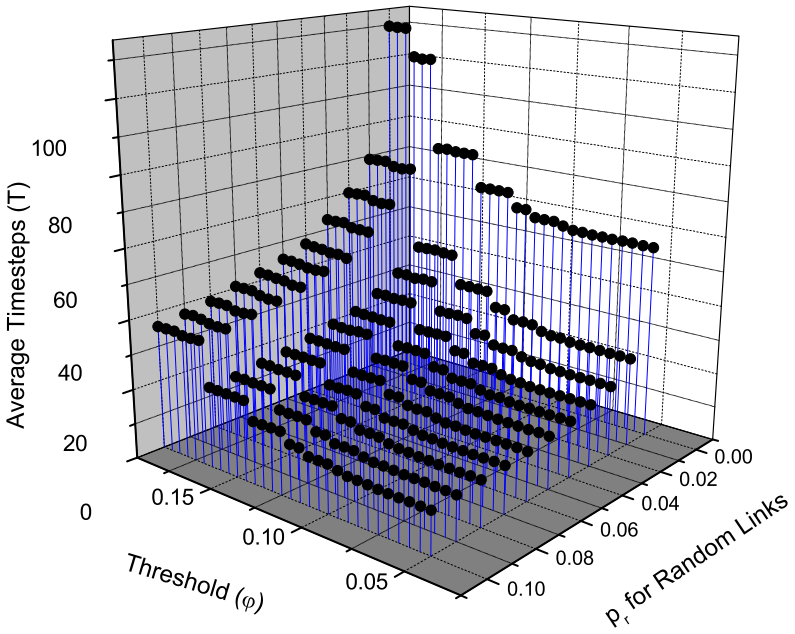


Figure 2.7: The reduction in global cascade time resulting from the increased number of long-ranged links. Similar behavior as in Fig. 2.6 is observed when radio range R is fixed. The graph is averaged over 1,000 simulations with synchronous updating, $N = 10^4$, and $R = 14.0$.

multi-hop transmissions in which messages reach their destination following a multi-hop route. The energy cost of such a transmission (for well-connected networks) can be approximated as

$$E_r \simeq cR^2(\frac{d}{R}) = cRd = E_l\frac{d}{R}, \tag{2.3}$$

where d is the distance between two nodes. Unlike in case of broadcast, each pair of nodes that has a long-ranged link between them will incur additional energy to communicate. Hence, the total energy consumption in communication for a successful global cascade with m local communications and n long-ranged communications is:

$$E = mE_l + nE_r \tag{2.4}$$

Since $n = p_rN$, then denoting the average length of long-range links as $\overline{d}$ and using

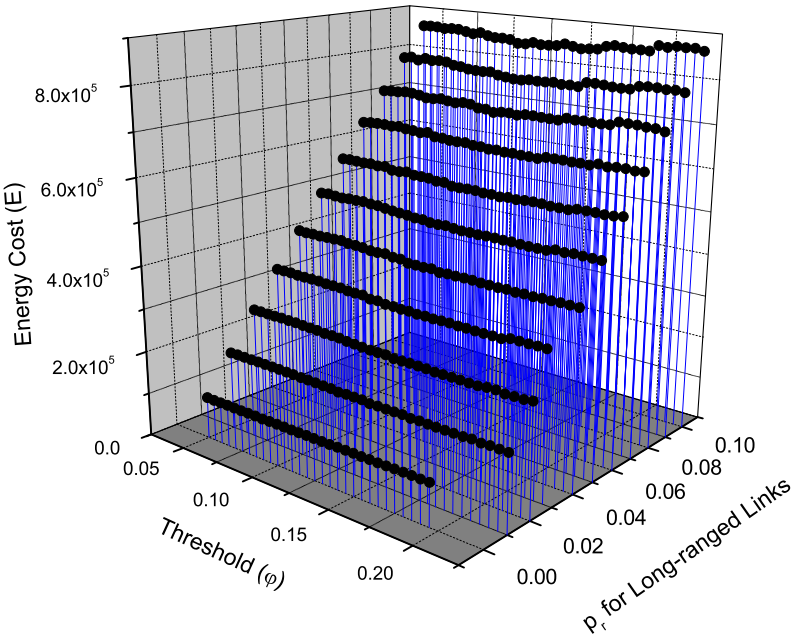


Figure 2.8: The energy cost of global cascades as a function of the threshold and the density of long-ranged links. The graph is averaged over 1,000 simulations with synchronous updating, $N = 10^4$, and $R = 14.0$.

an approximation $m \simeq N$ (the exact value of m is equal to the number of nodes participating in the global cascade), the energy cost can be rewritten as:

$$E \simeq cN(R^2 + p_r \overline{d} R) = NE_l \left(1 + \frac{p_r \overline{d}}{R}\right) . \tag{2.5}$$

The first term corresponds to the local communication energy, while the second term represents the additional energy needed for the long-ranged links. The ratio of the long-ranged link communication energy to the local communication energy is proportional to the probability of long-ranged links p_r and the average link length and inversely proportional to the radio radius R.

Graphs for the cascade window, after adding the long-ranged links, are similar to that of the model without long-range links [Fig. 2.1]. The most significant difference is that the lower boundary of the cascade window (which we referred to

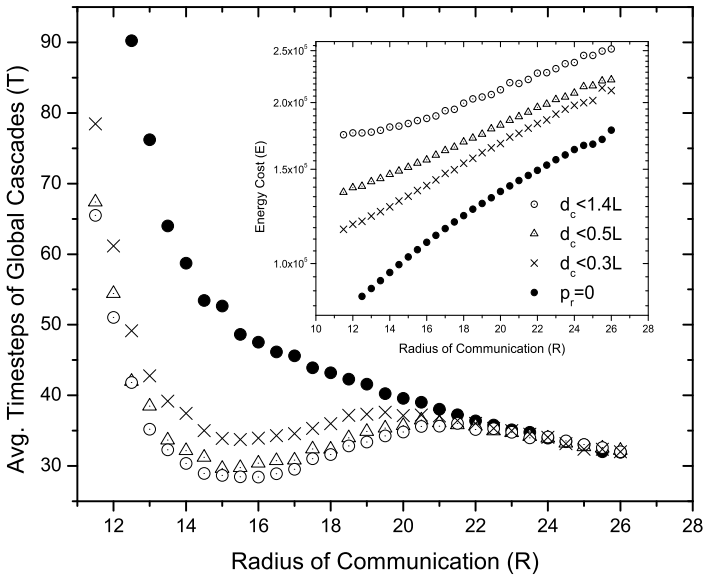


Figure 2.9: A comparison of average time and energy costs with different cutoff distance of long-ranged links. Different symbols are, $p_r = 0$ (solid circles); $p_r = 0.01$, $d_c = \sqrt{2}L$ (open circles); $p_r = 0.01$, $d_c = 0.5L$ (open triangles); $p_r = 0.01$, $d_c = 0.3L$ (cross). The network implemented the synchronous updating with the threshold fixed at $\varphi = 0.12$.

as phase transition I) shifts downwards by a small amount ($\sim 0.5 - 1.0$) toward smaller R. As we discussed above, a sufficiently large value of R is necessary for the entire graph to be well connected and to be able to support global cascades. By adding long-ranged links, it will be more likely that several small clusters would be connected via these long-ranged links to form a larger cluster, hence lowering the boundary of cascade windows. Figure 2.5 displays the phase diagram in terms of the probability of global cascades on the plane of the threshold φ and the probability of random long range links p_r, (φ, p_r), when the radio range R is fixed at 14.0. Different from the results of Ref. [34] (as a result of adding as opposed to re-wiring random links), the cascade window somewhat enlarges at a specific region of (φ, R) when p_r increases.

The advantage of the small-world links is that they can significantly decrease

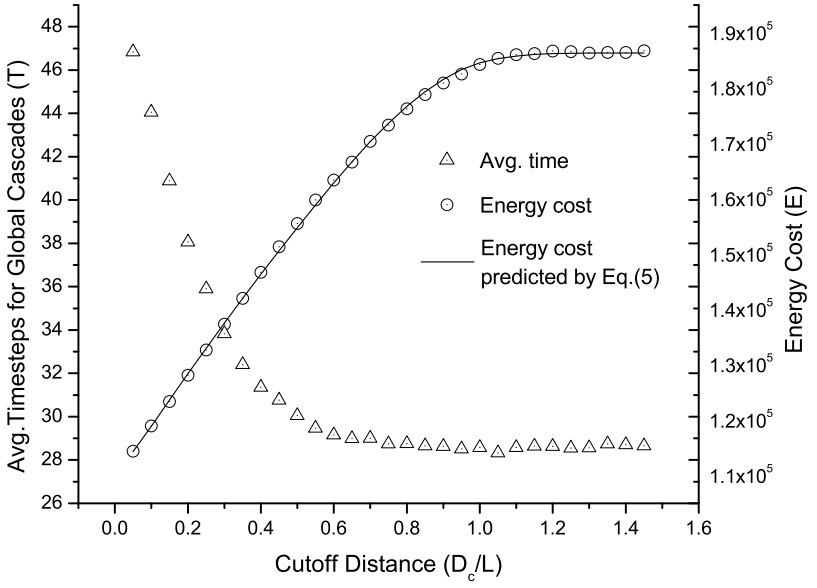


Figure 2.10: A dependence of the average time and energy costs on the cutoff distance when the threshold and radio range are both fixed. According to the shape of these two curves, we can optimize the system behavior by varying the cutoff distance d_c. Open triangles are the average time, open circles are the energy cost and solid line is the energy cost predicted by Eq. (2.5) with $R = 16$, $\varphi = 0.12$ and synchronous updating.

the global cascade time. In other words, alarms and messages propagate faster in small-world graphs than in the original random geometric graphs, as shown in Figs. 2.6 and 2.7. For fixed threshold, the time needed for a global cascade decreases monotonously as R increases in regular random geometric graphs. In contrast, in small-world graphs, the average time is much lower and reaches its optimal (minimum) value at $R \simeq 15.0$. The more long-ranged links are added to the network, the lower the average time is. Meanwhile, due to the long-range communications, the average energy cost for a successful global cascade is also increasing, linearly with p_r, in agreement with Eq. (2.5) as can be seen in Fig. 2.8.

The above observations can be used to develop a scheme to compensate the increase in energy cost caused by the added long-range links. In the discussion above,

the distribution of the distance of the long-range links was uniform. Next, we explore the effects of "suppressing" the occurrence of links with large spatial length. While one can implement, e.g., a power-law or an exponentially-tailed distribution for the spatial length of the added random links, here, for simplicity and to ease technical realizations, we use a sharp cutoff for $p_r(d)$,

$$p_r(d) = \begin{cases} p_{r0} & \text{if } d \leq d_c \\ 0 & \text{if } d > d_c, \end{cases} \tag{2.6}$$

where d_c is the cutoff distance for the long-ranged links. This distance may represent the range of a uni-directional antenna of special sensor nodes. Adding a small amount, $p_r N$, of such random links to the network will provide the long-range links in the network without the need to implement multi-hop routing.

Scaling d_c with the spatial size L of the system, when $d_c \geq \sqrt{2}L$, there is no restriction on the long-range link length. The average time and energy costs for global cascade under the restriction of long-range links is shown in Fig. 2.9.

It can be seen that even putting a strong restriction of the longest link distance ($d_c < 0.5L$) only slightly increases the average time for successful global cascades. However, under the same restrictions, the energy cost drops significantly, proportionally to the decrease in the average long-ranged link length, as predicted by Eq. (2.5). In particular, we explored the behavior of the average time (T) and the energy cost (E) vs. different cutoff distances (d_c) when R and φ are both fixed [Fig. 2.10]. The average time is close to its minimum when $d_c > 0.4L$ while the energy cost has still not saturated until $d_c > 1.0L$. Clearly, one can make a trade-off between the speed of message propagating and the energy cost with different cutoff distances of the long-range links. Depending on various applications and scenarios, we can choose different d_c to make the network respond faster to emergencies or be more efficient in terms of energy used.

CHAPTER 3
Naming Games in Two-dimensional and Small-world-connected Random Geometric Networks

Reaching agreement without global coordination is of fundamental interest in large-scale autonomous multi-agent systems. In the context of social systems, the objective is to understand and predict the emergence of large-scale population-level patterns arising from empirically supported local interaction rules between individuals (e.g., humans). Examples for such phenomena driven by social dynamics include the emergence and the evolution of languages [121, 120, 107] or opinion formation [29, 19, 46, 51, 85, 146, 80, 21, 9].

The creation of shared classification schemes in a system of artificial and networked autonomous agents can also be relevant from a system-design viewpoint, e.g., for sensor networks [92, 39]. Envision a scenario where mobile or static sensor nodes are deployed in a large spatially-extended region and the environment is unknown, possibly hostile, the tasks are unforeseeable, and the sensor nodes have no prior classification scheme/language to communicate regarding detecting and sensing objects. Since subsequent efficient operation of the sensor network inherently relies on unique object identification, the autonomous development of a common "language" for all nodes is crucial at the exploration stage after network deployment.

To this end, in this section we consider and slightly modify a simple set of rules, referred to as Language or Naming Games (NG), originally proposed in the context of semiotic dynamics [142, 76]. Such problems have become of technological interest to study how artificial agents or robots can invent common classification or tagging schemes from scratch without human intervention [142, 76]. The original model [142, 143, 141, 144] was constructed to account for the emergence of shared vocabularies or conventions in a community of interacting agents. More recently, a simplified version of the NG was proposed and studied on various network topologies by Baronchelli et al. [15, 13, 14], and by Dall'Asta et al. [41, 42] The advantage of studying a minimal model is that one can gain a deeper understanding of the spontaneous

self-organization process of networked autonomous agents in the context of reaching global agreement, and can extract quantitative scaling properties for systems with a large number of agents.

In the context of artificial agents, there are other possible scenarios when the NG algorithm, in addition to being interesting in its own merit in studying agreement dynamics on various networks, can also be particularly useful from a system-design viewpoint. That can be the case when one does not intend the outcome of the agreement process among many agents to be easily predictable. The actual process of electing a "leader" or coordinator among sensor nodes may actually be such a scenario. The leader must typically be a trusted node, with possible responsibilities ranging from routing coordination to key distribution [45]. Standard leader election (LE) algorithms [8, 65, 93, 105, 150] are essentially based on finding global extremum (e.g., maximum) through local communications [8, 65, 93]. Thus, the elections can be stolen by placing a node in the network with a sufficiently high ID (e.g., the largest number allowed by the number representation scheme of the sensor chips.) Along these lines, a possible application of the NG algorithm is autonomous key creation or selection for encrypted communication in a community of sensor nodes. Instead of having a centralized or hierarchial key management system with domain and area key distributors [45], group of sensor nodes can elect a key distributor or a security key for secure communications between group members.

In the following sections, we first briefly review recent results on the NG on various regular and complex networks. Then we put our focus on studying and analyzing different aspects of the behavior of the NG model, including the temporal and finite-size scaling of the opinions, the probability distribution of the agreement times and the cluster-size distribution in the NG on random geometrical graphs (RGGs). Further, we construct and present finite-size scaling for the agreement times in Small-World (SW)-connected RGGs.

3.1 Background and Prior Results on the Naming Game

In the simplified version of the NG, agents perform *pairwise* games in order to reach agreement on the name to assign to a *single* object. This version of the NG was

investigated on fully-connected (FC) (also referred to as mean-field or homogeneous mixing) [15, 13], on regular [14], on small-world (SW) [41, 98], and on scale-free networks [42, 16]. In the FC network, each agent has a chance to meet with all others and compare their current local vocabularies (list of "synonyms") before updating them. On regular networks, agents have only a limited and fixed number of neighbors on a one-, two-, etc., dimensional grid with whom they can interact/communicate. The communication in both cases is "local", in that *pairs of agents* are selected to interact and to update their vocabularies. The basic algorithmic rules of the NG are as follows [15, 14]. A pair of neighboring nodes (as defined by the underlying communication topology), a "speaker" and a "listener", are chosen at random.[1] The speaker will transmit a word from her list of synonyms to the listener. If the listener has this word, the communication is termed "successful", and both players delete all other words, i.e., collapse their list of synonyms to this one word. If the listener does not have the word transmitted by the speaker, she adds it to her list of synonyms without any deletion. The above rules are summarized in Fig. 3.1.

Among the above rules, the restriction to a single object [15, 13] strongly reduces the complexity of the model, compared to a more general case where the naming process of multiple objects can be performed simultaneously. From a linguistic viewpoint, this rather strong restriction is equivalent to preventing homonymy, and instead, treating all objects independently. This strong assumption can be more realistic for a system of artificial agents, where agents assign random numbers (e.g., chosen from 2^{31} integers) as "words" to new objects. In this case, the number of potential words can be far grater than the number of objects, and the probability that two players invent the same word for different objects (hence giving rise to homonymy) is negligible.

It was found that employing the above local rules (*pair-wise* interactions), after some time, the agents vocabularies converge to a unique vocabulary shared among all agents [15, 13, 14, 41]. The major differences between the NG on FC graphs and

[1]Note that on strongly heterogeneous (scale-free) graphs, the order whether the listener or the speaker is chosen first, strongly impacts the efficiency toward global agreement. Choosing the listener first at random will increase the chance for selecting a node (as a neighbor) with larger degree for speaker. In turn, hubs will be the most frequent speakers, giving rise to faster convergence to global agreement at a mildly elevated memory cost [42, 16].

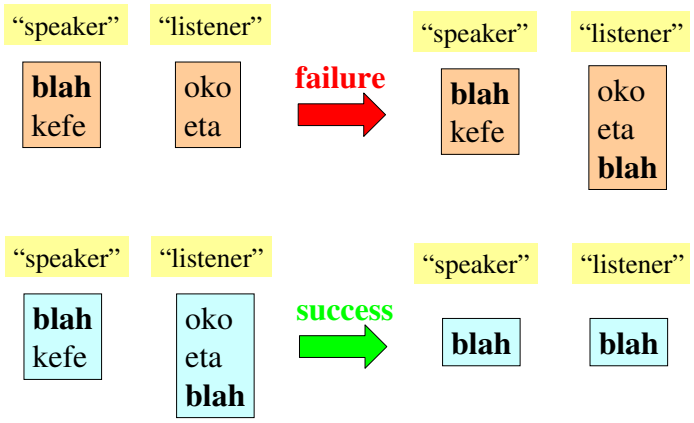


Figure 3.1: Schematic rules of the Naming Game [15] as described in the text. If the speaker has more than one word on her list, she randomly chooses one; if it has none, it generates one randomly.

on regular low-dimensional grids arise in the scaling of the the memory needed to develop the common language before convergence occurs, and in the scaling of the time t_c needed to reach global agreement. (The memory need in the present context is the typical value of the largest number of words an agent may posses throughout the evolution of the game [15, 14].) In the FC network, the convergence process to global agreement is fast [$t_c \sim \mathcal{O}(N^{1/2})$ for N agents], but large memory [$\mathcal{O}(N^{1/2})$] is needed per agent [15]. For a regular two-dimensional network (or grid), spontaneous evolution toward a shared dictionary is slow [$t_c \sim \mathcal{O}(N)$], but the memory requirement per agent is much less severe [$\mathcal{O}(1)$] [14]. When the NG is implemented on Watts-Strogatz [154] SW networks, the agreement dynamics performs optimally in the sense that the memory needed is small, while the convergence process is much faster than on the regular networks [$t_c \sim \mathcal{O}(N^{0.4})$, closer to that of the FC network] [41].

Sensor networks, which are motivating our study, are both spatial and random. As a large number of sensor nodes are deployed, e.g., from vehicles or aircrafts, they are essentially scattered randomly across large spatially-extended regions. Two

nodes are connected if they mutually fall within each others transmission range, depending on the emitting power, the attenuation function and the required minimum signal to noise ratio. Again we use random geometric graphs (RGGs) to capture the above scenario. RGGs, also referred to as spatial Poisson/Boolean graphs, are a common and well established starting point to study the structural properties of sensor network, directly related to coverage, connectivity, and interference. Further, most structural properties of these networks are discussed in the literature in the context of continuum percolation [111, 125, 40].

The common design challenge of these networks is to find the optimal connectivity for the nodes: If the connectivity of the nodes is too low, the coverage is poor and sporadic. If the node connectivity is too high, interference effects will dominate and result in degraded signal reception [61, 159, 88, 87, 25]. From a topological viewpoint, these networks are, hence, designed to "live" somewhere above the percolation threshold. This can be achieved by adjusting the density of sensor nodes and controlling the emitting power of the nodes; various power-control schemes have been studied along these lines [61, 87, 25]. In our work we consider RGGs in two-dimensions above the percolation threshold, as minimal models for the underlying network communication topology. Further, we consider RGGs with an added small density of "random" long-range links. The resulting structure resembles small-world (SW) networks [154, 115], also well studied in the context of artificial [79, 101] and social systems [153, 115].The focus of this work is to study the NG algorithm on these spatially-embedded random graphs.

3.2 Naming Games on Random Geometric Networks

3.2.1 Random Geometric Graphs

The same model of $2d$ random geometric graphs as in chapter 2 is implemented as the underlying structures capturing the essential features of ad hoc sensor networks. N nodes are uniformly random distributed in an $L \times L$ spatial area. When connecting adjacent nodes we apply identical radio range R for all nodes. Given the density of nodes $\rho = N/L^2$, and the radio range R of the nodes, the average connectivity of the graphs is then $\bar{k} = \rho \pi R^2$.

3.2.2 The Naming Game with Local Broadcast

We consider the Naming Game on random geometrical graphs. In the original context of the NG, agents try to reach agreement in finding a unique "word" for an object observed by them. In one of the above proposed potential applications, agents try to generate a shared unique key for encrypted communication. For simplicity, we will use the term "word" for the latter as well when describing the algorithm.

Motivated by communication protocols employed by sensor nodes, we modify the communication rules to make them applicable for sensor networks. Instead of pairwise communications, nodes will initiate *broadcast* (to all neighbors) in a continuous-time asynchronous fashion. The initial condition is such that the "vocabulary" of each node is empty at beginning. Then at every elementary time step, a node is chosen randomly out of N nodes (mimicking Poisson asynchrony for large N). This node (the "speaker") will broadcast a word from her list of "synonyms"; if her list of synonyms is empty, the speaker randomly invents a word; if she already has several synonyms, it randomly chooses one. Her neighbors (the "listeners") compare their vocabularies with the word transmitted by the speaker. If a listener has this word, she considers the communication a success, and she deletes all other words, collapsing her list of synonyms to this one word. If a listener does not have the word transmitted by the speaker, she adds it to her list of synonyms without any deletion. If *at least one* listener had the word transmitted, the speaker considers it (at least a partial) success, and (somewhat optimistically) collapses her list of synonyms to this one word. At every step, the "success" rate S is defined as the fraction of listeners who were successful (i.e., those that had the word transmitted by the speaker). From the above it is clear that one of the successful listeners, if any, has to report the outcome of the "word matching" to the speaker. In order to achieve that efficiently, in real sensor-network implementations one can employ the "lecture hall" algorithm [147]. In this paper time t is given in units of one "speaker"-initiated broadcast *per node*. The main difference between the above algorithm and the one in Refs. [15, 13, 14, 41] is the *broadcast* (instead of pairwise communications) and the underlying network (RGG in this paper) to capture the essential features of the NG in sensor networks.

When starting from empty vocabularies, agents invent words randomly. After time of $\mathcal{O}(1)$ [on average one speaker-initiated broadcast per node], $\mathcal{O}(N/(\overline{k}+1))$ different words have been created. Following the early-time increase of the number of different words $N_d(t)$, through local broadcasts, agents slowly reconcile their "differences", and eventually will all share the same word. First, a large number of small spatial clusters sharing the same word develop. By virtue of the slow coalescence of the interfaces separating the clusters, more and more of the small clusters are being eliminated, giving rise to the emergence of larger clusters, eventually leading to one cluster in which all nodes are sharing the same word. As suggested by Baronchelli et al. [14], this late-time process is analogous to coarsening, a well-known phenomenon from the theory of domain and phase ordering in physical and chemical systems [26]. Figure 3.2 shows snapshots of vocabularies of the nodes at different times. For later times, group of nodes which already share the same word, slowly coarsen, until eventually only one domain prevails. This behavior is also captured by Fig. 3.3(b), tracing the number of different words as a function of time $N_d(t)$, eventually reaching global agreement, $N_d = 1$.

3.2.3 Basic Scaling Considerations and Analogy with Coarsening

Before turning to the detailed discussion of our simulation results, we first sketch the framework of coarsening theory [26], applicable to the observed late-time dynamics of the NG on regular d-dimensional lattices [14]. Coarsening has also been observed in other models relevant to opinion formation and social dynamics [20, 85]. Unlike other minimalist (typically two-state) models often employed to study opinion formation [51], such as the one studied by Sznajd-Weron & Sznajd [146], the Voter model [95, 20], or the majority rule model [85], in the NG, each agent can be in an *unlimited* number of discrete states (corresponding to a chosen word). Further, at any instant before reaching global consensus, an agent can have different possible words for the object. Because of the potentially unlimited number of discrete states the agents can assume, the late-stage evolution of the NG resembles that of infinite-state (Q=∞) Potts model [132, 133, 73, 89, 130, 47, 48, 86, 139, 138, 6, 106, 158].

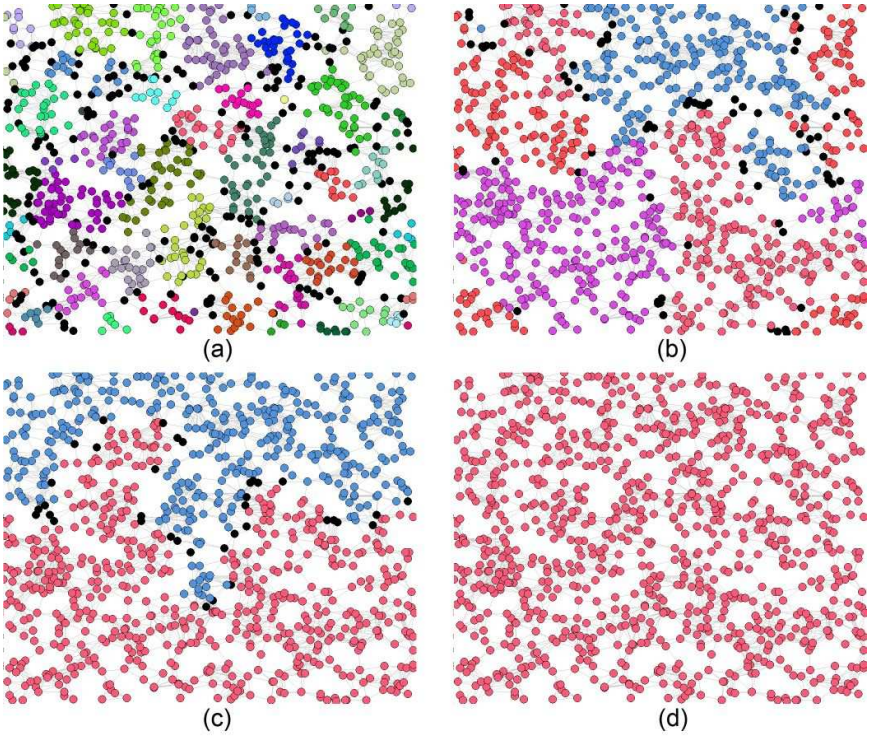


Figure 3.2: Snapshots of the time evolution of the contents of the agents' word lists during the process of reaching global agreement on RGG for $N = 1,000$ nodes at time (a) $t = 1$; (b) $t = 43$; (c) $t = 169$; (d) $t = 291$. The average degree is $\overline{k}\approx 12$. Initially, the word lists are empty for all agents. Time, as through the paper, is measured in units of "speaker"-initiated broadcasts *per node*. Different colors correspond to different words, with black indicating nodes with multiple words. After the early-time increase in the number of different words in the systems, small spatial clusters sharing the same word quickly form, then subsequently "coarsen" until eventually only one global cluster prevails.

While RGG is a random structure, it is embedded in two dimensions, and we also attempt to employ elementary scaling arguments from coarsening theory. According to Ref. [14], on regular d-dimensional lattices, the typical size of domains (each with already agreed upon one word) is governed by a single length scale $\xi(t) \sim t^{\gamma}$ with γ=1/2, analogous to that of domain formation in systems with a non-conserved order parameter [26]. Thus, in d dimensions the average domain size

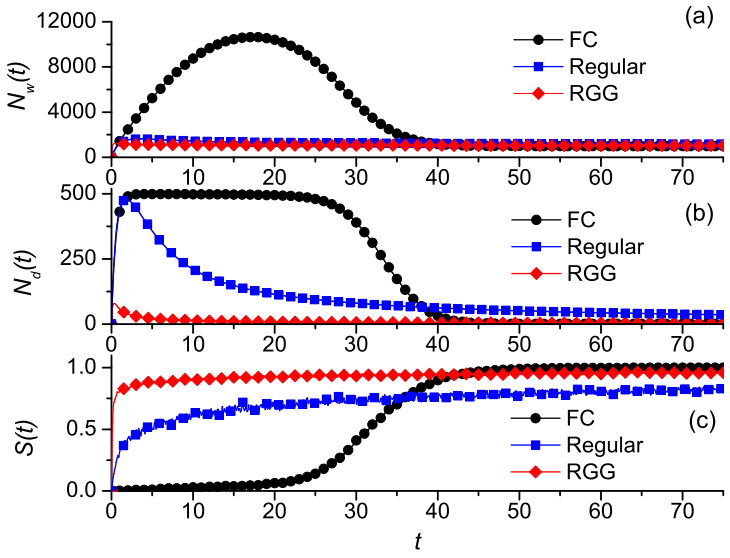


Figure 3.3: Time evolution of the relevant observables in the Naming Game in the fully-connected (FC), two-dimensional regular (with four nearest neighbors), and random geometric networks (RGG) for N=1024, averaged over $1,000$ independent realizations; (a) the total number of words in the system $N_w(t)$; (b) the number of different words $N_d(t)$; (c) the average success rate $S(t)$. The average degree of the underlying RGG is $\overline{k}\approx 12$. Data for the FC and $2d$ regular networks are reproduced by our simulations, following Refs. [15, 14], for comparison.

$C(t)$ follows

$$C(t) \sim \xi^d(t) \sim t^{d\gamma} \ . \tag{3.1}$$

and the total number of *different* words N_d at time t scales as the typical number of domains

$$N_d(t) \sim \frac{N}{\xi^d(t)} \sim \frac{N}{t^{d\gamma}} \ . \tag{3.2}$$

Further, the total number of words N_w (N_w/N being the average memory load per agent), at this late coarsening stage, can be written as the number of nodes N plus the number of nodes with more than one (on average, between one and two) words, separating the different domains. It is of order of typical number of domains times

the typical length of the interface of one domain, yielding

$$N_w(t) - N \sim \frac{N}{\xi^d(t)}\xi^{d-1}(t) \sim \frac{N}{\xi(t)} \sim \frac{N}{t^\gamma} \, . \tag{3.3}$$

Similarly, the "failure rate" for word matching, $1-S(t)$, (where $S(t)$ is the success rate) scales as the fraction of nodes at the interfaces separating domains with different words

$$1 - S(t) \sim \frac{1}{\xi(t)} \sim \frac{1}{t^\gamma} \, . \tag{3.4}$$

The main feature of the above power-law decays (up to some system-size dependent cut-offs) is that the number of different words N_d, the total number of words N_w, and the success rate $S(t)$ only depend on t through the characteristic length scale $\xi(t)$. Further, for the typical time t_c to reach global agreement or consensus, one has $\xi^d(t_c) \sim N$, i.e.,

$$t_c \sim N^{1/(d\gamma)} \, . \tag{3.5}$$

Unless noted otherwise (as in section 3.2.4.1 and 3.2.4.2), our notation, t_c, $N_d(t)$, $N_w(t)$, $C(t)$, and $S(t)$ refer to the *ensemble-averaged* values of these relevant observables.

3.2.4 Simulation Results

Relevant quantities measured in the simulations are the total number of words in the system $N_w(t)$ (corresponding to the total memory used by the agents for word allocation at time t), the number of different words $N_d(t)$, and the average size of domains/clusters $C(t)$. Figure 3.3 displays the time evolution of these three quantities for the RGG, compared to the fully connected (FC) and to the $2d$ regular networks. Here, for the comparison, we reproduced the corresponding data of Refs. [15, 14]. The behavior of the NG on RGG is qualitatively very similar to that of the NG on $2d$ regular graphs. After time of $\mathcal{O}(1)$, $\mathcal{O}(N/(\overline{k}+1))$ different words have been invented [Fig. 3.3(b)]. $N_w(t)$ also reaches its maximum in time of $\mathcal{O}(1)$ [Fig. 3.3(a)]. Focusing on the late-time behavior of the systems, plotting $N_w(t)/N-1$, $N_d(t)/N$, and $C(t)$ vs t on log-log scales, confirms the power-law decays associated with the underlying coarsening dynamics, predicted by Eqs. (3.3), (3.2), and (3.1), respec-

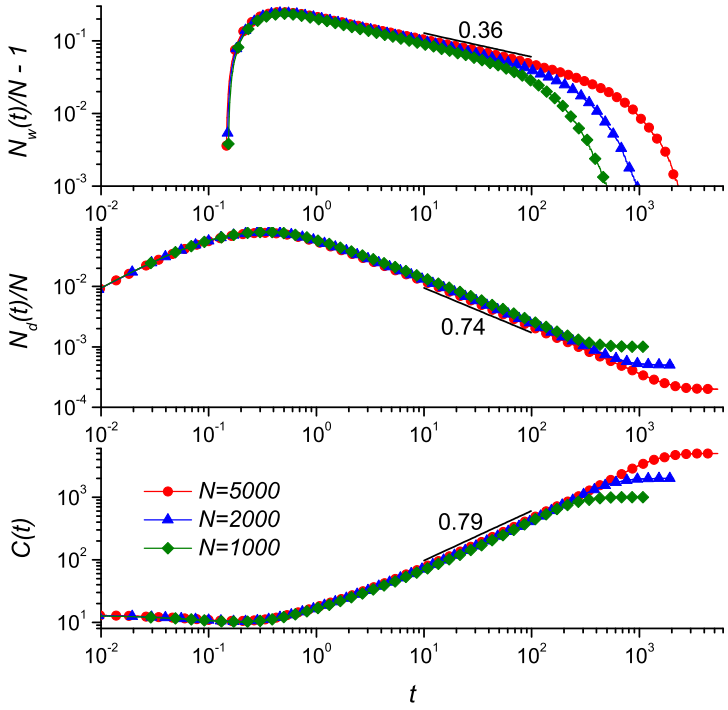


Figure 3.4: Time evolution of the relevant observables in the Naming Game in random geometric networks (RGG) for three system sizes on log-log scales, averaged over $1,000$ independent realizations; The average degree of the underlying RGGs is $\overline{k}\approx 12$.(a) the normalized total number of words in the system $N_w(t)/N-1$; (b) the normalized number of *different* words $N_d(t)/N$; (c) the average domain size $C(t)$. The straight line segments correspond to the best-fit power-law decays $N_w(t)/N-1\sim t^{-0.36}$, $N_d(t)/N\sim t^{-0.74}$, $C(t)\sim t^{0.79}$ for (a), (b), and (c), respectively.

tively [Fig. 3.4]. From the data for $C(t)$, we obtain $2\gamma=0.79\pm0.01$ [Fig. 3.4(c)], while from the data for $N_d(t)$ and $N_w(t)$, we extract $2\gamma=0.74\pm0.01$ and $\gamma=0.36\pm0.01$, respectively [Figs. 3.4(b,a)]. Based on our finite-size results, we can only conclude that the coarsening exponent is in the range $0.35<\gamma<0.40$ for the NG on two dimensional RGG. Different exponent values extracted from different observables for *finite* systems long hindered the precise determination of the coarsening exponent in the closely related large-Q Potts model [73, 89, 130]. There, employing advanced Monte

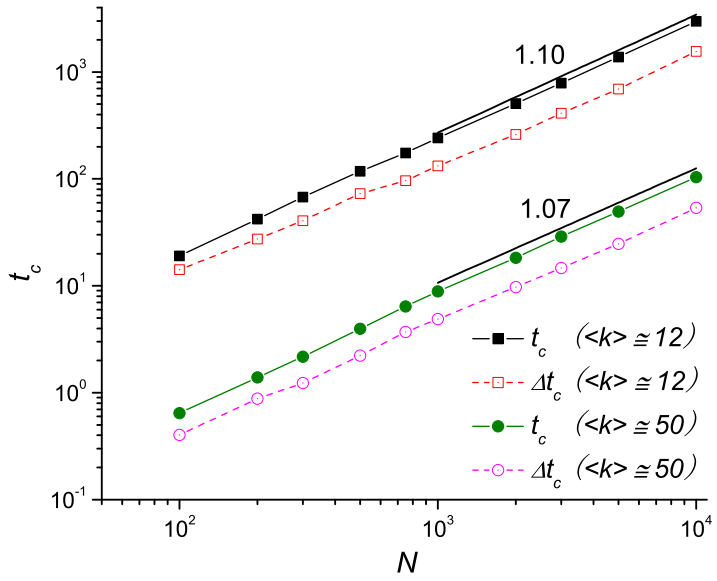


Figure 3.5: Average and the standard deviation of the convergence time t_c until global agreement is reached, as a function of the number of nodes on log-log scales, averaged over $1,000$ independent realizations of the NG on RGG. The average degree of the underlying RGGs is $\overline{k}\approx 12$ (squares) and $\overline{k}\approx 50$ (circles). The straight lines correspond to the best-fit power-laws with exponents 1.10, for both the average (solid squares) and the standard deviation (open squares) of RGGs with $\overline{k}\approx 12$, and 1.07 for those of RGGs with $\overline{k}\approx 50$, respectively.

Carlo renormalization (MCRG) schemes, it was shown that the coarsening exponent (within error) is 1/2 [130]. However, finite-size effects and very strong transients, in part due to "soft domain walls" and domain-wall intersections ("vertices") can produce values significantly smaller than 1/2 extracted from standard MC methods [73, 89, 130], such as ours.

Measuring the time to global agreement, averaged over $1,000$ independent runs (each on a different RGG network realization), we also obtained the scaling behavior of the agreement time, $t_c \sim N^{1.10\pm 0.01}$ and $t_c \sim N^{1.07\pm 0.02}$ for $\overline{k}\approx 12$ and $\overline{k}\approx 50$, respectively, as shown in Fig. 3.5. The corresponding scaling exponents both somewhat deviate from the one predicted by Eq. (3.5) with the exponent $1/(2\gamma)$. This

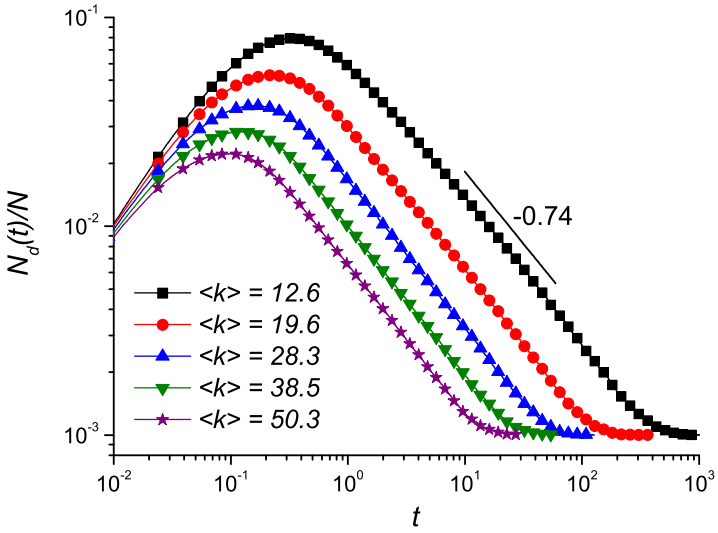


Figure 3.6: Time evolution of the (scaled) number of different words starting from an "empty" word list initial condition, for various average degree $\overline{k}$ on log-log scales. The number of nodes is N=1,000. The straight line segment indicates the asymptotic power-law decay as determined earlier [Figs. 3.4(b)], independent of the neighborhood size $\overline{k}$.

deviation is possibly due to strong finite-size effects, dominating the very late stage of the agreement dynamics.

For RGGs with many nodes, a relevant control parameter is the average number of neighbors (or average degree) $\overline{k}$. For sensor-network-specific implementations, as noted earlier, $\overline{k}$ can be adjusted by increasing either the density or the communication range of the nodes. We performed simulations of the NG for different average neighborhood size $\overline{k}$, as shown in Fig. 3.6. The results indicate that the scaling properties (in terms of N) of the time evolution of the agreement process do not change. The typical convergence times, however, are significantly reduced by increasing the neighborhood size. A closer examination of the convergence time reveals that, for fixed N, it scales as $t_c \sim \overline{k}^{-2.6}$, in the sparse-network limit ($\overline{k} \ll N$) in two-dimensional RGGs.

3.2.4.1 Agreement-time distributions

In addition to the average agreement (or convergence) time $\langle t_c \rangle$ (time until global agreement is reached), we also measured the standard deviation Δt_c [Fig. 3.5], and constructed the probability density (normalized histograms) $P(t_c, N)$ for N nodes [Fig. 3.7]. Since in this subsection, we analyze the full probability density of this observable, we use brackets for denoting the ensemble-averaged value of the convergence time, $\langle t_c \rangle$, while t_c alone denotes the stochastic variable, corresponding to a measurement in a single realization of the NG.

Up to the system sizes we could simulate, the standard deviation, within error, scales in the same fashion with the number of nodes as the average itself, $\Delta t_c \sim N^{1.10}$ ($\overline{k} \approx 12$) and $\Delta t_c \sim N^{1.07}$ ($\overline{k} \approx 50$) [Fig. 3.5]. [Suppressing large average convergence times and the corresponding large standard deviations (through modifying the network communication topology) will be addressed in the next section.]

Further, the *shape* of the histograms, for sufficiently large systems, remains invariant [Fig. 3.7]. Thus, introducing the scaled convergence time $x = t_c / \langle t_c(N) \rangle$, the corresponding scaled probability densities $p(x)$ for different system sizes collapse onto the same curve. [Fig. 3.7(b)]. The above findings indicates that the convergence-time distribution for the NG is governed by a *single* scale $\langle t_c \rangle$, hence can be written as

$$P(t_c, N) = \frac{1}{\langle t_c(N) \rangle} p(t_c / \langle t_c(N) \rangle) \; . \tag{3.6}$$

The distributions exhibit exponential tails for large arguments [Fig. 3.7(b) inset], a characteristic feature of opinion dynamics governed by coarsening [85, 158].

3.2.4.2 Cluster-size distributions

We also studied the probability distribution of the sizes of the clusters during the agreement dynamics $P(C, t)$ (the normalized histogram of the sizes of domains with different words at a given time) [Fig. 3.8(a)]. Similar to the previous subsection, we analyze the full probability density of this observable, hence we use brackets for denoting the ensemble-averaged value of the cluster size in the system at time t, $\langle C(t) \rangle$, while C alone denotes the stochastic cluster-size variable (sampled at an instant t in a single realization of the NG).

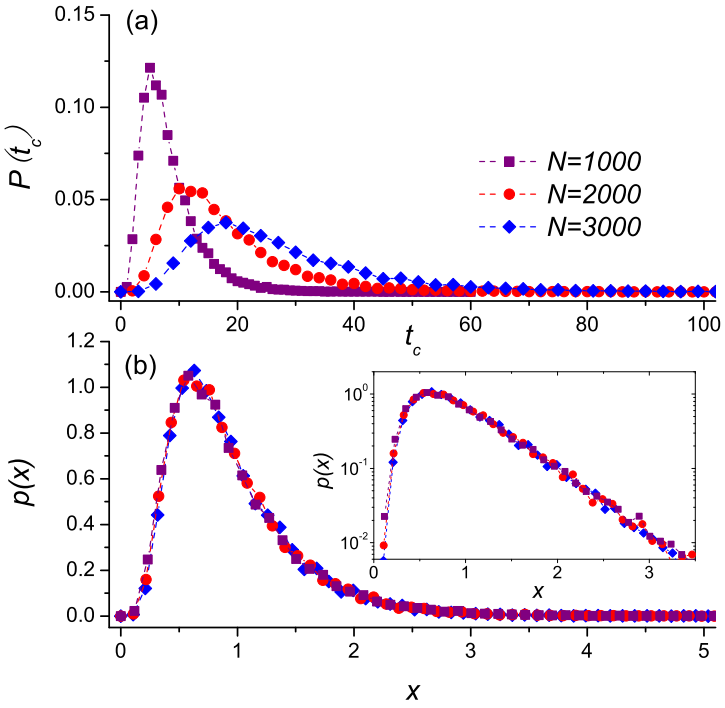


Figure 3.7: (a) Probability densities of the convergence time for three systems sizes. Data are gathered from $100,000$ independent realizations of the NG on RGG. The average degree of the underlying RGGs is $\overline{k}\approx 50$. (b) Probability densities for the scaled variable $x = t_c/\langle t_c(N)\rangle$ for the same data. The inset shows the same scaled histograms on log-lin scales.

Since the agreement process is governed by coarsening, one expects that this distribution exhibits dynamic scaling, i.e.,

$$P(C,t) = \frac{1}{\langle C(t)\rangle} p(C/\langle C(t)\rangle) \; . \tag{3.7}$$

Thus, $p(x)$, the distribution of the scaled cluster sizes $x = C/\langle C(t)\rangle$ remains invariant for different times. Our simulations confirm this picture, except for very early times (growth phase with initial domains forming) and for very late times (where finite-size effects dominate) [Fig. 3.8(b)]. The cluster-size distribution exhibit exponential-like tails for large arguments [47, 48, 86, 149], as can be seen in

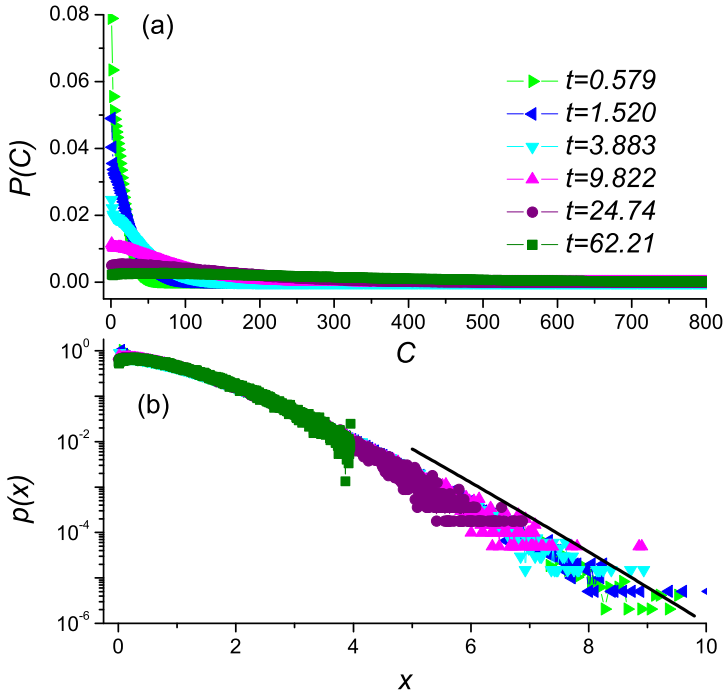


Figure 3.8: (a) Probability densities of the cluster size at different time of the agreement dynamics. Data are collected through $100,000$ independent realizations of the NG on RGG. The system size of the underlying RGGs is $N = 1000$ and the average degree is $\overline{k}\approx 12$. (b) Probability densities for the scaled variable $x = C/\langle C(t)\rangle$ for the same data as in (a) on log-lin scales.

Fig. 3.8(b).

3.3 Naming Games in Small-World-Connected Random Geometric Networks

In light of recent results on NG on one-dimensional SW networks [41], we now consider accelerating the agreement process by adding random long-range communication links between a small fraction of nodes of the RGG. Such networks have long been known to speed up the spread of local information to global scales [154, 153, 115, 82], with applications ranging from synchronization problems in distributed computing [79] to alarm-detection schemes in wireless sensor networks

[101]. For sensor networks, this can be implemented either by *adding* a small fraction of sensors equipped with long-range unidirectional antennas ("physical" long-range connections) or by establishing designated multi-hop transmission patterns ("logical" long-range connections) between certain nodes [64].

We construct the small-world-like RGG (SW RGG) as follows. We start with the original RGG (embedded in d dimensions, where d=2 in this paper). Then we *add* "long-range" links (or "shortcuts") between randomly chosen nodes in such a way that the total number of long-range links per node (the density of random links) is p. This SW construction differs slightly from the original Watts-Strogatz one [154] (also used by Dall'Asta et al. [41]), where random links are introduced by "rewiring" some of the original connections. The resulting network, however, has the same universal properties in the small-p, large-N limit [119], which is the center of our interest. Further, it is also motivated by actual implementations in sensor networks, where long-range "channels" are established in addition to the existing local ones.

3.3.1 Basic Scaling Considerations

Before presenting simulation results, using scaling arguments, one can obtain an order of magnitude estimate for the crossover time $t_\times$ present in the SW RGG and for the time to reach global agreement t_c [41]. In SW networks, embedded in d dimensions, the typical (Euclidean) distance between nodes with shortcuts scales as $l_{SW}{\sim}p^{-1/d}$ [119, 17, 18]. Starting from empty initial word lists word, for early times (following the creation of $\mathcal{O}(N/(\overline{k}+1))$) different words in the system), the system will exhibit coarsening, until the typical linear size of the growing domains, $\xi(t){\sim}t^\gamma$, becomes comparable to l_{SW}. (Here, both lengthscale measures are understood in terms of the underlying Euclidean metric.) After that time, the agreement process is governed by the presence of random long-range connections, yielding mean-field-like behavior. Hence the crossover from d-dimensional coarsening to mean-field-like dynamics occurs when $t^\gamma{\sim}p^{-1/d}$, yielding

$$t_\times \sim p^{-1/(d\gamma)} \ . \tag{3.8}$$

In a system of N agents, the above crossover is only displayed if the convergence time of the original system with no random links would exceed the above crossover time $N^{1/d\gamma} \gg p^{-1/d\gamma}$, which is equivalent to the condition for the onset of the SW effect $N \gg p^{-1}$ [119, 41]. Following the above system-size independent crossover time, the agreement dynamics is of mean-field like, and one can expect to observe a scaling behavior closer to that of FC networks [15]. In particular, the time to reach global agreement is expected to scale as [41]

$$t_c \sim N^{1/2} \, , \tag{3.9}$$

a significant reduction compared to that of the "pure" RGG with no long-range links where $t_c \sim N^{1.1}$.

3.3.2 Simulation Results

Simulating the NG on SW RGGs qualitatively confirms the above scaling scenario. Following the very early-time development of $\mathcal{O}(N/(\overline{k}+1))$ different words, the system of SW-networked agents, exhibits slow coarsening, with only small corrections to the behavior of the pure RGG [Fig. 3.9]. In fact, this early-time coarsening on SW RGGs is slightly slower compared to pure RGGs due to the effective pinning of interfaces near the shortcuts [41, 24, 31, 29]. In the NG on SW networks, however, the agreement process only slows down [41], but is not halted by "frozen" (metastable) disordered configurations [24, 29]. After a p-dependent crossover time [Eq. (3.8)], (when the typical size of the growing clusters becomes comparable to the SW length scale), an exponential convergence begins to govern the agreement process. This final-stage fast approach toward consensus sets in earlier for increasing values of the density of shortcuts p, yielding a significantly reduced convergence time compared to that of the NG on the "pure" RGG. The temporal behavior of the relevant observables for various values of p can be observed in Fig. 3.9.

Plotting the convergence time vs the density of long-range links, as shown in Fig. 3.10(a), suggests that (for sufficiently large but *fixed* N) the convergence time approaches an asymptotic power-law $t_c \sim p^{-s}$ with $s=0.79\pm0.01$ [41]. On the other hand, for fixed p and increasing N, the convergence time increases with N, $t_c \sim N^{\alpha_{\rm SW}}$,

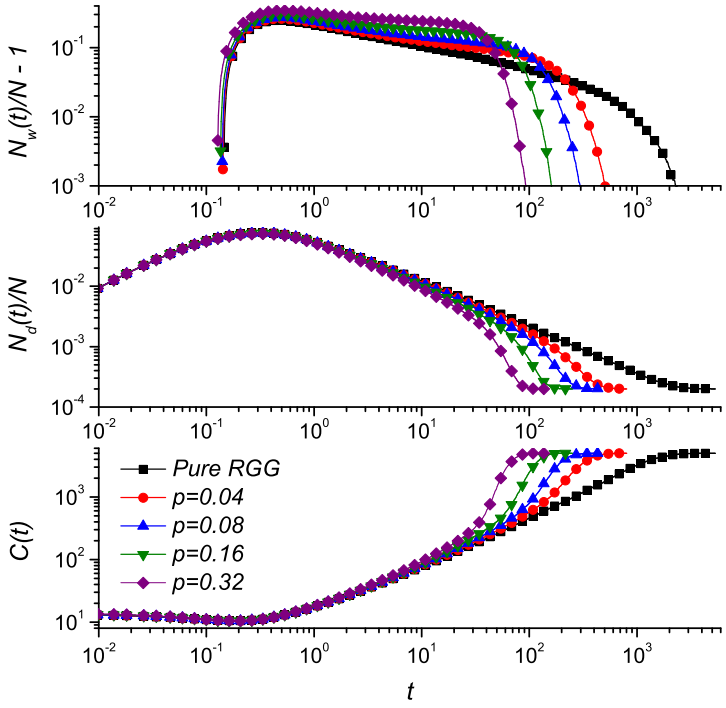


Figure 3.9: Time evolution of the (scaled) (a) total number of words, (b) number of different words and (c) the average cluster size for SW RGGs on log-log scales, starting from an "empty" word list initial condition, for various density of long-range links p, averaged over $1,000$ independent realizations of the NG on RGG. The number of nodes is N=$5,000$ with average degree $\overline{k}$≈12.

with $\alpha_{\rm SW}$=0.31±0.01 [Fig. 3.10(b)]. The agreement process is much faster than on a two-dimensional regular grid or RGG and is closer to the anticipated mean-field-like behavior [Eq. (3.9)] [41]. Thus, in the small-world regime ($Np \gg 1$) the convergence time depends on both the system size and density of random links, $t_c \sim N^{\alpha_{\rm SW}}/p^s$.

3.3.2.1 Finite-size scaling for the agreement time on SW-RGGs

In the pure-RGG limit ($Np \ll 1$), t_c only depends on N, $t_c \sim N^{\alpha_{\rm RGG}}$ with $\alpha_{\rm RGG}$≈1.10 [Fig. 3.10(b)] (since, essentially there are no shortcuts in the system). On the other hand, as seen above, in the SW-regime ($Np \gg 1$), the agreement time scales as

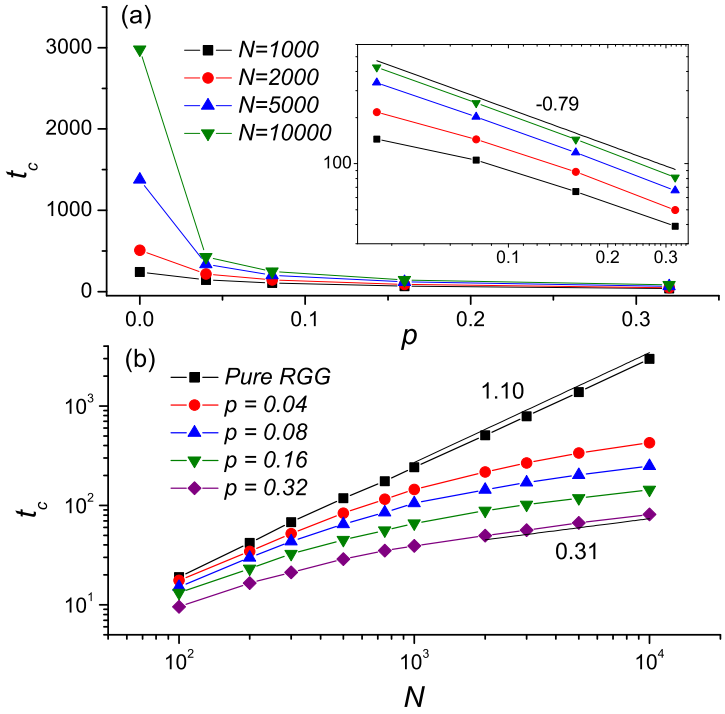


Figure 3.10: Average convergence time t_c for SW RGGs. (a) as a function of the density of shortcuts for various system sizes. The inset shows the same data on log-log scales. The straight lines corresponds to an estimate of the associated (asymptotic) power-law. (b) as a function of the number of nodes on log-log scales for various density of long-range links p. The curves shown are obtained by averaging over $1,000$ independent realizations of of the NG on RGG. The average degree of the underlying RGGs is $\overline{k}\approx 12$.

$t_c \sim N^{\alpha_{\rm SW}}/p^s$. One then can construct the full scaling behavior of $t_c(p,N)$, capturing the above two finite-size behaviors as limiting cases on SW-connected RGGs,

$$t_c(p,N) \sim \frac{N^{\alpha_{\rm SW}}}{p^s} f(Np) \ , \tag{3.10}$$

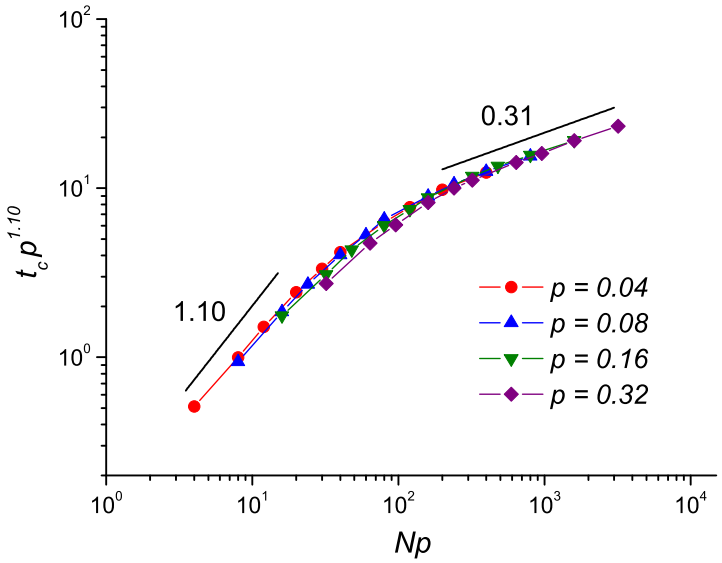


Figure 3.11: Scaled plot of data shown in Fig. 3.10, as suggested by the finite-size scaling argument [Eq. (3.14)]. The straight line segments correspond to the best-fit (asymptotic) power-law behavior of the scaling function $g(x)$ with exponents 1.10 and 0.31, for small and large arguments, respectively, as described in the text [Eq. (3.15)].

where $f(x)$ is a scaling function such that

$$f(x) \sim \begin{cases} x^s & \text{if } x \ll 1 \\ \text{const.} & \text{if } x \gg 1 \end{cases} . \tag{3.11}$$

The pure RGG limit ($Np \ll 1$) is recovered, provided that

$$t_c \sim (N^{\alpha_{\rm SW}}/p^s)(Np)^s \sim N^{\alpha_{\rm SW}+s} \sim N^{\alpha_{\rm RGG}}, \tag{3.12}$$

i.e.,

$$\alpha_{\rm RGG} = \alpha_{\rm SW} + s \, . \tag{3.13}$$

Our measured "phenomenological" exponents $\alpha_{\rm RGG} \approx 1.10$, $\alpha_{\rm SW} \approx 0.31$, and $s \approx 0.79$, satisfy the above proposed asymptotic scaling relation. For analyzing our data,

Eq. (3.10) can also be rewritten as

$$t_c(p,N) \sim \frac{(Np)^{\alpha_{\rm SW}}}{p^{s+\alpha_{\rm SW}}} f(Np) \sim \frac{1}{p^{\alpha_{\rm RGG}}} g(Np) \ , \tag{3.14}$$

where $g(x) = x^{\alpha_{\rm SW}} f(x)$. Thus, plotting $t_c p^{\alpha_{\rm RGG}}$ vs Np should yield data collapse, together with the asymptotic small- and large-argument exponents of $g(x)$, $\alpha_{\rm RGG}$ and $\alpha_{\rm SW}$, respectively [Fig. 3.11],

$$g(x) \sim \begin{cases} x^{\alpha_{\rm RGG}} & \text{if } x \ll 1 \\ x^{\alpha_{\rm SW}} & \text{if } x \gg 1 \end{cases} \ . \tag{3.15}$$

CHAPTER 4
The Naming Game in Social Networks: Community Formation and Consensus Engineering

Agent-based models and simulations provide invaluable frameworks and tools to gain insight into the collective behavior of social systems [52, 35, 7]. Opinion spreading and social dynamics [51, 28] on regular and random networks are examples of the latter. A large number of studies have investigated models of opinion dynamics [29, 19, 46, 62, 99, 85, 140, 146, 80, 21, 9, 70] and the dissemination of culture [10, 134, 108], while fundamental models for residential and ethnic segregation have also attracted strong interest [135, 160, 151, 97]. Most recently, researchers have also turned their focus to models where both the network topology and opinions change over time [80, 81]. With the availability of empirical data sets and cheap and efficient computing resources, one can implement stylized socio-economic models on empirical social networks, and evolve "artificial societies" [52] to study the collective properties of these systems.

Here, we continue to focus on the stylized model, the Naming Game (NG) [15]. By employing local communications, the NG is regarded as a minimal model which can capture generic and essential features of an agreement process in a wide range of complex systems, For example, in the context of a group of robots (the original application), the NG dynamics mimics the emergence of shared communication schemes (synthetic languages), while in the context of sensor networks, as we discussed in chapter 3, such an agreement process can describe to the emergence of a shared key for encrypted communications. In a system of human agents, the NG can be considered as a minimal model to describe the recent phenomenon of collaborative tagging or social bookmarking [33, 32, 58] on popular web portals like like Del.icio.us (http://del.icio.us), Flickr (www.flickr.com), CiteULike (www.citeulike.org), and Connotea (www.connotea.org). Another common example is the evolution and spread of coexisting dialects in everyday use (see, e.g., the geographical distribution of "Pop" vs "Soda" for soft drinks in the US [109]). In a broader context, the

NG can be employed to investigate the emergence of large-scale population-level patterns arising from empirically supported local interaction rules between individuals.

The common feature in the above examples and applications is that global agreement can emerge spontaneously (without global enforcement) purely as a result of local (e.g., pairwise agent-to agent) communications. The NG has been studied intensively on regular and random complex network models (see next Section). Here we investigate the evolution of the agreement process in the NG on empirical social graphs. It is well known that empirical social graphs exhibit strong community structure [57, 118, 122, 124, 123]. It is also known that in networks with community structure, reaching global agreement can be hindered [91, 27]. Here, we investigate the NG precisely from this viewpoint. Further, we analyze strategies to destabilize otherwise indefinitely coexisting clusters of opinions, to reach global consensus of a selected opinion. The later can also be considered as an abstract agent-based marketing approach.

In the following sections we present results for the NG on empirical social networks. In particular, we investigate the effect of communities in the underlying static social graphs on the agreement process (typically leading to indefinitely coexisting clusters of opinions). We are also interested in studying and analyzing node-selection strategies to facilitate the convergence to a global opinion.

4.1 Background, Model, and Prior Results on the Naming Game on Regular and Complex Network Models

In the simplified version of the NG, agents perform local communications in order to reach agreement on the name to assign to a *single* object. The model of NG has been introduced in chapter 3 in detail. Here we implement *pairwise* communications between agents, instead of local broadcast, i.e., each local interaction only involves *one* "speaker" and *one* "listener", as in the original model of Baronchelli et al [15]. In most cases, we considered initial conditions when all agents have an empty vocabulary. Then such an agent, when chosen to be a speaker, invents a *random* word to be transmitted to the listener. But we also look at initial conditions

with a single random word per agent.

This version of the NG has been studied on complete graphs (CGs), on regular, on small-world (SW), and on scale-free (SF) networks. In terms of the number of different words, the evolution of the game will go through stages of growth (due to unsuccessful communications) and stages of elimination (due to successful ones). In all of the above mentioned networks, starting from empty vocabularies, an early time explosion of words is followed by a slow elimination of all synonyms, except one; that is, agents come to global agreement on the naming of the object in question. In chapter 3 we also investigated the NG model on homogeneous random geometric graphs (RGGs). We found that the NG on RGG with homogeneous node density also leads to global consensus, facilitating an application to autonomous key creation for encrypted communication in a community of sensor nodes[102, 103].

It was found that the NG dynamics on the above networks will lead to global consensus among all agents, i.e., after some time, agents' vocabularies eventually converge to a unique word, the same for all agents [15, 13, 14, 41]. On the other hand, [42] found that on stylized network models with community structure (composed of fully connected cliques with a single link between cliques) the evolution of the NG runs into long-living meta-stable configurations, corresponding to different co-existing words (different for each clique). Here we study precisely this later scenario by implementing the NG on static empirical social graphs.

In passing we note that the issue of the emergence of meta-stable or frozen opinion clusters and fostering consensus have been discussed for models of opinion formation under bounded confidence [46, 62, 99]. In those models, however, community formation or opinion segmentation is the result of the agents' interaction being limited by bounded confidence: an agent can gradually adjust her opinion toward another one's only if their opinions were already sufficiently close to one another to begin with. As a result, opinion segmentation can emerge in networks with no community structure with low-confidence agents. In contrast, the NG dynamics does not require that agents' opinions are sufficiently close in order to potentially interact (i.e., their confidence is unbounded), and as mentioned earlier, the NG dynamics always lead to global consensus on networks *without* community structure.

Our motivation here, by studying the NG on empirical social graphs, is to directly study how the community structure of the underlying graphs affects the emergence of meta-stable or long-living opinion clusters.

4.2 The Naming Game on Empirical Social Networks

One of the most important feature of social graphs is their modularity: these networks typically consist of a number of communities; nodes within communities are more densely connected, while links bridging communities are sparse. Since the community structure of empirical networks is often not known a priori, detecting communities in large networks itself is a difficult problem [124]. A number of current methods for finding community structures utilize various forms of hierarchical clustering, spectral bisection methods [137, 117, 157], and iterative high-betweenness edge removal [118, 116, 57]. A different approach involves searching for the ground-states of generalized multi-state spin models (corresponding to different opinions) on these networks, such as the q-state Potts model [22, 127, 90, 55]. Also, recently a novel method has been developed to detect overlapping communities in complex networks [124].

The NG, as summarized in the chapter 3, in low-dimensional networks exhibits slow coarsening, while networks with small-world characteristic (small shortest path, such as in SW and SF networks) facilitate faster (and guaranteed) convergence to a global consensus among nodes. But in all cases, global consensus is reached, provided the network has no heterogeneous clustering or modularity (i.e., community structure).

4.2.1 Time Evolution of The Naming Game on High-school Friendship Networks

Here, we study the NG on networks which do exhibit strong community structure. The set of social networks (high-school friendship networks), on which we implemented the NG, were provided by the National Longitudinal Study of Adolescent Health (Add Health) [2]. The high-school friendship networks investigated

[2]This research uses the network-structure data sets from Add Health, a program project designed by J. Richard Udry, Peter S. Bearman, and Kathleen Mullan Harris, and funded by

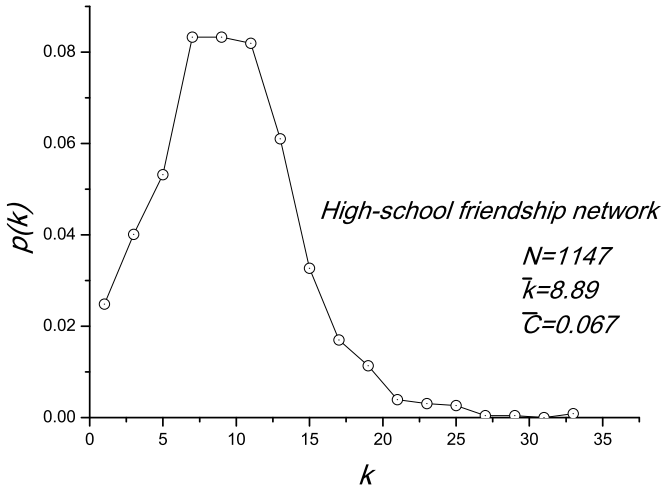


Figure 4.1: Degree distribution of a high-school friendship network. The network has total $N = 1147$ nodes (students) with average connectivity $\bar{k} = 8.86$ and clustering coefficient $\overline{C} = 0.067$.

here, were constructed from the results of a paper-and-pencil questionnaire in the AddHealth project [113]. Here, nodes represent students while the edges are for their mutual relations or friendships. Two students are considered to be friends (thus have a link between them) when one nominates the other as her/his friend and both of them participated in some activities, e.g., talked over the phone, spent the weekend together, etc., in the last seven days. (for this study, we considered the relationships reciprocal, and associated them with undirected links in the NG). These networks exhibit exponential degree distributions (no hubs), with an average degree of the order of 10. Fig. 4.1 shows a selected high-school friendship network with average connectivity $\overline{k} = 8.86$. For a baseline comparison we also constructed a Watts-Strogatz (WS) network [154] network with the same number of nodes N, average degree $\overline{k}$, and clustering coefficient $\overline{C}$ as the friendship network. The WS network has homogeneous clustering, hence, no community structure.

a grant P01-HD31921 from the National Institute of Child Health and Human Development, with cooperative funding from 17 other agencies. For data files contact Add Health, Carolina Population Center, 123 W. Franklin Street, Chapel Hill, NC 27516-2524, (addhealth@unc.edu, http://www.cpc.unc.edu/projects/addhealth/.

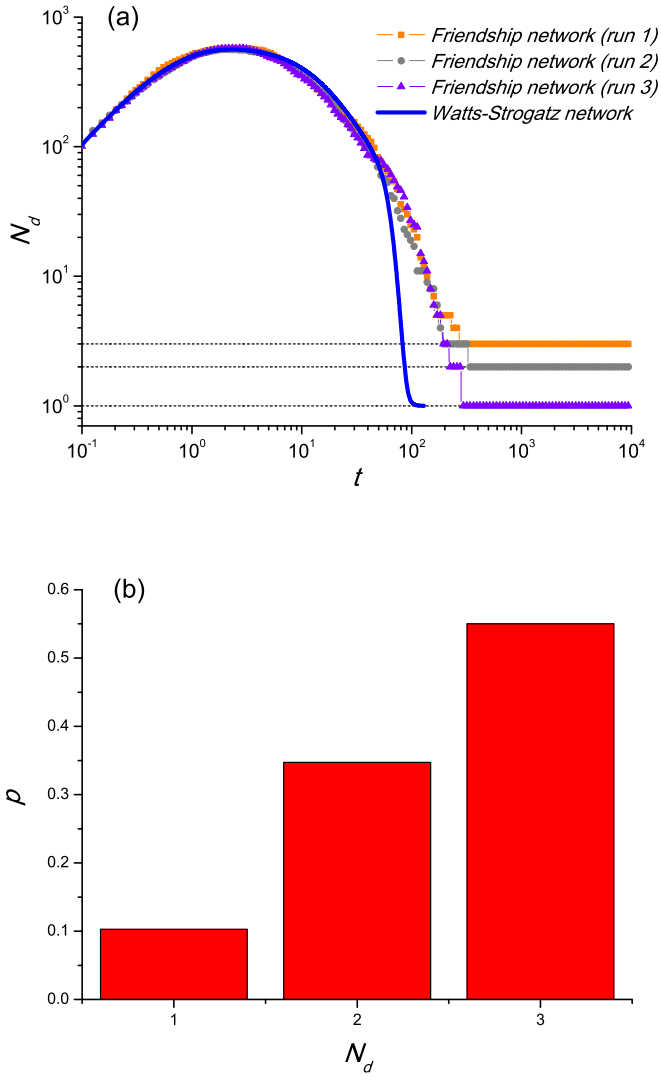


Figure 4.2: (a) Number of different words N_d vs time for a friendship network (thin lines) and for the Watts-Strogatz network (bold line). $N = 1,127$, $\overline{k} = 8.8$, and $\overline{C} = 0.067$ for both systems. Results for the WS network are averaged over 1000 independent realizations. For the high-school friendship network we show three individual realizations (thin lines), reaching different final states with $N_d = 1$, $N_d = 2$, and $N_d = 3$ (indicated with horizontal dashed lines). Note the log scales on both axes. (b) The probability (relative frequency) of final configurations with N_d different words (opinions) for the same high school friendship network as in (a) based on $10,000$ independent runs. Statistically, in this particular network, the most likely final configuration exhibits three opinions.

We selected a few networks with a large number of students (on the order of 1,000) from the available data set. Starting from an empty word list for all agents, both the friendship network and WS network show nearly identical early-time development of the number of different words N_d. However, the friendship-network simulations exhibit a long-time behavior very different from the ones discussed in earlier section, and also from the baseline reference, the NG on the WS network [Fig. 4.2(a)]. In the late stage of the NG, networked agents without community structure (including the WS network) *always* exhibit a spontaneous evolution toward a shared "dictionary" (or opinion), i.e., a global consensus is reached. In contrast, in the empirical high-school networks, consensus is rarely reached (for long but finite simulation times) [Fig. 4.2(a)]. For this particular high-school friendship network, performing $10,000$ independent runs of the NG with a fixed simulation time of $t = 10^4$ steps, 10%, 35%, and 55% of these runs, ended up with one, two, and three different words, respectively, in their final configurations [Fig. 4.2(b)]. Thus, in this network, the most likely (or typical) outcome of the NG is one with three different clusters of opinions. Snapshots taken from the typical evolution of the NG on this network are shown in Fig. 4.3. In analogy with domain formation in physical systems, we can regard these long-living configurations with coexisting multiple opinions as "meta-stable" ones.

The emergence of different long-living clustered opinions is not unexpected. In fact, the same high-school networks have been analyzed for community structures in a study of friendship segregation along racial lines among high-school students [113, 59]. For example, close to the final stage, the time-evolution of the NG on the particular network shown in Fig. 4.3(b) exhibits four communities. These four clusters of opinions correspond to segregation along the two-schools involved in the particular network, high-school (HS) – middle-school (MS) pair, and along racial lines, whites students – black students in each. Checking the race and school-grade attribute of the node information in the raw data, we confirmed that the four communities exhibited by the NG in Fig. 4.3(b) correspond to black HS (upper left), white HS (upper right), black MS (lower left), and white MS (lower right) students. Then, in the final state [Fig. 4.3(c)], only three communities remain;

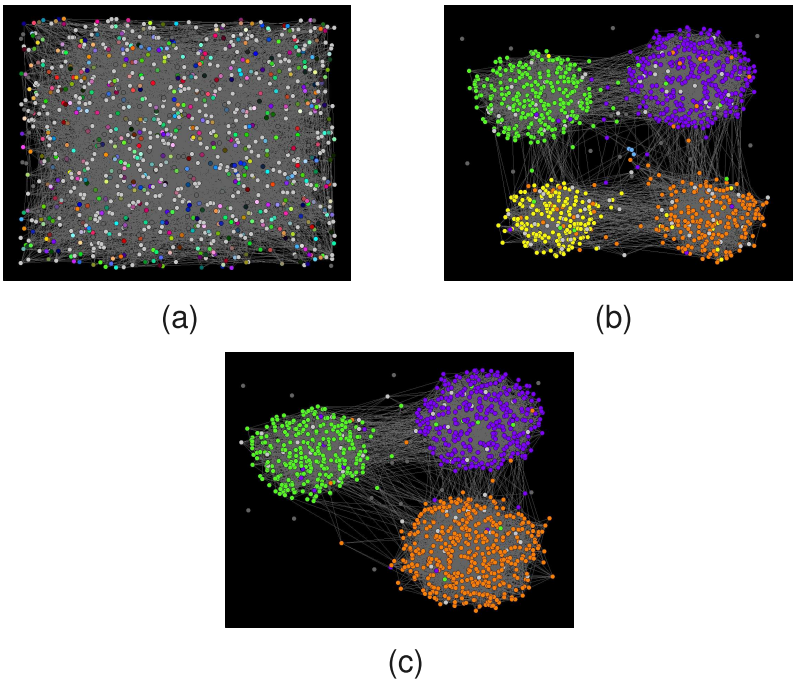


Figure 4.3: Snapshots of the time-evolution of the Naming Game on a high-school friendship network. Initially agents have an empty word list (no opinions). In the snapshots, different colors correspond to different words. In the very early stage of the game (a), "speakers" with no words has to create one randomly. After a slow but steady coarsening of opinions, in the final stages of the game, the system exhibit relatively long plateaus in the number of different opinions. The corresponding clusters, i.e., agents with the same opinions, can be regarded as communities. For the particular network shown here, in the next to final stage (b), the network exhibits four communities. Eventually, two of these communities coalesce, leading to a final configuration (c) with three communities.

opinions, segregated along the racial line coalesce in the middle-school portion of the students, simply indicating that racial segregation in friendships is weaker in this group, in this particular network set.

The data set of high-school friendship network from AddHealth is comprised of questionnaires from students of 84 schools, i.e., total 84 independent friendship networks consisting from 25 (the smallest) to 2250 (the largest) students. In order to show that the community structure and the existence of competing opinions in

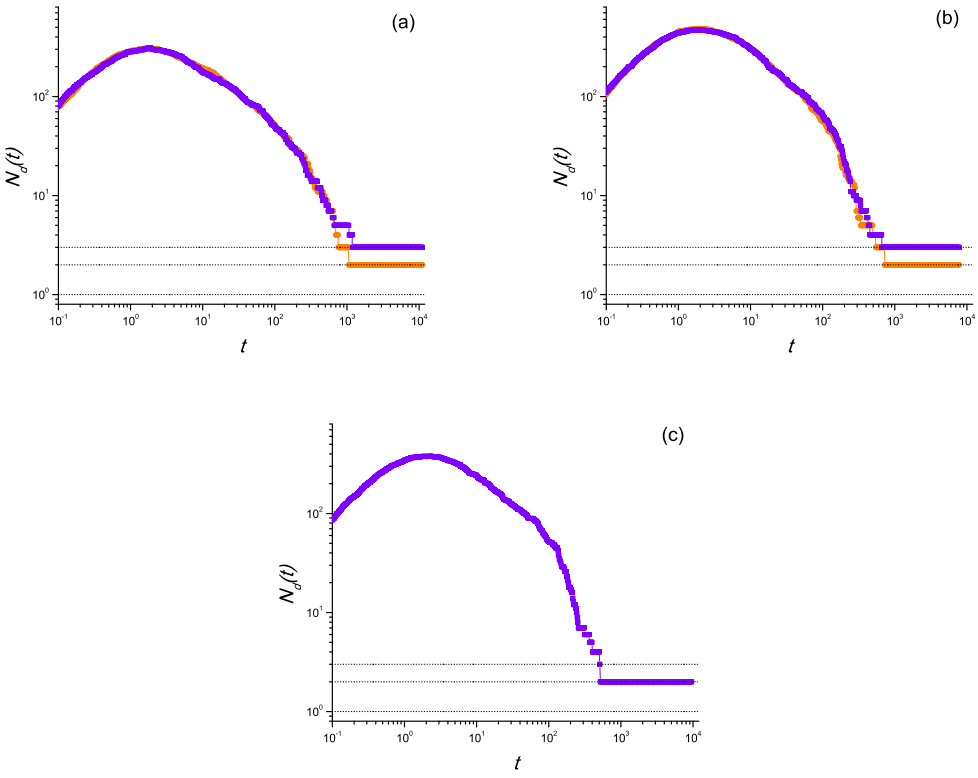


Figure 4.4: Number of different words N_d vs time for three different friendship networks. (a) Two individual runs on friendship network with $N = 900$, $\overline{k} = 3.66$, (b) two individual runs on friendship network with $N = 1,282$, $\overline{k} = 5.44$, and (c) one individual run on friendship network with $N = 1,047$, $\overline{k} = 5.26$. These runs reach different final stages with $N_d = 2$ or $N_d = 3$, respectively.

meta-stable state is not a special case on a particular network, we repeated the same simulations on several other friendship networks chosen randomly from the data set. Our results show it is not unusual to observe community structures in these networks [Fig. 4.4]. It further indicates that these networks are more likely to develop into small number of communities (usually two or three), if not reaching global consensus.

Admittedly, our objective is not to draw over-ambitious conclusions from a social science viewpoint. Instead, we are interested in how the evolution of the

NG (a stylized model for opinion formation) is affected by the community structure of the underlying graphs, such as the high-school friendship networks which are well-known to exhibit strong community structure [113, 59]. We demonstrated that the outcome of the NG, is strongly affected by the existence of communities in the underlying network. Conversely, at some coarse level, the long-living late-stage meta-stable clusters of words (opinions) reveal important aspects of the community structure of the underlying network. Thus, the NG, together with other stylized models for opinion formation, can not only be used as a tool to understand generic features of spontaneous agreement processes in a network of artificial or human agents, but can also be employed to extract relevant information on the community structure of complex networks [22, 127, 90, 55].

Because we address community structure from a dynamics-on-networks viewpoint over extended timescales, our results provide also the resolution of communities form fine to coarse as time evolves. The various stages through these scales are displayed through clearly identifiable plateaus, e.g., in the time series of the number of different opinions. These plateaus are essentially meta-stable configurations of several coexisting opinions/communities. So as we found, such a prototypical model for opinion formation is very sensitive to the underlying community structure, hence, we can employ it as a novel and sensitive tool to study community formation, structure, and the role of individuals to stablize or weaken the structure.

4.2.1.1 The Naming Game with One-word-per-agent Initial Conditions

Here, we show the behavior of the Naming Game (NG) with initial configurations where each agent has exactly one word (or opinion), different for each agent ($N_d(0) = N$). The rest of the rules of the NG are the same as described in above sections. Since each agent has a word initially, no new words will be invented, and there is no growth phase (N_d is monotonically decreasing as a function of time); pairwise communications will lead to the gradual elimination of existing opinions. At the end, only a few opinion clusters remain, again, reflecting the community structure of the underlying graphs. There is no significant difference in the *late-stage scaling behavior* between the empty-dictionary and the one-word-per-agent initial

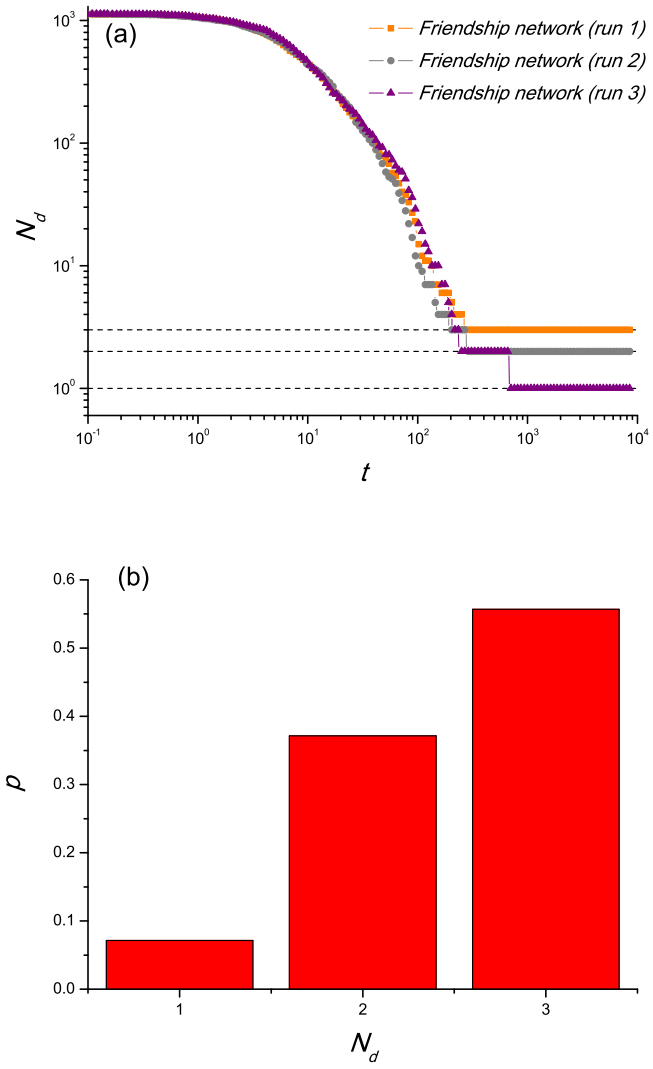


Figure 4.5: **The Naming Game on a friendship network where the simulations are initialized from configurations with a single word per agent (different for each agent). (a) Number of different words N_d vs time for a friendship network with $N = 1,127$, $\overline{k} = 8.8$, and $\overline{C} = 0.067$ (same friendship network as in Fig. 2 of the main article). Results are shown for three individual realizations, reaching different final states with $N_d = 1$, $N_d = 2$, and $N_d = 3$ (indicated with horizontal dashed lines). (b) The probability (relative frequency) of final configurations with N_d different words (opinions) for the same friendship network as in (a) based on $10,000$ independent runs.**

conditions: In the former case, it takes of $\mathcal{O}(1)$ time steps to reach the maximum of N_d, $N_d^{\max} \sim N/2$ [Fig. 4.2(a)], after which slow "opinion coarsening" begins. In the latter case, the number of different words initially starts from N and slowly begins to decay [Fig. 4.5(a).] The relative frequencies of the final configurations with one, two, and three words are very similar [Fig. 4.2(b) vs Fig. 4.5(b)], and the underlying opinion clusters (communities) exhibited this way are the same. Note that for the one-word-per-agent initial conditions, out of $10,000$ independent runs, we also recorded 3 runs with four surviving opinion clusters (N_d=4) after 10^4 time steps, corresponding to a 3×10^{-4} relative frequency. Since it is three orders of magnitudes smaller than the probability of other possible final configurations, it is not visible on the same scales and is not shown in Fig. 4.5(b).

4.2.1.2 The Naming Game with Finite-memory Agents

As was discussed by Dall'Asta et al [42] , the typical memory need (the maximum number of words in its list at any given time) of an agent of degree k is of $\sqrt{k}$. Limiting the agents' memory to a finite value L, in general, can slow down the consensus process [152]. For the particular friendship network we used here, $k_{\min} = 1$, $k_{\max} = 33$, and $\overline{k} = 8.8$. Here we considered a "first in - first out" finite-memory version of the NG: in case of an unsuccessful communication, if the memory of the listening agents is full, it drops the word from its list which has been there the longest, and adds the one just heard. All the other rules of the NG remain the same.

For comparison, here we show the agreement process in terms of number of different words (communities) vs time for both the Watts-Strogatz network and for the friendship network with the same number of nodes, average degree, and clustering coefficient [Fig. 4.6]. Since the average degree of these networks is rather small, there is very little variation in the late-stage agreement process and convergence times (the infinite-memory behavior is asymptotically approached as the memory length is of order $\sqrt{\overline{k}}$). The only exception is the L=1 case: here the listener simply replaces its current opinion with the one of the speaker's, hence the model becomes equivalent to the q-state voter model (with $q = \infty$) [67]. Unlike the NG with $L \geq 2$,

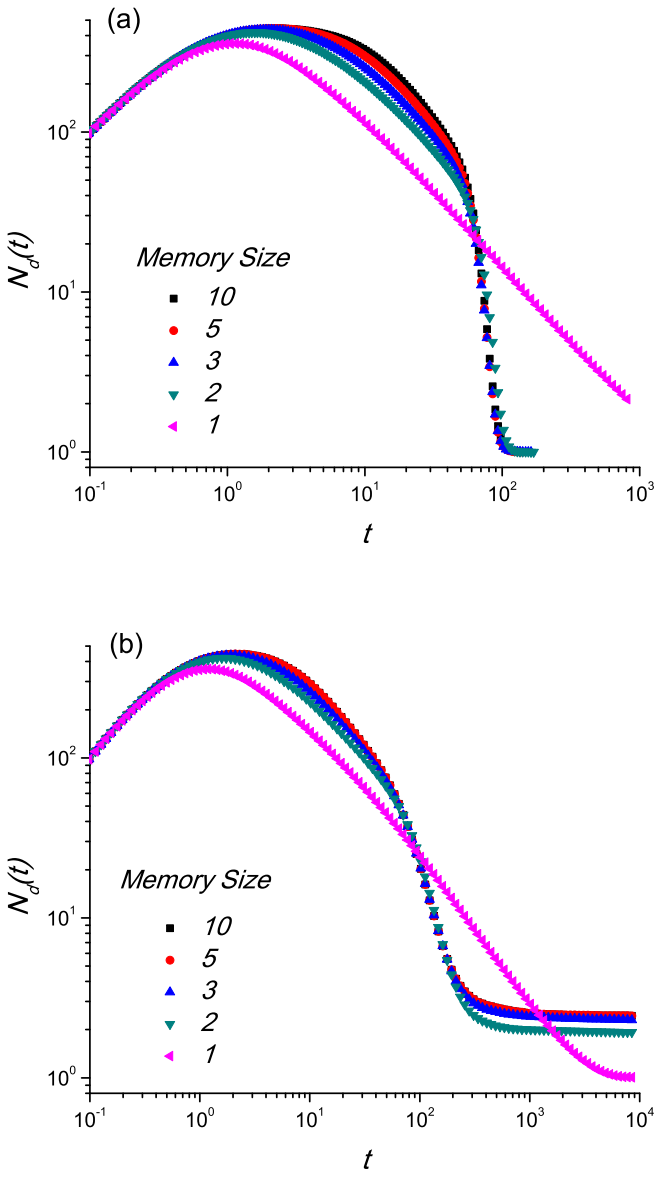


Figure 4.6: Naming Game with finite-memory agents. The number of different words vs time on (a) the Watts-Strogatz network; (b) a friendship network with the same number of nodes, average degree, and clustering coefficients ($N = 1,127$, $\overline{k} = 8.8$, and $\overline{C} = 0.067$). All curves represent averages over $10,000$ independent runs.

the voter model on random graphs has no tendency to form compact domains; the "interfaces" or boundaries separating domains stochastically disintegrate. While the voter model on finite networks eventually orders, it is the result of a large spontaneous fluctuation of ordered regions spanning the full system [31, 29]. Global consensus for the NG with L=1 (equivalent to the voter model) on networks with no community structure takes much longer than for $L \geq 2$ [Fig. 4.6(a)]. On networks with community structure (such as the friendship network), however, the tendency of the NG with $L \geq 2$ to form domains leads to the formation of compact long-living or meta-stable opinion clusters, hence global order is rarely reached. On the other hand, when L=1, there are no stable domain boundaries and the underlying community structure has little or no effect on reaching global consensus [Fig. 4.6(b)]. The NG with L=1 (voter model) orders on any finite network, essentially independently of the underlying community structure.

4.2.2 The NG on Large-scale Mobile Phone Network

We also investigated the evolution of the NG on another set of social network, a large-scale info-social graph (of the order of 4 million nodes), projected out of the call logs of a European mobile-phone provider [122]. Two nodes were connected with an undirected link if there had been at least one reciprocated pair of phone calls between them. The average connectivity of the cellular communication network is about $\overline{k} = 3.32$. However the scale-free degree distribution [Fig. 4.7] reveals there is a small fraction of "hub"-like individuals who approach massive amount of other people through the phone calls. The person that has the largest neighborhood size called 143 different people in the call log. Meanwhile, most of other people have very sporadic connections with others – they rarely talk with more than three persons on the phone. For a baseline reference, we constructed a randomized scale-free network of the same degree distribution and with the same number of nodes, by homogenously rewiring the majority of links of the original network. The simulation again shows the opinion segregation among people on the communication graph. At the same time, the randomized scale-free network, the NG always converges to a global consensus [Fig. 4.8].

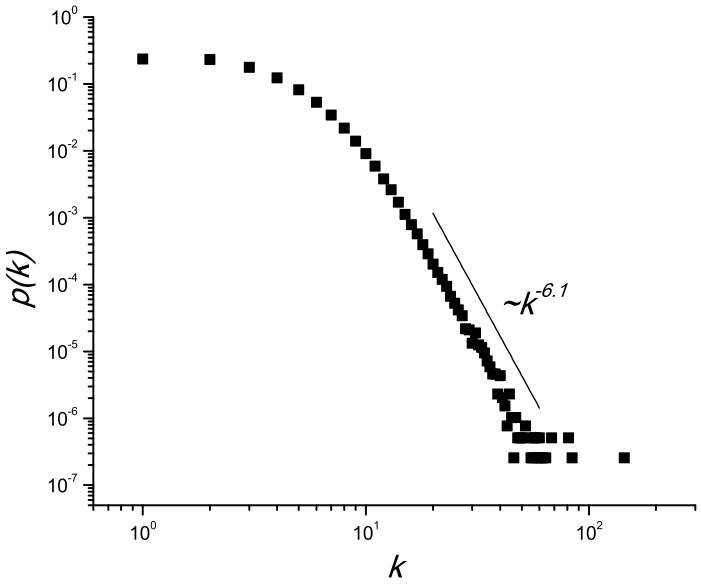


Figure 4.7: Degree distribution of mobile phone network. The network consists of $N = 3,906,441$ anonymous cell phone numbers (nodes). Edges connecting two numbers indicate there has at least one phone call between them found in the call log. The network has highly-skewed power-law degree distribution with $\alpha \sim 6.1$. The average neighborhood size is $\overline{k} = 3.32$, but the largest one has $k_{max} = 143$ neighbors. All nodes of the network are in one single connecting component, i.e., the size of the giant component equals to the network size.

4.2.3 The NG on Spatially Embedded Social Networks

The third social system we investigated is the geographic embedded population distribution data from LandScan Global Population Database [3]. It is a worldwide population database at grid cells of very fine resolutions, by distributing best available census counts to cells based on probability coefficients which, in turn, are based on road proximity (transportation networks, and in particular, the densities of roads), land slope (as most human settlements occur on flat to gently sloping terrain), land cover (from extremely low population density in desert, water, wet-

[3]LandScan has been developed as part of the Oak Ridge National Laboratory (ORNL) Global Population Project for estimating ambient populations at risk (http://www.ornl.gov/sci/landscan/).

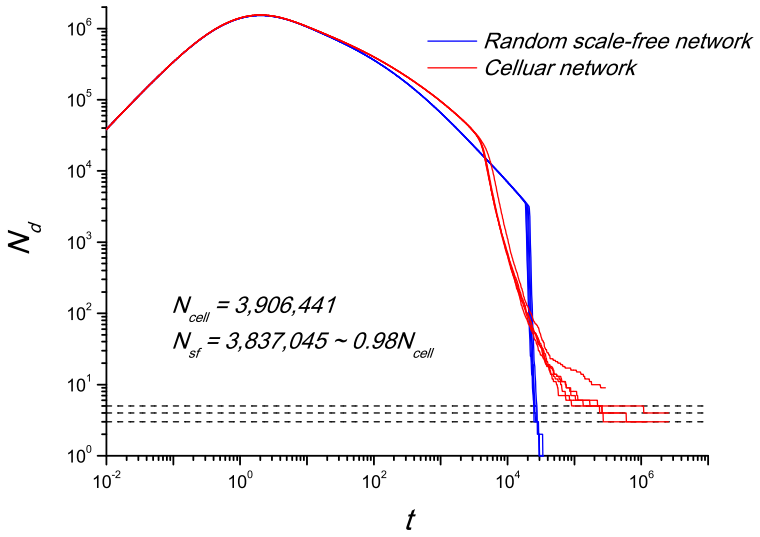


Figure 4.8: Time evolution of N_d for the mobile phone network and the random scale-free network with the same power-law degree distribution. Both networks have $\alpha \simeq 6.1$. Red lines are 8 individual runs of cellular network, and blue lines are 8 runs of random scale-free network. Three dash lines indicate positions of $N_d = 3$, $N_d = 4$, and $N_d = 5$ from bottom to top.

lands, etc., to high density in developed land cover), and nighttime lights. From the raw data of geographical population information, we construct a generalized version of random geometric graph to model a coarse-grained version of a social system, based on the idea that most social interactions occur within the limit of local geodesic distance. Nodes are randomly connected if they both fall inside the predefined neighborhood distance (the distance will be larger when the population density is low, and vice versa) of each other, until the designated average connectivity of about $\overline{k} \simeq 10$ [122] is reached [Fig. 4.9]. Due to the extreme fluctuation of population density, which can vary from over tens of thousands per square mile in metropolis region, to none in desert or wetland, the constructed network is a random spatial graph with strong inhomogeneous distribution of node density. Here the subset of U.S. population data is selected for the purpose of network construction. Population densities are much higher in east and west coast while leaving near

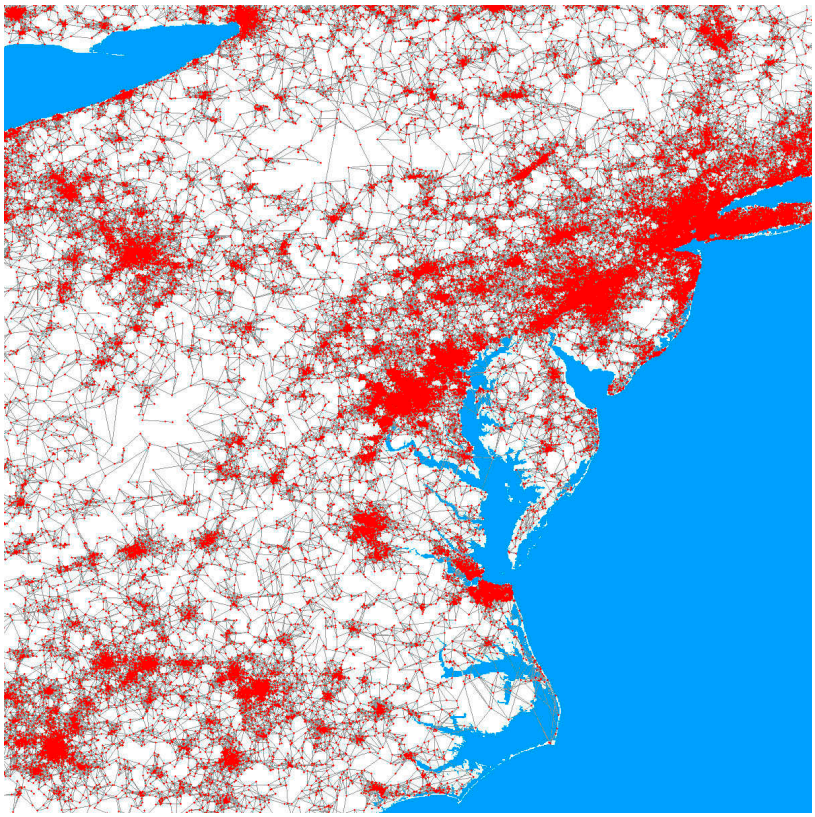


Figure 4.9: Random geometric graph constructed from North America geographical population data in LandScan global population data set. Population density are highly uneven hence people in city region usually have higher connectivity than those in rural areas.

vacuum in the mid-west – the inhomogeneity might lead to the spatial segregation of opinions or fads. The long-lasting of several coexisting opinions in the evolution of the NG somewhat illustrates the generic feature of opinion segregation at the global scales [Fig. 4.10], similar as in the geographical distribution of "Pop" versus "Soda" for soft drinks in the US [109].

4.3 Engineering Consensus in Social Networks

There are several ways to influence the outcome of social dynamics, e.g., to facilitate the outcome of a specific global opinion that one would prefer the system to achieve (preferred opinion for short). All methods essentially rely on "breaking

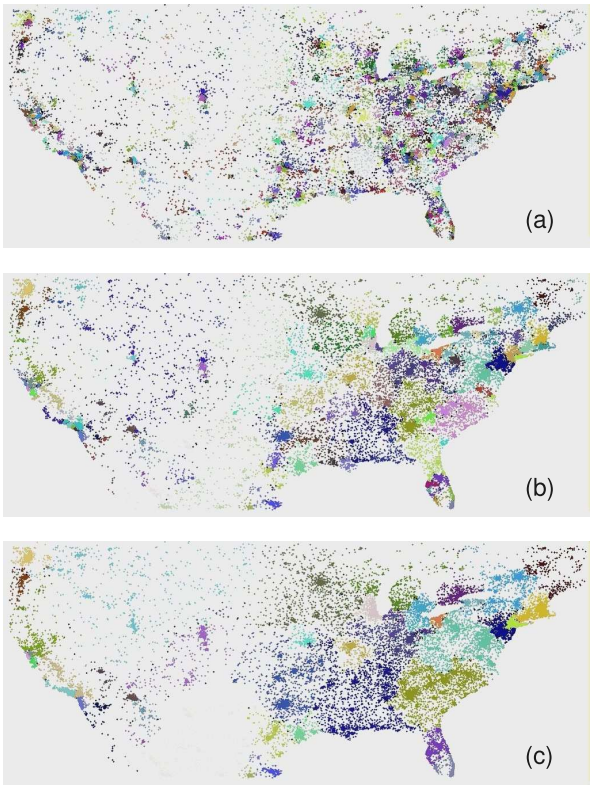


Figure 4.10: Snapshots of the time-evolution of the Naming Game on a social network based on geographical population data. The simulation starts from empty initial configuration. In the snapshots, different colors correspond to different words. In the very early stage of the game (a), a large number of different "words" are generated. Most of the words are eliminated in the coarsening stage (b). And finally in (c) configurations of long-living geographical segregated opinions are formed.

the symmetry" of the otherwise equivalent coexisting opinions. One possibility is to expose and couple many or all agents to an "external" global signal (analogous to mass media effects) [108, 27]. Alternatively, one can break the symmetry by choosing a small number of well-positioned "committed" agents who will stick to a preferred opinion without deviation. In the next subsection, we investigate this latter scenario first.

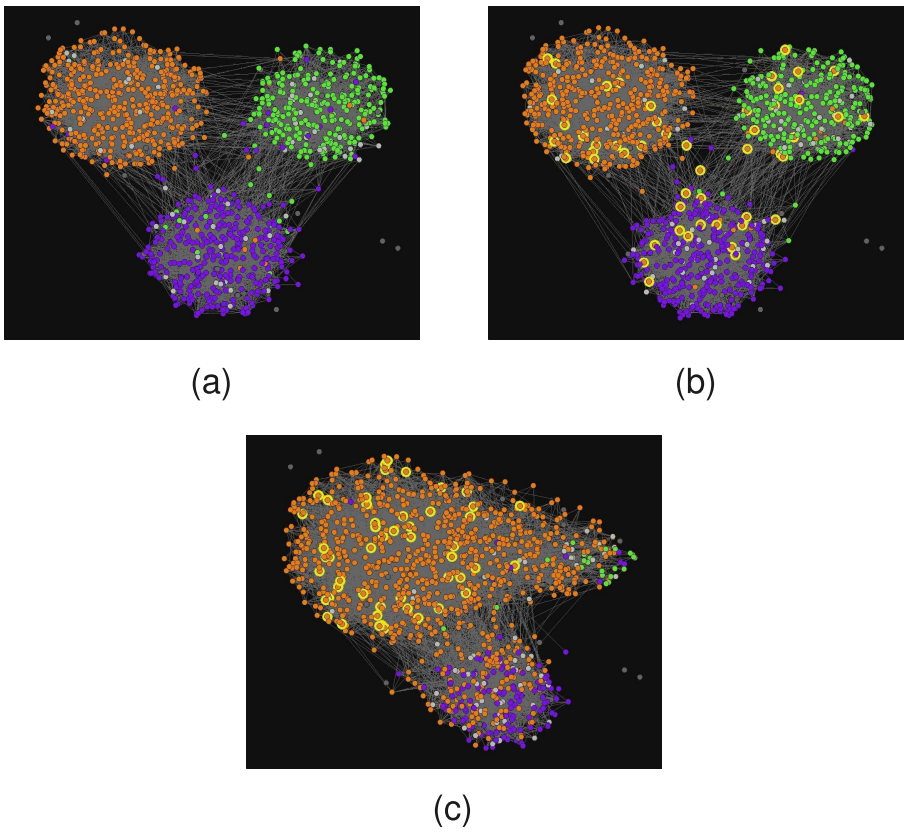


Figure 4.11: Snapshots of the Naming Game on a high-school friendship network with committed agents. The system is initialized from a state with three coexisting meta-stable communities [see (a)] with a small number of well-positioned committed agents (indicated with yellow core around the nodes as indicated in (b). Global consensus (i.e., a single opinion) is reached exponentially fast. Here we employed 50 committed agents, selected according to their degree ranking.

4.3.1 Committing Agents

In the simulations, by committed agents we mean an agent who has a fixed opinion which cannot be changed. In the context of the NG, a committed agent has a single word. As a listener, she does not accept any new word from their neighbors, but as a speaker, always transmits her word.

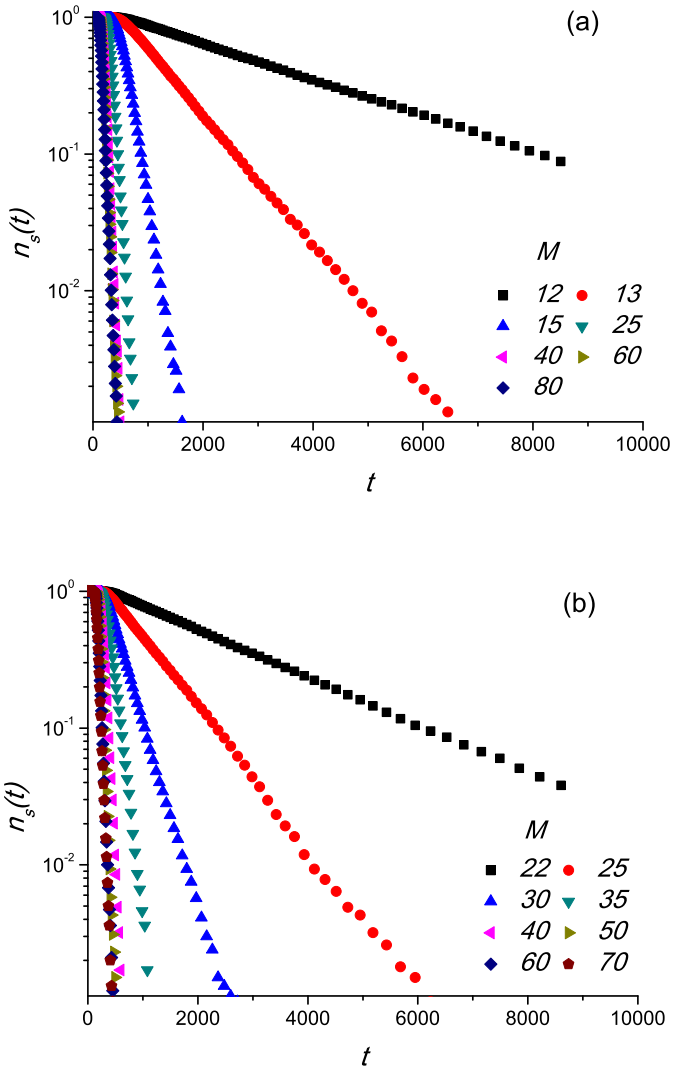


Figure 4.12: **Fraction of surviving runs as a function of time for varying number of committed agents M when agents are selected according to their (a) degree ranking and (b) (shortest-path) betweenness ranking. The total number of agents is $N = 1,127$. For the degree-based ranking selection method different symbols represent the fraction of surviving runs for 12, 13, 15, 25, 40, 60, and 80 committed agents, from top to bottom. In betweenness selection method the number of committed agents M ranges from 22, 25, 30, 35, 40, 50, 60, to 70, from top to bottom.**

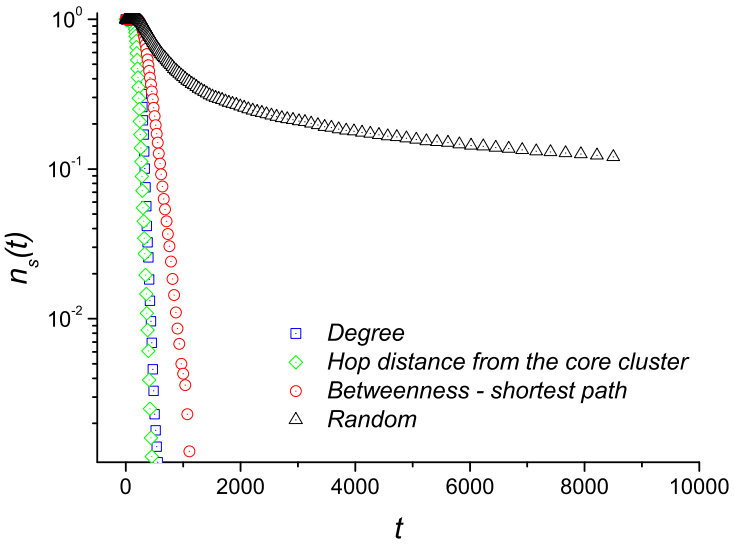


Figure 4.13: Fraction of surviving runs as a function of time for different strategies with the same number of committed agents on the same network ($M = 35$, $N = 1,127$, $f \simeq 0.031$.). The three strategies (selection of committed agents) shown here are based on degree ranking (squares), hop-distance proximity to the core cluster (diamonds), and shortest-path betweenness (circles). For comparison, the result of selecting committed agents randomly is also shown (triangles).

4.3.1.1 Committing Agents in High-school Friendship Networks

The NG on the particular high-school friendship network we investigated converges more likely to three communities at the end-stage [Fig. 4.2]. Of the three co-existing communities [Fig. 4.3(c)], we choose one community as the one representing the "preferred" opinion, and we "indoctrinate" selected committed agents with this opinion. Fig. 4.11 shows snapshots of the evolution of the NG with committed agents. Initiating the simulations from the final configuration of the original NG (exhibiting three meta-stable opinion clusters), introducing a small number of committed agents yields a relatively fast convergence to the global consensus of the selected opinion.

To quantify this phenomena we investigated the temporal behavior of this agreement process, in particular, its dependence on the method of selecting com-

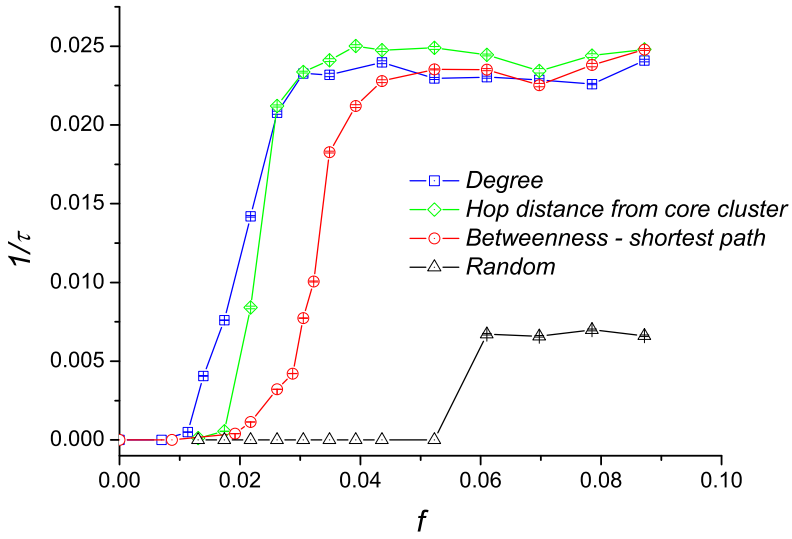


Figure 4.14: Convergence rate as a function of the fraction of committed agents $f{=}M/N$, for different selection methods of committed agents, including the degree ranking (squares), hop-distance proximity to core cluster (diamonds), and shortest-path betweenness (circles). For comparison, the result of selecting committed agents randomly is also shown (triangles).

mitted agents and on the number of these selected agents. Among the methods to select committed agents are selecting nodes with the highest degrees (nodes with the highest number of neighbors), with the highest betweenness (likely to bridge different communities), with hop-distance proximity to the core cluster (nodes outside, but no farther than two hops from the core cluster of "preferred" opinion), and for comparison, also selecting committed agents at random.

Our main observation is that once the number of committed agents is sufficient to induce global consensus, it happens *exponentially* fast, independently of the selection method. More precisely, we ran $10,000$ realizations of the NG with committed agents. The initial configuration here is the final multi-opinion meta-stable configuration of the original NG with no committed agents (with $N_d = 3$) [Fig. 4.11]. We kept track of the *fraction of surviving runs*, $n_s(t)$, defined as the fraction of runs that have not reached global consensus by time t, i.e., runs that have more than one

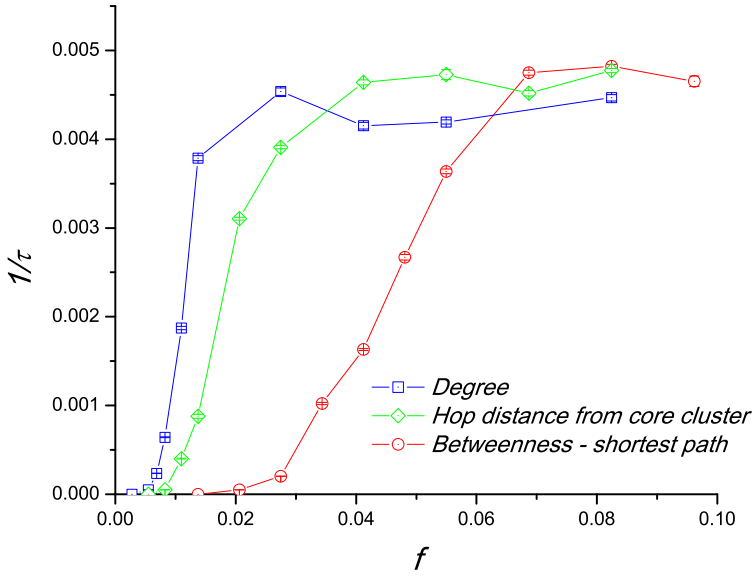


Figure 4.15: Convergence rate as a function of the fraction of committed agents f=M/N, for different selection methods of committed agents, including the degree ranking (squares), hop-distance proximity to core cluster (diamonds), and shortest-path betweenness (circles). This repeats the same simulations as in Fig. 4.14 but on a different high-school friendship network.

opinion at time t. (This quantity then can also be interpreted as the probability that a single run has not reached consensus by time t.)

We choose committed agents, to maximize their influence in reaching global consensus, according to their ranking in a number of graph theoretical measures. We selected the top M agents according to their degree, shortest-path betweenness centrality, [118, 116], hop-distance proximity to the preferred core opinion cluster, or at random, for reference. Figure 4.12 displays the fraction of surviving runs $n_s(t)$ for the degree and for the betweenness ranking for a number of different committed agents.

A common feature of all methods is that a very small fraction ($f = M/N$) of committed nodes is sufficient to induce global consensus. I.e., there seems to be a very low threshold in f_{c1}, such that for $f > f_{c1}$ the dynamics with committed

nodes leads to global agreement. Further, in this case, the fraction of surviving runs (fraction of runs with more than one opinion), $n_s(t)$, in the long-time regime, decays exponentially,

$$n_s(t) \propto e^{-t/\tau} \ . \tag{4.1}$$

The time scale of the exponential decay τ, of course, depends on the selection method and the fraction of committed nodes. The inverse time scale $1/\tau$, i.e., the rate at which global consensus is approached is, initially, an increasing function of the number of committed nodes, but it quickly *saturates* and essentially remains constant. This can be seen in Fig. 4.12, as the slopes of the exponential decays are becoming progressively steeper, up to a certain M, then they remain constant. Thus, there is second characteristic fraction of committed agents, such that for $f > f_{c2}$ the rate of reaching global consensus becomes essentially a constant (saturates).

These three features, (*i*) small threshold f_{c1} required for global consensus, (*ii*) exponential decay of $n_s(t)$ if $f > f_{c1}$ [Figs. 4.12 and 4.13], and (*iii*) saturation of the rate to reach consensus for $f > f_{c2}$ [Fig. 4.14], are exhibited by all selection method we considered here. Further, both characteristic values and the gap between them are very small, f_{c1}, f_{c2}, $f_{c2}-f_{c1} \ll 1$. These results are essentially summarized in Fig. 4.14. The convergence rate for the randomly selected committed nodes is also shown for comparison. On this particular social network, selecting a small number of the nodes with the highest degree works best, followed by the hop-distance proximity (to the core cluster) ranking. (We refer to a strategy as more efficient if the convergence rate $1/\tau$ is larger for the same fraction of committed agents.) For example, selecting the committed agents according to their degree ranking, $f_{c1} \approx 0.01$ and $f_{c2} \approx 0.03$ [Fig. 4.14]. Selecting committed agents just above this latter fraction is optimal, since the rate of convergence does not improve beyond this value.

Another high-school friendship network with $N = 900$ students and average connectivity $\overline{k} = 3.66$ [Fig. 4.4(a)] has also been investigated for the purpose of engineering global consensus. Similar features, such as small threshold f_{c1} required for global consensus, exponential decay of $n_s(t)$, and saturation of the rate to reach consensus for $f > f_{c2}$ are all present in the simulation [Fig. 4.15]

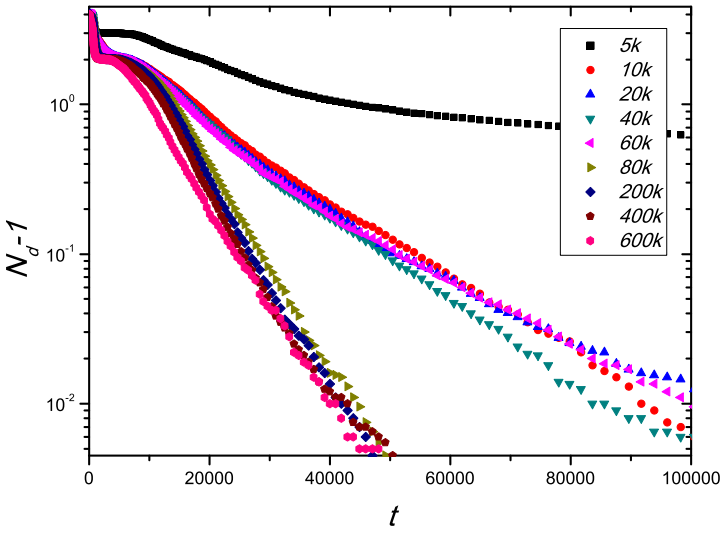


Figure 4.16: Convergence to single opinion for varying number of committed agents M when agents are selected according to degree ranking. The total number of agents is $N = 3,906,441$. Different symbols represent the average of $2,000$ runs with number of committed agents from $M = 5,000$ to $M = 600,000$, from top to bottom. Note here we measure $N_d - 1$ as a function of time t. As the simulation approaching final consensus it has the same shape as the fraction of surviving runs.

4.3.1.2 Committing Agents in A Mobile Phone Network

A more challenging task is to study the global consensus under human interventions on the mobile-phone network we discussed in above sections. Similar as on the high-school friendship network, the NG on the mobile-phone network as well develops into meta-stable state where three to five coexisting opinions reflect the same number of communities in the network [Fig. 4.8]. The difficulty lies in the size of the network – the mobile-phone network consists of nearly four million agents and tens of millions of links while the typical size of a high-school friendship network is only one thousand. Hence more efforts are need converting dissident agents, in both implementation-wise and simulation-wise. However, the study on such large-scale empirical social system usually gives more meaningful references to our real-world applications. Here we also look at the method of convincing and

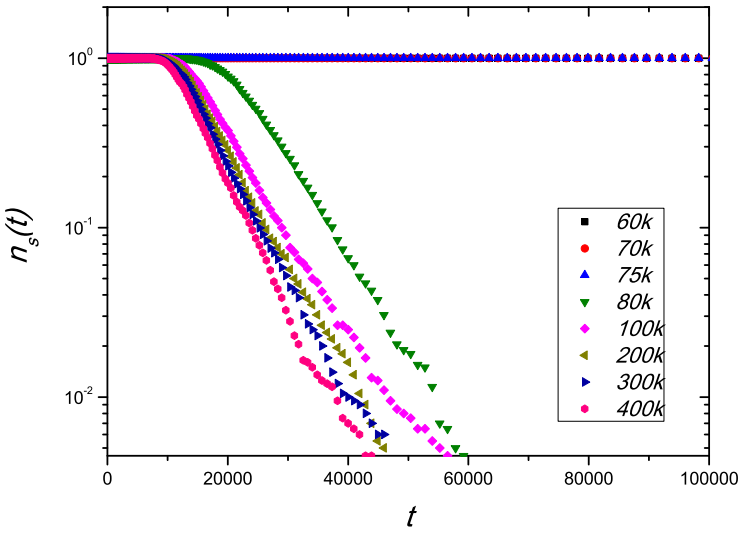


Figure 4.17: Fraction of surviving runs as a function of time for varying number of committed agents M when agents are selected according to activity ranking. The total number of agents is $N = 3,906,441$. Different symbols represent the average of $2,000$ runs with number of committed agents from $M = 60,000$ to $M = 400,000$ from top to bottom.

indoctrinating a small number of particular agents in order to for the network to reach global consensus.

An agent with large neighborhood size (high degree k) usually plays an important role in the spreading of opinions, as she is considered to have higher influences to her neighborhood than agents in the fringe with very limited number of neighbors. Therefore controlling those "hub"-like agents is expected to be an efficient way in facilitating the global consensus. Fig. 4.16 shows the exponential convergence to single opinion when the balance of coexisting opinions is broken by committing agents according to their degree ranking. Note here, when increasing the number of committed agents there exists a small threshold, where further increasing committed agents (to a fixed fraction) dose not lead to faster convergence. This threshold causes a small plateau in the convergence rate of committing agents in degree ranking [Fig. 4.20]. Possible explanation for this is that at the threshold the number of

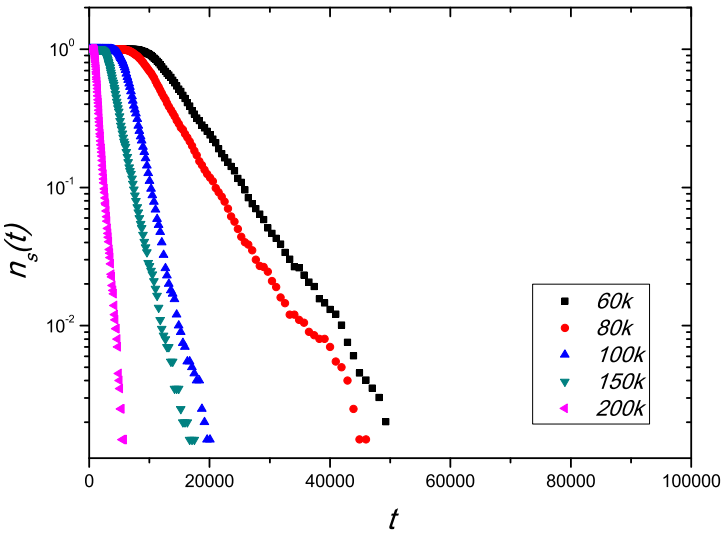


Figure 4.18: Fraction of surviving runs as a function of time for varying number of committed agents M when agents are selected according to degree ranking under the constraint where the hop-distance to core cluster must be greater than one. The total number of agents is $N = 3,906,441$. Different symbols represent the average of $2,000$ runs with number of committed agents from $M = 60,000$ to $M = 200,000$ from top to bottom.

committed agents is enough to convert one or two communities but still more efforts are needed to convert the remaining communities.

In the evolution of the NG, the local or global agreement is achieved by exchanging opinions among the agents. When an agent adopts a new opinion, or collapses his opinions to a single one, it is regarded as one *flip* of the state for that agent. By analyzing the flip number of agents throughout the evolution of the NG, we are able to classify a small proportion of nodes who change their states (opinions) more frequently than other agents. These nodes usually serve as the bridges between two or more major communities, and are exposed to multiple opinions simultaneously. They have higher chances of accepting multiple opinions than those sitting deep inside a community. Committing these bridging agents with high priority might be an effective way to infiltrate into other communities and eventually

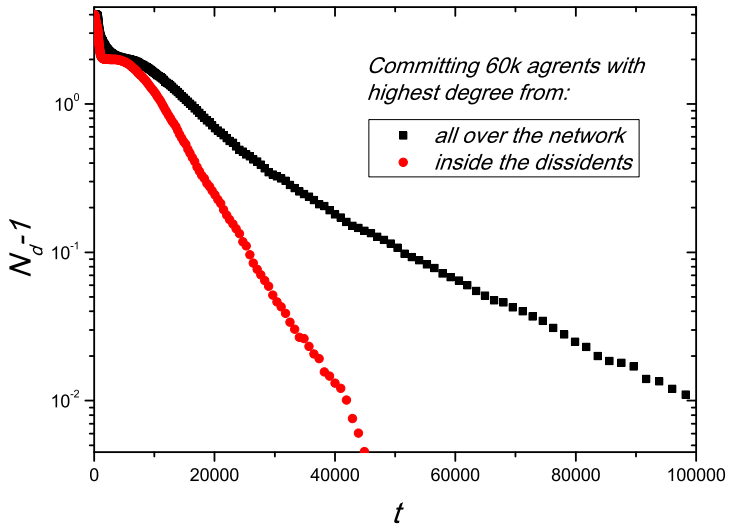


Figure 4.19: Comparison of opinion convergence rate between selecting highest degree nodes from all over the network (black squares), and selecting highest degree nodes from inside the dissidents (red circles), when $M = 60,000$, or $f = 0.0154$ agents are committed. Here we measure the number of different opinions ($N_d - 1$) instead of the fraction of surviving runs.

lead to the collapse of community structure. Therefore we explored the scenario of facilitating the global consensus by committing a small number of the nodes with the highest flip number. When sufficient fraction of agents are committed in this way, the system again exhibits exponentially fast decay to the state of global consensus [Fig. 4.17].

We consider the scenario of facilitating the global agreement toward a specific opinion (preferred opinion) by committing a small number of agents. In most cases, it is a waste of resources to indoctrinate agents who has already held the preferred opinion. Thus, selecting committing agents only from dissidents (whose opinions differ from the preferred one) is a natural choice if one wants to minimize the cost of the indoctrination. It is the same idea as selecting nodes with the hop-distance (to the preferred opinion) greater than one. To further optimize the strategy, we

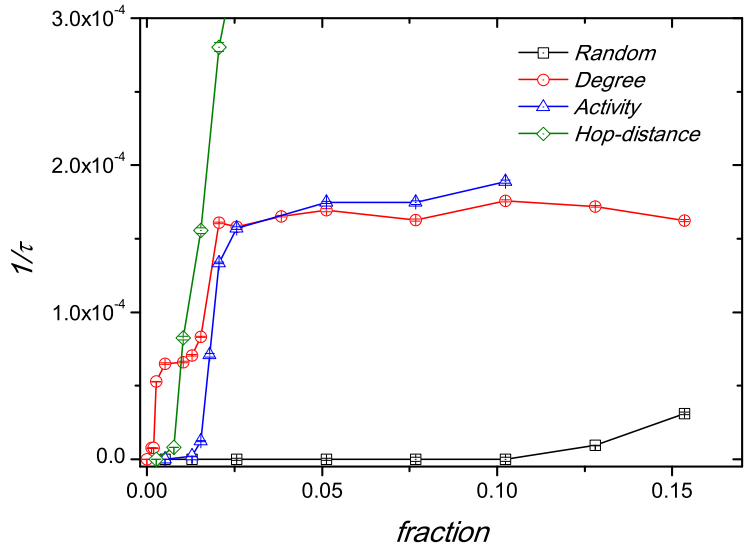


Figure 4.20: Convergence rate as a function of the fraction of committed agents f=M/N, for different selection methods of committed agents, including the degree ranking (circles), hop-distance proximity to core cluster (diamonds), and activity ranking (triangles). For comparison, the result of selecting committed agents randomly is also shown (squares).

also favor those nodes with higher connectivity or degree k. Fig. 4.18 shows the process of converging to the preferred opinion using the strategy. Comparing with other strategies such as degree ranking and activity ranking, it requires less number of committed agents to achieve the same rate of opinions collapsing [Fig. 4.19].

The summary of results for these strategies are shown in Fig. 4.20, along with the random selection for a baseline comparison. We again find the small threshold f_{c1} required for global consensus, exponential decay of $n_s(t)$ if $f > f_{c1}$, and saturation of the rate to reach consensus for $f > f_{c2}$ in all selection method we considered here. Global consensus is achieved when the symmetry is broken with certain amount of committed agents. The processes happen fast as the characteristic values of f_{c1} and f_{c2} and the gap between them are very small, f_{c1}, f_{c2}, $f_{c2}-f_{c1} \ll 1$. On this particular mobile-phone network, selecting a small number of the nodes according to the hop-distance proximity (to the core cluster) ranking works better than the

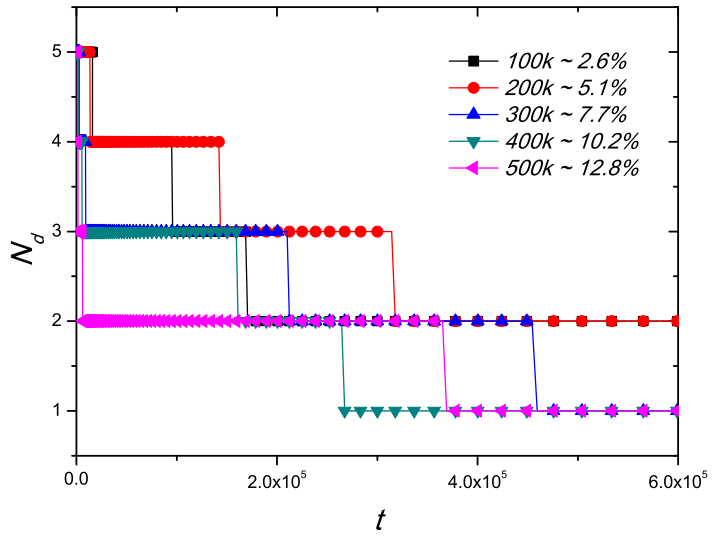


Figure 4.21: Global convergence initiated from a minority community. Different symbols represent the number of different words N_d as function of time t with different numbers of committed agents, from $M = 100,000$ to $M = 500,000$. The committed agents are selected according to the ranking of their hop-distance to the minority community.

highest degree ranking and activity ranking. This is due to the large imbalance of community sizes – the largest community consists of more than 90% of nodes with all remaining nodes evenly distributed to four small communities. Thus even selecting a small fraction of nodes with hop-distance greater than one might deplete the minority communities and lead to fast convergence.

The mobile-phone network provides a configuration with great disparity in community sizes. It is interesting to see whether it is possible for a small community to spread its opinion to the whole network once the symmetry is broken. Our results indicate that with proper positioning of committed agents this can also be achieved [Fig. 4.21].

In general, the optimal selection method will vary, depending on the community structure of the particular underlying network. However, because we changed the dynamics of the NG by breaking the symmetry of otherwise equivalent opinions,

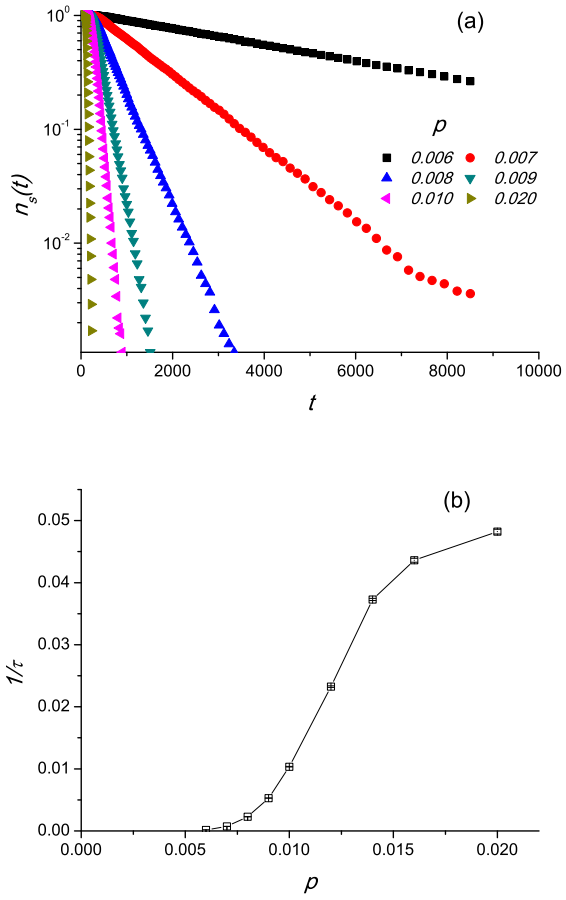


Figure 4.22: Global consensus under external influences in a high-school friendship network. (a) The fraction of surviving runs as a function of time for several values of the strength of external influence p (p is the probability that in a time step an agent will adopt the fixed externally and globally promoted opinion). (b) Convergence rate to global consensus as a function of the strength of external field p.

the exponential decay and the saturation of the convergence rate is expected to be generic for a large class of opinion formation models on networks with community structure.

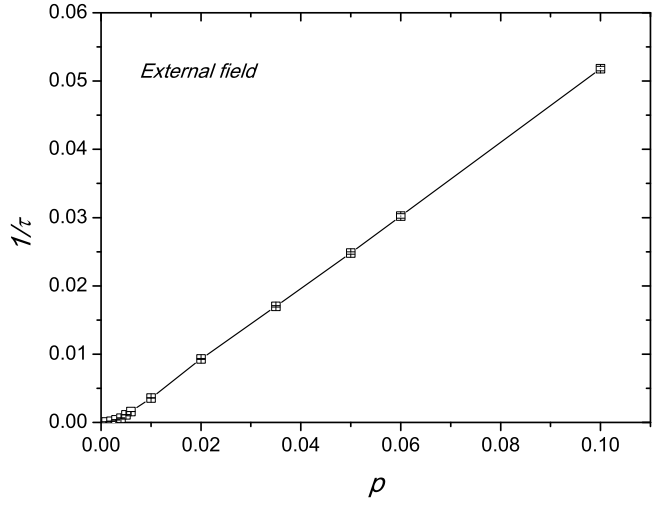


Figure 4.23: Convergence rate to global consensus as a function of the strength of external field p for mobile-phone network, where p is the probability that in a time step an agent will adopt the fixed externally and globally promoted opinion.

4.3.2 Global External Influence

As mentioned in the introductory paragraph of the section 4.3, another natural way of influencing the outcome of the competition among otherwise neutral and meta-stable opinions, is to couple all or a fraction of agents to a global external "signal" [mimicking a mass media effect [108, 27]]. For comparison, we implemented the NG with an external field (affecting all agents) corresponding to the selected opinion among the three meta-stable ones in the final stage of the NG. Then, similar to the committed-agent approach, we initialize the system with that final meta-stable state with co-existing opinions of the original NG. In the presence of mass media, an agent, when randomly chosen, with probability p will adopt the externally promoted opinion. Otherwise, the usual rules of the game are invoked (i.e., the node, as a speaker, initiates communication with a listener).

Our findings indicate that even for extremely small values of p, the fraction of surviving runs (the fraction of runs that have not reached global consensus) decays exponentially, ultimately leading to global order [Fig. 4.22(a)]. The rate of con-

vergence $1/\tau$ increases monotonically and smoothly with p [Fig. 4.22(b)]. Similar relation of the convergence rate and strength of external field p also exists in the mobile-phone network [Fig. 4.23]. For application oriented studies, one should associate a cost with the mass-media coupling, and a cost with committing an agent (e.g., finding these nodes and giving them incentives impossible to resists), then perform a relevant cost-benefit comparative analysis for the selection or optimal combination of two approaches.

CHAPTER 5
Summary and Conclusions

In our work, several stylized models of multiplex propagation and opinion dynamics have been analyzed on complex communication and empirical social networks, to gain better understanding of dynamical processes such as collective behaviors and components interactions in social and technological systems.

We studied cascade dynamics in threshold-controlled (multiplex) propagation on random geometric networks as a simple yet effective model of outliers cleansing and amplifying in wireless sensor networks. Hence, the local dynamics of cascading can serve as an efficient, robust, and reliable basic activation protocol for responding to various alarm scenarios and distinguishing between false (few outliers indicating alarm discovered) and real alarms (several outliers detected in close proximity of each other). We also found that the network modified by adding a few long-range communication links, resulting in a small-world network, changes the speed of the network's response. Hence, such construction can further enhance and optimize the speed of the network's response, while keeping energy consumption at a manageable level.

In addition to the information propagation model, here a prototypical agent-based model of opinion dynamics, the Naming Game, has also been carefully studied on Random Geometric Graphs and SW-connected RGGs embedded in two dimensions. While the underlying RGG communication topology is motivated by large-scale sensor networks, the NG on these networks captures fundamental features of agreement dynamics of spatially-embedded networked agent-based systems. We have found that, qualitatively similar to two-dimensional regular networks [14], the NG on RGG can be reasonably well described by the physical theory of coarsening. In particular, local clusters of nodes sharing the same word quickly form, followed by slow coarsening of these clusters in the late stage of the dynamics. The typical length scale grows as $\xi(t) \sim t^{\gamma}$ with the coarsening exponent estimated to be $0.35 < \gamma < 0.40$. Our simulation results also indicate that the average time to reach global agreement

is of $\mathcal{O}(N^{1.08\pm0.03})$ (for *fixed* average degree). The above results imply that, at least for the range of finite system sizes studied here (up to $N = 5,000$), the characteristic length scale in two-dimensional RGGs grows slower than 1/2. This deviation, in part, may very well be attributed to the effectively small system sizes that we could study. Similarly strong transients and finite-size corrections, due to the presence of "soft domain walls" and "vertices" (domain-wall intersections), also made the precise determination of the asymptotic coarsening exponent difficult in the two-dimensional large-Q (effectively $Q{=}\infty$) Potts models [132, 133, 73, 89, 130]. On the the other hand, based on our Monte Carlo studies, we cannot rule out the possibility that the deviation from the $\gamma{=}1/2$ coarsening exponent is the result of the inherent *local random* random structure of RGGs (in contrast to regular two-dimensional grids [14]).

While in this work we did not address the message complexity of the NG explicitly, one can make an order of magnitude estimate for the typical number of messages needed to reach global agreement on RGGs for an efficient implementation. (In sensor networks, this quantity is also relevant since it corresponds to the global energy consumption.) Once the coarsening process begins, nodes inside the clusters have reached agreement with all their neighbors, of which they are readily aware, hence, they no longer have to initiate broadcasts any longer. Thus, only these "active" nodes, found at the interfaces between these cluster (which have at least one neighbor with different words), will initiate broadcast for word matching. Using that the number of nodes at the interfaces scales as N/t^{γ} [Eq. (3.3)], and integrating this expression up to $t_c{\sim}N^{1.08\pm0.03}$, one finds that the total number of messages needed to be exchanged until global agreement is reached is of $\mathcal{O}(N^{1.68\pm0.05})$.

In an attempt to accelerate the agreement process by changing the communication topology between agents, we also studied the SW-connected version of the two-dimensional RGG. By adding a small density of shortcuts "on top" of the RGG, resulting in a SW-like network, the convergence time is strongly reduced and becomes of $\mathcal{O}(N^{0.31})$, similar to the behavior of NG on the Watts-Strogatz SW network [41].

Earlier works on NG have shown that this simple model for agreement dy-

namics and opinion formation always leads to global consensus on graphs with no community structure. We therefore further studied the Naming Game on social networks, which are known to have rich community structure. On empirical social networks, such as high-school friendship networks, the mobile-phone network, and the social network based on geographical population data, the Naming Game exhibits, in the late-stage of the dynamics, several meta-stable coexisting communities; these configurations, in effect, are the computationally observed final configurations.

In models for social dynamics, communities manifest themselves in the context in which distinct stylized opinions (e.g., religions, cultures, languages) have evolved and emerged over time. Clusters of nodes having reached consensus are part of a community, reflecting the inherent community structure of the underlying social networks. Thus, if at the late stages of the social dynamics on the networks, several communities persist (different opinions survive), they are the authentic signatures of the underlying community structure. The Naming Game, together with other similar models for opinion formation, can be employed to probe these properties of complex networks.

We then considered the task of destabilizing the coexisting meta-stable opinions (in order to reach consensus) by selecting the optimal number of committed agents with a preferred opinion, as an alternative to a global external signal (mass media effect). The results implied that a small number of committed agent is sufficient to facilitate an exponential decay toward global consensus of the selected opinion. Further, selecting more agent than a system-specific upper cut-off, yields no improvement in the convergence rate. Hence, there seems to be an optimal number of agents for this task, beyond which it does not pay off to invest in committing more agents. Selecting the committed agents according to their degree, betweenness, activeness, or hop-distance proximity to the core cluster of the preferred opinion, all displayed the above qualitative features. Further, they all significantly outperformed committing the same number of agents at random.

LITERATURE CITED

[1] L. A. Adamic. The small world web. In *Lecture Notes in Computer Science*, volume 1696, pages 443–454. Springer, New York, 1999.

[2] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. A survey on sensor networks. *IEEE Comm. Magazine*, pages 102–114, 2002.

[3] I.F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci. Wireless sensor networks: A survey. *IEEE Trans. Systems, Man and Cybernetics, B*, 38:393–422, 2002.

[4] R. Albert and A.-L. Barabási. Statistical mechanics of complex networks. *Rev. Mod. Phys.*, 74:47–97, 2002.

[5] R. Albert, H. Jeong, and A.-L. Barabási. Diameter of the world wide web. *Nature*, 401:130–131, 1999.

[6] P.A. Alemany and D. ben Avraham. Inter-particle distribution functions for one-species diffusion-limited annihilation, $A + A \to 0$. *Phys. Lett. A*, 206:18–25, 1995.

[7] M. Anghel, Z. Toroczkai, K.E. Bassler, and G. Korniss. Competition-driven network dynamics: Emergence of a scale-free leadership structure and collective efficiency. *Phys. Rev. Lett.*, 92:058701 [4 pages], 2004.

[8] D. Angluin. Local and global properties in networks of processors. In *Proceedings of the 12th ACM Symposium on Theory of Computing*, pages 82–93, 1980.

[9] T. Antal, P.L. Krapivsky, and S. Redner. Dynamics of social balance on networks. *Phys. Rev. E*, 72:036121, 2005.

[10] R. Axelrod. The dissemination of culture: A model with local convergence and global polarization. *The Journal of Conflict Resolution*, 41:203–226, 1997.

[11] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[12] M. Barahona and L.M. Pecora. Synchronization in small-world systems. *Phys. Rev. Lett.*, 89:054101, 2002.

[13] A. Baronchelli, L. Dall'Asta, A. Barrat, and V. Loreto. Strategies for fast convergence in semiotic dynamics. arXiv:physics/0511201, 2005.

[14] A. Baronchelli, L. Dall'Asta, A. Barrat, and V. Loreto. Topology induced coarsening in language games. *Phys. Rev. E*, 73:015102(R), 2006.

[15] A. Baronchelli, M. Felici, E. Caglioti, V. Loreto, and L. Steels. Sharp transition towards shared vocabularies in multi-agent systems. *J. Stat. Mech.*, page P06014, 2006.

[16] A. Baronchelli, V. Loreto, L. Dall'Asta, and A. Barrat. Bootstraping communications in language games: Strategy, topology, and all that. In *Proceedings of the 6th International Conference on the Evolution of Language, edited by A. Cangelosi, A.D.M. Smith, and K. Smith (World Scientific, Singapore, 2006)*, pages 11–18, 2006.

[17] A. Barrat. Comment on "small-world networks: Evidence for a crossover picture". arXiv:cond-mat/9903323, 1999.

[18] M. Barthelemy and L.A. Numes Amaral. Erratum: Small-world networks: Evidence for a crossover picture. *Phys. Rev. Lett.*, 82(25):5180–5180, 1999.

[19] E. Ben-Naim. Opinion dynamics: Rise and fall of political parties. *Europhys. Lett.*, 69:671–677, 2005.

[20] E. Ben-Naim, L. Frachebourg, and P. L. Krapivsky. Coarsening and persistence in the voter model. *Phys. Rev. E*, 53:3078–3087, 1996.

[21] I.J. Benczik, S.Z. Benczik, B. Schmitmann, and R.K.P. Zia. Lack of consensus in social systems. *Europhys. Lett.*, 82:48006, 2008.

[22] M. Blatt, S. Wiseman, and E. Domany. Superparamagnetic clustering of data. *Phys. Rev. Lett.*, 76:3251–3254, 1996.

[23] B. Bollobás. *Random graphs*. Cambridge University Press, 2nd edition, 2001.

[24] D. Boyer and O. Miramontes. Interface motion and pinning in small-world networks. *Phys. Rev. E*, 67:035102(R), 2003.

[25] J. Branch, G. Chen, and B. K. Szymanski. Escort: Energy-efficient sensor network communal routing topology using signal quality metrics. In *Proceedings of the International Conference on Networking (ICN 2005), Lecture Notes in Computer Science, Vol. 3420*, pages 438–448. Springer-Verlag, New York, 2005.

[26] A. Bray. Theory of phase-ordering kinetics. *Adv. in Phys.*, 43:357–459, 1994.

[27] J. Candia and K. I. Mazzitello. Mass media influence spreading in social networks with community structure. *J. Stat. Mech.: Theory Exp.*, page P07007, 2008.

[28] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *arXiv:0710.3256v1*, 2008.

[29] C. Castellano, V. Loreto, A. Barrat, F. Cecconi, and D. Parisi. Comparison of voter and glauber ordering dynamics on networks. *Phys. Rev. E*, 71:066107, 2005.

[30] C. Castellano, M. Marsili, and A. Vespignani. Nonequilibrium phase transition in a model for social influence. *Phys. Rev. Lett.*, 85:3536–3539, 2000.

[31] C. Castellano, D. Vilone, and A. Vespignani. Incomplete ordering of the voter model on small-world networks. *Europhys. Lett.*, 63:153–158, 2003.

[32] C. Cattuto, A. Baldassarri, V. D. P. Servedio, and V. Loreto. Vocabulary growth in collaborative tagging systems. *arXiv:0704.3316v1*, 2007.

[33] C. Cattuto, V. Loreto, and L. Pietronero. Semiotic dynamics and collaborative tagging. *Proc. Natl. Acad. Sci. USA*, 104:1461–1464, 2006.

[34] D. Centola, M.W. Macy, and V.M. Eguiluz. Cascade dynamics of multiplex propagation. *arXiv:0504165*, 2005.

[35] D. Challet, M. Marsili, and Y.-C. Zhang. *Minority games: Interacting agents in financial markets*. Oxford University Press, 2005.

[36] G. Chen, J. Branch, and B.K. Szymanski. Local leader election, signal strength aware flooding, and routeless routing. In *5th IEEE Int. Workshop Algorithms for Wireless, Mobile, Ad Hoc Networks and Sensor Networks, WMAN05*, 2005.

[37] J.E. Cohen, F. Briand, and C.M. Newman. *Community food webs: Data and theory*. Springer, New York, 1990.

[38] R. Cohen, S. Havlin, and D. Ben-Avraham. Efficient immunization strategies for computer networks and populations. *Physical Review Letter*, 91:247901, 2003.

[39] T. C. Collier and C. E. Taylor. Self-organization in sensor network. *Journal of Parallel and Distributed Computing*, 64:866–873, 2004.

[40] J. Dall and M. Christemsen. Random geometric graphs. *Phys. Rev. E*, 66:016121, 2002.

[41] L. Dall'Asta, A. Baronchelli, A. Barrat, and V. Loreto. Agreement dynamics on small-world networks. *Europhys. Lett.*, 73:969–975, 2006.

[42] L. Dall'Asta, A. Baronchelli, A. Barrat, and V. Loreto. Nonequilibrium dynamics of language games on complex networks. *Phys. Rev. E*, 74:036105, 2006.

[43] D. J. de S. Price. Network of scientific papers. *Science*, 149:510–515, 1965.

[44] D. J. de S. Price. A general theory of bibliometric and other cumulative advantage processes. *J. Amer. Soc. Inform. Sci.*, 27:292–306, 1976.

[45] B. DeCleene, L. Dondeti, S. Griffin, T. Hardjono, D. Kiwior, J. Kurose, D. Towsley, S. Vasudevan, and C. Zhang. Secure group communications for wireless networks. In *Proc. of MILCOM*, 2001.

[46] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch. Mixing beliefs among interacting agents. *Adv. Compl. Syst.*, 3:87–98, 2000.

[47] B. Derrida, C. Goudreche, and I. Yekutieli. Scale-invariant regimes in one-dimensional models of growing and coalescing droplets. *Phys. Rev. A*, 44(10):6241–6251, 1991.

[48] B. Derrida and R. Zeitak. Distribution of domain sizes in the zero temperature glauber dynamics of the one-dimensional potts model. *Phys. Rev. E*, 54(3):2513–2525, 1996.

[49] P. S. Dodds, R. Muhamad, and D. J. Watts. An experimental study of search in global social networks. *Science*, 301:827–829, 2003.

[50] S.N. Dorogovtsev and J.F.F. Mendes. Evolution of networks. *Adv. in Phys.*, 51:1079–1187, 2002.

[51] S. N. Durlauf. How can statistical mechanics contribute to social science? *Proc. Natl. Acad. Sci. USA*, 96:10582–10584, 1999.

[52] J. M. Epstein and R. Axtell. *Growing artificial societies: Social science from the bottom up.* MIT Press, Cambridge, MA, New York, 1996.

[53] Paul Erdős and Alfréd Rényi. The evolution of random graphs. *Magyar Tud. Akad. Mat. Kutato Int. Kozl.*, 5:17–61, 1960.

[54] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar. Next century challenges: Scalable coordination in sensor networks. In *ACM Mobicom*, pages 263–270, 1999.

[55] S. Fortunato and M. Barthelemy. Resolution limit in community detection. *Proc. Natl. Acad. Sci. USA*, 104:36–41, 2007.

[56] K. Fukushima. Cognition: A self-organizing multilayered neural network. *Biological Cybernetics*, 20:121–136, 1975.

[57] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, 99:7821–7826, 2002.

[58] S. Golder and B. A. Huberman. Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32:198–208, 2006.

[59] M. C. Gonzáles, H. J. Herrmann, J. Kertész, and T. Vicsek. Community structure and ethnic preferences in school friendship networks. *Physica A*, 379:307–316, 2007.

[60] M. Granovetter. Threshold models of collective behavior. *Am. J. Soc.*, 83:1420–1443, 1978.

[61] P. Gupta and P. K. Kumar. The capacity of wireless networks. *IEEE Trans. Inf. Theor.*, IT-46:388–404, 2000.

[62] R. Hegselmann and U. Krause. Opinion dynamics and bounded confidence models, analysis and simulation. *J. Artif. Soc. Soc. Simul.*, 5(3):paper 2, 2002.

[63] W. Heinzelman, N.B. Abu-Ghazaleh, and S. Tilak. A taxonomy of wireless micro-sensor network models. *Mobile Comp. Comm. Rev.*, 6:23–36, 2002.

[64] A. Helmy. Small worlds in wireless networks. *IEEE Comm. Lett.*, 7:490–492, 2003.

[65] D. S. Hirschberg and J. B. Sinclair. Decentralized extrema-finding in circular configurations of processors. *Communications of the ACM*, 23:627–628, 1980.

[66] P.W. Holland and S. Leinhardt. Transitivity in structural models of small groups. *Comparative Group Studies*, 2:107–124, 1998.

[67] M. Howard and C. Godrèche. Persistence in the voter model: Continuum reaction-diffusion approach. *J. Phys. A: Math. Gen.*, 31:L209–L215, 1998.

[68] `http://en.wikipedia.org/wiki/Social_network`. Date Last Accessed 09/26/2009.

[69] `http://www.facebook.com/group.php?gid=14436512661`. Date Last Accessed 10/12/2009.

[70] W.-S. Jung, H.-T. Moon, and H. E. Stanley. Dynamics of clustered opinions in complex networks. *J. Econ. Interac. Coord.*, 3:81–88, 2008.

[71] Victor M. Eguiluz K. Suchecki and M. San Miguel. Conservation laws for the voter model in complex networks. *Europhysics Letters*, 69:228–234, 2005.

[72] Victor M. Eguiluz K. Suchecki and M. San Miguel. Voter model dynamics in complex networks: Role of dimensionality, disorder, and degree distribution. *Phys. Rev. E*, 72:036132, 2005.

[73] K.K. Kaski, J. Nieminen, and J.D. Gunton. Domain growth and scaling in the q-state potts model. *Phys. Rev. B*, 31:2998–3002, 1985.

[74] S.A. Kauffman. Metabolic stability and epigenesis in randomly connected genetic nets. *Journal of Theoretical Biology*, 22:437–467, 1969.

[75] S.A. Kauffman. *The origins of order: Self-organization and selection in evolution*. Oxford University Press, 1993.

[76] S. Kirby. Natural language from artificial life. *Artificial Life*, 8:185–215, 2002.

[77] K. Klemm, V.M. Eguíluz, R. Toral, and M. San Miguel. Nonequilibrium transitions in complex networks: A model of social interaction. *Phys. Rev. E*, 72(2):026120, 2003.

[78] G. Korniss, 2009. Private communications.

[79] G. Korniss, M. A. Novotny, H. Guclu, Z. Toroczkai, and P. A. Rikvold. Suppressing roughness of virtual times in parallel discrete-event simulations. *Science*, 299:677–679, 2003.

[80] B. Kozma and A. Barrat. Consensus formation on adaptive networks. *Phys. Rev. E*, 77:016102, 2008.

[81] B. Kozma and A. Barrat. Who's talking first? consensus or lack thereof in coevolving opinion formation models. *Phys. Rev. Lett.*, 100:158701, 2008.

[82] B. Kozma, M. B. Hastings, and G. Korniss. Diffusion processes on power-law small-world networks. *Phys. Rev. Lett.*, 95:018701 [4 pages], 2005.

[83] B. Kozma, M.B. Hastings, and G. Korniss. Roughness scaling for edwards-wilkinson relaxation in small-world networks. *Phys. Rev. Lett.*, 92:108701, 2004.

[84] B. Kozma, M.B. Hastings, and G. Korniss. Diffusion processes on power-law small-world networks. *Phys. Rev. Lett.*, 95:018701, 2005.

[85] P. L. Krapivsky and S. Redner. Dynamics of majority rule in two-state interacting spin systems. *Phys. Rev. Lett.*, 90:238701 [4 pages], 2003.

[86] P.L. Krapivsky and E. Ben-Naim. Domain statistics in coarsening systems. *Phys. Rev. E*, 56(4):3788–3789, 1997.

[87] W. Krause, I. Glauche, R. Sollacher, and M. Greiner. Impact of network structure on the capacity of wireless multihop ad hoc communication. *Physica A*, 338:633–658, 2004.

[88] B. Krishnamachar, S. B. Wicker, and R. Béjar. Phase transition phenomena in wireless ad hoc networks. In *Symposium on Ad-Hoc Wireless Networks, IEEE Globecom, San Antonio, Texas, November (2001)*, 2001.

[89] S. Kumar, J.D. Gunton, and K.K. Kaski. Dynamical scaling in the q-state potts model. *Phys. Rev. B*, 35:8517–8522, 1987.

[90] J. M. Kumpula, J. Saramaki, K. Kaski, and J. Kertész. Limited resolution in complex network community detection with potts model approach. *Eur. Phys. J. B*, 56:41–45, 2007.

[91] R. Lambiotte and M. Ausloos. Coexistence of opposite opinions in a network with communities. *J. Stat. Mech.: Theory Exp.*, page P08026, 2007.

[92] Y. Lee, T. C. Collier, C. E. Taylor, and E. E. Stabler. The role of population structure in language evolution. In *Proceedings of the 10th International Symposium on Artificial Life and Robotics*, 2005.

[93] G. LeLann. Distributed systems, towards a formal approach. In *IFIP Congress Proceedings*, pages 155–160, 1977.

[94] J. Leskovec and E. Horvitz. Planetary-scale views on an instant-messaging network. *arXiv:0803.0939v1*, 2008.

[95] T. M. Liggett. *Interacting particle systems*. Springer-Verlag, New York, 1985.

[96] T. M. Liggett. *Stochastic interacting systems*. Springer-Verlag, New York, 1999.

[97] M. Lim, R. Metzler, and Y. Bar-Yam. Global pattern formation and ethnic/cultural violence. *Science*, 317:1540–1544, 2007.

[98] B.-Y. Lin, J. Ren, H.-J. Yang, and B. H. Wang. Naming Game on small-world networks: The role of clustering structure. e-print arXiv:physics/0607001, 2006.

[99] J. Lorenz. Continuous opinion dynamics under bounded confidence: A survey. *Int. J. Mod. Phys. C*, 18:1819–1838, 2007.

[100] A.J. Lotka. The frequency distribution of scientific productivity. *Journal of the Washington Academy of Sciences*, 16:317–323, 1926.

[101] Q. Lu, G. Korniss, and B. K. Szymanski. Threshold-controlled global cascading in wireless sensor networks. In *Proceedings of the Third International Conference on Networked Sensing Systems (INSS 2006)*, 2006.

[102] Q. Lu, G. Korniss, and B. K. Szymanski. Naming Games in spatially-embedded random networks. In *Proceedings of the 2006 American Association for Artificial Intelligence Fall Symposium Series, Interaction and Emergent Phenomena in Societies of Agents (AAAI Press, Menlo Park, CA 2006)*, pages 148–155, 2006.

[103] Q. Lu, G. Korniss, and B.K. Szymanski. Naming Games in two-dimensional and small-world-connected random geometric networks. *Phys. Rev. E*, 77:016111, 2008.

[104] Q. Lu, G. Korniss, and B.K. Szymanski. The Naming Game in social networks: Community formation and consensus engineering. *J. Econ. Interac. Coord.*, 4:221–235, 2009.

[105] N. Malpani, J. Welch, and N. Vaidya. Leader election algorithms for mobile ad-hoc networks. In *Proc. Fourth International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, pages 96–103, 2000.

[106] T. Masser and D. ben Avraham. Kinetics of coalescence, annihilation, and the q-state potts model in one dimension. *Phys. Lett. A*, 275:382–385, 2000.

[107] F. A. Matsen and M. A. Nowak. Win-stay, lose-shift in language learning from peers. *Proc. Natl. Acad. Sci. USA*, 101:18053–18057, 2004.

[108] K. I. Mazzitello, J. Candia, and V. Dossetti. Effects of mass media and cultural drift in a model for social influence. *Int. J. Mod. Phys. C*, 18:1475–1482, 2007.

[109] A. McConchie. The great pop vs. soda controversy. http://popvssoda.com:2998, Last Date Accessed 10/18/2009.

[110] I. McCulloh, J. Lospinoso, and K.M. Carley. Probability mechanics in communications networks. In *Proceedings of the 12th International Conference on Applied Mathematics*. World Science Engineering Academy and Society, 2007.

[111] R. Meester and R. Roy. *Continuum percolation*. Cambridge University Press, 1996.

[112] S. Milgram. The small world problem. *Psychology Today*, 2:60–67, 1967.

[113] J. Moody. Race, school integration and friendship segregation in america. *American Journal of Sociology*, 107:679–716, 2001.

[114] Jacob L. Moreno. *The Autobiography of J. L. Moreno M.D.* Harvard University, Boston, Massachuestts, 1985.

[115] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003.

[116] M. E. J. Newman. Detecting community structure in networks. *Eur. Phys. J. B.*, 38:321–330, 2004.

[117] M. E. J. Newman. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E*, 74:036104, 2006.

[118] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, 2004.

[119] M. E. J. Newman and D. J. Watts. Renormalization group analysis of the small-world network model. *Phys. Lett. A*, 263:341–346, 1999.

[120] M. A. Nowak and N. L. Komarova. Evolution of universal grammar. *Science*, 291:114–118, 2001.

[121] M. A. Nowak, J. B. Plotkin, and C. Krakauer. The evolutionary language game. *J. Theor. Biol.*, 200:147–162, 1999.

[122] J.-P. Onnela, J. Saramaki, J. Hyvonen, G. Szabó, D Lazer, K. Kaski, J. Kertész, and A.-L. Barabási. Structure and tie strength in mobile communication networks. *Proc. Natl. Acad. Sci. USA*, 104:7332–7336, 2007.

[123] G. Palla, A.-L. Barabási, and T. Vicsek. Quantifying social group evolution. *Nature*, 446:664–667, 2007.

[124] G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435:814–818, 2005.

[125] M. Penrose. *Random geometric graphs*. Oxford University Press, 2003.

[126] P. Rämö, J. Kesseli, and O. Yli-Harja. Perturbation avalanches and criticality in gene regulatory networks. *J. Theor. Biol.*, 242:164, 2006.

[127] J. Reichardt and S. Bornholdt. Detecting fuzzy community structures in complex networks with a potts model. *Phys. Rev. Lett.*, 93:218701, 2004.

[128] Garry Robins, Pip Pattison, Yuval Kalish, and Dean Lusher. An introduction to exponential random graph (p*) models for social networks. *Social Networks*, 29(2):173–191, May 2007.

[129] T. Rohlf, N. Gulbahce, and C. Teuscher. Damage spreading and criticality in finite random dynamical networks. *Phys. Rev. Lett.*, 99(24):248701, 2007.

[130] C. Roland and M. Grant. Monte carlo renormalization-group study of domain growth in the potts model on a triangular lattice. *Phys. Rev. B*, 41(7):4663–4668, 1990.

[131] M.L. Sachtjen, B.A. Carreras, and V.E. Lynch. Disturbances in a power transmission system. *Physical Review E*, 61:4877, 2000.

[132] S.A. Safran, P.S. Sahni, and G.S. Grest. Kinetics of ordering in two dimensions. i. model systems. *Phys. Rev. B*, 28:2693, 1983.

[133] P.S. Sahni, J.J. Srolovitz, G.S. Grest, M.P. Anderson, and S.A. Safran. Kinetics of ordering in two dimensions. ii. quenched systems. *Phys. Rev. B*, 28:2705, 1983.

[134] M. San Miguel, V. M. Eguiluz, R. Toral, and K. Klemm. Binary and multivariate stochastic models of consensus formation. *Computing in Science & Engineering*, pages 67–73, 2005.

[135] T. C. Schelling. Dynamic models of segregation. *Journal of Mathematical Sociology*, 1:143–186, 1971.

[136] B. Schwikowski, P. Uets, and S. Fields. A network of protein-protein interactions in yeast. *Nature Biotechnology*, 18:1257–1261, 2000.

[137] J. Scott. *Social network analysis: A handbook*. Sage, London, 2nd edition, 2000.

[138] C. Sire and S. N. Majumdar. Coarsening in the q-state potts model and the ising model with globally conserved magnetization. *Phys. Rev. E*, 52:244–254, 1995.

[139] C. Sire and S. N. Majumdar. Correlations and coarsening in the q-state potts model. *Phys. Rev. Lett.*, 74:4321–4324, 1995.

[140] V. Sood and S. Redner. Voter model on heterogeneous graphs. *Phys. Rev. Lett*, 94:178701, 2005.

[141] L. Steels. A self organizing spatial vocabulary. *Artificial Life*, 2:319–332, 1995.

[142] L. Steels. The synthetic modeling of language origins. *Evolution of Communication*, 1:1–34, 1997.

[143] L. Steels. The origins of ontologies and communication conventions in multi-agent systems. *Autonomous Agents and Multi-Agent Systems*, 1:169–194, 1998.

[144] L. Steels and A. McIntyre. Spatially distributed Naming Games. In *ECAL 97*, 1997.

[145] S.H. Strogatz. Exploring complex networks. *Nature*, 410:268–276, 2001.

[146] K. Sznajd-Weron and J. Sznajd. Opinion evolution in closed community. *Int. J. Mod. Phys. C*, 11:1157–1165, 2000.

[147] B.K. Szymanski and G. Chen. Computing with time: From neural networks to sensor networks. *The Computer Journal*, 51(4):511–522, 2008.

[148] B.K. Szymanski and B. Yener (eds). *Advances in pervasive computing and networking*. Springer, New York, NY, 2004.

[149] K. Tafa, S. Puri, and D. Kumar. Kinetics of domain growth in systems with local barriers. *Phys. Rev. E*, 63(4):046115, 2001.

[150] S. Vasudevan, B. DeCleene, N. Immerman, J. Kurose, and D. Towsley. Leader election algorithms for wireless ad-hoc networks. In *Proc. of the 3rd DARPA Inf. Survivability Conf. and Exposition (DISCEX- III)*, 2003.

[151] D. Vinkovic and A. Kirman. A physical analogue of the schelling model. *Proc. Natl. Acad. Sci. USA*, 19:19261–19265, 2006.

[152] W.X. Wang, B.Y. Lin, C.L. Tang, and G.R. Chen. Agreement dynamics of finite-memory language games on networks. *Eur. Phys. J. B*, 60:529–536, 2007.

[153] D. J. Watts. Networks, dynamics, and the small-world phenomenon. *Am. J. Soc.*, 105:493–527, 1999.

[154] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393:440–442, 1998.

[155] D.J. Watts. A simple model of global cascades on random networks. *Proc. Natl. Acad. Sci. USA*, 99:5766–5771, 2002.

[156] A.L. Wilkes and N.J. Wade. Bain on neural networks. *Brain and Cognition*, 33:295–305, 1997.

[157] F. Wu and B. A. Huberman. Finding communities in linear time: A physics approach. *Eur. Phys. J. B.*, 38:331–338, 2004.

[158] M. Wykes, Q. Lu, G. Korniss, B. K. Szymanski, and Z. Rácz. Consensus-time distributions in models for opinion formation. 2006. to be published.

[159] F. Xue and P. R. Kumar. The number of neighbors needed connectivity of wireless networks. *Wireless Networks*, 10:169–181, 2004.

[160] J. Zhang. A dynamic model of residential segregation. *Journal of Mathematical Sociology*, 28:147–170, 2004.

[161] X. Zhu, M. Gerstein, and M. Snyder. Getting connected: Analysis and principles of biological networks. *Genes Dev.*, 21:1010–1024, 2007.