Helsinki University of Technology
Department of Biomedical Engineering and Computational Science Publications
Teknillisen korkeakoulun Lääketieteellisen tekniikan ja laskennallisen tieteen laitoksen julkaisuja
Espoo, 2009 REPORT A11

SOCIAL NETWORKS:
MODELING STRUCTURE AND DYNAMICS

Riitta Toivonen

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission of the Faculty of Information and Natural Sciences, Helsinki University of Technology, for public examination and debate in Auditorium E at Helsinki University of Technology (Espoo, Finland) on the 20th of February, 2009, at 12 noon.

Helsinki University of Technology
Faculty of Information and Natural Sciences
Department of Biomedical Engineering and Computational Science

Teknillinen korkeakoulu
Informaatio- ja luonnontieteiden tiedekunta
Lääketieteellisen tekniikan ja laskennallisen tieteen laitos

Distribution:
Helsinki University of Technology
Department of Biomedical Engineering and Computational Science
P.O.Box 9203
FI-02015 TKK
FINLAND

Tel. +358 9 451 4842
Fax +358 9 451 4833
http://www.becs.tkk.fi

Online in PDF format: http://lib.hut.fi/Diss/2009/isbn9789512297672/

E-mail: Riitta.Toivonen@gmail.com

ISBN 978-951-22-9766-5 (printed)
ISBN 978-951-22-9767-2 (PDF)
ISSN 1797-3996

Picaset Oy
Helsinki 2009

ABSTRACT OF DOCTORAL DISSERTATION	HELSINKI UNIVERSITY OF TECHNOLOGY P. O. BOX 1000, FI-02015 TKK http://www.tkk.fi

Author Riitta Toivonen

Name of the dissertation

Social networks: modeling structure and dynamics

Manuscript submitted 16th December 2008	Manuscript revised 3rd February 2009
Date of the defence 20th February 2009	
☐ Monograph	☒ Article dissertation (summary + original articles)

Faculty	Information and Natural Sciences
Department	Biomedical Engineering and Computational Science (BECS)
Field of research	Social networks
Opponent(s)	Dr. Fredrik Liljeros
Supervisor	Prof. Kimmo Kaski
Instructor	Dr. Jari Saramäki

Abstract

The study of networks of social interaction can be seen to originate from the work of Jacob Moreno in the 1920's. At the turn of the millennium new actors entered the field, researchers with a background in physics and computer science, who brought with them a new set of tools that could be used to collect and analyse large sets of data. Analysis of large scale social network data from various sources has increased our knowledge of the common features of various social networks, observed in networks of acquaintance and collaboration alike. The quantification and modeling of a particular feature of social networks, namely the tendency of individuals to form densely connected groups with relatively few links to individuals outside the group (called *communities* in complex networks theory), has taken large steps in recent years. Modeling these structures and their effect on social dynamics is a highly topical issue, relevant for fields such as spreading of epidemics or rumors and formation of opinions, with applications such as prevention of epidemics and marketing.

This thesis aims to increase our understanding of the structure of large scale social networks, and of dynamics unfolding in such networks, in several ways: 1) In order to answer a need for social network models that generate realistic structures at large scale, we introduce a model based on simple local mechanisms leading to community structure. 2) A thorough comparative study of models for social networks assesses the adaptability of the models to fit real social network data, and their success at reproducing prominent structural features of social networks. In discussing in detail two major approaches to modeling social networks, this study may promote the understanding between researchers from the two 'schools of thought'. 3) We study models of competing options, with focus on perhaps the most important feature of social network structure, namely communities, that had been largely lacking in earlier research.

Keywords social networks, network models, community structure, social dynamics, models of competing options

ISBN (printed) 978-951-22-9766-5	ISSN (printed) 1797-3996
ISBN (pdf) 978-951-22-9767-2	ISSN (pdf)
Language English	Number of pages 18 + 52 + appendices 78

Publisher Department of Biomedical Engineering and Computational Science

Print distribution Department of Biomedical Engineering and Computational Science

☒ The dissertation can be read at http://lib.tkk.fi/Diss/2009/isbn9789512297672/

VÄITÖSKIRJAN TIIVISTELMÄ	TEKNILLINEN KORKEAKOULU PL 1000, 02015 TKK http://www.tkk.fi

Tekijä Riitta Toivonen

Väitöskirjan nimi

Sosiaaliset verkot: rakenteen ja dynamiikan mallinnus

Käsikirjoituksen päivämäärä 16.12.2008	Korjatun käsikirjoituksen päivämäärä 3.2.2009
Väitöstilaisuuden ajankohta 20.2.2009	
☐ Monografia	☒ Yhdistelmäväitöskirja (yhteenveto + erillisartikkelit)

Tiedekunta	Informaatio- ja luonnontieteiden tiedekunta
Laitos	Lääketieteellisen tekniikan ja laskennallisen tieteen laitos (BECS)
Tutkimusala	Sosiaaliset verkot
Vastaväittäjä(t)	FT Fredrik Liljeros
Työn valvoja	Prof. Kimmo Kaski
Työn ohjaaja	TkT Jari Saramäki

Tiivistelmä

Sosiaalisten vuorovaikutusverkkojen tutkimuksen voidaan katsoa alkaneen Jacob Morenon työstä 1920-luvulla. Vuosituhannen vaihteessa verkkotutkimukseen liittyneet fysiikan ja tietotekniikan asiantuntijat toivat mukanaan joukon uusia työkaluja, joilla voidaan kerätä ja tutkia

Matkapuhelimet, sähköposti ja verkostoitumissivustot tarjoavat hyvin laajoja sosiaalista vuorovaikutusta koskevia aineistoja. Näiden aineistojen analyysi on lisännyt tietämystämme rakenteellisista samankaltaisuuksista erityyppisissä sosiaalisissa verkoissa, jotka voivat perustua jokapäiväiseen vuorovaikutukseen tai yhteistyöhön. Sosiaalisissa verkoissa muodostuu tyypillisesti tiheitä ryhmiä, joiden välillä on niukasti kytkentöjä. Tällaisia ryhmiä kutsutaan kompleksisten verkkojen teoriassa *yhteisöiksi*. Yhteisöjen rakenteen kvantifiointi ja mallinnus on erittäin ajankohtainen tutkimusaihe. Se on edellytyksenä yhteisöjen vaikutuksen selvittämiselle verkossa tapahtuviin prosesseihin, kuten tartuntatautien tai huhujen leviämiseen ja mielipiteiden muotoutumiseen. Tutkimuksen tulokset ovat sovellettavissa esimerkiksi epidemioiden hallintaan ja markkinointiin.

Tämä väitöskirja pyrkii lisäämään ymmärrystämme laajan mittakaavan sosiaalisten verkkojen rakenteesta ja niissä tapahtuvista prosesseista seuraavin tavoin: 1) Olemme kehittäneet lokaaleihin mekanismeihin perustuvan, yhteisörakennetta tuottavan verkkomallin, joka vastaa tarpeeseen tuottaa todenmukaisia sosiaalisen verkon rakenteita laajassa mittakaavassa. 2) Perusteellinen vertaileva tutkimuksemme sosiaalisten verkkojen malleista selvittää näiden mallien sovittumiskykyä ja niiden tuottaman rakenteen yhdenmukaisuutta havaintoaineiston kanssa. Tutkimuksemme kokoaa yhteen ja vertailee kahta eri lähestymistapaa noudattavia malleja, mikä saattaa osaltaan edistää ymmärrystä kahden ’koulukunnan’ välillä. 3) Kilpailevia vaihtoehtoja koskeva tutkimuksemme keskittyy sosiaalisten verkkojen kenties tähdellisimpään rakenteelliseen ominaisuuteen, yhteisörakenteeseen, jota ei ole juurikaan huomioitu aiemmassa tutkimuksessa.

Asiasanat sosiaaliset verkot, verkkomallit, yhteisörakenne, sosiaalinen dynamiikka, kilpailevien vaihtoehtojen mallit

ISBN (painettu) 978-951-22-9766-5	ISSN (painettu) 1797-3996
ISBN (pdf) 978-951-22-9767-2	ISSN (pdf)
Kieli englanti	Sivumäärä 18 + 52 + liitteet 78

Julkaisija Lääketieteellisen tekniikan ja laskennallisen tieteen laitos

Painetun väitöskirjan jakelu Lääketieteellisen tekniikan ja laskennallisen tieteen laitos

☒ Luettavissa verkossa osoitteessa http://lib.tkk.fi/Diss/2009/isbn9789512297672/

Preface

This thesis has been prepared during the years 2005–2008 in the Department of Biomedical Engineering and Computational Science (former Laboratory of Computational Science) at the Helsinki University of Technology. and it concludes my Doctor of Technology degree.

I was very lucky to have a supervisor with such an enthusiastic and bright attitude as Dr. Jari Saramäki. He has provided excellent guidance throughout my work and helped me beyond his share. I am also grateful to my other supervisor, Prof. Kimmo Kaski, who has provided an excellent environment for this work to take place, together with continuous encouragement and a warm-hearted attitude.

It was through Jukka-Pekka Onnela that I learned about the complex networks research at LCE, which lead me to decide to do a PhD here, and I owe much to him in my work. Special thanks go to the ever-dependable Lauri Kovanen - thanks for the excellent work and all the great music - and to Mikko Kivelä, who never overlooked an unclear argument. Thanks to Jörkki Hyvönen for the speedy code, reliable admin work and all those jokes; to Jenni Hulkkonen for making sure I left work now and then while finalizing my thesis; and to the rest of the gang, Ville-Petteri Mäkinen, Jussi Kumpula and Riku Linna, for useful software and interesting discussions.

I would also like to thank my collaborators Xavi Castelló, Victor Eguíluz, and Maxi San Miguel at the Institute for Cross-Disciplinary Physics and Complex Systems (IFISC) in Spain for sharing their expertise in physics-based sociodynamics, and for accommodating me on my work-filled but enjoyable visits to the Mediterranean. The many PhD students at IFISC, who came from various regions of the Spanish-speaking world, made my stays particularly memorable with their company and with our numerous shared discussions on cultures, ethics, and global issues.

Cheers to all the rest of my friends at LCE/BECS: Margareta Segerståhl the mad scientist, Kaija Virolainen the dancer, Eeva Lampinen, Senja Kojonen, Linda Kumpula, Jari Kätsyri, Mikko Viinikainen, Ville Lehtola the politician, Jaakko Riihimäki with whom we enjoyed literary discussions, and Elina Parviainen who wielded the rubber chicken.

Espoo, February 2009

Riitta Toivonen

Contents

List of publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I R. Toivonen, J.-P. Onnela, J. Saramäki, J. Hyvönen, and K. Kaski: *A Model for Social Networks*. Physica A, 371(2), 851 (2006).

II R. Toivonen, L. Kovanen, M. Kivelä, J.-P. Onnela, J. Saramäki, and K. Kaski: *A comparative study of social network models: network evolution models and nodal attribute models.* Helsinki University of Technology, Technical report A10, ISBN 978-951-22-9764-1 (2009)

III R. Toivonen, J. Kumpula, J. Saramäki, J.-P. Onnela, J. Kertész, and K. Kaski: *The role of edge weights in social networks: modelling structure and dynamics.* In Noise and Stochastics in Complex Systems And Finance, edited by J. Kertész, S. Bornholdt, and R.N. Mantegna. In *Proceedings of SPIE*, Vol. 6601, 66010B-1 (2007)

IV X. Castelló, R. Toivonen, V. M. Eguíluz, J. Saramäki, K. Kaski, and M. San Miguel: *Anomalous lifetime distributions and topological traps in ordering dynamics.* Europhysics Letters, 79, 66006 (2007)

V X. Castelló, R. Toivonen, V. M. Eguíluz, M. San Miguel: *Modelling bilingualism in language competition: the effects of complex social structure*. In *Proceedings of the 4th Conference of the European Social Simulation Association (ESSA 07)*. IRIT Editions (2007).

VI X. Castelló, R. Toivonen, V. M. Eguíluz, L. Loureiro-Porto, J. Saramäki, K. Kaski, M. San Miguel: *Modelling language competition: bilingualism and complex social networks.* In *The evolution of language; Proceedings of the 7th International Conference (EVOLANG7).* Edited by A.D.M. Smith, K. Smith, and R. Ferrer-Cancho, World Scientific Publishing Co. (2008)

VII R. Toivonen, X. Castelló, J. Saramäki, V. M. Eguíluz, K. Kaski, and M. San Miguel: *Broad lifetime distributions for ordering dynamics in complex networks.* Physical Review E, 79, 016109 (2009).

Author's contribution

The research reported in this thesis is a result of collaboration between Riitta Toivonen and the other authors of the included publications. Toivonen was the principal author of Publications I, II, III, and VII. In publications II and VII, she had the main responsibility in developing the ideas, in the analysis of the results, and in writing and revising the manuscript. She also implemented the computer programs and performed the numerical analyses entirely in Publications I and VII, and for a large part in Publications II, III, IV, and VI, and partly in V, and verified all numerical findings in Publication IV. As first author of publications I, II, III, and VII, she is also responsible for their written material, and as second author, she contributed significantly to the writing of Publication IV.

List of Abbreviations

ABMs	agent based models
NAMs	nodal attribute models
NEMs	network evolution models
ND	node deletion
LD	link deletion
RN	randomized network
WNC	weighted network with communities
WRNC	weight-randomized network with communities
BPDA	Boguñá-Pastor-Satorras-Díaz-Guilera-Arenas network model
DEB	Davidsen-Ebel-Bornholdt network model
KOSKK	Kumpula-Onnela-Saramäki-Kaski-Kertész network model
MVS	Marsili-Vega-Redondo-Slanina network model
TOSHK	Toivonen-Onnela-Saramäki-Hyvönen-Kaski network model
Váz	Vázquez network model
WPR	Wong-Pattison-Robins network model
ERGM	exponential random graph model
VM	Voter model of competing options
AB-model	Castelló-Eguíluz-San Miguel model of competing options
AS-model	Abrams-Strogatz model of competing options
mAS-model	microscopic Abrams-Strogatz model

List of Symbols

N	system size (in a network, the number of nodes)
N_{LC}	largest component size
R_{LC}	relative largest component size
N_0	initial system size
L or E	number of links in a network
t	time
i, j, k, l	indices of nodes or links
ν_i, v_i, or i	node i
l_{ij} or e_{ij}	link between nodes i and j
$\mathcal{N}(\nu_i)$	neighborhood of node i
k_i	degree of node i
c_i	clustering coefficient of node i
$k_{nn,i}$	average degree of nearest neighbors of node i
w_{ij}	weight of link e_{ij}
s_i	strength of node i (weight-sum of the links connected to i)
l_{ij}	shortest path length (minimum network distance) between nodes i and j
o_{ij}	overlap (of the neighborhoods of the nodes at the ends) of link e_{ij}
E_i or T_i	number of triangles around node i
r	assortativity coefficient
$\langle\ \rangle$	average over all nodes (or links) within the network
$\sum$	sum
$p(k)$	probability density distribution of degrees
$P(k)$	cumulative probability distribution of degrees
$P(t)$ or $f(t)$	fraction of runs still alive at time t
$U[0,k]$	uniform probability distribution of integers between 0 and k
σ_A	local density of nodes in state A
ι_A	local influence of nodes in state A
ρ	interface density
τ	survival time
f_{min}	fraction of agents in the minority state

Chapter 1

Introduction

Social networks are a hot topic of our age. The importance of networks is hyped in business and social life, as well as in the function of systems ranging from economical to biological. Social and economical systems are generally seen through a network-shaped lens, and even making friends in college is no longer called just making friends, but has been tagged *networking*. Social networking sites, such as Facebook, mySpace, and LinkedIn, abound on the web, allowing people to communicate with their friends, benefit from their network, and display their friendships for all the world to see. Such sites provide information on social networks on a scale that was not even dreamed of a couple of decades ago, providing data on social networks as large as millions of users (1; 2), complemented with even more precise information on the patterns of human interaction based on mobile phone calls (3; 4; 5). Such data sets allow us to make discoveries about the structure of social networks and their dynamics at a new scale.

It has long been known that different networks of social interaction have certain structural features in common (6; 7). Perhaps the most fundamental of these is the tendency of an individual's acquaintances also to be acquainted with one another. This phenomenon is called *clustering*, or *transitivity* (6), and it is seen in social networks as a higher prevalence of triangles than expected by chance (6). Friends tend to link with friends, eventually forming tight groups with many internal connections, called *communities* in complex networks theory (7; 8; 9; 10; 11; 12; 13; 14; 15; 16).

The structural universals observed in social networks are likely to be caused by fundamental processes of human interaction. We can search for these processes using agent based models (ABMs) of social network evolution. ABMs in general are based on the idea that simple and predictable local interactions can generate global patterns. A beautiful example of the emergence of group coordination from simple rules obeyed by each agent is the movement of a flock of birds, modeled in 1987 by Craig Reynolds (17). In his model, each bird reacts to the movement of only the birds closest to it, but the flock remains coherent. Similarly, in social networks, collective behavior such as the formation of communities, or the coordination of conventions (18; 19), can emerge from the individual actions of agents. In other words, they are *complex systems*. ABMs are used in this thesis both in generating network structure, and in simulation of social dynamics in a given network.

Simulation complements the two traditional foundations of science, theory and experimentation, and provides a way to perform virtual experiments in order to explain system level phenomena that depend on local interactions. The agent based approach does not always lend itself to analytical calculations, but typically relies on computer simulation. It can often be employed where analytical derivations are not feasible, making it an invaluable tool in the analysis of social dynamics, where the interactions are often so complex that analytical approaches fail or are too cumbersome. A shift away from analytical approaches towards agent based models can be seen in computational sociology, where modeling social processes as interactions among variables has in recent decades given way to modeling interactions among adaptive agents influencing one another - a shift “from factors to actors” (19).

It is generally acknowledged that individual decisions can be influenced by group pressure. In fact, I refused for a long time to join the happy gang of Facebook users, holding on to the fear that my personal information (*our* personal information as a network of friends) could somehow be misused by some suspicious third party. But my friends were already passing invitations to parties and events through Facebook, and if I didn’t join, I’d miss out. Finally, I succumbed to group pressure and created an account (eventually adding loads of superfluous information). This illustrates the importance of *social pressure*, or the influence of peers - friends and family, colleagues, or some other peer group - on individual choices. In particular, *communities* such as groups of close friends can have a decisive impact on the choices of the individuals within them. Social dynamics based on peer pressure is one of the two central themes in the work presented in this thesis. We will discuss models that concern the forming of opinions in society, based on the assumption that individual choices are dominated by the influence of acquaintances. Our research on social dynamics builds on the other theme, which concerns the emergence of the universal structural features of social networks from the local interactions of nodes. But before rushing on to details, let us take a look at the history of social networks research.

Networks of social interaction have been the subject of both empirical and theoretical study for several decades, starting with the work of Jacob Moreno and his colleagues in the 1920’s (7; 20; 21; 22), and leading to such famous concepts as Milgram’s six degrees of separation (23) and Granovetter’s strength of weak ties (24). The theory of graphs in general dates further back, to the work of Leonhard Euler in the 18th century[1]. A leap towards the field that eventually became 'complex networks' was taken in 1959 by the mathematicians Erdős and Rényi (25), as they began to consider *randomness* in graphs. Their network model, in which every pair of nodes has an equal probability of having a link between them, came to be known as the Erdős-Renyi (ER) random graph. The stochastic nature of the Erdős-Renyi networks made them seem more realistic in modeling interactions in economy, biology and society than the earlier fixed, deterministic networks, and they were long used as models of real world networks. Even today, the ER random graph is often used as a baseline, or a null model, as it is a network with basically no structure.

The idea of random networks was adopted by mathematical sociologists, who incorporated sociological hypotheses about the causes of link formation between actors in a network. Frank and Strauss (26) discussed in 1986 the first random networks that included *dyadic dependence*, i.e. in which links were not independent of all other links. Allowing for dependency enabled the inclusion of motifs such as triangles and stars. These so-called Markov random graphs of Frank and Strauss were generalized to $p*$ models (27; 28; 29; 30), also called exponential random graph models (ERGM) (31). Later ERGM models included more complicated dependence assumptions, such as that the probability of a link between two agents depends on their number of mutual friends (32). The philosophy behind ERGM models is to make inferences about to which extent nodal attributes and local structural features explain the global structures observed in empirical networks. Although the local structural dependencies can be thought to reflect processes at play in network evolution, this approach essentially excludes the evolutionary aspect of networks. A class of actor-oriented models proposed by Snijders in 1996 focused on network evolution. This focus on evolutionary mechanisms is shared by a vast number of later models belonging to the field that came to be called *complex networks* (22; 33; 34).

A new group of actors with a background mainly in statistical physics began to participate in research on social networks at the turn of the millennium. In a seminal paper from 1998, Watts and Strogatz showed that adding random links upon a regular structure could reproduce a feature observed in many real world networks - that they exhibit both high *clustering* (the tendency for the friends of an individual being acquainted as well) and short path lengths (only a few links need to be traversed in order to get from one node to another). Although this combination of properties was present already in the Markov random graphs (26), the paper by Watts and Strogatz boosted research on complex networks by making the physics community aware of the topic. In another seminal work in 1999, Barabási and Albert noticed that the number of links on web pages (their *degree*)

[1]Any introductory book on graph theory will recount how graph theory originated in Euler's insightful albeit disappointing solution to the search for the best route for a Sunday stroll on the bridges of idyllic Königsberg.

followed an unexpected distribution that could not possibly correspond to a random graph, and they decided to try an approach based on network evolution. They succeeded in reproducing the degree distribution with a growing network model in which a node gains links in proportion to the number of links it already has, based on the same principle as a 1955 model of city growth by Simon (35) that also successfully explains the WWW degree distribution (36). The resulting degree distribution in the Barabási-Albert (BA) model is *scale-free*, and in it some nodes have an immensely high degree. Whereas in the Erdős-Renyi networks all nodes are essentially equal, and no dramatic structure arises, here we were now dealing with *complex networks* - networks describing a system in which some global phenomenon *emerges* as a result of individual decisions. The evolutionary aspect of networks was seen to be a very useful component in explaining observed structure.

This so-called emergence of the scale-free distributions in a network was an exciting discovery, and it resulted in an explosion of network research by physicists. A plethora of network models appeared, many of them focusing on the scale-free degree distribution in the footsteps of Barábasi and Albert, proposing new mechanisms that might lead to scale-freeness, such as vertex-copying (37), and soon adding features such as high clustering coefficients or assortativity to emulate social networks (38; 39; 40; 41; 42; 43; 44) and link weights to imitate flows of traffic or materials in transport networks (45; 46). Immense data sets that contain information on the structure of social networks were gathered from the newly appearing electronic databases (47; 48; 49; 50; 51; 3; 4; 1; 2). In these networks, universal features were discovered that are particular to social networks, and that laid them apart from technological and transport networks and the Web.

Along with the wave of excitement about the structure of complex networks that took place at the turn of the millennium, interest was also rekindled in a field called *social dynamics*. The idea of using agent based models to study social dynamics had been around for several decades. The economist Schelling had done pioneering work in the 1970's on agent based models concerning social segregation, which will be discussed in Chapter 3. The physicist Galam had begun at around the same time to speak for the use of methods from physics in the study of social systems (52), although until recent years he mostly faced fervent opposition. The themes that together make up the field of social dynamics vary from the natural and social sciences to economy and marketing (34). Examples from social sciences include studies on the evolution of language (53) and diffusion of culture (54), and on the possibility of cooperation between people faced with social dilemmas (55; 56). On the commercial side, applications include the improvement of viral marketing strategies (57).

When I began working on this thesis, studies of social dynamics had for a large part been carried out either by employing so-called mean field calculations that assume that everyone interacts with everyone, or by assuming the interaction network to be a regular lattice or a purely random Erdős-Renyi network. Often the new network models - the scale-free BA networks and the small-world networks - were also adopted as a proxy of social networks. What was missing however were models that would take into account the very essence of social networks, their clustered structure. The time was ripe for asking the next question: what about communities? How could they be modeled, and how would they affect dynamics that depend on the network structure, such as rumor spreading, fashions, and the forming of opinions? These are the kind of questions that are the topic of this thesis.

This introductory part of the thesis is further divided into two parts, which deal with the dual aspects of the study of social networks: modeling structure and dynamics. Chapter 2 discusses the modeling of the structure of social networks. The central question here is the emergence of structural properties, such as communities, out of local rules. We propose a model for social networks, and carry out a comparative study of a class of stochastic network models that are based on a variety of assumptions about how social ties are formed. Chapter 3 in turn deals with social dynamics, discussing our research on the competition of two options in a networked population, again focusing on community structure. In Chapter 4, I summarize the results obtained in this thesis, and discuss future directions.

Chapter 2

Modeling the structure of social networks

At the simplest, the network representation of social interactions consists solely of the structure or *topology* of interaction: do these two individuals interact or not? The ties between individuals who together form a social network can be defined by 1) acquaintance, or 2) participation in common activities. A useful proxy of acquaintance networks can be obtained for example by observing mobile phone calls between individuals (3; 58; 5). The latter type of networks are exemplified by collaboration networks, such as the network of scientists who co-authored a paper (48), actors who appeared on the same cast (50), executives sitting on the same board (59), or jazz musicians playing in the same band (49). While these networks are obviously very different in content, there are many similarities in their structure. The introduction discussed two typical characteristics of social networks: clustering and community structure. Another notion concerning social ties is that popular (highly social, or actively collaborating) people are often acquainted with other popular people, while people with fewer friends tend to group among similar individuals. This concept has also been empirically verified (60), and has been given the name *assortativity*. One of the typical features observed across various social networks, and indeed complex networks in general, has found its way into the public imagination. By far the most frequent question that people will ask me upon hearing that I study social networks is one inspired by the famous phrase 'six degrees of separation': Is it true that any two persons in the world are linked through at most six intermediate friends? Although this bewildering idea does not hold true to the letter, the typical separation between two people in a given social network can indeed be only a handful of steps.[1] The six-degrees concept

[1] In a 2007 study by Jure Leskovec and Eric Horvitz, the largest distance observed between any pair of 240 million Microsoft Messenger users worldwide was 29 (1). 78 percent of all pairs of individuals were linked through at most 7 steps. The actual distances could be shorter, because the

illustrates the fact that individuals in social networks are interconnected through relatively short paths.

Why are such features universally observed in different types of social networks? Do they arise as a result of the interactions between individuals? Modeling social networks can help us answer these questions. We can pose hypotheses about the mechanisms with which social networks form and evolve, and test whether these mechanisms lead to the observed structures. Short path lengths have been seen to arise easily whenever 'long-distance' links are present, i.e. when the possibility of random global connections is present in network evolution process. The emergence of communities seems to be a far more complicated issue. The work presented in Publications I, II, and III addresses this question.

The characterization, analysis and classification of networks relies on measurements that are capable of expressing their most relevant topological features (61; 22). We begin this chapter by reviewing commonly used measures for social networks in Section 2.1. Section 2.2 presents an agent based network model by the author and colleagues that was one of the earliest large scale network models with community structure. Finally, in Section 2.3 we discuss and compare recently developed agent based models for the study of social network structure.

2.1 Characterization of social networks

In order to discuss the structure of networks, we need to be familiar with the relevant measures and concepts, which will be reviewed in this section. Social contacts can be represented by a *complex network* in which nodes represent individuals and links represent the ties between them. Complex networks in general fall into four main types: *weighted digraphs* (directed graphs), *unweighted digraphs*, *weighted graphs*, and *unweighted graphs* (61). A digraph can be transformed into a graph by a symmetry operation, and a weighted (di)graph into a (di)graph by thresholding. In the context of social networks, directed graphs can be thought to depict for example the networks of phone calls, messages sent, or favors done between individuals. The underlying network of social contacts can nevertheless often be meaningfully analysed as consisting of mutual ties, and depicted by an undirected graph. In this work, we will only deal with undirected graphs, both unweighted and weighted, and generally refer to them using the terms *network* and *weighted network*. The concepts from graph theory and complex networks theory used in this work are defined below, based on references (61) and (62).

users will have links through other media except Messenger; on the other hand, only a fraction of all people are Messenger users. Milgram arrived at approximately six steps in his famous experiment in which people living in states like Kansas or Nebraska were asked to pass on a letter to a prominent person in a large city such as Boston or New York, through individuals they knew on a first name basis (23).

Graph. A graph G consists of a set $V(G)$ of *vertices* or *nodes*, a set $E(G)$ of *edges* or *links*, and a relation that associates with each edge two nodes (not necessarily distinct) called its *endpoints*. Each node can be identified by an integer $i = 1, 2, ..., N$. If multiple links are not allowed, each link can be identified by an an unordered pair $\{i, j\}$ that represents a connection between the nodes i and j. A graph that does not contain multiple links between nodes i and j or self-links $\{i, i\}$ is called *simple*. We consider only simple graphs in this work, and also use the notation l_{ij} to denote the link $\{i, j\}$.

Weighted graph. A *weighted graph* additionally includes a mapping $\omega : E(G) \to \mathbb{R}$ which associates a weight with each link.

Adjacency, neighbors. Nodes i and j are said to be *adjacent*, or *neighbors*, if the link set $E(G)$ contains link $\{i, j\}$. The link set of a graph G without multiple links can be represented by an *adjacency matrix* A, in which the elements $a_{ij} = a_{ji} = 1$ if $\{i, j\} \in E(G)$, and $a_{ij} = a_{ji} = 0$ otherwise.

Neighborhood. The *neighborhood* $\mathcal{N}(i)$ of node i consists of the nodes adjacent to i.

Dyad. A *dyad* is a pair of nodes, not necessarily adjacent.

Path. A *path* is a simple graph whose nodes can be ordered so that two vertices are adjacent if and only if they are consecutive in the list. The *length* l of a path is its number of links.

Geodesic path. A *geodesic path* or a *shortest path* between nodes i and j is a path of minimal length (not necessarily unique) containing i and j.

Subgraph. A *subgraph* of a graph G is a graph H such that $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$ and the assignment of endpoints to links in H is the same as in G. We then write $H \subseteq G$ and say that "*G contains H*".

Connectedness. A graph G is said to be *connected* if each pair of nodes in G belongs to a path; otherwise, G is *disconnected*. Similarly, two non-adjacent nodes i and j said to be connected if they belong to a path.

Component. A *component* of a network is a maximal connected subgraph. In this work, we often consider the *largest component* of a graph. The size of the largest component is denoted by N_{LC}.

k-Clique. A *k-clique* is a set of k pairwise adjacent nodes. A *triangle* is a 3-clique.

Degree. The number of neighbors of a node i is called its *degree* k_i. An isolated node has degree zero. The degree distributions $p(k)$ of large social networks are often highly skewed, with some nodes having very high degrees.

Clustering coefficient. A measure of local triangle density, the (unweighted) *clustering coefficient* c_i (61) (Fig. 2.1), describes the extent to which the neighbors of node i are "acquainted with one another": if none of them are adjacent, $c_i = 0$, while if all of them are adjacent, $c_i = 1$. For a node i with degree k_i and belonging

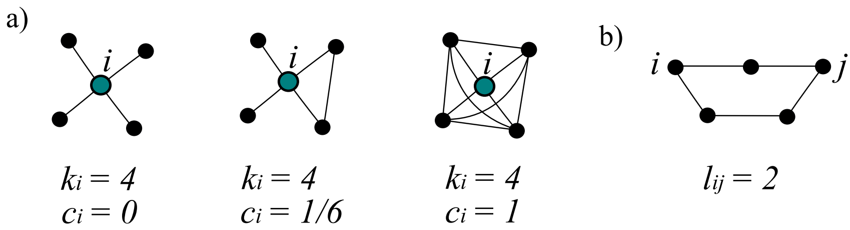

Figure 2.1: (a) Clustering coefficient $c_i = \frac{T_i}{k_i(k_i-1)/2}$ of node i with degree k_i and participating in T_i (undirected) triangles. (b) Shortest path length l_{ij} between nodes i and j.

to T_i triangles, the clustering coefficient is defined as

$$c_i = \frac{T_i}{k_i(k_i-1)/2}, \tag{2.1}$$

where the denominator $k_i(k_i-1)/2$ expresses the maximum possible number of triangles to which i could belong given its degree. The clustering coefficient is not defined for nodes with degree $k < 2$. The average clustering coefficient, averaged over all nodes with $k \geq 2$ in the network, is denoted $\langle c \rangle$. $c(k)$ denotes the average clustering coefficient of nodes having degree k. The curve $c(k)$ is called the *clustering spectrum*.

Note that a high average clustering coefficient is not always an indication of modular structure. For example, a regular 2-dimensional lattice with each node having eight nearest neighbors has a high clustering coefficient, although its structure is homogeneous.

Assortativity.

Social networks typically show a positive correlation (also called assortativity) between the degrees of adjacent nodes (degree-degree correlations), 'popular people know other popular people'. Networks with negative correlations between degrees of adjacent nodes, which is typical for networks designed for the flow of information or traffic, are called dissortative. One way of quantifying this effect is using the Pearson correlation coefficient, also called the *assortativity coefficient* r (47):

$$r = \frac{\sum_e k_i k_j / L - \left[\sum_e \frac{1}{2}(k_i + k_j)\right]^2 / L^2}{\sum_e \frac{1}{2}(k_i^2 + k_j^2)/L - \left[\sum_e \frac{1}{2}(k_i + k_j)\right]^2 / L^2},$$

where L is the total number of links in the network, and $\sum_e$ denotes summing over all links. A positive value of the assortativity coefficient signifies that the nodes with a large number of ties are connected to one another more likely than would be expected by chance, and nodes with a small number of ties are connected more likely with one another. A negative value signifies that mostly nodes with small degree are connected to the large connectors, which are not directly linked between themselves. Assortativity can also be quantified using the measure *average nearest neighbor degree* $\langle k_{nn}(k)\rangle$, found by taking all nodes with degree k, and averaging the degrees of their neighbors. If the curve $\langle k_{nn}(k)\rangle$ plotted against k has a positive trend, nodes with high degree typically also have high-degree neighbors, hence the network is assortative. In part, assortativity could be explained by the fact that a social network typically contains communities of different sizes, and the average degree of the individuals in each is likely to depend on community size. Hence, connected individuals would tend to have similar degree.

Overlap. Several measures from network sociology describe the overlap of the neighborhoods of two nodes. The predecessor of such measures is the Jaccard coefficient dating from 1901 (63), which does not concern networks but deals with the overlap of features of two actors. The dyad-wise shared partners (DSP) measure (32) simply counts the common neighbors of all dyads in the network. The edge-wise shared partners (ESP) measure (32) is similar but only takes into account connected dyads. Another definition from the same family of measurements, presented in (4), examines the fraction of all possible triangles between two adjacent nodes i and j based on their degree, taking into account that part of their degree is spent on the mutual link. This measure, called *overlap* O_{ij}, varies between 0 and 1 and is defined as

$$O_{ij} = \frac{n_{ij}}{(k_i - 1) + (k_j - 1) - n_{ij}}, \tag{2.2}$$

where n_{ij} is the number of neighbors common to both nodes i and j, and k_i and k_j are their degrees (see Fig. 2.2). Overlap is defined for edges with at least one end having degree $k > 1$. Within a cluster, adjacent nodes tend to share many neighbors, and thus overlap is high, while edges between communities will often have low or zero overlap values.

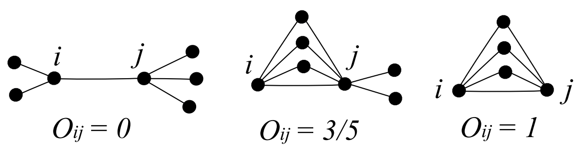

Figure 2.2: Overlap O_{ij}.

Communities. A particular feature of social networks is that they are organized into groups of densely interlinked individuals, or communities. Figure 2.3(a) illustrates the community concept with a well-known example, the *Zachary's Karate Club* network (64), which is a social network of friendships between 34 members of a karate club at a US university in the 1970. Two communities can be discerned by eye in the graph, each of which has certain leader individuals that are linked to most of the others in the community. One of the subgroups eventually broke apart from the club due to internal discord. This small network with relatively clear community structure is often used as a benchmark test for community detection methods.

A large variety of algorithms exist for detecting communities in a network, along with a plethora of often implicit definitions of what communities are (7; 16; 8; 9; 10; 65; 11; 12; 13; 14; 15; 66; 67; 68; 69). Perhaps the simplest possible measure of community structure is the number of cliques, or fully connected subgraphs, of different sizes in the network. Other definitions are less strict. A local deterministic method for detecting communities, called *clique percolation* (14), allows some links to be missing from a group of nodes, and defines communities as overlapping chains of smaller cliques. Radicchi et al. provided a precise definition of the intuitive idea that a community is a subnetwork in which internal connections are denser than external connections (11). For a node i in subgraph V, they use the term *in-degree* to signify the number of links from i to other nodes within V, and *out-degree* to signify the number of links from i to nodes not in V. Note that these terms do not refer to directed networks. They then define that a subgraph V is a community *in the strong sense* if the in-degree exceeds out-degree for *every* node within the community, i.e. if $k_i^{in}(V) > k_i^{out}(V) \quad \forall i \in V$ (see Fig. 2.3(d) for an example). The *weak* definition requires that in-degree exceeds out-degree only *on average*: a subgraph V is a community in the weak sense if $\sum_{i \in V} k_i^{in}(V) > \sum_{i \in V} k_i^{out}(V)$ (Fig. 2.3(d)).

Many heuristic algorithms are based on the intuitive idea that communities have a large number of internal connections compared to the number of links leading to nodes outside the community. A popular criterion for the partition of a network into communities is *modularity* (13), which favors grouping together subsets of nodes that are densely connected and between which links are sparse[2]. This criterion has, however, been shown to be unable to detect small communities (70). The search continues for valid definitions of communities and for reliable and fast methods for their detection.

[2]Modularity Q is defined as $Q = \frac{2}{K}\sum_{s=1}^{m}(l_{ss} - [L_{ss}])$, where K is the degree sum of the network, m is the number of communities, l_{ss} is the number of links in community s, $[L_{ss}] = K_s^2/2K$ is the expected number of links within community s for a random network with the same degree sequence, and K_s is the sum of degrees within s.

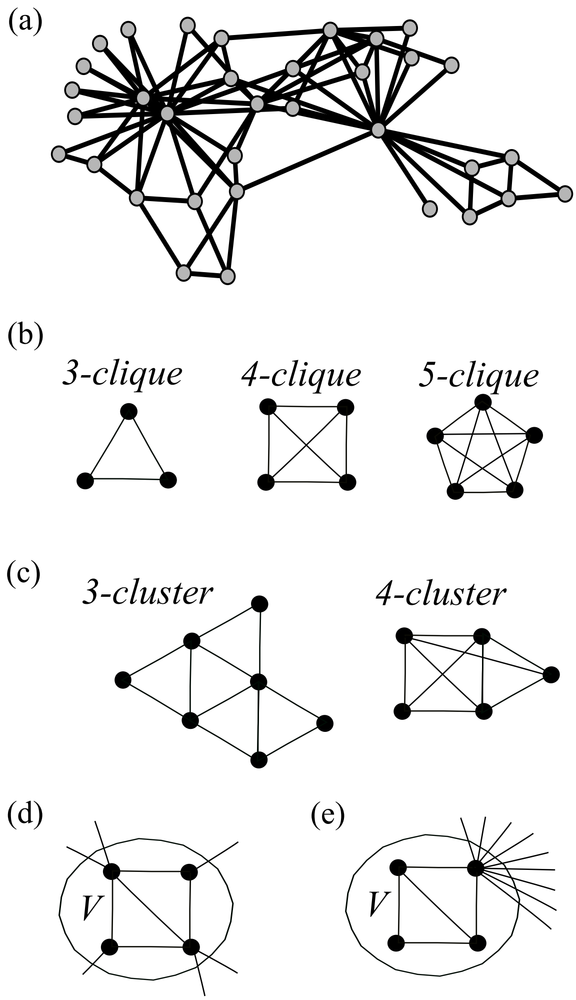

Figure 2.3: (a) Community structure in the *Zachary's Karate Club* network (64), visualized using Himmeli (71). (b) Cliques (fully connected subgraphs). (c) Two examples of k-clusters (k-clique-communities) as defined in the k-clique-percolation method (14): a 3-cluster with 7 nodes and a 4-cluster with 5 nodes. (d) and (e) Examples of communities in the strong sense (d) and weak sense (e) as defined by Radicchi et al. (11).

Table 2.1: Properties of several large empirical social acquaintance networks.

Network	N	L	<k>	<c>	<r>	<l>	$P(k)$	$c(k)$
MSN (1)	1.8×10^8	1.34×10^9	14.9	0.137		6.6	pow law with exp cutoff	$k^{-0.37}$
MCG (4)	3.9×10^6	6.5×10^6	3.3	0.26	0.23	14.5	pow law with exp cutoff	k^{-1}
MCG (58)	2.5×10^6	5.4×10^6	4.3				pow law	
lastfm-fin (72)	8003	1.7×10^4	4.2	0.31	0.22	7.4	lognormal	
email (51)	1133	5.5×10^3	9.6	0.22	0.08	3.6	exponential	

2.2 The TOSHK model for social networks and further developments

It has long been known that social networks are characterized by short path lengths (23) as well as a high prevalence of transitivity (6), measured by triangle count or the average clustering coefficient. Recently gathered empirical data on large scale social networks, such as those based on communication via mobile phone (4; 5) and Microsoft Messenger (1), has revealed among other things that the degree distributions are surprisingly broad, often characterized as power laws with exponential cutoff (4; 1) (Table 2.2). Moreover, in contrast to technological or biological networks, social networks tend to have positive degree degree correlations, i.e. they are assortative (60; 47). Importantly, social networks tend to consist of tightly clustered groups of nodes. Table 2.2 lists properties of a few of the recently obtained large scale empirical data on social networks. The Microsoft Messenger network (MSN) (1) is based on communication in Messenger. The mobile call graph (MCG) by Onnela et al. (4) consists of the largest component of the network of reciprocated pairs of phone calls. The other mobile phone call data is from Lambiotte et al. (58). The *lastfm-fin* network is the friendship network collected from the web site *www.last.fm* by the author and colleagues. The properties of the *email* network are calculated for the largest component of a network based on email communication at a Spanish university (51).

Despite long-time efforts in the analysis and modeling of social networks, when this work was begun four years ago there was still a substantial lack of models for large scale social networks. Apart from a spatial model by Wong et al, published in 2005, the author is aware of no other models feasible in large scale that would have produced community structure. Some of the early social network models presented by physicists were designed to produce high average clustering coefficients (39; 41; 40), but even they did not pay attention to community structure. High clustering had already been achieved by an earlier dynamical model based on triangle formation (the DEB model, (40)) but it did not seem to produce much community structure. Another model that produced high clustering (the MVS model, (41)) in turn produced only relatively weakly assortative networks. These models are discussed in Section 2.3 and in Publication II. Another universal feature of social networks that was only addressed by a handful of models is assortativity (47; 42). In order to respond to the need for more realistic models for social networks, the author and colleagues set out to design a model that could produce community structure and assortative networks in large scale. Our model is discussed in Publication I, and further developments in Publication III.

In developing the TOSHK model (73) for social networks we aimed at reproducing many of the features observed in empirical social networks, while keeping the model as simple as possible. Most importantly, the model should produce networks structured into communities, i.e. densely connected subgroups with few connections between them. It was also required that the degree distribution should have a broad tail, and that the networks should exhibit high average clustering coefficients and assortativity, and that average path lengths should grow slowly with network size in accordance with the small world phenomenon.

A growing model was selected to enable analytical derivations of some of the network characteristics. A growing model can be motivated as a model for social networks in several contexts. For example, in a network of co-authorship based on publication records, new links form but old ones remain. Similarly, in online social networking systems people rarely remove links, and new users keep joining the network. The growth mechanisms of the TOSHK model are selected to imitate the way people might join an already established social network. The model is not intended to simulate the evolution of a social network ab initio. The algorithm grows by adding at each time step a new node that links to the network via two processes (Fig. 2.4): (1) linking to one or more initial contacts selected uniformly randomly, and (2) possibly linking to one or more neighbors of the initial contact. Following a random edge is likely to lead to a node with a large number of links, which implies that the local search causes the new node to link preferentially to high degree nodes. However, the preference is not exactly linear in degree, both because the edge that is followed is not uniformly randomly selected, and because the positive degree-degree correlations in the network imply that the neighbors of small degree nodes also tend to have small degree. Roughly speaking, the neighborhood connections contribute to the formation of communities, while the new node acts as a bridge between communities if more than one initial contact was chosen.

The local nature of the second process gives rise to high clustering, assortativity and community structure. The TOSHK model showed that local attachment could indeed produce community structure. Very large cliques are not observed, however, if the maximum number of triangle formation steps from an initial contact is kept small. In the comparative study of Publication II, the TOSHK model did not fare particularly well, partly because adaptability was restricted by our choice of keeping to the uniformly random distribution for the number of links from each initial contact. With this choice, low link density forced the number of triangle formation steps to be very low.

Further research has shown that including link weights enables the formation

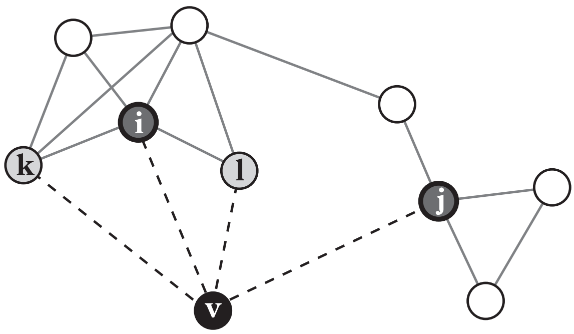

Figure 2.4: Growth process of the TOSHK network. A new node v links to one or more randomly chosen initial contacts (here i, j) and possibly to some of their neighbors (here k, l). Figure taken from Publication I.

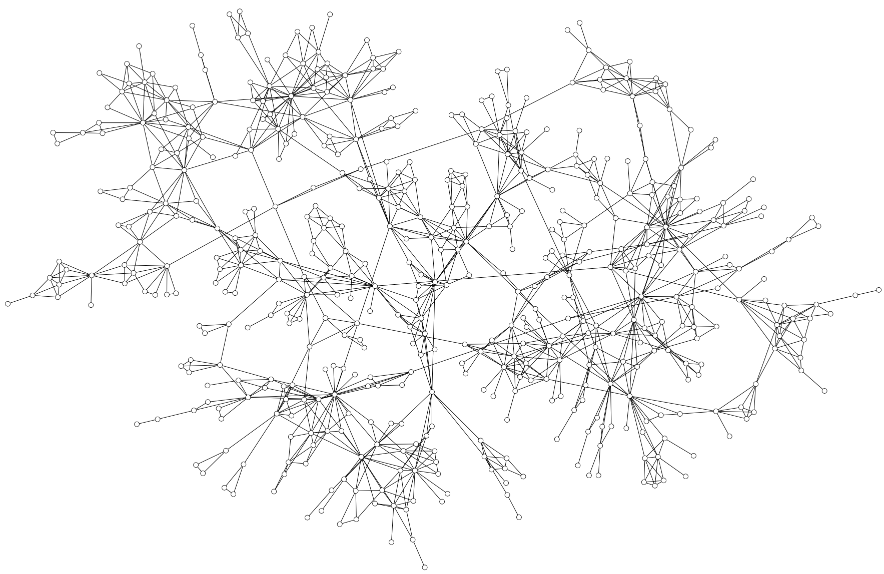

Figure 2.5: A network with $N = 500$ nodes produced by the TOSHK model shows communities of various sizes. Figure taken from Publication I.

of much clearer community structure, demonstrated by the KOSKK model (74; 75). In the KOSKK model, internal links within communities are strong, and weak links connect the communities to one another, in agreement with Granovetter's weak ties hypothesis (24). This weighted model has been employed by the author to study how the correlations between link weights and topology affect opinion formation dynamics; this is discussed in Section 3.3.

2.3 Modeling approaches

One of the major approaches to modeling social networks, the exponential random graph models (ERGMs), were mentioned in the introduction. ERGMs can be broadly described as probabilistic models of network structure. ERGM models can be used to pose questions about correlations between structural features. For example, if structures of type A are present more often than would be expected by random, are also structures of type B present more often? ERGM models can also identify characteristics of the agents (actor attributes) that could explain observed network structures. For example, does homophily by race explain the network structure to a significant extent?

Our work takes an alternative approach with focus on network evolution. We mainly focus on models that we categorize in Publication II as network evolution models (NEMs) in which the network evolves according to a specified set of (mainly local) rules concerning the addition and deletion of nodes and/or links in the network (Fig. 2.7). Network evolution models attempt to answer the question of whether the universal properties of social networks can be modeled with simple local rules that the individuals follow. How could the structures observed in real networks emerge from the actions of individuals? Publication II presents NEMs and ERGMs alongside each other for the first time in the same paper, which we hope will promote understanding and discussion between researchers following each of the two approaches. NEMs can be further divided between *growing models*, in which links and nodes are simply added until the network has the desired number N of nodes, and *dynamical models*, in which the steps for adding and removing ties on a fixed set of nodes are repeated until the structure of the network no longer statistically changes.

A third category, *nodal attribute models* (NAMs), consists of models in which link probabilities depend only on nodal attributes, typically via homophily (76), the tendency for like to interact with like. This category also includes any ERGM models that do not incorporate structural dependencies.

Figure 2.6 places in these categories the models that we study in Publication II. We include two nodal attribute models (WPR (77) and BPDA (78)). Three of the network evolution models are dynamical (DEB (40), MVS (41), and KOSKK (74)), and two are growing (Váz and TOSHK). All of the network evolution models we study are based on a combination of triadic closure (24) and global connections. There are also other models that fall into the category of network evolution models but are based on different ideas than triadic closure, such as the networked Seceder (79) model, in which each individual seeks to differ as much from the average as possible. Models in which the network topology co-evolves together with the nodal attributes have also been proposed (see for example (80; 81; 82)). The descriptions of the various network models in Publication II will hopefully serve as a useful reference.

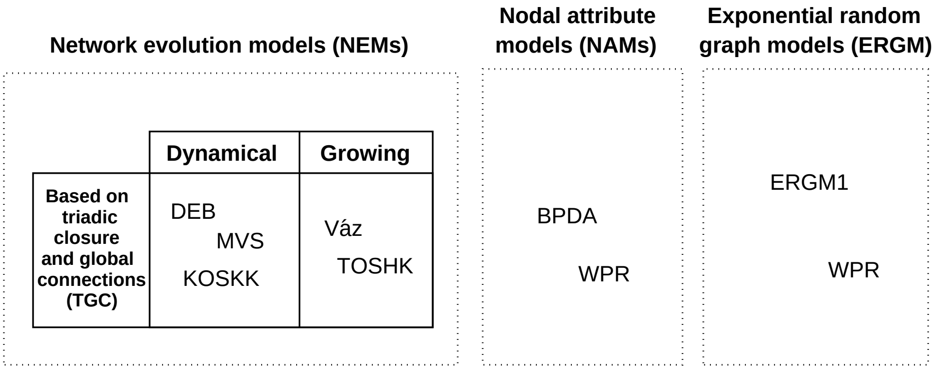

Figure 2.6: Categories of social network models: Network evolution models (NEMs), nodal attribute models (NAMs), and exponential random graph models (ERGMs).

The network structures produced by the models are examined in detail. Although some were designed mainly to produce high clustering, in order to get a better picture of the models we also consider several other characteristics. The models are compared systematically by unifying some of their average properties, and then comparing the resulting higher order statistics such as degree distributions, clustering spectra, geodesic path lengths, and community structure. Assessing the adaptability of the NEMs to data in this extent had not been done before, although for the ERGMs it has been common practice. This is partly explained by the different approaches - the ERGMs attempt to make inferences based on specific data sets, whereas the NEMs attempt to test whether general structural characteristics can be produced by an assumed network evolution mechanism.

We find that many of the NEMs based on triadic closure and global connections produce degree distributions and clustering spectra that match empirical data fairly well, but not very high assortativity nor very clearly clustered structure. The NEM that includes edge weights, KOSKK, is an exception in that it generates very clear community structure. On the other hand, the nodal attribute models successfully produce highly clustered and assortative networks and a structure of loosely connected, relatively dense clusters, but not very realistic degree distributions nor clustering spectra. High average clustering coefficients arise in both types of models by design.

To complement the comparison of the models, we compare in detail the different mechanisms for creating and deleting links in the selected dynamical NEMs (Fig. 2.7). The triangle formation step of linking two nodes at geodesic distance two can be implemented in different ways, two of which are compared here (T1 and T2, see Fig. 2.7). Link deletion (LD) refers to deleting randomly chosen links; node deletion (ND) implies that all of the links of a node are removed at

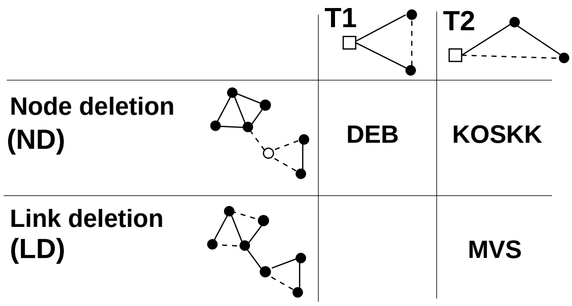

Figure 2.7: Methods of link addition and deletion in dynamical network evolution models (dynamical NEMs) based on triadic closure and global connections.

the same time. This could be interpreted as an individual leaving the network and a newcomer without any links taking its place. We implement all combinations as variants of the simplest dynamical NEM (the DEB model). Details in the employed mechanisms affect the resulting distributions of degree and clique sizes as well as assortativity.

Many of the models produce structures surprisingly close to empirical observations. Even models with only two parameters are able to reproduce many features of empirical social networks such as high clustering, assortativity, sometimes even a very reasonable distribution of clique sizes. On the other hand, it must be admitted that none of the models are able to faithfully adapt to all the selected features of the data, nor match all potentially relevant higher order structures.

Chapter 3

Modeling dynamics of competing options

This chapter presents an overview of a subset of the vast field of social dynamics that relates to peer pressure and competing options in society. We begin in Section 3.1 by highlighting the usefulness of agent based modeling in the study of social systems, and then briefly review various types of models of competing options. Section 3.2 discusses the role of the interaction network in social dynamics. Section 3.3 focuses on the work by the author and colleagues on specific dynamical models that can be interpreted as opinion formation models or language competition. These are explored in complex social networks, with focus on the effects of the mesoscopic structure of the interaction network on the dynamics.

3.1 Overview of social dynamics

Agent based modeling allows us to perform thought experiments, and can give us insights that would otherwise be difficult to obtain. As an example, let us consider one of the early agent based models from sociology, presented by the Nobel prize-winning economist Thomas Schelling in 1971, that concerns residential segregation. Individuals often favor living among others belonging to their own ethnic group. The various motivations include attachment to group identity and group culture, as well as stereotypes and expectations that people of the same ethnic would be more likely to provide mutual support and be more welcoming. For modeling the effect of such preferences on the distribution of members of different ethnic groups in residential areas, the detailed causes of the preferences are unimportant. Schelling's simple agent based model of housing tests the outcome of such preferences (83). Schelling simplified the problem by representing the city with a checkerboard of households, either black or white or empty, and let individual households move to free locations according to their preferences.

If too many of the neighbors of a household were of a different color, it would relocate to a new site. What Schelling observed was that even a mild preference for residing among the same ethnic group could be amplified by the dynamics of relocation, and cause highly segregated residential areas to form. Simply the desire to avoid becoming a small minority could lead to segregation, giving the impression of racist attitudes.

Another interesting example of applied social dynamics comes from a political scientist dealing with the study of cooperation. Robert Axelrod (84; 85) has experimented extensively with agent based models of social dilemmas, situations in which the individually optimal choice is not best for the common outcome. Such situations can be modeled by a game called the Prisoner's dilemma (PD), in which two players have to make a choice either to cooperate or to defect, without knowing what the other player will do. For each individual player, defection is the best option regardless of what the other person chooses, but both players will end up collectively worse off if they choose to follow their optimal strategies. Axelrod has called the Prisoner's dilemma game the "E. coli" of social sciences, because it can be used to model a large variety of situations, ranging from live-and-let-live strategies in trench warfare to success in personal relations (84; 85).[1] In order to determine which kind of strategies would be most successful in repeated interactions of two players, Axelrod arranged two computer tournaments in which contestants were asked to send in strategies for the iterated PD game, that would be played against one another. The entry that won both tournaments, called TIT-FOR-TAT, employed a strategy that combined reciprocity and retaliation - starting out nice, but thereafter retaliating for any defection, and responding to cooperation with cooperation. One might argue that everyday experience or empirical studies could have told us that reciprocity and punishment for noncooperation are useful in promoting cooperation, and that they are also widely used in various social situations. But simulation helped in identifying a simple and effective implementation of these concepts. Another benefit of simulation here is related to validation: Among the strategies sent to the contest, the very simple TIT-FOR-TAT strategy was generally superior to more complicated and less forgiving strategies. Later, by employing genetic algorithms for generating a vast number of random strategies and playing them against one other, Axelrod was able to validate that combinations of reciprocity and retaliation similar to those employed in TIT-FOR-TAT are in fact generally very efficient strategies, and the success of TIT-FOR-TAT was not dependent on human factors and expectations behind the strategies submitted to the tournaments.

The dynamics studied in this thesis are based on the phenomenon of social influence or peer pressure. The fact that group opinion influences individual decisions has been verified in many experiments that often amusingly demonstrate the

[1] Various social situations can also be described by other dilemma games, see for example (86).

willingness of people to make ridiculous claims or do silly things in order to avoid differing from others. Apparently we humans have a strong tendency to think that what a large number of people are doing must be reasonable. Broadly speaking, models in which agents choose between options and are affected by peer pressure can be categorized as opinion formation models. They are typically very general and abstract, and are therefore applicable to many different situations. A recent review of social dynamics (87) divides the research activity concerning such processes into three major branches, namely opinion dynamics, cultural dynamics and language dynamics. The distinction is often subtle and arbitrary, and in fact we will discuss a model in terms of opinion formation that was inspired by language competition. The options from which the agents choose can be thought of as opinion, but often equally, one could imagine any competing options - response to a political question, set of cultural features, or a correspondence between objects and words.

In the real world we are often faced with discrete choices. Voting for one of a limited number of candidates, or buying computer with either a Windows, Linux or Mac operating system are examples of such choices. Our opinions on complex issues, such as whether to accept the use of nuclear energy, could be uncertain and vary over time based on many factors. Such opinions could be considered as a continuous variable (until a choice has to be made). In mathematical modeling, opinions are represented by numbers, either discrete or continuous. Here, we will focus on discrete opinions.

Everyday experience confirms that the opinions of those individuals with whom you have lately discussed an issue easily affect individual opinions. Generally, in opinion formation models, the agents choose an option (opinion) from a small set of variables, based on the influence of their peers. As agents interact, they generally tend to become more alike. With repeated interactions, agents begin to form homogeneous groups, eventually leading either to consensus or to a fragmentation of society in which homogeneous groups exist that no longer interact. This process of consensus formation, or the reaching of agreement, is the focus of many opinion formation models, and also a central topic in this thesis.

Some of the opinion formation models that have received quite a lot of attention in the physics literature include Voter type models (88; 87), majority rule models (89; 90), the Sznajd model (91), and bounded confidence models (87; 92). In Voter type models (88; 87), some of which are discussed in more detail in section 3.2, an individual is likely to adopt the opinion of the majority of its acquaintances, but can also occasionally be convinced by the minority. Majority models (89; 90) are based on the idea that as people discuss a topic in groups, the participants will be influenced by the majority opinion of that temporary group. The Sznajd model dynamics (91) in turn rests on the idea that two acquainted individuals who agree on an issue will be able to convince their friends on that issue as well. All of these models are based on different aspects of the the persuasive power of agreeing groups. Bounded confidence models add the element that individuals are less likely to interact with others that are too different from themselves. Many variants of each of these types of models exist. This work focuses on variants of the Voter model.

One of the above examples concerned a large number of agents that were free to move about in a lattice (checkerboard) representing geographical space. The second concerned a single pair of agents, repeatedly interacting with one another. In the rest of this chapter, we will focus on dynamical models in which agents are part of a network of interactions that is thought to be unchanging (either a lattice or a complex network), and are repeatedly interacting with their network neighbors. The following section discusses the role of this interaction network in social dynamics.

3.2 The role of the interaction network

Little is known to date about how the mesoscopic structure of social networks affects the processes taking place in them. Some studies have focused on macroscopic structural features that have been observed in real social networks, such as the small-world phenomenon (18; 93; 94; 95) or the skewed degree distribution (96; 97; 98), but the mesoscopic structure of social networks has received little attention. Until very recently, such analyses in fact have not been possible due to a lack of data and models of the community structure in large scale social networks. Hence, at the time when the work leading to this thesis was begun, no studies on opinion formation models or other social dynamics in networks with community structure existed. A handful of studies have appeared in the past couple of years that deal with the effect of community structure on dynamics[2]. We will review their findings here.

A 2007 study by Lambiotte et al (106) examines a two-state majority model, in which agents meet in groups of three mutually acquainted individuals, and all three adopt the current majority opinion of the group. The authors posed the question of whether clusters of individuals can hold different opinions indefinitely if the clusters are not very strongly interconnected. So as to enable analytical derivations, they represented the networked social structure by only two cliques of equal size that share a fraction of their nodes. Their finding was that for large cliques (with clique size tending to infinity), there is indeed a limit for the fraction of shared nodes ($\nu = N_0/N$, where N_0 is the number of nodes shared by the two cliques) below which each community will hold on to its opinion, and no system-wide consensus will be reached (106). This study suggests a manner of

[2]Dynamics sensitive to community structure have also been employed for *identifying* communities. Such dynamics include various spin systems such as in the Ising model (99; 100), the Potts model (101; 102), and models of random walk (103; 104; 105).

incorporating the community structure

An extensive study on mobile communication networks by Onnela et al. published the same year verified in large scale networks the Granovetter hypothesis (24) which states that tie strengths between agents in social networks correlate with the overlap of their neighborhoods, and explored the effect of this coupling on the diffusion of information (3). It was seen that information spreads rapidly within communities but passes to other communities with difficulty. In the early stages of simulated spreading, the number of nodes that had received the information rose rapidly each time that the information reached a new community, but plateaus between such steep rises showed that the information was not frequently passed on to a new community. The synchronization of oscillators coupled via a complex network is known to progress analogously, such that in networks with mesoscopic structure, synchronization takes place first within highly interconnected local structures, and synchronized domains expand via intercommunity connections (107; 108; 109; 110; 111; 112).

An interesting study by Lozano, Arenas and Sánchez from 2008 explored the effect of community structure on cooperation (113). The authors employed the same prototypical model for social dilemmas that Axelrod worked with, namely the Prisoner's dilemma (PD). In the network-based formulation of the game, at each time step each agent interacts locally with all of its neighbors using its selected strategy, (C)ooperate or (D)efect, obtaining a score that sums up all the interactions. The agents learn by imitation, adopting the strategy (C or D) of the neighbor that scored highest. This dynamics involving locality and imitation has been generally seen to promote cooperation, indicated by a large fraction of the population using the C strategy, due to clusters of cooperators that can outcompete defectors (114). Using two real world social networks as substrates for the dynamics, the authors of (113) identified features of the community structure, related both to the internal structure of the communities and to their interconnections, that affect cooperation levels in the system as a whole.

The work on social simulation presented in this thesis deals with similar questions. In particular, we pose the question of how the community structure of social networks affects the prevalence of different opinions among the agents. This question is approached using various models of social network structure. Section 3.2 discusses the employed models and networks, and reviews the findings.

3.3 Models of competing options

This section deals with the work of the author and colleagues on models of competing options. The models are motivated by language competition and opinion formation, which are discussed first. Section 3.3 introduces a few concepts and measures that will be used in the following discussions. Then, we introduce models inspired by language competition that can be seen as variants of the prototypical Voter model, and finally a weighted model that takes into account the intensity of each interaction.

It is impossible to know precisely how individuals form their opinions or influence others. Therefore any modeling of social agents involves a great simplification of the problem. Defining realistic microscopic models for opinion formation is a difficult task. This is not to say that simple models could not capture the essence of some forms of social interaction, as demonstrated by the two examples that began this chapter - the Prisoner's dilemma seems to depict appropriately a multitude of social situations. Simplification is not harmful but instead beneficial, as long as the essential factors of the interaction are taken into account. We have worked under the hypothesis that peer pressure is relevant to the actual processes of opinion formation in society, and chosen the models according to the principle of simplicity.

In all the agent based models discussed in this section, an agent will change its option with a probability that depends on the options held by its neighbors (peer pressure). The probabilities depend on the fraction of neighboring agents holding each option, and the intensity of each interaction, depicted by link weight. However, in most cases link weights are ignored, and all network neighbors are considered to have an equal influence upon the agent. In the following, to denote the state of a node, the concepts of option, opinion, language, and state are used interchangeably depending on the context. The models we discuss here concern discrete options, labeled A, B, and AB.

Characterizations of the dynamics

Let us first define a few measures and concepts that will be useful for discussing the dynamics in the following sections.

Local density. The fraction of first neighbors in state A (B, AB) of an agent is called the *local density* of A (B, AB), and denoted by σ_A (σ_B, σ_{AB}).

Interface density. The degree of ordering in a system can be characterized by the fraction of links joining agents in different states. This fraction is called the *interface density* ρ. The interface density decreases as homogenized domains grow in size, and eventually disappears if one of the options wins over. In a regular network topology, the interface density indicates the average size of domains, and in complex networks it can be used to describe domain growth approximately, such that low interface density implies a high degree of ordering. The interface density is used to study the formation of domains in individual realizations of the stochastic dynamics, and the average behavior of the system is described by the average interface density $\langle\rho\rangle$, where the average is taken over an ensemble of realizations starting from different random initial conditions.

Absorbing state. The system of interacting agents has reached an *absorbing state* when agents can no longer change their state and the dynamics halts. In the dynamics treated in this work, absorbing states are those in which all agents are in the same state A or B, because in the absence of neighbors in a different state the probability of a node changing its state becomes zero. A system with all agents in the AB state is not in an absorbing state, because AB-agents can spontaneously change their state to A or B in the dynamics described later in this section.

Coarsening. The term *coarsening* signifies the formation and growth of homogeneous domains. It is indicated by a decrease in interface density, because this corresponds to a growth in the average domain size (preceding any decrease related to random fluctuations that eventually lead to an absorbing state in finite systems).

Metastable states. *Metastable states* in physics and chemistry are described by the Encyclopedia Britannica (115) as "a particular excited state of an atom, nucleus, or other system that has a longer lifetime than the ordinary excited states and that generally has a shorter lifetime than the lowest, often stable, energy state. A metastable state may thus be considered a kind of temporary energy trap or a somewhat stable intermediate stage of a system the energy of which may be lost in discrete amounts." Here, the term is used for a dynamical system to describe states that last particularly long, but where the system has not yet reached an absorbing state. In the following dynamics, we will encounter various types of metastable states, *dynamical* and *trapped*, which will be described later.

With the concepts clarified, let us move on to discuss models of competing options.

Background in language competition

The languages and cultures of the world are in a constant state of flux. Yet, it has periodically taken place in history that one culture and language becomes dominant over others and practically supersedes all others. Even today, English is becoming the new lingua franca, and of the roughly 6000 languages spoken in the world, Between 50 to 90 percent are estimated to become extinct by the end of the 21st century. Although the causes for such cycles of the homogenization and fragmentation of culture and language are varied, there might also exist some fundamental properties of the system (here, the cultures of the world) that drive it towards order and eventually again into disorder. The drive towards homogeneity could be in part caused by the tendency of individuals, who are in contact with one another, to become alike. On the other hand, isolated groups of people tend to develop different views and cultures. Although obviously also other factors than interactions at the individual level are at play in the homogenization of opinions, culture or language, the models presented here make a simplification and focus on individual level interactions.

We present first a model of competing options concerning language endangerment (116) that has inspired many other models, including those studied in this thesis. In an article published in Nature in 2003, Abrams and Strogatz analysed the decay of minority languages during the 20th century in 42 regions of Europe and South America, including languages such as Quechuan (threatened by Spanish) and Welsh (threatened by English). They employed a system of differential equations to describe how the fraction of the population speaking each of the two languages changed over time. In their model, the probability of adopting language A increases with the fraction x of the population speaking language A, because the speakers are motivated to adopt a language spoken by many others. The rest of the population, the fraction $y = 1 - x$, speak language B. The deviation from a linear dependence on x and y is described by an by an exponent a, that was unexpectedly found to be roughly constant across cultures, about $a = 1.31 \pm 0.25$. Furthermore, the benefits of learning to speak each language, such as increased access to education or jobs, are incorporated in the model through a parameter s, called the *prestige* or *social status* of a language. The transition probabilities between languages A and B are thus represented by the equations $p_{B\to A}(x, s) = cx^a s$ of and $p_{A\to B}(x, s) = c(1 - x)^a(1 - s)$, and the fraction of the population speaking language A changes as

$$dx/dt = y\, p_{B\to A}(x, s) - x\, p_{A\to B}(x, s). \tag{3.1}$$

The data were surprisingly well fitted by this very simple model. The Abrams-Strogatz (AS) model does not take into account spatial or social structure however, and all speakers are assumed to learn only one of the two competing languages. Later modifications have added these features.

The original AS-model predicts that the language with smaller prestige always dies out. It is natural to pose the question of whether it is possible to prevent the extinction of the less prestigious language, and which methods could be employed to that end. For example, Patriarca and Leppänen (117) demonstrated with an analytical model that if each language is only influential within a particular region, e.g. due to political or geographical factors, two languages can coexist despite one of them having lower status. Could the structure of social networks also aid in preserving a language? How does it affect the dynamics of language competition? This is explored in agent based versions of the AS-model and its variants that take into account the structure of social interaction.

The Voter model and the microscopic Abrams-Strogatz model

In order to incorporate social or geographical constraints, such that not everyone is in contact with everyone else, Stauffer et al. reformulated the AS-model of language competition as agent-based (118). In this formulation, called the *microscopic Abrams-Strogatz model* (mAS), each agent holds one of two options,

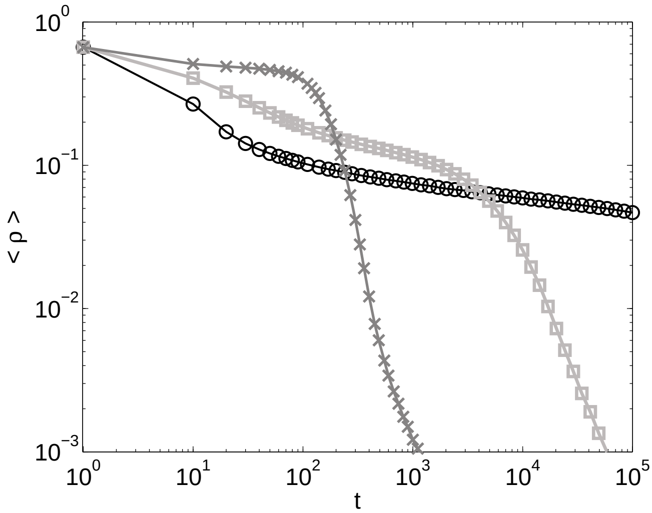

Figure 3.5: The social influence *si* model in community networks with weight-topology correlations (black circles), in weight-randomized networks with community structure (light gray squares), and in fully randomized networks (gray crosses). Figure taken from Publication III.

Chapter 4

Summary of results and discussion

This thesis has dealt with modeling the structure of social networks, as well as with models of competing options that form a subset of social dynamics.

The author and colleagues produced a useful comparative analysis of social network models (Publication II), categorizing them and pointing out similarities and differences in the underlying mechanisms in the models and in the resulting network structures. The comparison shows that the mechanisms of triadic closure, or linking to friends of friends, explains much of the structure of social networks, although in many models it alone fails to produce large enough clusters. Homophily based on social or spatial attributes is seen to successfully produce community structure, but when used alone, it produces networks in which high degree nodes have unrealistically high clustering coefficients.

In order to answer a clear need for large scale models of social networks, we have proposed a new model (TOSHK) (Publication I) based on simple mechanisms of random attachment combined with link formation within the local neighborhood (connecting to "friends of friends"). This model is seen to produce many of the universal structures observed across different social networks. Interest in the TOSHK model was expressed by many researchers who asked for its source code.

The TOSHK model was immediately useful as a substrate for studying the dynamics of competing options. It showed that the community structure of networks can have a profound influence on dynamics of competing options such as formation of opinions in a networked population (Publication IV). A dynamical model of three competing options, in which the intermediate option describes indecision between the two opposing states, was seen to develop in the TOSHK topology trapped metastable states that survive at all time scales. The relative isolation of groups of nodes corresponding to community structure was seen to enable an opinion held by the minority to persist for a very long time against the influence of the rest of the network.

In a more detailed study of the same dynamics (Publication VII), employing test networks with extremely clear community structure, we characterized the minimal criteria of the topologies that produce broad lifetime distributions for the metastable states: community structure alone was seen not to be sufficient, as demonstrated by randomly connected cliques with equal size and equal number of outward connections. Instead, producing a broader than exponential lifetime distribution was seen to require heterogeneity in the *dynamical robustness* of the communities, a concept defined by the authors to describe the resistance of network substructures against changing their state under outside influence. A weighted variant of the three-state model of competing options (Publication III) showed that the correlations between topology and interaction strength in social networks may further increase the chances of communities holding on to a minority opinion.

In the dynamics of competing options, we have worked under the hypothesis that peer pressure is relevant to the actual processes of opinion formation in society, and chosen the models according to the principle of simplicity. This accords with the guidelines for using ABMs in sociological modeling, given by Macy and Willer (19). They suggest to *start it simple*, stating that "a model that is as complex as the phenomenon it attempts to represent is useless. Complications should be added one at a time, once full understanding of the simpler case is reached." The simplicity principle was followed during the course of the work, attempting to figure out the behavior of simpler models before adding more features.

Macy and Willer also encourage to "*Test external validity*. If a model has been successfully used to test a hypothesis, and shown to be robust, researchers need to think of ways in which the results could be tested in laboratory or natural conditions." The work of the author and colleagues on opinion dynamics has been useful for generating initial hypotheses on the effect of communities on the formation of opinions, and could provide ideas for experimentation. For example, it would be very interesting and informative to test the assumptions of opinion formation models in a laboratory setting. Aggregate outcomes such as election results have been collected and studied; and at the individual level, it is also known that peer pressure affects individual decisions. The intermediate level between these extremes - collecting data on group influence, and interaction between groups forming opinions, is still waiting experimental work.

An experiment can be imagined in which individuals are each given a different set of information on an issue on which they would later need to make a choice, and decide upon either action A or action B. Some participants would be given mostly facts against A and for B, while others would receive information that mostly supports the choice A instead of B. Participants would be asked at frequent intervals about their current opinion on the matter, and the opinions of those that they have discussed with. Finally, the participants should vote on the final action to be taken. In order to study explicitly the effect of group structure, the experiment could either monitor any groups that form naturally during the experiment, or employ a predefined network of interaction with community structure. Experiments such as these would provide experimental validation or disvalidation of our models, and suggest more appropriate ones.

There is a clear need for verification of the models of social dynamics against empirical data. This has been successfully achieved in some fields, such as pedestrian dynamics, while it is largely lacking in others, such as opinion formation. Through discussion with experts in the relevant fields, we need to consider the fundamental questions of where the models are applicable, and how they should be modified to make them appropriate. Considering the AB-model, for example, it could be argued that the chosen language among bilingual speakers is a property of the link instead of an the agent (everyday experience shows that a bilingual person can use either one of his languages with different acquaintances).

Identifying general classes into which the various models of social dynamics fit would be beneficial to the research field. Thus far, the understanding of the general behavior of families of opinion formation models has been incremented through small studies of model variants, each of them a piece in the bigger picture.

The excitement in studying networks showed by physicists, who have largely been ignorant of the great amount of research that sociologists have done on networks, has often been (partly deservedly) ridiculed. However, there could be great benefits in cooperation. What the "new science of networks" can provide to the study of social systems is a set of tools and algorithms that can be used to analyse large sets of data. Often researchers with a background in physics and mathematics have developed algorithms that can determine in large scale some network characterizations that were originally developed by social scientists. Established measures can also find new applications; one of the benefits of network methods is indeed the wide applicability of relatively simple techniques across various fields. For example, centrality measures were developed by sociologists to describe the status of individuals in a social network. Similar measures found a highly successful application in the Google web search engine, in which a particular centrality measurement called page rank is used to identify among billions of web sites the most popular and most cited pages (133). Another field in which network research has provided fruitful insights is epidemics. Heterogeneities in contact rates have been seen to have a large effect on the early stage of the HIV/AIDS epidemic (134). Woolhouse et al. found that the 80/20 rule that a small fraction of hosts is responsible for a large fraction of all infections applies to both vector-borne parasites and sexually transmitted pathogens alike, and suggested assessment of whether degree-based interventions could be implemented for higher cost-effectiveness in prevention of their spreading (135). Analytical studies confirm the efficiency of degree based immunization strategies (136). Other applications of network research range from viral marketing (57) to communication protocols for large distributed systems that scale well with increasing system size (137).

Some of the research on social dynamics has had direct implications to human safety and policies. Particularly useful results have been obtained in crowd dynamics (138); modeling the turbulent flow of pedestrians in an extremely dense crowd during a religious ceremony attended by more than a million people, Helbing et al. were able to recommend maneuvers in the organization of the event that helped the flow to become smoother, and likely saved lives by preventing trampling accidents. Agent based modeling has also provided insights on pedestrian traffic concerning various phenomena such as the formation of paths across a campus lawn, the paths of people walking in opposing directions in a corridor, or the packing of a crowd attempting to exit a building in case of a fire. While the benefits of opinion dynamics thus far are not equally direct, there is hope that they will find important applications in the future. Although the most immediate applications are likely to be in marketing (for example, e-commerce sites are very interested in making use of the friendship networks of their customers for informing potential customers of their products), the same method seems to be useful in spreading information about beneficial causes, as exemplified by the popularity of applications that disseminate information on environmental and social causes on Facebook.

References

[1] **J. Leskovec and E. Horvitz**, "Planetary-scale views on a large instant-messaging network", in "WWW '08: Proceeding of the 17th international conference on World Wide Web", 915–924, ACM, New York, NY, USA, 2008, longer version available online at arXiv:0803.0939 (2008).

[2] **A. Traud, E. Kelsic, P. Mucha, and M. Porter**, "Community structure in online collegiate social networks", arXiv:0809.0690 (2008).

[3] **J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, and A.-L. Barabási**, "Structure and tie strengths in mobile communication networks", *Proceedings of the National Academy of Sciences (PNAS) (USA)*, 104, 7332 (2007).

[4] **J.-P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, M. A. de Menezes, K. Kaski, A.-L. Barabási, and J. Kertész**, "Analysis of a large-scale weighted network of one-to-one human communication", *New Journal of Physics*, 9, 179 (2007).

[5] **M. Seshadri, S. Machiraju, A. Sridharan, J. Bolot, C. Faloutsos, and J. Leskovec**, "Mobile call graphs: Beyond power-law and lognormal distributions", *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, Las Vegas, Nevada, USA*, 596–604 (2008).

[6] **J. A. Davis**, "Clustering and hierarchy in interpersonal relations: Testing two graph theoretical models on 742 sociograms", *American Sociological Review*, 35, 843 (1970).

[7] **J. Scott**, *Social network analysis: A Handbook, 2nd ed.*, Sage, London (2000).

[8] **M. Newman**, "Detecting community structure in networks", *European Physical Journal B*, 38, 321 (2004).

[9] **L. Danon, A. Díaz-Guilera, J. Duch, and A. Arenas**, "Comparing community structure identification", *Journal of Statistical Mechanics*, P09008 (2005).

[10] **L. Danon, J. Duch, A. Arenas, and A. Díaz-Guilera**, "Community structure identification", in "Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science", 93–113, World Scientific (2007).

[11] **F. Radicchi, C. Castellano, F. Cecconi, V. Loreto, and D. Parisi**, “Defining and identifying communities in networks”, *Proceedings of the National Academy of Sciences (PNAS) (USA)*, 101, 2658 (2004).

[12] **A. Arenas, L. Danon, A. Díaz-Guilera, P. Gleiser, and R. Guimerá**, “Community analysis in social networks”, *European Physical Journal B*, 38, 373 (2004).

[13] **M. Newman and M. Girvan**, “Finding and evaluating community structure in complex networks”, *Physical Review E*, 69, 026113 (2004).

[14] **G. Palla, I. Derényi, I. Farkas, and T. Vicsek**, “Uncovering the overlapping community structure of complex networks in nature and society”, *Nature*, 435, 814 (2005).

[15] **G. Palla, A.-L. Barabási, and T. Vicsek**, “Quantifying social group evolution”, *Nature*, 446, 664 (2007).

[16] **S. Fortunato and C. Castellano**, “Community structure in graphs”, in R. A. Meyers (editor), “Encyclopedia of Complexity and System Science”, Springer (2008).

[17] **C. W. Reynolds**, “Flocks, herds, and schools: A distributed behavioral model”, *Computer Graphics (SIGGRAPH ’87 Conference Proceedings)*, 21(4), 25 (1987).

[18] **A. Baronchelli, L. Dall’Asta, A. Barrat, and V. Loreto**, “The role of topology on the dynamics of the naming game”, *European Physical Journal Special Topics*, 143, 233 (2007).

[19] **M. W. Macy and R. Willer**, “From factors to actors: Computational sociology and agent-based modeling”, *Annual Reviews of Sociology*, 28, 143 (2002).

[20] **L. Freeman**, “Social network analysis: Definition and history”, in A. E. Kazdan (editor), “Encyclopedia of Psychology”, 350–351, Oxford University Press (2000).

[21] **D. J. Watts**, “The ’new’ science of networks”, *Annual Reviews of Sociology*, 30, 243 (2004).

[22] **M. Newman**, “The structure and function of complex networks”, *SIAM Review*, 45, 167 (2003).

[23] **J. Travers and S. Milgram**, “An experimental study of the small world problem”, *Sociometry*, 32, 425 (1969).

[24] **M. Granovetter**, “The strength of weak ties”, *American Journal of Sociology*, 78, 1360 (1973).

[25] **P. Erdős and A. Rényi**, “On the evolution of random graphs”, *Publicationes Mathematicae Debrecen*, 6, 290 (1959).

[26] **O. Frank and D. Strauss**, “Markov graphs”, *Journal of the American Statistical Association (JASA)*, 81, 832 (1986).

[27] **C. J. Anderson, S. Wasserman, and B. Crouch**, “A p^* primer: logit models for social networks”, *Social Networks*, 21, 37 (1999).

[28] **D. Strauss and M. Ikeda**, “Pseudolikelihood estimation for social networks”, *Journal of the American Statistical Association (JASA)*, 85(409), 204 (1990).

[29] **S. Wasserman and P. Pattison**, “Logit models and logistic regressions for social networks, i. an introduction to Markov graphs and p^*”, *Psychometrika*, 61, 401 (1996).

[30] **P. J. Carrington, J. Scott, and S. Wasserman** (editors), *Models and Methods in Social Network Analysis*, Cambridge University Press (2005).

[31] **G. Robins, P. Pattison, Y. Kalish, and D. Lusher**, “An introduction to exponential random graph (p^*) models for social networks”, *Social Networks*, 29, 173 (2007).

[32] **D. Hunter**, “Curved exponential family of models for social networks”, *Social Networks*, 29, 216 (2007).

[33] **T. Evans**, “Complex networks”, *Contemporary Physics*, 45, 455 (2004).

[34] **L. da F. Costa, O. N. O. Jr., G. Travieso, F. A. Rodrigues, P. R. V. Boas, L. Antiqueira, M. P. Viana, and L. E. C. da Rocha**, “Analyzing and modeling real-world phenomena with complex networks: A survey of applications”, *http://arxiv.org/abs/0711.3199* (2007).

[35] **H. A. Simon**, “On a class of skew distribution functions”, *Biometrika*, 42, 425 (1955).

[36] **S. Bornholdt and H. Ebel**, “World wide web scaling exponent from Simon’s 1955 model”, *Physical Review E*, 64(3), 035104(R) (2001).

[37] **P. Krapivsky and S. Redner**, “Network growth by copying”, *Physical Review E*, 71, 036118 (2005).

[38] **A. Vázquez**, “Growing networks with local rules: Preferential attachment, clustering hierarchy, and degree correlations”, *Physical Review E*, 67, 056104 (2003).

[39] **P. Holme and B. Kim**, “Growing scale-free networks with tunable clustering”, *Physical Review E*, 65, 026107 (2002).

[40] **J. Davidsen, H. Ebel, and S. Bornholdt**, “Emergence of a small world from local interaction: Modeling acquaintance networks”, *Physical Review Letters*, 88, 128701 (2002).

[41] **M. Marsili, F. Vega-Redondo, and F. Slanina**, “The rise and fall of a networked society: A formal model”, *Proceedings of the National Academy of Sciences (PNAS) (USA)*, 101, 1439 (2004).

[42] **M. Catanzaro, G. Caldarelli, and L. Pietronero**, “Assortative model for social networks”, *Physical Review E*, 70, 037101 (2004).

[43] **K. Klemm and V. M. Eguíluz**, “Highly clustered scale-free networks”, *Physical Review E*, 65(3), 036123 (2002).

[44] **K. Klemm and V. M. Eguíluz**, “Growing scale-free networks with small-world behavior”, *Physical Review E*, 65(5), 057102 (2002).

[45] **A. Barrat, M. Barthélemy, R. Pastor-Satorras, and A. Vespignani**, “The architecture of complex weighted networks”, *Proceedings of the National Academy of Sciences (PNAS) (USA)*, 101, 3747 (2004).

[46] **A. Barrat, M. Barthélemy, and A. Vespignani**, “Weighted evolving networks: Coupling topology and weight dynamics”, *Physical Review Letters*, 92, 228701 (2004).

[47] **M. Newman**, “Assortative mixing in networks”, *Physical Review Letters*, 89, 208701 (2002).

[48] **M. Newman**, “The structure of scientific collaboration networks”, *Proceedings of the National Academy of Sciences (PNAS) (USA)*, 98, 404 (2001).

[49] **P. Gleiser and L. Danon**, “Community structure in jazz”, *Advances in Complex Systems*, 6(4), 565 (2003).

[50] **L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley**, “Classes of small-world networks”, *Proceedings of the National Academy of Sciences (PNAS) (USA)*, 97, 11149 (2000).

[51] **R. Guimerà, L. Danon, A. Díaz-Guilera, F. Giralt, and A. Arenas**, “Self-similar community structure in a network of human interactions”, *Physical Review E*, 68, 065103(R) (2003).

[52] **S. Galam**, “Sociophysics: a personal testimony”, *Physica A*, 336, 49 (2004).

[53] **S. Kirby, H. Cornish, and K. Smith**, “Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language”, *Proceedings of the National Academy of Sciences (PNAS) (USA)*, 105, 10681 (2008).

[54] **D. Strang and S. A. Soule**, “Diffusion in organizations and social movements: From hybrid corn to poison pills”, *Annual Reviews of Sociology*, 24, 265 (1998).

[55] **M. A. Nowak**, “Five rules for the evolution of cooperation”, *Science*, 314, 1560 (2006).

[56] **P. Kollock**, “Social dilemmas: The anatomy of cooperation”, *Annual Reviews of Sociology*, 24, 183 (1998).

[57] **J. Leskovec, L. A. Adamic, and B. A. Huberman**, “The dynamics of viral marketing”, *ACM Transactions on the Web*, 1 (2007).

[58] **R. Lambiotte, V. D. Blondel, C. de Kerchove, E. Huens, C. Prieur, Z. Smoreda, and P. V. Dooren**, “Geographical dispersal of mobile communication networks”, *Physica A*, 387, 5317 (2008).

[59] **S. Battiston and M. Catanzaro**, “Statistical properties of corporate board and director networks”, *European Physical Journal B*, 38, 345 (2004).

[60] **M. Newman and J. Park**, “Why social networks are different from other types of networks”, *Physical Review E*, 68, 036122 (2003).

[61] **L. da F. Costa, F. A. Rodrigues, G. Travieso, and P. R. V. Boas**, “Characterization of complex networks: A survey of measurements”, *Advances in Physics*, 56, 167 (2007).

[62] **D. West**, *Introduction to Graph Theory (2nd Edition)*, Prentice Hall (2000).

[63] **P. Jaccard**, “Distribution de la flore alpine dans le bassin des dranses et dans quelques régions voisines”, *Bulletin de la Société Vaudoise des Sciences Naturelles*, 37, 241 (1901).

[64] **W. W. Zachary**, “An information flow model for conflict and fission in small groups”, *Journal of Anthropological Research*, 33, 452 (1977).

[65] **M. Girvan and M. Newman**, “Community structure in social and biological networks”, *Proceedings of the National Academy of Sciences (PNAS) (USA)*, 99, 7821 (2002).

[66] **E. Ravasz, A. L. Somera, D. A. Mongru, Z. N. Oltvai, and A.-L. Barabási**, “Hierarchical organization of modularity in metabolic networks”, *Science*, 297, 1551 (2002).

[67] **J. P. Bagrow**, “Evaluating local community methods in networks”, *Journal of Statistical Mechanics*, P05001 (2008).

[68] **J. M. Hofman and C. H. Wiggins**, “Bayesian approach to network modularity”, *Physical Review Letters*, 100(25), 258701 (2008).

[69] **D. Garlaschelli and M. I. Loffredo**, “Maximum likelihood: Extracting unbiased information from complex networks”, *Physical Review E*, 78, 015101(R) (2008).

[70] **J. Kumpula, J. Saramäki, K. Kaski, and J. Kertész**, “Limited resolution and multiresolution methods in complex network community detection”, *Fluctuation and Noise Letters*, 7, L209 (2007).

[71] **V. Mäkinen**, Himmeli, a free software package for visualizing complex networks, available at *http://www.artemis.kll.helsinki.fi/himmeli* (2005).

[72] **R. Toivonen, L. Kovanen, M. Kivelä, J.-P. Onnela, J. Saramäki, and K. Kaski**, *A comparative study of social network models: Network evolution models and nodal attribute models*, Helsinki University of Technology, Technical report A10, ISBN 978-951-22-9764-1 (2009).

[73] **R. Toivonen, J.-P. Onnela, J. Saramäki, J. Hyvönen, and K. Kaski**, “A model for social networks”, *Physica A*, 371(2), 851 (2006).

[74] **J. Kumpula, J.-P. Onnela, J. Saramäki, K. Kaski, and J. Kertész**, “Emergence of communities in weighted networks”, *Physical Review Letters*, 99, 228701 (2007).

[75] **R. Toivonen, J. Kumpula, J. Saramäki, J.-P. Onnela, J. Kertész, and K. Kaski**, “The role of edge weights in social networks: modelling structure and dynamics”, in J. Kertész, S. Bornholdt, and R. Mantegna (editors), “Noise and Stochastics in Complex Systems And Finance”, Proceedings of SPIE Vol. 6601, 66010B-1 (2007).

[76] **M. McPherson, L. Smith-Lovin, and J. Cook**, “Birds of a feather: Homophily in social networks”, *Annual Reviews of Sociology*, 27, 415 (2001).

[77] **L. H. Wong, P. Pattison, and G. Robins**, “A spatial model for social networks”, *Physica A*, 360, 99 (2006).

[78] **M. Boguña, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas**, “Emergence of clustering, correlations, and communities in a social network model”, *Physical Review E*, 70, 056122 (2004).

[79] **A. Grönlund and P. Holme**, “Networking the seceder model: Group formation in social and economic systems”, *Physical Review E*, 70, 036108 (2004).

[80] **M. Zimmermann, V. Eguíluz, and M. S. Miguel**, “Coevolution of dynamical states and interactions in dynamic networks”, *Physical Review E*, 69(6), 065102 (2004).

[81] **J. Poncela, J. G.-G. nes, L. M. Floría, A. Sánchez, and Y. Moreno**, “Complex cooperative networks from evolutionary preferential attachment”, *PLoS ONE*, 3(6), e2449 (2008).

[82] **D. H. Zanette**, “Coevolution of agents and networks in an epidemiological model”, (2007).

[83] **T. Schelling**, “Dynamic models of segregation”, *Journal of Mathematical Sociology*, 1, 143 (1971).

[84] **R. Axelrod**, *The Evolution of Co-operation*, Penguin Books (1990).

[85] **R. Axelrod**, *The complexity of cooperation: Agent Based Models of Competition and Collaboration*, Princeton University Press (1997).

[86] **B. Skyrms**, *The Stag Hunt and the Evolution of Social Structure*, Cambridge University Press (2003).

[87] **C. Castellano, S. Fortunato, and V. Loreto**, “Statistical physics of social dynamics”, *Reviews of Modern Physics (in press) (arXiv:0710.3256)* (2009).

[88] **R. Holley and T. Liggett**, "Ergodic theorems for weakly interacting infinite systems and the Voter model", *Annals of Probability*, 3, 643 (1975).

[89] **S. Galam**, "Minority opinion spreading", *European Physical Journal B*, 25, 403 (2002).

[90] **P. L. Krapivsky and S. Redner**, "Dynamics of majority rule in two-state interacting spin systems", *Physical Review Letters*, 90, 238701 (2003).

[91] **K. Sznajd-Weron**, "Sznajd model and its applications", *Acta Physica Polonica B*, 36, 2005 (2005).

[92] **G. Deffuant, F. Amblard, and G. Weisbuch**, "Modelling group opinion shift to extreme : the smooth bounded confidence model", in "Proceedings of the European Social Simulation Association (ESSA) Conference", (2004).

[93] **D. Zanette**, "Dynamics of rumor propagation on small-world networks", *Physical Review E*, 65, 041908 (2002).

[94] **J. M. Kleinberg**, "Navigation in a small world", *Nature*, 406, 845 (2000).

[95] **C. Castellano, D. Vilone, and A. Vespignani**, "Incomplete ordering of the Voter model on small-world networks", *Europhysics Letters*, 63, 153 (2003).

[96] **R. Pastor-Satorras and A. Vespignani**, "Epidemic spreading in scale-free networks", *Physical Review Letters*, 86(14), 3200 (2001).

[97] **F. Santos, J. Pacheco, and T. Lenaerts**, "Evolutionary dynamics of social dilemmas in structured heterogeneous populations", *Proceedings of the National Academy of Sciences (PNAS) (USA)*, 103, 3490 (2006).

[98] **V. Sood and S. Redner**, "Voter model on heterogeneous graphs", *Physical Review Letters*, 94, 178701 (2005).

[99] **S.-W. Son, H. Jeong, and J. D. Noh**, "Random field Ising model and community structure in complex networks", *European Physical Journal B*, 50, 431 (2006).

[100] **Y. Fu and P. W. Anderson**, "Application of statistical mechanics to np-complete problems in combinatorial optimisation", *J. Phys. A: Math. Gen.*, 19, 1605 (1986).

[101] **J. Reichardt and S. Bornholdt**, "Detecting fuzzy community structures in complex networks with a Potts model", *Physical Review Letters*, 93, 218701 (2004).

[102] **M. Blatt, S. Wiseman, and E. Domany**, "Superparamagnetic clustering of data", *Physical Review Letters*, 76, 3251 (1996).

[103] **H. Zhou**, "Distance, dissimilarity index, and network community structure", *Physical Review E*, 67, 061901 (2003).

[104] **K. A. Eriksen, I. Simonsen, S. Maslov, and K. Sneppen**, "Modularity and extreme edges of the internet", *Physical Review Letters*, 90(14), 148701 (2003).

[105] **I. Simonsen**, “Diffusion and networks: A powerful combination!”, *Physica A*, 357, 317 (2005).

[106] **R. Lambiotte, M. Ausloos, and J. Holyst**, “Majority model on a network with communities”, *Physical Review E*, 75, 030101(R) (2007).

[107] **Y. M. Vega, M. Vázquez-Prada, and A. F. Pacheco**, “Fitness for synchronization of network motifs”, *Physica A*, 343, 279 (2004).

[108] **A. Arenas and A. Díaz-Guilera**, “Synchronization and modularity in complex networks”, *European Physical Journal B*, 143, 19 (2007).

[109] **A. Arenas, A. Díaz-Guilera, and C. Pérez-Vicente**, “Synchronization reveals topological scales in complex networks”, *Physical Review Letters*, 96, 114102 (2006).

[110] **E. Oh, K. Rho, H. Hong, and B. Kahng**, “Modular synchronization in complex networks”, *Physical Review E*, 72, 047101 (2005).

[111] **T. Zhou, M. Zhao, G. Chen, G. Yan, and B.-H. Wang**, “Phase synchronization on scale-free networks with community structure”, *Physics Letters A*, 368, 431 (2007).

[112] **A. Arenas, A. Díaz-Guilera, J. Kurths, Y. Moreno, and C. Zhou**, “Synchronization in complex networks”, *Physics Reports*, 469, 93 (2008).

[113] **S. Lozano, A. Arenas, and A. Sánchez**, “Mesoscopic structure conditions the emergence of cooperation on social networks”, *PLoS one*, 3, e1892 (2008).

[114] **M. A. Nowak and R. M. May**, “Evolutionary games and spatial chaos”, *Nature*, 359, 826 (1992).

[115] Definition of metastable state in *Encyclopædia Britannica Online*, retrieved September 23, 2008,
http://www.britannica.com/EBchecked/topic/378009/metastable-state.

[116] **D. M. Abrams and S. H. Strogatz**, “Modelling the dynamics of language death”, *Nature*, 424, 900 (2003).

[117] **M. Patriarca and T. Leppänen**, “Modeling language competition”, *Physica A*, 338, 269 (2004).

[118] **D. Stauffer, X. Castelló, V. M. Eguíluz, and M. S. Miguel**, “Microscopic Abrams-Strogatz model of language competition”, *Physica A*, 374, 835 (2007).

[119] **P. Clifford and A. Sudbury**, “A model for spatial conflict”, *Biometrika*, 60(3), 581 (1973).

[120] **K. Binder**, “Ising model”, in M. Hazewinkel (editor), “Encyclopaedia of Mathematics”, Kluwer Academic Publishers (2001).

[121] **K. Suchecki, V. M. Eguíluz, and M. S. Miguel**, “Conservation laws for the Voter model in complex networks”, *Europhysics Letters*, 69, 228 (2005).

[122] **L. Frachebourg and P. L. Krapivsky**, “Exact results for kinetics of catalytic reactions”, *Physical Review E*, 53, R3009 (1996).

[123] **K. Suchecki, V. M. Eguíluz, and M. S. Miguel**, “Voter model dynamics in complex networks: Role of dimensionality, disorder, and degree distribution”, *Physical Review E*, 72, 036132 (2005).

[124] **C. Castellano, V. Loreto, A. Barrat, F. Cecconi, and D. Parisi**, “Comparison of Voter and Glauber ordering dynamics on networks”, *Physical Review E*, 71, 066107 (2005).

[125] **D. Vilone and C. Castellano**, “Solution of Voter model dynamics on annealed small-world networks”, *Physical Review E*, 69, 016109 (2004).

[126] **W. S.-Y. Wang and J. W. Minett**, “The invasion of language: emergence, change and death”, *TRENDS in Ecology and Evolution*, 20, 263 (2005).

[127] **X. Castelló, V. M. Eguíluz, and M. S. Miguel**, “Ordering dynamics with two non-excluding options: bilingualism in language competition”, *New Journal of Physics*, 8, 308 (2006).

[128] **I. Dornic, H. Chaté, J. Chave, and H. Hinrichsen**, “Critical coarsening without surface tension: the universality class of the Voter model”, *Physical Review Letters*, 87, 045701(4) (2001).

[129] **V. Spirin, P. Krapivsky, and S. Redner**, “Freezing in Ising ferromagnets”, *Physical Review E*, 65, 016119 (2001).

[130] **C. Castellano and R. Pastor-Satorras**, “Zero temperature Glauber dynamics on complex networks”, *Journal of Statistical Mechanics*, P05001 (2006).

[131] **D. Boyer and O. Miramontes**, “Interface motion and pinning in small-world networks”, *Physical Review E*, 67, 035102 (2003).

[132] **R. Toivonen, X. Castelló, V. M. Eguíluz, J. Saramäki, K. Kaski, and M. S. Miguel**, “Broad lifetime distributions for ordering dynamics in complex networks”, *Physical Review E*, 79, 016109.

[133] **S. Brin and L. Page**, “The anatomy of a large-scale hypertextual web search engine”, *Computer Networks*, 30, 107 (1998).

[134] **R. Anderson, G.Medley, R. May, and A. Johnson**, “A preliminary study of the transmission dynamics of the human immunodeficiency virus (hiv), the causative agent of aids”, *IMA Journal of Mathematics Applied in Medicine and Biology*, 3, 229 (1986).

[135] **M. E. J. Woolhouse, C. Dye, J.-F. Etard, T. Smith, J. D. Charlwood, G. P. Garnett, P. Hagan, J. L. K. Hii, P. D. Ndhlovu, R. J. Quinnell, C. H. Watts, S. K. Chandiwana, and R. M. Anderson**, "Heterogeneities in the transmission of infectious agents: Implications for the design of control programs", *Proceedings of the National Academy of Sciences (PNAS) (USA)*, 94, 338 (1997).

[136] **R. Cohen, S. Havlin, and D. ben Avraham**, "Efficient immunization strategies for computer networks and populations", *Physical Review Letters*, 91(24), 247901 (2003).

[137] **A.-M. Kermarrec, L. Massoulié, and A. J. Ganesh**, "Probabilistic reliable dissemination in large-scale systems", *IEEE Transactions on Parallel and Distributed Systems*, 14(3) (2003).

[138] **D. Helbing, A. Johansson, and H. Z. Al-Abideen**, "Dynamics of crowd disasters: an empirical study", *Physical Review E*, 75, 046109 (2007).