COMPUTATIONAL ANALYSIS OF COMPLEX SYSTEMS: APPLICATIONS TO POPULATION DYNAMICS AND NETWORKS

By

Ferenc Molnar

A Dissertation Submitted to the Graduate

Faculty of Rensselaer Polytechnic Institute

in Partial Fulfillment of the

Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major Subject: MULTIDISCIPLINARY SCIENCE

Examining Committee:

_____

Gyorgy Korniss, Dissertation Adviser

_____

Boleslaw K. Szymanski, Dissertation Adviser

_____

Peter R. Kramer, Member

_____

Thomas Caraco, Member

Rensselaer Polytechnic Institute
Troy, New York

November 2014
(For Graduation December 2014)

CONTENTS

LIST OF TABLES

LIST OF FIGURES

ACKNOWLEDGMENT

First of all, I would like to express my thanks to my advisor, Prof. Gyorgy Korniss, for his support and guidance in our research work and in the writing of this dissertation. I could always depend on him for advice and enjoyed friendly conversations with him on any topic, not just science. I am particularly grateful for all the opportunities he gave me to attend several international conferences. I always enjoyed his company and I hope that we can continue working and having fun together in the future.

I would like to thank each member of my doctoral committee for their education and advice. My co-advisor, Prof. Thomas Caraco introduced me to population biology and showed me intriguing connections between physics and biology. He guided my research in this area, which I could not have completed without him. Prof. Boleslaw Szymanski, my co-advisor in network science, always gave me his full support and attention. I particularly enjoyed our discussions on graph algorithms, and I am most grateful for introducing me to all his colleagues at IBM and ARL. Prof. Peter Kramer taught me valuable lectures on probability theory and stochastic modeling, as well as the essence of rigorous mathematical modeling. His lectures were always inspirational and engaging, and I am most delighted to have learned from such an excellent teacher.

I must express my thanks to all my friends and colleagues in our lab. I really enjoyed their company, and learned much from them. Especially, I would like to thank Sameet Sreenivasan, Andrea Asztalos, Pramesh Singh, David Hunt, Panagiotis Karampourniotis, Noémi Derzsy, and Tao Jia. Our stimulating discussions on each other's research topics helped me to widen my insights in many areas of network science.

I must also thank Katie Bahran and Diane Torres from the Social Cognitive Networks Academic Research Center, and Nicole McQuade and Ashleigh Cioffi from the physics department. They helped me with many administrative issues and travel arrangements. Their cheerful smile always gave me a pleasant feeling every time I visited their offices.

Lastly and most importantly, I would like to thank my parents, Dr. Ferenc Molnár and Eleonóra Narancsik, for their love and continuing support throughout my entire life, and especially during my years at RPI.

This work was supported in part by grant No. FA9550-12-1-0405 from the U.S. Defense Advanced Research Projects Agency (DARPA), by the National Science Foundation (NSF) Grant No. DEB-0918413, DEB-0918392, DMR-1246958, and by the Army Research Laboratory (ARL) under Cooperative Agreement Number W911NF-09-2-0053. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies either expressed or implied of the National Science Foundation, Defense Advanced Research Projects Agency, Army Research Laboratory, or the U.S. Government.

ABSTRACT

In most complex evolving systems, we can often find a critical subset of the constituents that can initiate a global change in the entire system. For example, in complex networks, a critical subset of nodes can efficiently spread information, influence, or control dynamical processes over the entire network. Similarly, in nonlinear dynamics, we can locate key variables, or find the necessary parameters, to reach the attraction basin of a desired global state. In both cases, a fundamental goal is finding the ability to efficiently control these systems.

We study two distinct complex systems in this dissertation, exploring these topics. First, we analyze a population dynamics model describing interactions of sex-structured population groups. Specifically, we analyze how a sex-linked genetic trait's ecological consequence (population survival or extinction) can be influenced by the presence of sex-specific cultural mortality traits, motivated by the desire to expand the theoretical understanding of the role of biased sex ratios in organisms.

We analyze dynamics within a single population group, as well as between competing groups. We find that there is a finite range of sex ratio bias that can be maintained in stable equilibrium by sex-specific mortalities. We also find that the outcome of an invasion and the ensuing between-group competition depends not on larger equilibrium group densities, but on the higher allocation of sex-ratio genes.

When we extend the model with diffusive dispersal, we find that a *critical patch size* for achieving positive growth only exists if the population expands into an empty environment. If a resident population is already present that can be exploited by the invading group, then any small seed of invader can advance from rarity, in the mean-field approximation, as long as the local competition dynamics favors the invader's survival.

Most spatial models assume initial populations with a uniform distribution inside a finite patch; a simple, but not a cost-efficient approach. We show, using a novel application of simulated annealing, that a specific, non-trivial shape of spatial distribution can minimize the total cost of successful invasion, i.e., the cost

of ecological restoration. Further, our approach can be generalized to essentially any reaction-diffusion model with diffusive spreading.

In the second part of the dissertation we conduct an extensive study of minimum dominating sets (MDS) in complex networks; particularly, in scale-free networks. MDS is the smallest subset of nodes in a network that can reach every other node as nearest neighbors, thus it provides a key subset of nodes that play critical role in controllability and observability of social, biological, and technological networks. Continued interest in network control, monitoring and influencing of complex networks motivates our research of understanding the properties and practical application-related issues of the MDS.

Our study of the scaling behavior reveals that the size of MDS always scales linearly with network size, as long as the power-law degree exponent γ of the degree distribution is larger than 2. However, when $\gamma < 2$, a domination transition occurs, allowing the MDS size to become $O(1)$, leading to easily dominated networks, under certain structural conditions.

Motivated by practical applicability in large networks, we develop a new dominating set selection method, derived from probabilistic node selection techniques, which can select small dominating sets without complete network topology information. We also show that the effectiveness of our method, as well as the effectiveness of other heuristics of dominating set selection, strongly depends on the assortativity of networks.

Finally, we conduct a numerical study to analyze the fraction of nodes that remain dominated, after the network is damaged, and some nodes are removed. We find that dominating sets optimized for small size are particularly vulnerable to damage; a significant amount of "domination coverage" may be lost if key dominator nodes are deleted. However, we also find that increasing the redundancy of dominating sets by adding a few well-picked nodes can successfully increase the post-damage dominated fraction of the network. Based on this idea, we develop two algorithms to build dominating sets with flexible balance between size and damage resilience.

CHAPTER 1
INTRODUCTION

Invasion phenomena and finding critical subsets of a system that can initiate global change are of significant interest in multiple areas across scientific disciplines. In biology, a group of organisms arriving in a critical quantity can take over the habitat of a resident population [1–3]. While competition dynamics governs invasion on a local scale, it can also drive the evolution of genetic and cultural traits on larger time scales. The classical analogue in statistical physics is the description of magnetism using the Ising model [4], where a critical patch of uniformly parallel spins can spread across the system, changing all spins to the same state.

Similarly, in complex networked systems, a critical subset of nodes can spread information to all other nodes [5], observe the entire network [6], or drive dynamical processes to a desired global state [7]. The existence of tipping points in spreading dynamics on networks [8] can be interpreted as another form of critical behavior. Indeed, finding these critical nodes is one of the main focuses of network science, because they are essential in several problems, including prevention of the spread of diseases in social networks [9], monitoring infrastructural networks [6], and gathering or dispersing information efficiently [10, 11].

Finding critical limits, either as critical thresholds of order parameters in systems governed by nonlinear dynamics, or as the optimal subset of nodes in a network, represent an underlying desire to find a way to efficiently *control* these systems.

Portions of this chapter previously appeared as: F. Molnár Jr., C. Caragine, T. Caraco, and G. Korniss, "Restoration Ecology: Two-Sex Dynamics and Cost Minimization," PLoS One **8,** e77332 (2013).

Portions of this chapter previously appeared as: F. Molnár Jr., S. Sreenivasan, B. K. Szymanski, and G. Korniss, "Minimum Dominating Sets in Scale-Free Network Ensembles," Sci. Rep. **3,** 1736 (2013).

Portions of this chapter previously appeared as: F. Molnár Jr., N. Derzsy, É. Czabarka, L. Székely, B. K. Szymanski, and G. Korniss, "Dominating Scale-Free Networks Using Generalized Probabilistic Methods," Sci. Rep. **4,** 6308 (2014).

Portions of this chapter to appear as: F. Molnár Jr., N. Derzsy, B. K. Szymanski, and G. Korniss, "Building Damage-Resilient Dominating Sets in Complex Networks against Random and Targeted Attacks," (under review).

The study of critical phenomena in nonlinear dynamics and networks belongs to complex systems analysis [12]. There is no single definition for a complex system, however its defining qualities include that they are composed of multiple parts (agents, nodes, dynamical variables, etc.), and the interactions between them give rise to collective behavior and emergent phenomena, which cannot be explained using a classical bottom-up approach (i.e., by understanding the constituent parts individually without their interactions) [13].

The study of complex systems cuts across the boundaries of traditional scientific disciplines. Mathematical descriptions and model solutions of one system are often applicable to problems in other fields. For example, herding and flocking behavior can be observed in organic as well as inorganic systems [14]; reaction-diffusion models can describe animal coloration patterns [15] as well as chemical precipitation patterns [16]; cellular automata models can describe the spreading of forest fire [17] as well as diseases [18]; multi-scale stochastic models can describe climate [19] as well as micromechanics [20]. Complex networks are, naturally, an integral part of complex systems. Network theory has been successfully applied in several fields, including statistical physics [21, 22], biology [23–25] and economics [26].

In this dissertation we study two distinct topics of complex systems. The first topic explores invasion dynamics and *critical cluster* behavior in the context of a sex-structured population dynamics model, thus provides contributions to theoretical ecology as well as statistical physics. Specifically, our study contributes a theoretical exploration of how an extraordinary sex ratio's ecological consequence, population persistence or extinction, can be affected by interactions with a culturally inherited mortality trait. Further, we contribute the notion of the critical initial patch distribution, which ensures positive growth at minimum cost, with potential generalizations to any reaction-diffusion system.

The second topic of this work provides a comprehensive research of dominating sets in complex networks, with emphasis on scale-free networks. We contribute empirical and graph-theoretical descriptions of the scaling behavior of the size of minimum dominating sets (MDS). Further, we explore alternative methods for finding MDS approximations based on probabilistic node selection. Finally, we propose new

algorithms for building dominating sets that ensure maximal network domination, even when the network suffers considerable damage.

1.1 Population Dynamics and Sex Ratio

Invasion phenomena, and the spreading of species in new environments, are of significant interest in biological sciences. The uncontrolled spread of invasive species can cause not only the disruption of an ecosystem, but also significant economic losses. For example, zebra mussel (*Dreissena polymorpha*) accidentally introduced to the Great Lakes now threatens water plants, boating and fishing [27]. Kudzu (*Pueraria montana*), an invasive plant in southern United States causes significant ecological damage by growing over native flora and blocking sunlight [28]. It has been estimated that damages and control costs caused by invasive species exceed a hundred billion dollars annually in the United States [29].

In other cases, invasion is the desired goal. Ecological restoration aims to replenish an ecosystem's biodiversity [30, 31] by reintroducing species to their former habitat. Success depends on both ecological and economic criteria [32]; ecosystem managers must not only evaluate the impact of restoration on the target species, but also the total cost of deploying the control agent [33].

Consider an example where restoration failed. Historical records indicate that Canada lynx (*Lynx canadensis*) were found in New York State (NYS), but were seen only rarely during most of the 20th century [34]. Between 1989 and 1992, no fewer than 80 lynx were captured in Canada and released in the Adirondack Mountains of NYS. Each animal carried a radio-collar, so that survival and dispersal could be monitored. The lynx rapidly dispersed; mortality during dispersal was high. Lynx population density grew too low for successful reproduction, and the species is now considered extirpated in NYS [34].

Interest in controlling various species, by either preventing or promoting spreading, motivates our research of invasion dynamics. However, there is much debate in ecology on the modeling approach to take. We can either model one particular species and attempt to capture its specific details and complexity of its interactions with the environment, or we can build an abstract, strategic model from "general

principles" [35]. Some argue that ecological models must be developed alongside data and observations [36] in order to build quantitative and predictive models [36, 37]. Computational models, such as cellular automata could capture complexity successfully [38]. On the other hand, simple, mean-field type models offer mathematical tractability [39]; idealized models are successful in summarizing theoretical concepts and providing conceptual frameworks for more complex models [40], as long as their predictions are not misinterpreted [39].

We study a population dynamics model built from "general principles" in our work, because it allows us to analyze the role of one of the most basic, yet not fully understood properties of organisms, the *sex ratio*. In population biology, understanding the effects and evolution of sex ratios remains a significant topic [41–47]. Since Hamilton's seminal work [48] we know that Fisher's principle for frequency-based selection [49] does not apply when sex ratios are influenced by sex-linked genes, and sex ratios may diverge significantly from unity. However, the consequences of such sex ratios are largely unclear, as they depend on many external factors, and it remains unknown whether they provide competitive or evolutionary advantage. Our specific goal is to study the role of biased sex ratios in invasion dynamics and resource competition.

Sex ratios can be often biased by the presence of cultural traits that result in between-sex mortality difference [50]. It is certainly observable in human cultures that favor sons over daughters [51–53], e.g., the "missing women" in Asia [54]. In other species, cultural traits in habitat choice, tool use or foraging behavior can cause between-sex differences, but their relationships to specific mortality rates are unknown [55–58]. We aim to study how a cultural trait influencing male mortality might regulate ecological consequences of sex-ratio evolution.

In Chapter 2, we provide a comprehensive study of a sex-detailed population dynamics model. We contribute analytical results on the limits of sex ratio bias that can be sustained by sex-specific mortality traits. We also show that during the competition of resident and invader groups the survival outcome depends on higher allocation of genes controlling the sex ratio. Further, we find that although genetic sex-ratio traits may never coexist in a stable equilibrium, cultural mortality traits can coexist, under certain conditions.

At the end of Chapter 2, and throughout Chapter 3, we study spatial effects by extending our model with one- and two-dimensional space, and simple assumptions of diffusive dispersal. This allows us to study the fundamental phenomenon of critical cluster size (critical radius) [1–3, 59–62] and its dependence on sex ratios and mortality traits. In addition, we propose a novel approach to generalize the spatial distribution of the initial population patch, allowing us to find a distribution that minimizes the cost of restoration, while ensuring survival. Due to the generality of our approach, our results are applicable to any reaction-diffusion system with various growth dynamics.

1.2 Dominating Sets in Complex Networks

Complex network research has expanded significantly in the past two decades. Seminal works in network theory [21, 22, 63–69] revealed that many social, biological and technological networks have complex structure, such as dense clustering, heavy-tail degree distributions (in particular, power-law), and small diameter ("small-world"), that cannot be described by classical random graph models [70]. Numerous models, explanations and implications of complex network structures have been studied in detail [71].

One particular topic that has received significant attention is the controllability of dynamical processes on networks [72–81]. Network controllability refers to the ability to drive all nodes to a desired state by a subset of controller nodes, and thus it is strongly dependent on the dynamical process that takes place on the network. Similarly, network observability refers to the ability to deduce the state of all nodes from a few observations. However, in many applications, the dynamical process in question can be simplified, and simple nearest-neighbor interactions can be assumed (i.e., nodes can influence or observe their nearest neighbors only).

This reduces controllability to a purely topological problem and leads to the research of dominating sets in complex networks. By definition, a dominating set is a subset of nodes in a network, such that every node is either in the set (a dominator), or adjacent to a dominator. The smallest cardinality dominating set is called the *minimum dominating set* (MDS), and its size is defined as the domination

number $\gamma(G)$ of graph G. Dominating sets play significant role in several problems, including network controllability [7,73,82] and observability [6], selecting high-impact nodes in protein networks [23,83], social influencing [10,11], optimal sensor placement for disease outbreaks [9], resource allocation [84], bank network supervision [85], and network search [86].

The MDS is an important construct, because the inclusion of members into the dominating set often comes at a certain non-zero cost. For example, in the case of network sensor placement, if placing a sensor has a non-zero cost, and if each sensor can eavesdrop on all its neighbors, then the MDS nodes define the lowest cost placement that allows all nodes to be monitored. Continued interest in network control, detection and efficient spreading or curbing of network flows motivates our interest in understanding the properties of MDS on stylized network models, as well as real networks.

Our first contribution is a comprehensive empirical and theoretical study of the scaling of MDS size in scale-free networks, presented in Chapter 5. Understanding the scaling tells us the expected costs of implementing network domination. We also show, both theoretically and empirically, that scale-free networks may become easily dominated when a *domination transition* occurs, under certain structural conditions.

In most applications that utilize dominating sets, the main goal is to minimize the size of the selected dominating set, hence minimizing the cost of domination. However, finding the exact MDS in a general graph is one of the well-known NP-hard problems of graph theory. In addition, it has been shown that finding a sublogarithmic approximation (one that is smaller than $c \log N$ times the true MDS size, where N is the network size and $c > 0$) is also NP-hard [87]. Therefore, the only alternative is to develop heuristics and approximation algorithms for finding small dominating sets.

A commonly used method for finding an MDS approximation is a simple greedy algorithm [88–90]. It provides a theoretically optimal approximation to the MDS (specifically, it provides a factor of $1 + \log N$ approximation, in the worst case), and runs in time linear in the number of edges, which makes it ideal for practical applications. We also use this method in our study of MDS scaling behavior.

Another approach is to solve the binary integer programming equivalent of the MDS problem. In this case finding the solution remains NP-hard (since the result would be the exact MDS); but solvers such as Branch and Bound (first introduced in [91], used in e.g. [92]) can provide intermediate results that can only be used as approximate solutions. However, obtaining even one approximation requires the solution of at least one linear programming relaxation of the integer problem, which has at least $O(N^2)$ run time (complexity of processing the input), exponential run time in the worst case, and yields no significantly better approximation to the MDS size in finite waiting time than the greedy algorithm, according to our experiments.

Several methods developed recently for finding small dominating sets try to exploit certain structural features of the network. For example, hybrid genetic algorithms have been proposed to find small dominating sets [93, 94]. In addition, various algorithms have been developed to approximate special types of dominating sets, such as the minimum connected dominating set (where the induced subgraph of dominator nodes has a single connected component), applicable in wireless communication and sensor networks [5, 95–98], or the minimum independent dominating set (where none of the dominators are adjacent) [99–101]. It is also a notable area of research to develop faster exponential algorithms to find the exact MDS [102–105]. While these are significant in theoretical computer science, exponential algorithms cannot be applied to large real networks due to their lack of scalability.

In all methods mentioned above, the common theme is the assumption of the availability of global connectivity information (i.e., adjacency matrix, or equivalent information). However, in case of real large-scale networks (over tens of millions of nodes and edges) this is highly unlikely, and the cost of collecting this information could negate the benefits of efficient control by finding small dominating sets. Thus, our goal is to develop new dominating set selection strategies that rely only on limited local information, and are inherently distributable. We address this issue in Chapter 6.

Attacks on complex networks, fault tolerance, and defense strategies against damage of nodes and edges have gained significant interest in network science [106–109]. Networks with scale-free topologies have been found to be resilient against random

node damage, but vulnerable to targeted removal of high degree nodes [110–112]. Research has also focused on improving the robustness of these networks against various combinations of attacks [113–115], and on studying the dynamically progressing effects of an initial damage, such as cascading failures [116, 117].

While the topological effects of damage are well-understood, little is known about the effects on network domination. If nodes from the dominating set are deleted, some nodes in the remaining network may no longer be dominated (i.e., we may lose control over them). However, if the topology is relatively intact, then efficient control over the remaining network is still desirable. Dominating sets optimized for small size (minimum cost) may not have sufficient redundancy to provide complete domination "coverage." On the other hand, adding a small amount of well-picked nodes to the dominating set may significantly improve post-damage domination, at an acceptable cost. We explore these questions in Chapter 7, and contribute two new methods for selecting dominating sets with flexible balance between resilience and set size.

Mixing patterns are a fundamental property of complex networks [118, 119], usually measured by *assortativity*. A network is considered assortative if its nodes tend to connect to other nodes that have similar number of connections, while in a disassortative network the high degree nodes are adjacent to low degree nodes. Investigating the behavior of dominating sets with respect to assortativity is essential for deeper understanding of the network domination problem. Several studies conducted on real-world networks have shown that social systems are assortative, while technological ones exhibit disassortative behavior [119]. Social psychology studies have shown that humans are more likely to establish a connection with individuals from the same social class, or with whom they share common interests, such as education or workplace. This tendency, named homophily [120], also governs the attachment rules in real-life social systems, and it is reflected in the mixing patterns of these networks, which are of significant importance in dynamical processes on social networks. Specific connectivity schemes affect influence propagation and epidemic spread [9, 121], and are also responsible for Web page ranking [122] and internet protocol performance [123].

We contribute a new method for adjusting the assortativity in an existing network topology, described in Chapter 4. Using this method we investigate the effects of network assortativity on various dominating sets in Chapter 6 and its effects on domination stability in Chapter 7.

Finally, in Chapter 8, we provide a short proof of concept, using the linear threshold model [8, 124–127] of opinion dynamics, to illustrate the applicability of MDS in social influencing.

CHAPTER 2
CULTURAL EFFECTS ON ECOLOGICAL CONSEQUENCES OF EXTRAORDINARY SEX RATIOS

Since Fisher's classic insight [49], sex-ratio evolution [41–43] and the impact of a given sex ratio on ecological dynamics [44–47] have remained central issues in population biology. Fisher noted [49] that neither sex should be rarer at evolutionary equilibrium, a consequence of frequency-dependent selection. That is, equal investment of reproductive effort in the two sexes — commonly implying a sex ratio close to unity (1:1 ratio of males and females) — can be evolutionarily stable [128].

Hamilton [48] studied sex ratios departing significantly from unity, emphasizing that Fisher's argument of frequency-dependent selection does not apply when a sex-linked gene controls sex ratio at birth. In particular, if a gene governing sex ratio occurs in the heterogametic sex only (i.e., sex-linked genes, carried only by females in the ZW sex-determination system, and males in the XY system), the gene's fitness depends only on the number of heterogametic offspring produced. The frequency of such a gene may advance rapidly, endangering population persistence [129, 130]. That is, a biased sex ratio controlled by sex-linked genes can leave members of the more common sex without mates; the consequent "marriage squeeze" [44] may lead to population decline [48, 131]. Equivalently, an Allee effect (reduced per-capitum growth rate at low population densities [132–134]; dependent on the density of each sex) can limit the degree of sex-ratio bias, for given total density, capable of averting direct decline to extinction [45, 133–135]. Our study supposes that an extraordinary sex ratio's ecological consequence, population persistence or extinction, depends on interaction with a culturally inherited trait.

Cultural traits may enforce a between-sex mortality difference [50], leading to population densities biased toward one sex. In certain human cultures, infanticide and neglect increase female mortality [51, 52, 54]; Laland *et al.* [53] assume that these

Portions of this chapter previously appeared as: F. Molnár Jr., T. Caraco, and G. Korniss, "Extraordinary Sex Ratios: Cultural Effects on Ecological Consequences," PLoS One **7,** e43364 (2012).

cultural traits are transmitted vertically, i.e., parent to offspring. In other species, vertical cultural transmission clearly causes between-sex differences in habitat choice, tool use, or foraging behavior, but their relationships to sex-specific mortality rates are unknown [55–58]. Our models explore how a cultural trait influencing male mortality might regulate ecological consequences of sex-ratio evolution.

We treat sex ratio as a sex-linked genetic trait, and restrict cultural transmission to the vertical case [136]. Our two-sex population dynamics assumes competition for a growth-limiting resource; competition generates a strong Allee effect. Within a group, each female carries the same sex-ratio allele, and each male experiences the same mortality rate; parameters differ between groups. Resource competition is preemptive; each group has the same niche [137–139].

Our approach assumes pairwise competition between resident and invader groups, where *group* refers to population structure, not the level of selection. In Sober's [140] terminology, we associate properties driving selection with groups, and associate the objects of selection with individuals — individual females in this case. The resident group (sex ratio, male mortality culture) rests at ecological equilibrium, and we ask if a rare, different group can invade the resident. Our results for invasion, extinction, and (cultural) coexistence indicate how resource competition, cultural variation, and sex-ratio evolution interact.

2.1 Model

2.1.1 Generic Assumptions

In birds (and butterflies) sex determination follows the ZW system. W is the sex-determining chromosome; females are ZW, and males are ZZ [141]. Our model assumes that the W chromosome carries an allele fixing the sex ratio among that female's offspring. The sex linkage means that a female inherits her mother's sex ratio, and the sex-ratio gene never occurs in males. Hence, the fitness of the sex-ratio allele (and of any gene on the W chromosome) is advanced only through production of daughters [48]. To focus our discussion accordingly, we model the "female ratio," the proportion of a female's offspring born female. Females of a single group carry the same sex-ratio allele.

The assumption of sex-linkage might seem restrictive. However, in a number of bird species, individual females shed Z-chromosome and W-chromosome bearing eggs non-randomly [142, 143]. The observed variation in sex ratio among females may reflect facultative plasticity [144], but could generate some of the population-dynamic consequences of sex ratio that we model.

In addition to the genetic sex-ratio trait in females, all members of a given group share a vertically transmitted cultural norm that governs male behavior, which, in turn, fixes the male mortality rate for that group. Females of different groups share the same mortality rate. Hence, for simplicity, we assume a female adopts her mother's culture. If both parents belong to the same group, their son faithfully acquires the parental culture. When parents of different cultures (groups) mate, a son acquires one or the other culture, each with probability $^1/_2$.

Note, that our key assumption about sex ratio is the linkage to the heterogametic chromosome, not the specific sex-determination system (i.e., ZW). For example, mammals (including humans) follow the XY system, where males are the heterogametic sex (XY). In this case we could equivalently model a "male ratio" trait linked to the Y chromosome that a male inherits from his father, and cultural traits influencing female mortalities. Due to this symmetry, our model is applicable to a wide range of species, where sex is determined by specific sex chromosomes, and exhibit cultural mortality traits. In particular, our model is applicable to humans, due to several Y-linked genes [145], as well as cultural traits influencing female mortality [51–54].

To address competition between groups, we envision a resident group (a single female ratio and a single male mortality rate) at ecological equilibrium in a resource-limited environment. We then introduce (via demic/genetic migration) a small inoculum of an invader group. The resident and the rare invader differ in female ratio, and ordinarily differ in male mortality. The competitive dynamics proceeds to ecological equilibrium. If the rare female-ratio allele has positive growth, it will drive change in culture. Since individuals mate randomly, extinction of a group's female-ratio allele need not always imply loss of the associated cultural trait. However, loss of a cultural mortality trait implies that the associated female-ratio allele has been excluded competitively.

2.1.2 Mathematical Model

Based on our generic assumptions above we can formulate the model mathematically. We consider two-sex population growth with two female ratio/male mortality groups; the groups allow us to model resident-invader differences. When a female of group i ($i = 1, 2$) reproduces, the resulting offspring is female with probability θ_i, and male with probability $(1 - \theta_i)$, independently of the group of the male with whom she mates. θ_i is the female ratio for group i, transmitted faithfully from mother to daughter. Different groups, by definition, differ in female ratio. All females have the same mortality rate, μ_{f}.

A male's group specifies his mortality rate, μ_i ($i = 1, 2$). If male mortality exceeds the rate for females, $\mu_1, \mu_2 > \mu_{\mathrm{f}}$. But we do not exclude the case where the female mortality exceeds one or both male rates. If both parents belong to the same group, each male offspring has that group's mortality rate, acquired by vertical cultural transmission. If a male's parents belong to different groups, the male acquires mortality rate μ_i with probability $1/2$.

F_i and M_i represent the global density of females and males, respectively, of group i. All individuals require the same resources, so that population growth at larger densities will self-regulate. The preceding assumptions imply the following dynamics under homogeneous mixing (or "mean-field"):

$$\begin{aligned}
\partial_t F_1 &= \theta_1 (1 - N) F_1 (M_1 + M_2) - \mu_{\mathrm{f}} F_1 \\
\partial_t M_1 &= (1 - N) \left[(1 - \theta_1) F_1 \left(M_1 + \frac{M_2}{2} \right) + (1 - \theta_2) F_2 \left(\frac{M_1}{2} \right) \right] - \mu_1 M_1 \\
\partial_t F_2 &= \theta_2 (1 - N) F_2 (M_1 + M_2) - \mu_{\mathrm{f}} F_2 \\
\partial_t M_2 &= (1 - N) \left[(1 - \theta_2) F_2 \left(\frac{M_1}{2} + M_2 \right) + (1 - \theta_1) F_1 \left(\frac{M_2}{2} \right) \right] - \mu_2 M_2, \quad (2.1)
\end{aligned}$$

where $N = F_1 + M_1 + F_2 + M_2$ is total global density; $0 \leq N \leq 1$. Males encounter females as a mass-action process, modeling random mating [135, 136]; more complicated assumptions about pair formation suggest different "marriage functions" [47]. The fraction of matings that reproduce successfully equals the unoccupied fraction of the environment, $(1 - N)$. Below, we take group 1 as the resident, and identify group 2 as the (initially rare) invader.

If only a single group occupies the environment, the equations reduce to those studied by Tainaka *et al.* [130]:

$$\begin{aligned} \partial_t F &= \theta\,(1-M-F)\,FM - \mu_{\rm f} F \\ \partial_t M &= (1-\theta)\,(1-M-F)\,FM - \mu_{\rm m} M\ . \end{aligned} \tag{2.2}$$

The authors focused on the symmetric case, $\mu_{\rm f} = \mu_{\rm m}$. An important feature of this model is that the cubic dynamics produces a strong Allee effect [133, 134]. That is, there exists a threshold for the initial population density, below which growth is necessarily negative, and extinction must follow [135]. The single-group model serves as the starting point of our analysis. In particular, initial conditions of our competition dynamics will depend on the stable, non-trivial fixed point of the single-group model (corresponding to positive equilibrium densities for females and males of group 1).

2.2 Analysis of Dynamics

In this section we provide a complete description of the fixed points (stationary solutions) of Eqs. (2.1) and (2.2) analytically, and we use both analytic and numerical methods to analyze their local stability.

Since this study employs extensive numerical integration, we justify our choice of an ordinary differential equation (ODE) solver. Equations (2.1) are strongly coupled and may become stiff, a challenge to the solver. Speed is another important factor, because we want to map the entire parameter space of the model, which requires a very large amount of computation. We choose the explicit fourth-order Runge-Kutta method [146], which gives the precision we require. We also utilize adaptive time stepping to avoid problems with any potential stiffness, and to increase integration speed when the slopes of the densities are small. Since we are interested in stationary solutions of the equations, the stopping condition for the integration specifies that all numerical derivatives are smaller than a predetermined limit:

$$\frac{\Delta A}{\Delta t} < \epsilon,\ A \in \{F_1, M_1, F_2, M_2\} \tag{2.3}$$

In our ODE numerical integrations, we set the stopping condition at $\epsilon = 10^{-8}$.

2.2.1 Stability of the Resident Group

The first step of our analysis is establishing the necessary conditions for an ecologically stable resident population. A stable resident occupies the habitat alone, at a real, positive fixed point where self-regulation limits growth, governed by Eqs. (2.2).

In general, the system has three fixed points. There is one trivial solution at zero density, corresponding to extinction:

$$(F^0, M^0) = (0, 0), \tag{2.4}$$

and a pair of nonzero fixed points, which can be obtained by first manipulating the two stationary state equations, Eqs. (2.2), to write a simple quadratic equation for the stationary total density:

$$N(1-N) = \frac{\mu_{\rm f}}{\theta} + \frac{\mu_{\rm m}}{1-\theta}\,, \tag{2.5}$$

yielding solutions

$$N^{\pm} = \frac{1 \pm \sqrt{D}}{2}, \tag{2.6}$$

with

$$D(\mu_{\rm f}, \mu_{\rm m}, \theta) = 1 - 4\left(\frac{\mu_{\rm f}}{\theta} + \frac{\mu_{\rm m}}{1-\theta}\right)\,. \tag{2.7}$$

Therefore, the nonzero fixed points are:

$$(F^{\pm}, M^{\pm}) = \left(\frac{\mu_{\rm m}}{1-\theta}\cdot\frac{1}{1-N^{\pm}}\,,\ \frac{\mu_{\rm f}}{\theta}\cdot\frac{1}{1-N^{\pm}}\right)\,. \tag{2.8}$$

The stability of these fixed points can be easily analyzed by linearizing Eqs. (2.2); the "−" fixed point is unstable, and the "+" one is stable. The nonzero fixed points, hence a stable positive equilibrium (an ecologically stable resident), exist only if $D > 0$, for which the necessary condition is:

$$\sqrt{\mu_{\rm f}} + \sqrt{\mu_{\rm m}} < 1/2\,. \tag{2.9}$$

The biological significance of the structure of the above solutions is two-fold [130]. First, for $D > 0$, the system exhibits the Allee effect. Unless the (initial) population density is sufficiently high ($N(0) > N^-$), the zero fixed point attracts the dynamics, and the population goes extinct. Second, provided that $\sqrt{\mu_\mathrm{f}} + \sqrt{\mu_\mathrm{m}} < 1/2$, there is only a finite interval $\theta_{\mathrm{c}1}(\mu_\mathrm{f}, \mu_\mathrm{m}) < \theta < \theta_{\mathrm{c}2}(\mu_\mathrm{f}, \mu_\mathrm{m})$, where $D > 0$, i.e., where the population can persist at equilibrium; see Fig. 2.1. These boundaries, functions of the culturally transmitted mortality rate, are given by:

$$\theta_{\mathrm{c}1,2}(\mu_\mathrm{f}, \mu_\mathrm{m}) = \frac{(1 + 4\mu_\mathrm{f} - 4\mu_\mathrm{m}) \pm \sqrt{(1 + 4\mu_\mathrm{f} - 4\mu_\mathrm{m})^2 - 16\mu_\mathrm{f}}}{2} \, . \tag{2.10}$$

Note, that as we increase either male or female mortalities, the range of possible θ values capable of maintaining an extant resident population shrinks, see Fig. 2.1. Essentially, Fig. 2.1 shows a cross-section of the parameter space where $D > 0$. For a complete picture, Fig. 2.2 shows the entire region of the parameter space satisfying $D > 0$.

We can also identify the optimal sex ratio, corresponding to the maximum total density. Between the two critical θ values, at

$$\theta^* = \frac{1}{1 + \sqrt{\mu_\mathrm{m}/\mu_\mathrm{f}}} \, , \tag{2.11}$$

the global maximum density is:

$$N^{\max} = N^+(\theta^*) = \frac{1 + \sqrt{1 - 4(\sqrt{\mu_\mathrm{f}} + \sqrt{\mu_\mathrm{m}})^2}}{2} \tag{2.12}$$

where the female to male density ratio is $F^*/M^* = \sqrt{\mu_\mathrm{m}/\mu_\mathrm{f}}$.

As a check, we can use the analytical stationary densities to quantify our numerical integration's accuracy. We performed 5000 test runs with randomly chosen parameters that obey $D(\mu_f, \mu_m, \theta) > 0$. For $\epsilon = 10^{-8}$ as the stopping condition, we find that the absolute difference of the numerically computed fixed point was only $9.5 \times 10^{-7} \pm 14\%$ from the analytical value, with 95% confidence. This accuracy suffices for our work.

As we noted earlier, the population must not only have parameters satisfying

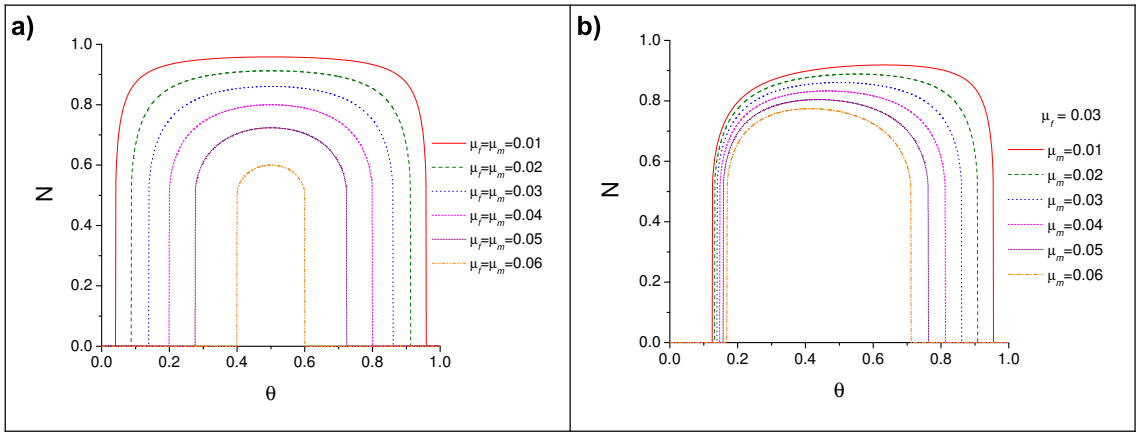

Figure 2.1: Stationary total population density as a function of female ratio at birth, for various mortality rates. (a) identical female and male mortality rates; (b) female and male mortality rates differ.

$D > 0$, but also the initial growth must overcome a strong Allee effect [134], in order to reach stable, positive equilibrium. The Allee effect defines a separatrix on the phase map of initial female and male densities, separating the two attraction basins of the stable fixed points. Below the separatrix (closer to zero density) extinction always results, independently of other parameters, since the growth rate is negative. Above the separatrix the population grows to self-regulated equilibrium. To find this threshold numerically, at given model parameters, we sweep over initial female density values, and at each female density we conduct a search using bisection for the initial male-density threshold value, numerically integrating Eqs. (2.2) until they converge to a stationary value (zero or nonzero). Using this method we can determine the threshold value with arbitrary precision.

Figure 2.3 displays the Allee-threshold for various parameter combinations. In Fig. 2.3(a), where $\mu_{\rm f} = \mu_{\rm m}$, an unbiased female ratio ($\theta = 0.5$) allows the lowest total population density before extinction due to the Allee effect ensues. When the sexes have the same mortality, unbiased sex allocation also maximizes total population density at positive equilibrium [130].

Figure 2.3(b) verifies that increasing female mortality, $\mu_{\rm f}$, for given θ and $\mu_{\rm m}$, expands the region where the Allee effect leads to extinction. Not surprisingly, increasing male mortality produces a parallel effect. Mortality-rate asymmetry and

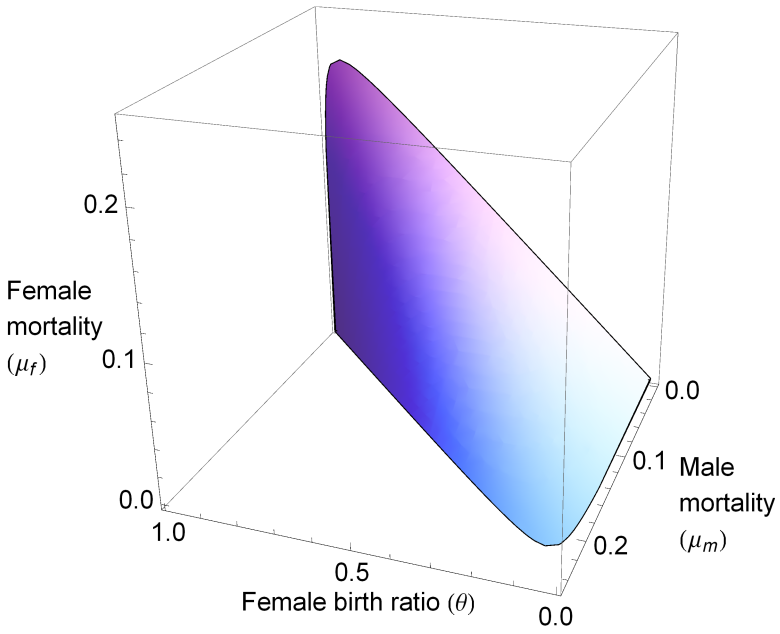

Figure 2.2: Region of the parameter space where the resident is persistent. Parameter space region defined by $D(\mu_f, \mu_m, \theta) > 0$. Choosing parameters from the indicated domain always results in a stable nonzero population, given sufficiently high initial densities.

biased female ratios distort the shape of the thresholds in Fig. 2.3, but the same general patterns emerge.

For a resident population, we have specified how existence of a positive equilibrium depends on the interaction of female ratio at birth and sex-specific mortalities. We have also shown that initial conditions (given existence of a positive equilibrium) required to avert extinction due to the Allee effect depend on the same parameters. A practical consequence is that we must choose initial densities for numerical integration carefully, so that when the competitive dynamics results in extinction, we can clearly identify the reason as either the Allee effect or exclusion.

2.2.2 Dynamics of Two Interacting Groups

Given a complete picture of single-group dynamics, we now focus on between-group interactions to quantify how population consequences of female-ratio evolution can be affected by male mortality. Equations (2.1) govern the dynamics of the

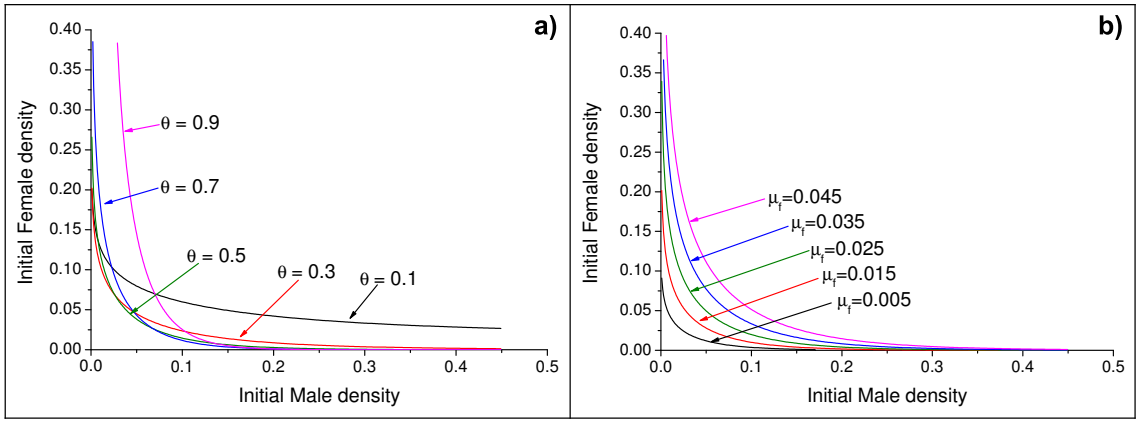

Figure 2.3: Allee threshold of the resident. Survival/extinction threshold defined by the Allee-effect, at various female ratios (a) and various female mortalities (b). Other parameters: (a): $\mu_{\rm f} = \mu_{\rm m} = 0.02$**; (b):** $\theta = 0.5$**,** $\mu_{\rm m} = 0.02$**.**

system, which has nine fixed points. One is the trivial fixed point where all densities vanish, and we discuss the rest as follows.

We can easily identify fixed points directly related to those of the single-group case; there are two symmetric pairs of nonzero fixed points that we would find in each single group model, while the other group is excluded (having zero density). Identically, these densities are given by Eq. (2.8); the "−" solution is always unstable, while the "+" one is stable, but only if the male mortality rate of the allele with the higher female ratio is less than twice the male mortality rate of the other allele. For example, when allele 1 is excluded, these two fixed points are

$$\begin{aligned} (F_1^{\pm}, M_1^{\pm}) &= (0\ ,\ 0) \\ (F_2^{\pm}, M_2^{\pm}) &= \left(\frac{\mu_2}{1-\theta_2} \cdot \frac{1}{1-N^{\pm}}\ ,\ \frac{\mu_{\rm f}}{\theta_2} \cdot \frac{1}{1-N^{\pm}} \right)\ , \end{aligned} \tag{2.13}$$

where $N^{\pm}$ is given by Eq. (2.6) with $\mu_{\rm m} = \mu_2$ and $\theta = \theta_2$. Further, the "+" solution above is stable provided that $\mu_2/\mu_1 < 2$ and $\theta_2 > \theta_1$. We obtain the symmetric case by interchanging the extant allele/culture with that excluded:

$$\begin{aligned} (F_1^{\pm}, M_1^{\pm}) &= \left(\frac{\mu_1}{1-\theta_1} \cdot \frac{1}{1-N^{\pm}}\ ,\ \frac{\mu_{\rm f}}{\theta_1} \cdot \frac{1}{1-N^{\pm}} \right) \\ (F_2^{\pm}, M_2^{\pm}) &= (0\ ,\ 0)\ , \end{aligned} \tag{2.14}$$

where $N^{\pm}$ is given by Eq. (2.6) with $\mu_{\rm m} = \mu_1$ and $\theta = \theta_1$. The "+" solution above is stable provided that $\mu_1/\mu_2 < 2$ and $\theta_1 > \theta_2$. We refer to stable fixed points of this type [i.e., where the densities of the extant allele are given by the "+" solution in Eqs. (2.13) and Eqs. (2.14)] as type-I fixed points.

It is important to emphasize that if both groups are present initially, then competitive exclusion leaves one group extinct, and one extant, if $\frac{1}{2} < \mu_2/\mu_1 < 2$. Therefore only one of the "+" solutions is locally stable, and it belongs to the group with the *greater female ratio*. In other words, when male mortality rates imply a type-I fixed point, the greater female ratio always excludes the lesser ratio.

The four remaining fixed points (again, forming two pairs by symmetrically exchanging group 1 with group 2) are qualitatively distinct from those discussed above. At these fixed points only one female-ratio allele remains extant, but male mortality traits "coexist" (i.e., male cultural dimorphism). That is, the population is genetically uniform, in that all females carry the same female ratio allele, but the (male) population is culturally dimorphic; father to son transmission maintains the culture of the group whose females have been excluded competitively [see Eqs. (2.1)]. We refer to the corresponding fixed points as type-II fixed points.

Consider a stable fixed point of this sort, when $F_1 = 0$. The remaining equations for the stationary state then become

$$\begin{aligned} 0 &= (1-N)\,(1-\theta_2)\,F_2\left(\frac{M_1}{2}\right) - \mu_1 M_1 \\ 0 &= \theta_2\,(1-N)\,F_2\,(M_1+M_2) - \mu_{\rm f} F_2 \\ 0 &= (1-N)\,(1-\theta_2)\,F_2\left(\frac{M_1}{2}+M_2\right) - \mu_2 M_2\ , \end{aligned} \tag{2.15}$$

where the overall density is now $N = M_1 + M_2 + F_2$. After some tedious algebra, we again find a simple quadratic equation for the overall density:

$$N(1-N) = \frac{\mu_{\rm f}}{\theta_2} + \frac{2\mu_1}{1-\theta_2}\ , \tag{2.16}$$

which has the solutions

$$\tilde{N}^{\pm} = \frac{1 \pm \sqrt{\tilde{D}}}{2} \tag{2.17}$$

with

$$\tilde{D}(\mu_{\rm f},\mu_1,\theta_2) = 1 - 4\left(\frac{\mu_{\rm f}}{\theta_2} + \frac{2\mu_1}{1-\theta_2}\right) . \tag{2.18}$$

The fixed points follow from Eqs. (2.15) after some further elementary manipulations:

$$\begin{aligned}(F_1^{\pm}, M_1^{\pm}) &= \left(0\ ,\ \frac{\mu_2/\mu_1 - 2}{\mu_2/\mu_1 - 1}\cdot\frac{\mu_{\rm f}}{\theta_2}\cdot\frac{1}{1-\tilde{N}^{\pm}}\right) \\ (F_2^{\pm}, M_2^{\pm}) &= \left(\frac{2\mu_1}{1-\theta_2}\cdot\frac{1}{1-\tilde{N}^{\pm}}\ ,\ \frac{1}{\mu_2/\mu_1 - 1}\cdot\frac{\mu_{\rm f}}{\theta_2}\cdot\frac{1}{1-\tilde{N}^{\pm}}\right)\end{aligned} \tag{2.19}$$

The necessary conditions are $\mu_2/\mu_1 > 2$, $\theta_2 > \theta_1$, and positivity of the discriminant,

$$\tilde{D}(\mu_{\rm f},\mu_1,\theta_2) = 1 - 4\left(\frac{\mu_{\rm f}}{\theta_2} + \frac{2\mu_1}{1-\theta_2}\right) > 0 . \tag{2.20}$$

The preceding condition holds if

$$\sqrt{\mu_{\rm f}} + \sqrt{2\mu_1} < 1/2. \tag{2.21}$$

The "$-$" solution in Eqs. (2.19) is always unstable, while the "$+$" solution can be stable, if the necessary conditions above are satisfied. In that case, there is a finite range of $\tilde{\theta}_{\rm c1}(\mu_{\rm f},\mu_1) < \theta_2 < \tilde{\theta}_{\rm c2}(\mu_{\rm f},\mu_1)$ where $\tilde{D}(\mu_{\rm f},\mu_1,\theta_2) > 0$, so that cultural coexistence persists. The boundaries of this coexistence region are given by:

$$\tilde{\theta}_{\rm c1,2}(\mu_{\rm f},\mu_1) = \frac{(1+4\mu_{\rm f}-8\mu_1) \pm \sqrt{(1+4\mu_{\rm f}-8\mu_1)^2 - 16\mu_{\rm f}}}{2} . \tag{2.22}$$

Within this regime, the overall population density is maximal at

$$\theta_2^* = 1/(1+\sqrt{2\mu_1/\mu_{\rm f}}), \tag{2.23}$$

where the overall female to male density ratio is $F_2^*/(M_1^* + M_2^*) = \sqrt{2\mu_1/\mu_{\rm f}}$. Interestingly, at the stable fixed point in Eq. (2.19) the male density ratio is $M_1/M_2 = \mu_2/\mu_1 - 2$; hence the relative abundances of the male cultural trait values do not depend on the female ratio.

Analogously, one can obtain fixed points of the same form as in Eqs. (2.19) by choosing $F_2 = 0$ and simply interchanging indices 1 and 2 in all respective expressions,

leading to:

$$\begin{aligned}(F_1^{\pm}, M_1^{\pm}) &= \left(\frac{2\mu_2}{1-\theta_1}\cdot\frac{1}{1-\bar{N}^{\pm}}\ ,\ \frac{1}{\mu_1/\mu_2-1}\cdot\frac{\mu_{\rm f}}{\theta_1}\cdot\frac{1}{1-\bar{N}^{\pm}}\right)\\ (F_2^{\pm}, M_2^{\pm}) &= \left(0\ ,\ \frac{\mu_1/\mu_2-2}{\mu_1/\mu_2-1}\cdot\frac{\mu_{\rm f}}{\theta_1}\cdot\frac{1}{1-\bar{N}^{\pm}}\right)\end{aligned} \tag{2.24}$$

with

$$\bar{N}^{\pm} = \frac{1\pm\sqrt{\bar{D}}}{2} \tag{2.25}$$

and

$$\bar{D}(\mu_{\rm f},\mu_2,\theta_1) = 1 - 4\left(\frac{\mu_{\rm f}}{\theta_1}+\frac{2\mu_2}{1-\theta_1}\right)\ . \tag{2.26}$$

Note, that similarly to the type-I fixed points, only one of the "+" solutions in either Eqs. (2.19) or (2.24) can be stable at one time, and again, it belongs to the group with the *greater female ratio*. In other words, when male mortality rates imply a type-II fixed point, females with the greater female ratio exclude the other group's females, while males from both groups coexist.

To summarize, the model does not permit equilibrium coexistence of female ratio alleles, but can permit equilibrium diversity in cultural traits governing male mortality. Also note, as is clear from the necessary conditions, that of type-I and type-II fixed points, *only one* can be stable at a time. In Fig. 2.4 we illustrate the flow in the mean-field dynamics for a set of parameters when both type-I and type-II fixed points exits, but in the presence of co-occurring males of the other allele, only type-II is stable.

2.3 Invading a Stable Resident Group

Having obtained the nine fixed points for the two-group model analytically, we approach the stability analysis, and the analysis of possible outcomes of an invasion numerically. Analytical study of the system's stability proves difficult, due to the number of variables and parameters (4 variables and 5 parameters). To be as thorough as possible, we performed numerical integration systematically to span a significant region of the five-dimensional parameter space. The range and step of the parameters in our numerical scheme can be found in Table 2.1.

Each run begins with a stationary resident population, with allele 1 and cultural

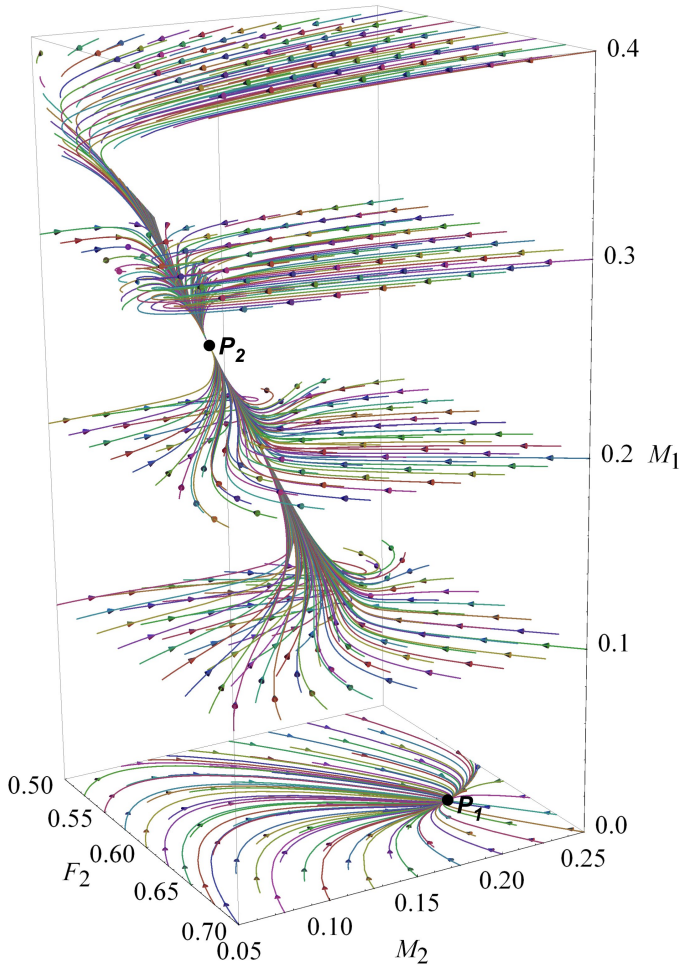

Figure 2.4: Mean-field density flows in the (M_1, F_2, M_2) space (restricted to $F_1 \equiv 0$) with type-I fixed point P_1 (stable if $M_1 = 0$) and type-II fixed point P_2 (stable if $M_1 > 0$) for $\mu_f = 0.02$, $\mu_1 = 0.01$, $\mu_2 = 0.04$; $\theta_1 = 0.4$, $\theta_2 = 0.6$.

trait μ_1. If model parameters allow a stable positive equilibrium, we choose initial densities accordingly. We then introduce the invaders, with female-ratio allele 2 and cultural trait μ_2. For each set of parameters (in each series) we performed two runs, one with infinitesimal initial density of invaders (10^{-4}) and one with high invader density (0.45).

To portray the results, we generate a number of "4D" plots. Each shows a table containing 2D plots with the results of each run; the axes of each 2D plot are values of the same two cultural parameters (μ_1 and μ_2, all with the same range).

Table 2.1: Parameter regions and step sizes for numerical integration. Each set of parameters identifies two runs: one with high (0.45) and one with low (10^{-4}) initial invader density.

Parameter	Lower bound	Upper bound	Step
Series 1			
θ_1	0.01	0.99	0.01
θ_2	0.01	0.99	0.01
μ_1	0.01	0.04	0.005
μ_2	0.01	0.04	0.005
μ_f	0.01	0.04	0.005
Series 2			
θ_1	0.1	0.9	0.1
θ_2	0.1	0.9	0.1
μ_1	0.001	0.1	0.001
μ_2	0.001	0.1	0.001
μ_f	0.01	0.04	0.01

Another two parameters (female ratios θ_1 and θ_2) vary across the rows and columns of the tables (the 4D plots). We produce as many tables as required by the range of the fifth parameter (female mortality μ_{f}). In each 2D plot, one pixel represents the final stationary densities of the female ratio alleles. The pixel's location corresponds to the parameters for which it was computed; resident and invader allele densities are shown on different color channels. This way, we can visually compare all the results simultaneously, simplifying the analysis greatly. Figure 2.5 shows one 4D plot; the associated parameter ranges produce the full set of the model's outcomes.

In the following sections we investigate the necessary and sufficient conditions for successful (pairwise) genetic invasion of the resident female ratio, and the necessary conditions for cultural "coexistence."

2.3.1 Invasion to Exclusion of the Resident

Figure 2.6(a) shows an example of successful invasion leading to exclusion of both the resident allele and resident culture. Following introduction of the invading group, the resident density drops quickly, and the successful allele (females) and successful culture (observed in males) advance to become the new resident group. As a numerical check, we note that both infinitesimal and high invader densities always result in identical final densities.

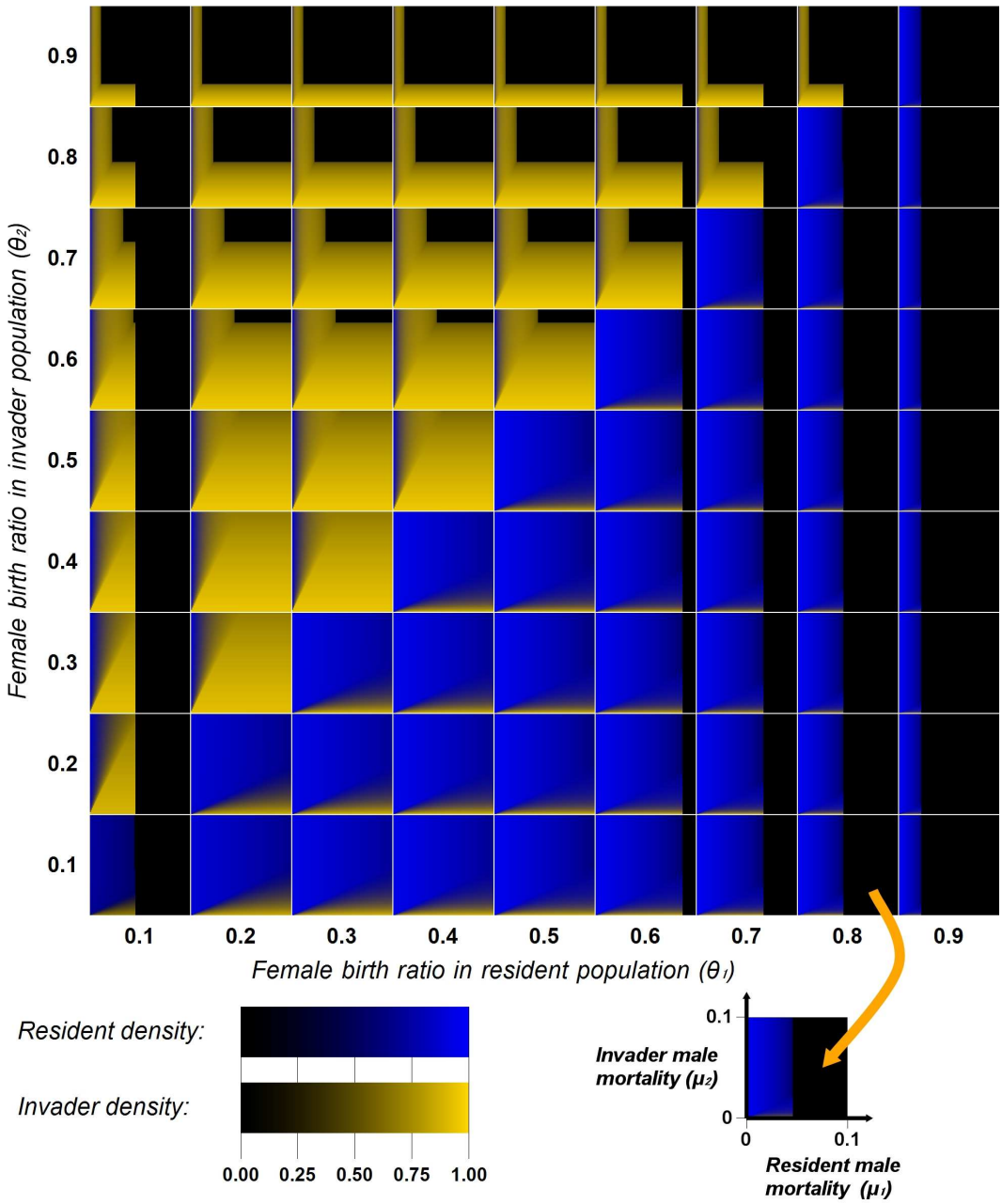

Figure 2.5: Stationary population densities. The stationary resident (group 1) is invaded by group 2, initially at an infinitesimal density (10^{-4}). Large axes indicate common parameters in rows and columns; every tile has the same axes, scaled as indicated in the bottom right corner. Color scales use independent color channels to show resident and invader densities independently. Fixed parameter: $\mu_f = 0.02$.

Our numerical results on Fig. 2.5 show immediately that female ratios determine the outcome of invasion; a successful invader in pairwise competition has the greater female ratio. That is, successful invasion always requires $\theta_2 > \theta_1$, and $\theta_2 < \theta_1$ assures that the resident resists invasion. Furthermore, successful invasion by a female-ratio allele assures that the associated cultural trait (with value μ_2) advances from rarity.

Notice, that for a successful invasion the greater female ratio is only a necessary, but not a sufficient condition. It only determines which sex-ratio allele excludes the other. Since the female-ratio allele is sex-linked, dependence of invasion on $(\theta_2 - \theta_1)$ simply recalls Hamilton [48]. But in our model, the ecological effect of invasion also depends on the culturally transmitted trait. Achieving both genetic and cultural exclusion of the resident (F_1=0, M_1=0, i.e., type-I fixed point for the invader) further requires $\sqrt{\mu_\mathrm{f}} + \sqrt{\mu_2} < 1/2$ and $\mu_2/\mu_1 < 2$.

In addition to the necessary conditions for the type-I fixed point, there is an additional condition that guarantees the survival of the invader after it has excluded the resident allele and cultural trait. That is, the invader alone must also be stable, for which the sufficient condition is $\theta_{\mathrm{c}1}(\mu_\mathrm{f}, \mu_2) < \theta_2 < \theta_{\mathrm{c}2}(\mu_\mathrm{f}, \mu_2)$ [see Eq. (2.10)], given by the positivity requirement of the corresponding discriminant [Eq. (2.7)]. We explore the outcomes of cases when the invader does not satisfy the conditions above in the following sections.

2.3.2 Invasion to Cultural Coexistence

Recall that Eqs. (2.1) do not have fixed points where differing female ratios co-occur. The model, however, does allow for cultural coexistence, where males of both groups co-occur, but females of only one group remain extant.

Figure 2.6(c) displays an example where the resident culture, but not the resident allele, persists after successful invasion. The invader has the greater female ratio ($\theta_2 > \theta_1$), and therefore excludes the resident allele competitively. However, the final equilibrium state is a type-II fixed point where the resident's male-mortality trait persists via father-to-son cultural transmission. The ratio of males at dynamic equilibrium is $M_1/M_2 = \mu_2/\mu_1 - 2$. Note that the competitively driven increase in female ratio produces a decrease in total population density (females plus males) at equilibrium [Fig. 2.6(c)].

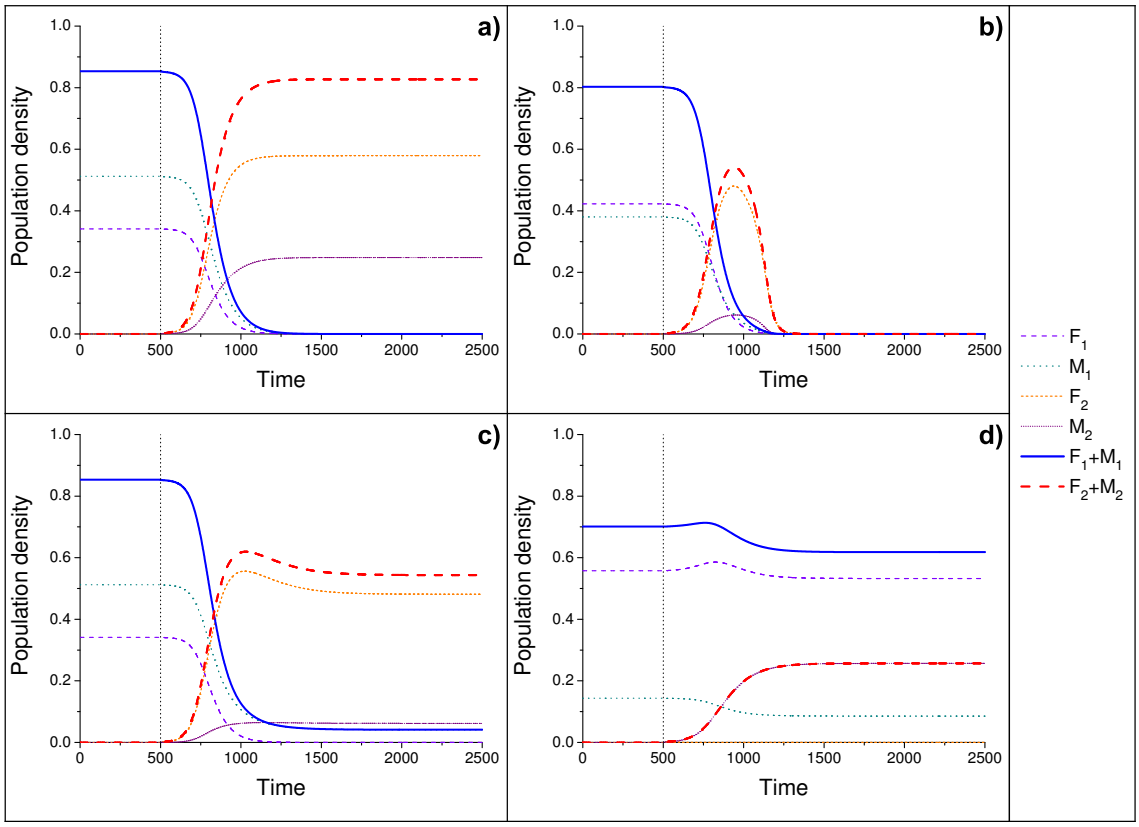

Figure 2.6: Population-density time series. (a) successful invasion; (b) invasion followed by extinction; (c) coexistence of resident males with the invader allele; (d) coexistence of invader males with resident allele. The vertical dotted line indicates the time when the invader was added to the system, at 10^{-3} density (both males and females). Common parameter: $\mu_f = 0.03$. Individual parameters: (a) $\theta_1 = 0.4$, $\theta_2 = 0.7$, $\mu_1 = 0.03$, $\mu_2 = 0.03$; (b) $\theta_1 = 0.4$, $\theta_2 = 0.7$, $\mu_1 = 0.05$, $\mu_2 = 0.08$; (c) $\theta_1 = 0.4$, $\theta_2 = 0.7$, $\mu_1 = 0.03$, $\mu_2 = 0.08$ (here, F_1=0 in the final equilibrium); (d) $\theta_1 = 0.7$, $\theta_2 = 0.4$, $\mu_1 = 0.05$, $\mu_2 = 0.01$ (here, F_2=0 in the final equilibrium).

In order to achieve a cultural coexistence, where resident females are excluded (F_1=0, i.e., type-II fixed point), the following necessary conditions must be met: $\sqrt{\mu_f} + \sqrt{2\mu_1} < 1/2$ and $\mu_2/\mu_1 > 2$. The latter condition is the opposite of what we have for a type-I fixed point, and indeed it decides whether a type-I fixed point (complete exclusion) or a type-II fixed point (cultural coexistence) is the conclusion, as long as all other conditions are met.

Notice, that for a type-II fixed point, the necessary conditions do not include the stability requirement for neither the invader nor the resident group standing alone. In essence, type-II fixed points open the possibility for groups to survive together, while they could not sustain themselves alone. This enables a scenario where we can save an unstable resident population from extinction by introducing another allele and/or culture to the group. We explore these possibilities in Section 2.4.

While the sufficient conditions for the survival of either resident or invader alone are not required in a type-II fixed point, we do have a similar sufficient condition for the cultural coexistence: the invaders' female ratio must fall into a finite interval, $\tilde{\theta}_{c1}(\mu_f, \mu_1) < \theta_2 < \tilde{\theta}_{c2}(\mu_f, \mu_1)$ [see Eq. (2.22)], given by the positivity requirement of the corresponding discriminant [Eq. (2.20)].

By the symmetry of the equations, a type-II stable fixed point also exists with F_2=0. That is, the resident population resists invading females ($\theta_2 < \theta_1$), but the introduced cultural trait advances from rarity and persists at equilibrium; see Fig. 2.6(d). The ratio of males at this equilibrium is $M_2/M_1 = \mu_1/\mu_2 - 2$. Put simply, we can exchange the resident-invader roles of the two groups, and reach the same dynamic equilibrium. Necessary conditions for this case are $\sqrt{\mu_f} + \sqrt{2\mu_2} < 1/2$ and $\mu_2/\mu_1 < 1/2$.

Figure 2.5 includes cases of equilibrium cultural coexistence. For example, in the tile where $\theta_1 = 0.2$ and $\theta_2 = 0.4$, the sharp change in color along the line $\mu_2 = 2\mu_1$ indicates the condition for cultural coexistence. When this condition is not satisfied, the culture associated with the lower female birth ratio always declines to extinction.

2.3.3 Invasion to Extinction

Given the competitive advantage of increased female allocation in our model, evolution of the sex-linked trait might threaten population persistence. Our model's dynamics includes a case where successful invasion of a stable resident is followed by extinction of the entire population. We observe this result in numerical experiments where the invader has both the greater female ratio ($\theta_2 > \theta_1$) and the greater male mortality rate, so that the sufficient conditions for both type-I and type-II fixed

points fail to hold. The greater female ratio drives invasion, but the invader's combined genetic-cultural demography does not satisfy the condition for a stable, positive equilibrium. Hence, the successful invader would not advance from rarity absent the resident group.

Figure 2.5 shows an example of invasion to extinction; note the black region of the tile where $\theta_1 = 0.4$ and $\theta_2 = 0.7$. For a particular mortality-rate combination, Fig. 2.6(b) depicts the time-dependent densities for a case of invasion to extinction. The necessary conditions for invasion are met ($\theta_2 > \theta_1$, $\sqrt{\mu_\mathrm{f}} + \sqrt{\mu_2} < 1/2$, and $\mu_2/\mu_1 < 2$). However, the sufficient condition for the invader's survival does not hold: $\theta_2 > \theta_{c2}$. Hence the invader grows when rare and excludes the resident, but the invader cannot persist. Essentially, the invading female ratio allele increases its initial density by "exploiting" males of the resident group while competing for resources with resident females. After some time the density of the resident females reaches zero. The reduced density of females means that the production of males (both resident and invader) is reduced. Consequently, the invading group, once occupying the environment alone, cannot maintain a positive equilibrium density, and a "marriage squeeze" takes the population to extinction.

Given this result, one can envision a stable population where immigration or mutation introduces new alleles over a lengthy time scale. If a new allele has a higher female ratio than the current resident, it will advance. A series of allelic substitutions might increase the female ratio continuously. Our model does not prevent the female ratio from surpassing the threshold defined by Eq. (2.10), where the population begins to decline to extinction — recalling Hamilton's [48] comment on sex linkage and sex-ratio evolution.

2.4 Rescuing an Unstable Resident Group

In the previous section we have explored scenarios where the resident group was assumed to be initially in an ecologically stable state. However, invasion of an unstable, declining resident population is also a possible scenario. While invasions ending in type-I fixed points (where the invaders exclude the resident sex-ratio allele and the cultural trait) are the trivially identical in this case, type-II fixed points

(coexistence of cultural traits) open the possibility to save the resident population from extinction by introducing a suitable invader.

Recall that a resident group has two conditions for stability. First, when $\sqrt{\mu_f} + \sqrt{\mu_m} < 1/2$, the discriminant in the fixed-point solutions is positive. This allows for a finite range of sex ratios, $\theta_{\rm c1}(\mu_{\rm f}, \mu_{\rm m}) < \theta < \theta_{\rm c2}(\mu_{\rm f}, \mu_{\rm m})$, where the resident is ecologically stable [see Eq. (2.10)]. Violation of either conditions makes the resident unstable, where any initial densities rapidly reduce to zero. However, if an invader is introduced while the resident population is still sufficiently large, extinction may be averted.

2.4.1 Saving Both Genetic and Cultural Traits

The only possible way to save both males (expressing the cultural trait) and females (expressing the female ratio allele) is to reach a type-II fixed point, where F_2=0. In this case, the invader's male-mortality trait persists via father-to-son cultural transmission, while resident females benefit from the presence of additional males, providing sufficient mating partners. Indirectly, resident males also benefit, because they are sustained at a positive equilibrium density.

In case of a stable resident, the invader must have a smaller sex ratio ($\theta_2 < \theta_1$) to ensure that the resident females remain extant, and invader females go extinct. However, this is not a requirement here, because we do not need to actually introduce invader females to the resident group. Since our goal is to exclude them eventually, it is sufficient to introduce only a new cultural trait, with male-only invaders. Further, the introduction of the new cultural trait does not have to come from immigration; spontaneous emergence of a new trait with the right mortality rates is sufficient to avert extinction.

For the newly introduced male mortality trait, the necessary and sufficient conditions for a type-II fixed point must still hold. In particular, the invader male mortality must be less than half that of the resident ($\mu_2 < \mu_1/2$), otherwise the invader mortality trait would not be sustainable; $\sqrt{\mu_f} + \sqrt{2\mu_2} < 1/2$ must hold to have a positive discriminant for type-II fixed points; and finally, we have a maximum

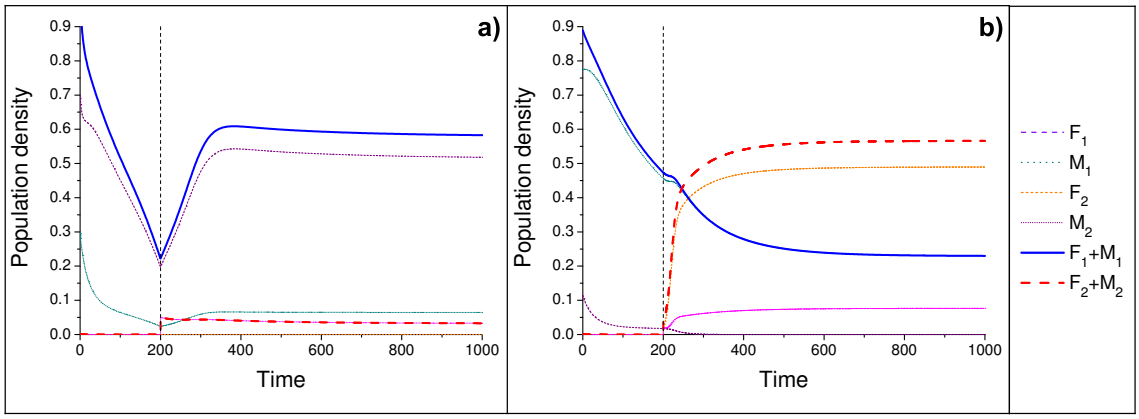

Figure 2.7: Population-density time series. (a) saving the resident allele and mortality trait; (b) saving the resident mortality trait only. The vertical dotted line indicates the time when the invader was added to the system, at 0.05 density for (a) males only, (b) males and females. Parameters: (a) $\theta_1 = 0.8$, $\theta_2 = 0.5$ (not used), $\mu_1 = 0.05$, $\mu_2 = 0.02$, $\mu_f = 0.03$ (here, F_2=0 in the final equilibrium); (b) $\theta_1 = 0.2$, $\theta_2 = 0.8$, $\mu_1 = 0.01$, $\mu_2 = 0.05$, $\mu_f = 0.05$ (here, F_1=0 in the final equilibrium).

allowed invader mortality:

$$\mu_2 < \frac{\theta_1 - \theta_1^2 - 4\mu_f + 4\theta_1\mu_f}{8\theta_1}. \tag{2.27}$$

The formula above is derived from Eq. (2.22).

Figure 2.7(a) shows an example of successful rescue of a declining resident population. After the introduction of the new mortality trait, the population shows positive growth immediately, and soon reaches a new, stable positive equilibrium.

It is important to note that we do not have conditions on the resident male mortality for a type-II fixed point. Indeed, if we have sufficient invader males with conditions listed above, then resident males may have *any large* mortality rates; invader males alone can sustain the reproduction of females, ensuring a stable positive equilibrium.

We also need to emphasize the importance of $\mu_2 < \mu_1/2$ condition in order to avert extinction. If the resident mortality is too high, it is not sufficient for the invader to have only slightly lower mortality, it must have less than half that of the

resident (i.e., has to live and reproduce at least twice as long). This is the result of our reproduction rules in the model; male offsprings of mating between F_1 females and M_2 males only attain the invader's trait (and thus replenish M_2 males) with probability $^1/_2$.

Finally, we realize that introduction of new cultural/mortality traits can have significant effects on ecological consequences, as well as sex-ratio evolution. We have seen that over a long time scale, repeated allelic substitutions can lead to ever-increasing sex ratios, and thus eventual extinction, which is not prevented by our model. However, when the sex ratio becomes too large and unsustainable, we can equivalently say that the male mortality is too high for that particular sex ratio. Extinction can be averted, and thus the evolution of ever-increasing sex ratios can be delayed, by introduction of a new cultural trait with significantly lower mortality.

2.4.2 Saving Only the Cultural Trait

If our goal is limited to ensuring the presence of the resident male mortality trait in a positive equilibrium (but not the female sex-ratio allele), then we can aim for a type-II fixed point where the invader females exclude the resident females ($F_1 = 0$). Naturally, saving both traits is still a possible solution (see the previous section), but the $F_1 = 0$ case extends the range of parameter space where the population (or at least, its cultural/mortality trait) can be saved.

Due to symmetry, we can exchange the roles of resident and invader, for the purpose of establishing the necessary conditions of our desired outcome. Resident males coexisting with invader males and females is identical to having a resident group resisting an invading female allele, but coexisting with its males [see Section 2.3.2 and Fig. 2.6(d)]. However, the restriction that the resident (here, the invader) has to be stable if left alone in the environment can be lifted; it is now possible to invade the unstable resident with an unstable invader, which together can still coexist and survive.

From Section 2.3.2, the necessary conditions for a type-II fixed point with $F_1 = 0$ are: $\theta_2 > \theta_1$, $\sqrt{\mu_\mathrm{f}} + \sqrt{2\mu_1} < 1/2$ and $\mu_2/\mu_1 > 2$. In addition, the sufficient condition for the range of θ_2 has to hold: $\tilde{\theta}_{\mathrm{c}1}(\mu_\mathrm{f}, \mu_1) < \theta_2 < \tilde{\theta}_{\mathrm{c}2}(\mu_\mathrm{f}, \mu_1)$ [see Eq. (2.22)].

Notice, that these conditions do not refer to the stability of the resident (since resident stability depends on θ_1 being in the correct range, not θ_2), again emphasizing that for a stable coexistence of mortality traits the resident or the invader could be unstable if left alone in the environment.

Figure 2.7(b) shows a scenario where only the resident male mortality trait is saved in coexistence with the invader. As the invader is introduced, their density quickly rises, while the density of the resident females drop to zero even faster than before introduction. The system quickly reaches equilibrium, where resident males have a positive density. Also note that both resident and invader groups have parameters that would make them unstable alone.

2.4.3 Density Threshold of Intervention

There is a density threshold for the resident population, above which it is not too late to introduce the invader, and save the resident cultural/mortality trait, with or without the females carrying the sex-ratio allele. Accordingly, there is a limited time, before the declining population density drops below this threshold. Essentially, the invader population has to be introduced (with the necessary initial density levels) such that the dynamics is still within the attraction basin of the type-II fixed point that we want to reach. However, finding the edge of this attraction basin (i.e., the separatrix), which is essentially the Allee threshold for the type-II coexistence, is rather difficult, as it depends on the initial densities of both resident and invader groups (at the time of introducing the invader), thus it resides in a four-dimensional phase space. Nevertheless, it is possible, if necessary, to find this threshold numerically, with arbitrary precision. Note, that in our experiments [Fig. 2.7], given the right parameters, we used simple trial-and-error to find the time when introduction of the invader resulted in the desired coexistence, and not eventual extinction.

2.5 Spatial Invasion

2.5.1 Invading to Open Habitat: The Critical Radius

Ecological invasion often has a distinctly spatial character [1, 2]. Therefore, we extend our model beyond the assumption of homogeneous mixing, and introduce

spatial detail by analyzing the model's reaction-diffusion extension.

Equations (2.1) assume that each individual encounters any potential mate at the same average rate, i.e., homogeneous mixing, which is a strong simplification for most organisms. Mating encounters ordinarily occur more frequently between nearby, than between distant pairs. Spatially structured mating can be especially important during ecological invasion, because introduced invaders often cluster locally [1–3, 147, 148]. To consider these effects, we assume a two-dimensional habitat with local mating and random mobility of individuals. We generalize Eqs. (2.1) as a reaction-diffusion system [15] by replacing the homogeneous global densities with the corresponding local densities ($F_i(\mathbf{x})$, $M_i(\mathbf{x})$) at location $\mathbf{x}$. To model dispersal, we add a diffusion term ($D_{\text{diff}}\nabla^2 F_i(\mathbf{x})$ and $D_{\text{diff}}\nabla^2 M_i(\mathbf{x})$ for group i) to the respective equation of motion. Therefore, our spatial model is described by the following equations:

$$\begin{aligned}
\partial_t F_1 &= D_{\text{diff}}\nabla^2 F_1 + \theta_1 (1-N) F_1 (M_1 + M_2) - \mu_{\text{f}} F_1 \\
\partial_t M_1 &= D_{\text{diff}}\nabla^2 M_1 - \mu_1 M_1 + \\
&\quad + (1-N)\left[(1-\theta_1) F_1 \left(M_1 + \frac{M_2}{2}\right) + (1-\theta_2) F_2 \left(\frac{M_1}{2}\right)\right] \\
\partial_t F_2 &= D_{\text{diff}}\nabla^2 F_2 + \theta_2 (1-N) F_2 (M_1 + M_2) - \mu_{\text{f}} F_2 \\
\partial_t M_2 &= D_{\text{diff}}\nabla^2 M_2 - \mu_2 M_2 + \\
&\quad + (1-N)\left[(1-\theta_2) F_2 \left(\frac{M_1}{2} + M_2\right) + (1-\theta_1) F_1 \left(\frac{M_2}{2}\right)\right]. \quad (2.28)
\end{aligned}$$

To integrate the spatial model numerically, we discretize the partial differential equations (PDEs) to ODE equations (based on the Method of Lines technique [149]) on a rectangular grid of size 400×400 (representing an area of 100×100 units), using Neumann boundary conditions. We integrate the resulting ODEs using an explicit Euler time stepping, for which we use a sufficiently small time step ($\Delta t = 0.01$). These parameters allow us to use diffusion coefficients as large as 2.5 without producing finite-size effects, or instability. For the spatial model, we define global equilibria with the stopping condition $\epsilon = 10^{-6}$.

Note, that our spatial, but deterministic reaction-diffusion equations still maintain an essential (local) "mean-field character" (in the statistical physics sense

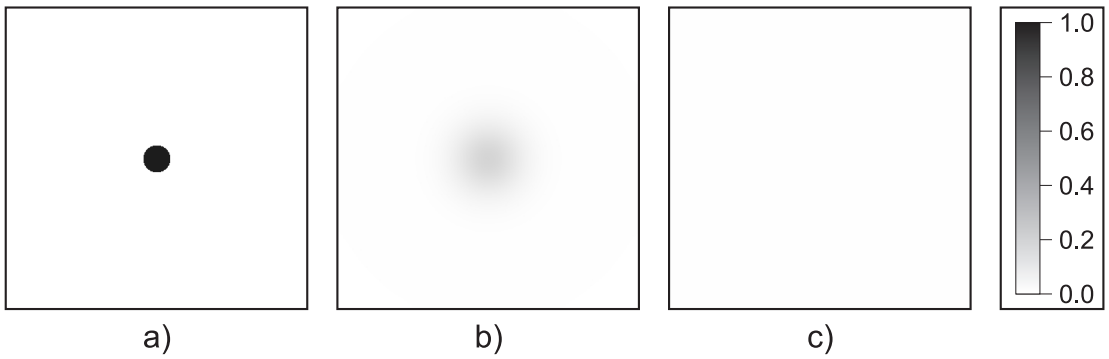

Figure 2.8: Unsuccessful spatial persistence of a single group attempting to spread in an open (empty) habitat. The initial radius is less than the critical radius ($R_0 = 4.5 < R_c = 5.1$). Simulation time: (a) $t = 100$, (b) $t = 150$, (c) $t = 300$. Parameters: $\theta = 0.5$, $\mu_f = 0.02$, $\mu_m = 0.03$, $D_{\text{diff}} = 1.0$.

and terminology) in that all correlation functions are still factorized into products of concentrations [150, 151]. A stochastic, spatial individual-based model, or its Langevin-type stochastic reaction-diffusion analogue (not addressed in this work) may, in principle, lead to different behaviors [152, 153]. For example, the region of persistence in the case of a single-group two-sex population becomes significantly narrower in a stochastic lattice-based model [130].

Successful invasion in spatial environments ordinarily requires that an initial invader cluster have some minimal size for further growth [2, 3, 147, 154]. This criterion may be due to an Allee effect [154] or inherent geometrical constraints on cluster expansion [3]. For systems exhibiting the Allee effect under homogeneous mixing, one can specify this minimal cluster size as the critical radius (R_c) required for spatial invasion. Assuming radially symmetric growth, one expects $R_c \sim \sqrt{D_{\text{diff}}}$, where D_{diff} is the diffusion coefficient [154]. For simplicity, we take D_{diff} as a constant across all individuals. The first goal of our spatial analysis was to confirm this scaling relationship for the critical radius when a single group is introduced in an open (unoccupied) habitat.

For spatial invasion in an open habitat, individuals diffusing away from the perimeter of the invader cluster encounter mate densities too low for population increase, given the Allee effect (i.e., extinction is stable). A small invader cluster

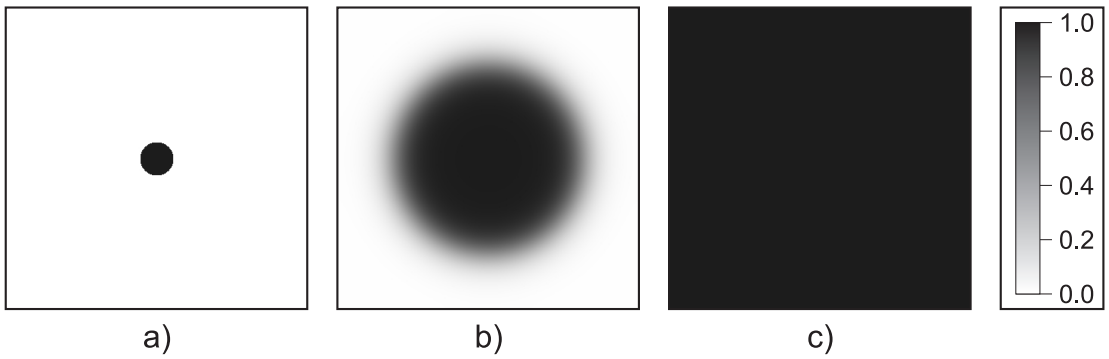

Figure 2.9: Successful spatial persistence of a single group spreading in an open (empty) habitat. The initial radius is greater than the critical radius ($R_0 = 5.5 > R_c = 5.1$). Simulation time: (a) $t = 100$, (b) $t = 300$, (c) $t = 600$. Parameters: $\theta = 0.5$, $\mu_f = 0.02$, $\mu_m = 0.03$, $D_{\text{diff}} = 1.0$.

can shrink as a result. A cluster size exceeding the critical radius generates interior densities sufficient to drive cluster expansion. The critical radius depends on both density inside the cluster and the diffusion coefficient. Therefore, calculating a critical radius demands specifying initial densities within the circular cluster. We noted that as we chose densities closer to, but exceeding, the Allee threshold of the homogenous-mixing case, the critical radius increased. Therefore, a reasonable (deterministic) choice is the stationary density of the non-spatial model, which we can calculate, given the female ratio and sex-specific mortality rates [see Eq. (2.8)].

We found the critical radius using bisection method with the initial interval of $R \in [1, 20]$. At each step, a simulation runs with a particular initial radius, until all densities at all grid points come to a stationary state (where all time derivatives are less than $\epsilon = 10^{-6}$). In this final state either all grid points have the positive, stationary densities of the non-spatial model, or all have zero densities. The resolution of the grid (4 cells/unit distance) and the discretization of a circle on a rectangular grid allow us to measure non-integer radii. Time evolution of a shrinking ($R<R_c$) and a successfully growing, invading population ($R>R_c$) are illustrated in Figs. 2.8 and 2.9, respectively.

We obtained the critical radius for various diffusion coefficients, at certain fixed set of parameters [Fig. 2.10]. As anticipated [154], the results confirm that the critical radius is proportional to the square root of the diffusion coefficient.

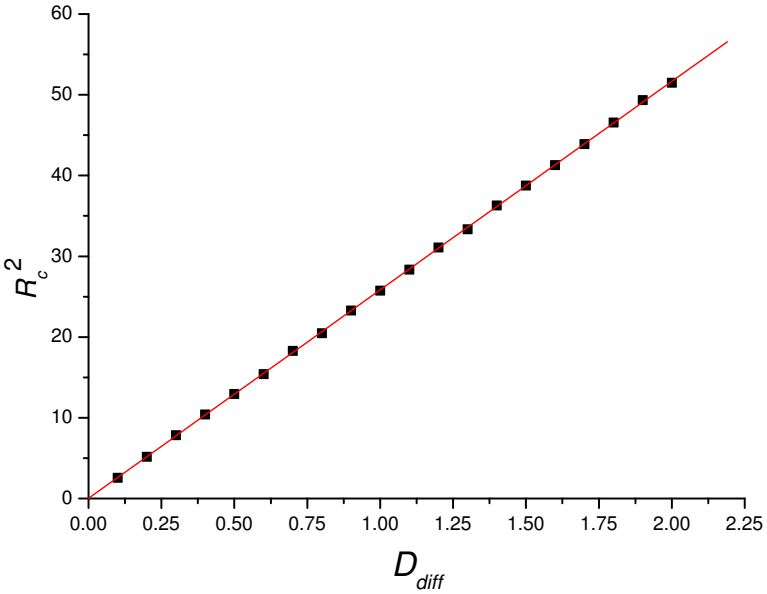

Figure 2.10: Scaling of the critical radius in single group dynamics. The line shown is a fitting using minimum least squares, with a correlation value of 0.99997**. Parameters:** $\theta = 0.5$, $\mu_f = 0.02$, $\mu_m = 0.03$.

2.5.2 Spatial Invasion of a Resident Population

Our goal here is to ascertain if there is a critical radius for successful invasion when invaders can mate with residents in an occupied habitat. We initiate simulations differently than in the open-habitat case. Here, every grid point is initialized to the stationary density of the resident group. Then, we introduce the invader within a circle of a given radius, at a small density. The simulation runs until all grid points come to a stationary state (where all time derivatives are less than 10^{-6}).

We find that no matter how small we set the invader density and cluster radius, the result is always identical to the homogeneously mixed case. That is, the allele with the higher female ratio persists, and the ecological impact of the winning female ratio depends on the male mortality rates. Male cultural traits may coexist (type-II fixed point), or both females and males of the lower female-ratio group go extinct (type-I fixed point). Figure 2.11 shows a scenario where the invader has the same parameters as the open-habitat invasion in Fig. 2.8. However, here the result

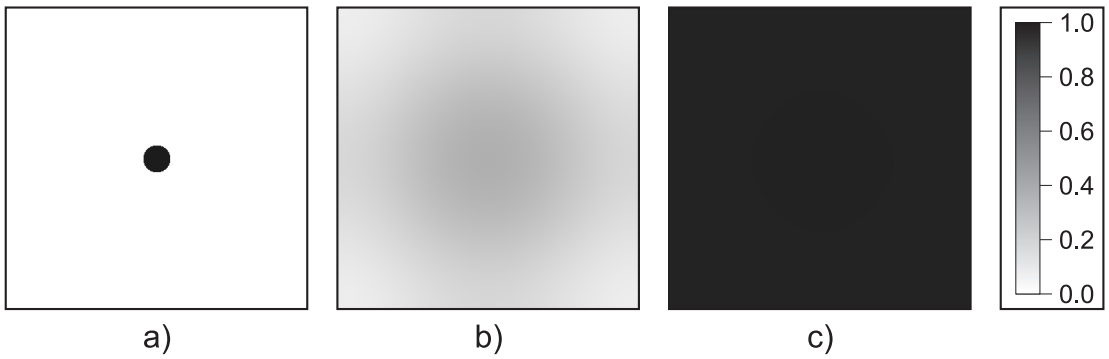

Figure 2.11: Spatial invasion of a stable resident population. The initial radius is less than the critical radius ($R_0 = 4.5 < R = 5.1$) for invasion into an open habitat. For clarity, only the invaders' density is shown. Simulation time: (a) $t = 100$, (b) $t = 450$, (c) $t = 820$. Parameters: $\theta_1 = 0.3$, $\theta_2 = 0.5$, $\mu_f = 0.02$, $\mu_1 = \mu_2 = 0.03$, $D_{\text{diff}} = 1.0$.

is different, because of the presence of the resident population. The invader can (effectively) exploit the resident population as mates, enabling the invader to spread successfully and eventually exclude the resident.

We understand the absence of a critical radius in the resident-occupied environment by considering cases where even an infinitesimal invader density can completely exclude the resident in the homogenously mixed case. In the worst-case scenario (for the invading allele and culture), we introduce only a small density of invaders at only a single grid point, with a high diffusion rate. Then, diffusion spreads the invader to all grid points, making its density extremely small, but greater than zero. However, this is enough for successful invasion at every grid point, independently of other locations, as we noted in the model with global mixing. If we introduce a greater density of invaders, with slower diffusion, then the invader can quickly overtake the local area before spreading out as a diffusive front. The eventual result will be the same. Hence we conclude that there is no critical radius for invasion with diffusion, if a resident population already occupies the habitat.

2.6 Discussion and Conclusions

Our population dynamics differs from models for gene-culture coevolution, where different alleles and cultural traits directly affect each other's evolution [53]. Our model's cultural trait directly influences the resident's population density and the invader's growth rate when rare; female ratios and male mortalities interactively drive the invader's dynamics. We do not assume functional dependence between the genetic and cultural traits. Rather, we evaluate consequences of the feasible range of male-mortality rate combinations for the entire range of female ratio combinations (resident and invader).

Most models of sex ratio evolution, whether analyzed as evolutionarily stable sex allocation [128, 131] or developed with population-genetic detail [43], assume that a parent is related symmetrically to female and male offspring. Hamilton [48] noted that sex-linked inheritance of a gene for sex ratio breaks this symmetry, and extraordinary sex ratios can evolve as a consequence. Frank [155] summarizes effects of asymmetric relatedness to offspring by sex, and cites several studies where this asymmetry is correlated with strongly biased investment in the sexes [156]. Our results specify how the degree of bias can interact with a between-sex mortality difference to influence the population dynamic consequences of sex ratio evolution.

Tainaka *et al.* [130] and Nitta *et al.* [157] developed spatially detailed models to study how sex ratio might affect population persistence. For successful mating, their model requires that at least one fertile individual of each sex occupy a site neighboring an empty site (where the offspring is placed). At the scale of individuals, the dynamics is the simplest generalization of the contact process [158–160] that can capture both two-sex reproduction and preemptive competition [1, 2, 148]. Given female and male mortality rates, they find the sex ratio maximizing population density, and note that sex ratios differing too much from this singular value lead to population extinction [130]. Compared to the mean-field result, the extinction effect due to biased sex ratio sharpens in simulation of the stochastic, lattice-based model; the range of sex ratios producing population persistence becomes quite narrow. Since mating pairs form locally, biasing the sex ratio rapidly diminishes the chance that an open site will be neighbored by one individual of each sex. So, demographic

stochasticity may lead to extinction once sex ratio is biased, and genetic drift may permit biased sex ratios to evolve even when bias is selectively disfavored [161].

Our study generalizes the model of Tainaka *et al.* [130] by including between-sex differences in mortality and detailing outcomes of competition between different female ratios. Our model limits expression of sex ratio to the heterogametic sex, so that stronger bias in sex allocation has a competitive advantage. Our results elucidate the ecological effects of interaction among the degree of sex ratio bias and sex-specific mortality for competitive/cultural invasion and demographic stability. In the simplest case, an introduced female allocation and associated cultural trait, male mortality, invades and excludes the resident allele and culture. Complete exclusion requires only that the invaders have the higher female allocation and that their male mortality rate is lower than twice that of the resident males. If the invader's male mortality rate is large enough to exceed this limit, but the difference in female allocation remains, the resident culture (but not the resident allele) survives and coexists with the invader's culture.

Our analysis also identified an interesting invasion-to-extinction scenario. A group with the greater female allocation and greater male mortality (compared to the demographically stable resident) cannot invade an empty environment. Yet it invades and excludes the resident, and then goes extinct, because of its high female ratio. Since the invaders can mate with the residents, they effectively exploit the resident group in the early phase of invasion and, when sufficiently numerous, drive the resident extinct. Thereafter, a marriage squeeze leaves the invader declining to extinction. This type of outcome, where sex ratio and an Allee effect can push a population to extinction, may have applications in the management of pest populations [134]. Evolutionarily, the demographic consequences of sex ratio bias may favor suppression of sex-ratio distorters [48], and may promote (or be tolerated by) clonal reproduction [162].

CHAPTER 3
ECOLOGICAL RESTORATION AT MINIMUM COST

Ecological restoration aims to replenish an ecosystem's biodiversity, often responding to human-induced losses of indigenous species [30, 31]. When ecosystem managers reintroduce a species to its former habitat, the restoration effort's success is ordinarily defined by combined ecological and economic criteria [32]. Similarly, optimizing biological control programs may integrate impact on the target species with costs of deploying the control agent [33].

We envision restoration of a single species whose population dynamics depends on the density of each sex [163]. Before evaluating costs, we must identify those spatial distributions of the initial population that assure successful restoration. Suppose that we initiate restoration with a single spatial cluster, within which individuals are distributed at a uniform density. Then we must find the "critical cluster" size, the minimal area the species must occupy to sustain positive population growth. Analysis of the critical-cluster criterion has advanced understanding of spatial systems in both physics [59–61] and ecology [1–3, 62]. However, if ecosystem managers can vary initial densities according to location, non-uniform spatial distributions might reduce restoration cost. Given multiple initial population distributions assuring sustained population increase, the most preferred option should *minimize* cost. Our study investigates how the minimum cost of successful restoration depends directly on spatial pattern, and how the optimal pattern depends on sex ratio, and on sex-specific mortality rates.

In this context, we model a species' restoration as a spatially detailed, one-dimensional, two-sex reaction-diffusion system; specifically, we utilize the model that we have thoroughly analyzed in the previous chapter, as it provides all the necessary details we need, and we can take advantage of existing results. Here, we optimize the initial densities and spatial distributions of the sexes to minimize

Portions of this chapter previously appeared as: F. Molnár Jr., C. Caragine, T. Caraco, and G. Korniss, "Restoration Ecology: Two-Sex Dynamics and Cost Minimization," PLoS One **8,** e77332 (2013).

the cost of restoring the species to its positive, stable (homogeneous) equilibrium density throughout a habitat. Generally, the prototype of such models is the Fisher-Kolmogorov equation [49, 164]

$$\partial_t u = D\nabla^2 u + \alpha u(1-u) \ , \tag{3.1}$$

which describes the dynamics of single-sex populations with logistic growth and diffusive dispersal. Our model extends the basic reaction-diffusion framework by including sex-structured dynamics [130, 165, 166], where an Allee effect [154, 167] generates an unstable fixed point between extinction and the habitat's carrying capacity.

In many natural and managed populations, per-capitum growth is reduced as density becomes small (Allee effects [132–134]), induced by various behavioral, ecological and genetic mechanisms [168]. For example, low population density may diminish individual reproduction by reducing mate encounters, making prey capture more difficult, or by leaving individuals more susceptible to their own predators [169, 170]. In addition, these effects will be amplified if dispersal into unoccupied habitat reduces local population density and generates negative population growh, as we have shown in Section 2.5, thwarting restoration [154].

Generically, the cost of restoration can be defined as:

$$C = c^* \int_\Omega u(r, t=0)dr, \tag{3.2}$$

where Ω represents the extent of the habitat, and $u(r,t)$ is the population density at location r. Without loss of generality, we can consider a constant *per-capitum* cost ($c^* = 1$). Mathematically, the restoration cost, which we seek to minimize, is a functional of the initial population's spatial distribution. The constraint requiring population persistence cannot be expressed analytically, since that would require solving the model's partial differential equations exactly. Therefore, we utilize numerical techniques to obtain the desired spatial distribution and the corresponding minimum cost.

In the previous chapter (in Section 2.5) we briefly analyzed the two-sex dynamics with diffusive dispersal, and found that the critical radius (hence, critical cluster size) for an initial population's successful invasion of a two-dimensional habitat is proportional to the square root of the diffusion coefficient, which was the same for both sexes. This is the starting point here; we extend our study by considering the dependence of the critical radius on all (non-symmetric) parameters. While our reaction-diffusion model includes numerous simplifications for detailed application to particular species, nevertheless it exhibits the essential ecological characteristics of more complex two-sex models. Hence, implications of our results for restoration will likely hold across a wide range of specific models.

We organize the rest of this chapter as follows. First, we recall the main features of our sex-structured population dynamics from Section 2.5, and outline the analytic and numerical methods we employ in this work. Then, we conduct a systematic study of the restoration cost in multiple stages, considering different initial spatial population distributions in each stage. Starting with a simplified, single-sex version of our model, we derive an unstable, aperiodic stationary solution for the PDE, and use it as initial distribution, resulting in a single-sex cost. We then continue with the sex-structured model and analyze the way restoration cost depends on model parameters, given a simple, homogeneous initial distribution of individuals inside a cluster. We refer to this initial setup as "rectangular", for its shape on a density vs. location plot. We analyze this setup first with the constraint of equal cluster sizes for both sexes; later we relax this constraint. In the third stage we allow any possible shape of initial population distribution and study how the cost can be reduced as a result. Finally, we discuss the minimum costs found in each stage, and conclude which approach yields the most economical restoration.

3.1 Model and Methods

Recall that our model, as introduced in Section 2.1.2 (for a single group) has three key parameters: the sex ratio θ at birth, and sex-specific mortality rates μ_m and μ_f, for males and females, respectively. We assume that females and males disperse independently by homogeneous diffusion (similarly to Section 2.5), and that

males encounter females as a mass-action process, equivalent to random mating [135]. The fraction of matings leading to successful reproduction is proportional to the unoccupied fraction of the environment, $(1 - m - f)$ [130, 165, 166]. That is, the population grows in a self-regulated manner. Hence, we have:

$$\begin{aligned} \partial_t f &= D_f \nabla^2 f + \theta\,(1 - m - f)\, fm - \mu_{\mathrm{f}} f \\ \partial_t m &= D_m \nabla^2 m + (1 - \theta)\,(1 - m - f)\, fm - \mu_{\mathrm{m}} m\,, \end{aligned} \tag{3.3}$$

where $f(x,t)$ and $m(x,t)$ denote the local densities of females and males, respectively. Diffusion rates are described by coefficient D_f for females and D_m for males. In principle, by extending the above deterministic reaction-diffusion equations with appropriate noise terms [171–173], the resulting stochastic partial differential (or Langevin-type) equations [174] can capture the relevant macroscopic features of the underlying spatial, stochastic individual-based model [153, 175, 176]. However, the rigorous derivation of such stochastic partial differential equations can be rather challenging [177–179].

We can transform Eq. (3.3) to interpret it as a single-sex model by making the equations symmetric. To do so, we must let $\theta = 0.5$, restrict $\mu_f = \mu_m =: \mu$, $D_f = D_m =: D$ and use the same initial density distributions for both males and females. In this way, the two densities behave identically over time: $f(x,t) = m(x,t) = u(x,t)$, described by the following equation:

$$\partial_t u = D\nabla^2 u + \frac{1}{2}u^2(1 - 2u) - \mu u. \tag{3.4}$$

This transformation bridges the single-sex and two-sex models; using the constants in Eq. (3.4), we can directly compare results between models without rescaling parameters. Note that the (cubic) local dynamics also retains an Allee effect.

We discriminate successful restoration from extinction by numerically integrating the model until it has converged to a global equilibrium. We use second-order finite difference discretization for spatial derivatives and explicit Euler method for integration over time [149], with a sufficiently small time-step. Integration stops when all time-derivatives at all spatial coordinates are less than 10^{-6}. Finally, the

cost is obtained by multiplying the initial density and the initial cluster size, and summed for males and females.

The prerequisite for analyzing the cost of restoration is to ensure the local stability of successful restoration, i.e., a positive stable fixed point of local dynamics. The necessary stability condition for the two-sex model [Eq. (3.3)] is:

$$1 - 4\left(\frac{\mu_f}{\theta} + \frac{\mu_m}{1-\theta}\right) > 0, \tag{3.5}$$

which we have derived in Section 2.2.1. Similarly, we can find the necessary stability condition for the single-sex model [Eq. (3.4)]:

$$\mu < \frac{1}{16}. \tag{3.6}$$

These conditions provide us guidelines for selecting proper model parameters when evaluating costs in the following sections.

3.2 Unstable Stationary Solutions

Our first approach to cost minimization selects suitable unstable stationary solutions of the PDE model as initial population distributions [180]. They are good candidates for initial spatial distributions, because they are "critical" solutions, in the sense that given a small perturbation, they transform to a stable, spatially homogeneous solution (positive perturbations result in positive homogeneous population densities and successful restoration, negative perturbations result in zero densities and extinction). The cost associated with each stationary solution is found by numerical integration of the density profile (i.e., the area under the curve).

Partial differential equations, such as Eqs. (3.3) and (3.4) have infinitely many unstable stationary solutions [180]. We cannot find these solutions directly, but we can derive analytical formulas for the relationship between the density and its spatial derivative. For the single-sex model [Eq. (3.4)] we have the following definitions:

$$\frac{\partial u}{\partial x} = v \tag{3.7}$$

$$\frac{\partial v}{\partial x} = \frac{-1}{D}\left(\frac{1}{2}u^2(1-2u) - \mu u\right). \tag{3.8}$$

Using v, we change variables to write a first-order differential equation:

$$v(u)\frac{\partial v}{\partial u} = \frac{-1}{D}\left(\frac{1}{2}u^2(1-2u) - \mu u\right). \tag{3.9}$$

By separating variables, we obtain the following analytical solution:

$$v(u) = \pm\sqrt{\frac{3u^4 - 2u^3 + 6\mu u^2 + 12DE}{6D}}, \tag{3.10}$$

where E is a free parameter. The phase diagram for this equation is depicted in Fig. 3.1(a), and its contents are summarized as follows.

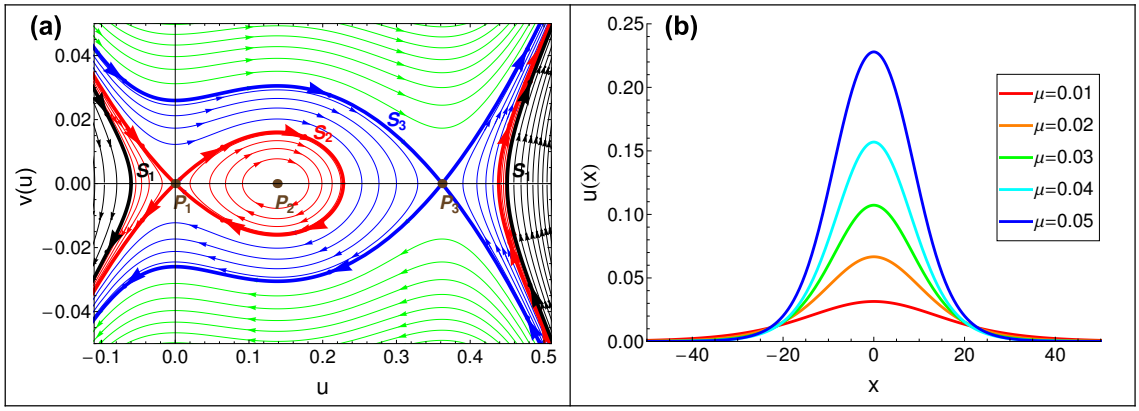

Figure 3.1: Stationary solutions in the single-sex model. (a) Phase plot of stationary solutions, described by Eq. (3.10). $D = 1.0$, $\mu = 0.05$. **The dots indicate fixed points, the thick lines indicate separatrices. Different curves correspond to different** E **parameters, however, values of** E **were not chosen uniformly, for aesthetic reasons. (b) The stationary solutions found by integrating along the** S_2 **separatrix, for multiple mortality rate parameters;** $D = 1.0$**;** x **is distance form the habitat's center.**

The fixed points P_i ($i \in \{1, 2, 3\}$) correspond to *homogeneous* stationary solutions. Naturally, these are also fixed points of the original equations [Eq. (3.4)], but here they are only special cases of stationary solutions that do not vary spatially. Hence $v(u) = 0$. P_1 and P_3 are saddle points (stable equilibrium nodes of the local dynamics) corresponding to extinction ($u = 0$) and persistence ($u > 0$), respectively. P_2 is a center (unstable fixed point of the local dynamics) corresponding to the

unstable equilibrium due to the local dynamics' strong Allee effect.

Curves in Fig. 3.1(a) correspond to *inhomogeneous* stationary solutions that can be classified by the value of the free parameter E in Eq. (3.10). Separatrices S_1, S_2, and S_3 correspond to the following values (with the same subscripts, see Appendix A.1 for details):

$$E_1 = \frac{1}{48D}(u_2^* - 2\mu)(u_2^* - 6\mu) \tag{3.11}$$

$$E_2 = 0 \tag{3.12}$$

$$E_3 = \frac{1}{24D}(u_3^*)^2(u_3^* - 6\mu), \tag{3.13}$$

where u_1^*, u_2^*, and u_3^* are the equilibrium densities of local dynamics, i.e., the values of u corresponding to P_1, P_2, and P_3, respectively:

$$u_1^* = 0, \tag{3.14}$$

$$u_{2,3}^* = \frac{1}{4} \mp \sqrt{\frac{1}{16} - \mu}. \tag{3.15}$$

The closed elliptical curves around P_2 represent periodic stationary solutions, and they are the only ones of interest, because all other curves extend to infinitely large negative or positive densities, neither having biological meaning.

Spatially periodic stationary solutions may offer candidate initial population distributions. In principle, if minimum densities within each period were close to zero, then we could select a segment of the solution, one period in length between two density minima, and apply it as an initial spatial distribution. However, in our case, as the minimum value of u goes to zero, the period of the solutions goes to infinity, and the curves converge to the S_2 separatrix, which corresponds to an aperiodic stationary solution. The exact aperiodic shape of $u(x)$ can be found by numerical integration along S_2, depicted in Fig. 3.1(b). Since $u(x)$ converges to zero rapidly, we can use it as an initial population distribution by taking its central segment above an arbitrary small (biologically meaningful) density threshold. Because of the fast convergence to zero, the length of the segment will be finite. In our study, we

use 10^{-5} for the density threshold. Then, for every D and μ combination, we have an exact shape, and an exact cost value defined as twice the area under the curve (counting both males and females), denoted by C_{stat}. We will compare these costs with those found by the other two methods described in the following subsections.

As an interesting observation, we note that contrary to intuition, the period length of the stationary solution does not converge to zero as the solution curves approach P_2, see further details in Appendix A.2.

For the sex-structured model [Eq. 3.3] we cannot derive unstable stationary solutions analytically. Instead, we study the scaling of restoration cost through a numerical analysis of the critical cluster size and density, which are directly related to the minimum cost of successful restoration.

3.3 Critical Cluster Size and Minimum Cost

Criticality of an initial cluster's size occurs when density reduction due to dispersal exactly balances the net effect of local natality and mortality. In two dimensions the expanding population front's speed is reduced in proportion to the curvature of the cluster; therefore, it affects the critical cluster size. To avoid confusing effects of curvature with other parameters' impact on the critical cluster size and, hence, the minimum cost, we restrict our study to one dimension.

3.3.1 Rectangular Setup

We begin our analysis of the critical cluster size by assuming a *rectangular* initial setup, i.e., homogeneous initial spatial distribution inside the cluster with a specific population density. This is the most obvious choice, for its mathematical simplicity, and its plausible applicability (e.g., an animal population surrounded by a fence before release can be modeled with a uniform spatial distribution). In a rectangular setup there exists a critical cluster size, the smallest spatial extent such that the given density achieves sustained positive growth. Symmetrically, for a given cluster size, there is a critical initial density, the lowest density assuring sustained population growth. In both cases, the critical limit also corresponds to the minimum cost, since the cost is proportional to both cluster size and density.

The exact, parameter-dependent values of these critical limits cannot be derived analytically. Instead, we use bisection method on a wide initial range of possible values, accomplished by testing each value for successful restoration vs. extinction.

We have four parameters describing the rectangular setup: f_0, m_0, l_f, l_m, which represent female density, male density, length of space occupied by females, and length of space occupied by males, respectively, at the initiation of restoration. Note, that both female and male clusters are centered symmetrically. We shall refer to l_f and l_m as the cluster sizes of the initial population. The cost of the initial state is defined simply as:

$$C_{\text{rect}} = f_0 \times l_f + m_0 \times l_m, \tag{3.16}$$

and it is minimized by using the critical cluster sizes l_m^* and l_m^*, for males and females, respectively. Note, that l_m^* and l_m^* are themselves dependent on initial densities f_0 and m_0, as well as model parameters D_f, D_m, μ_f, μ_m, and θ. Therefore, we systematically study dependence of critical clusters on all parameters.

As an initial step, we analyze the dependence of critical cluster size on the diffusion coefficients. We anticipate that the critical cluster size, i.e., critical length is proportional to the square root of the diffusion coefficient, as we have observed similar scaling in two-dimensions (Section 2.5). Our aim here is to show the same behavior in the one-dimensional reaction-diffusion system, and to ask whether it holds when male and female diffusion coefficients and cluster sizes differ.

Figure 3.2(a,b) shows that as long as we employ the same cluster size for males and females, the critical cluster size has the expected scaling behavior $l^* \sim \sqrt{D}$ [154] with respect to both male and female diffusion coefficients, even if one sex has a fixed diffusion coefficient. Note, that it is sufficient to study the dependence on the diffusion of one sex (here, males) while the other is fixed (here, females), because of the symmetric construction of the model. Further, Fig. 3.2(c,d) indicates that fixing the cluster size of one sex while measuring the critical cluster size of the other sex with respect to diffusion coefficients yields non-trivial scaling. However, we observe that a higher dispersal rate results in a larger critical cluster size, and, hence, a larger restoration cost.

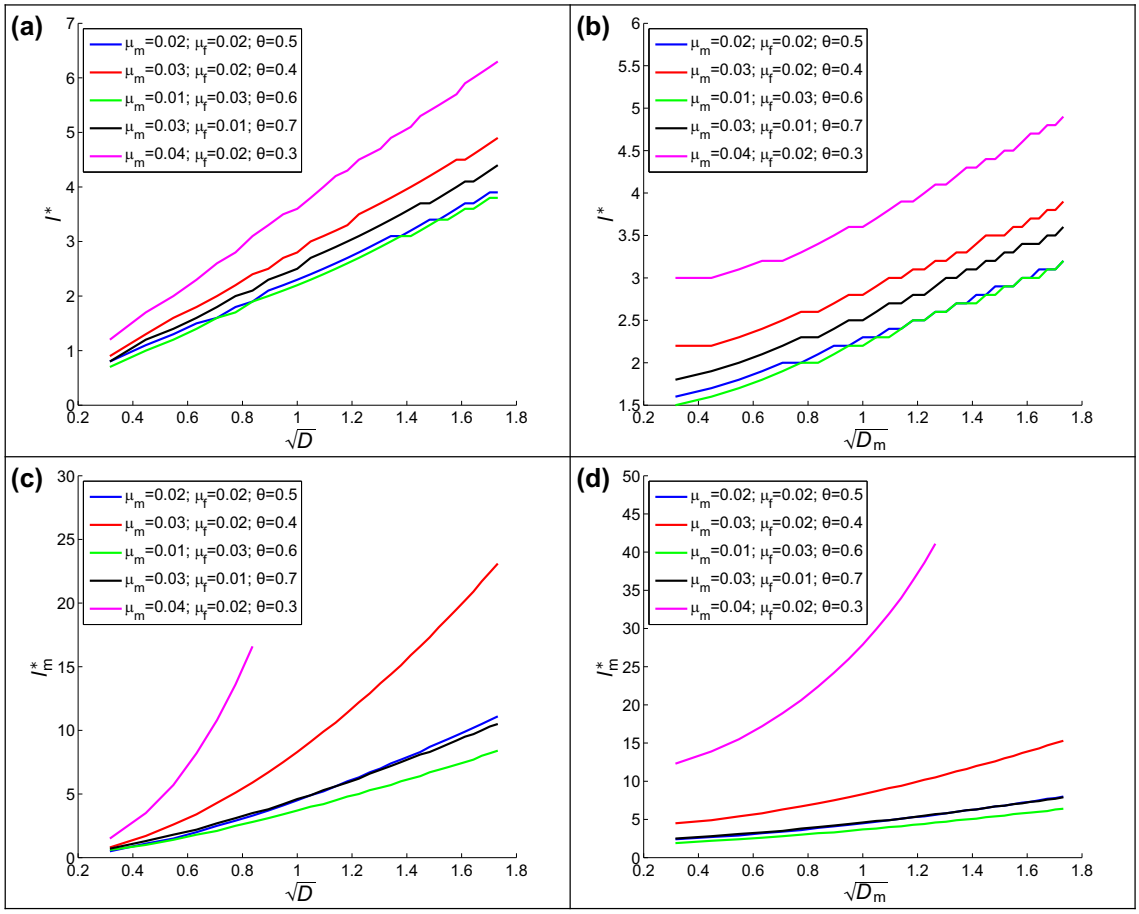

Figure 3.2: Scaling of critical cluster sizes vs. diffusion coefficients, at various parameter values. (a) $l_f^* = l_m^* = l^*$, $D_f = D_m = D$; **(b)** $l_f^* = l_m^* = l^*$, $D_f = 1.0$; **(c)** $l_f^* = 1.0$, $D_f = D_m = D$; **(d)** $l_f^* = 1.0$, $D_f = 1.0$. **In every case, initial (spatially homogeneous) densities are fixed:** $f_0 = m_0 = 0.45$.

3.3.2 Equal Cluster Sizes for Males and Females

Continuing our analysis of critical cluster sizes, we now assume that the critical cluster sizes are equal for both sexes ($l_f^* = l_m^* =: l^*$); we relax this constraint later. Even with the equal cluster-size constraint, initial densities for males and females within that cluster may, in principle, differ. In practice, a density difference could be implemented for most dioecious species. To find the best choice of initial density values (with respect to minimizing cost), we aim to relate them to model parameters, taken as given for the focal population.

Naturally, the initial densities must exceed the Allee threshold (the unstable fixed-point densities); otherwise, the population can never achieve positive growth. Figure 3.3(a) shows that as the initial density is lowered, the critical cluster size increases, and goes to infinity as we approach the Allee threshold.

Scaling of the cost, however, is non-trivial. We can always find the minimum cost at a density somewhere in the vicinity of the positive stable fixed point of the system, but always slightly below it. We understand this by considering the dynamics just after initial introduction. If the population starts from its stationary density (the stable, positive fixed point), then the local densities can only decrease, due to diffusive dispersal. However, if the initial density is lowered slightly, then the population has a chance to grow locally (in particular, at the center of the cluster) before the effects of diffusion reach it, while the eventual spread through the habitat remains the same. In essence, the cost is slightly lowered by handing over some of the spreading effort to growth dynamics. However, as Fig. 3.3(b) and Fig. 3.4 show, this advantage in cost-reduction is very small. We can conclude that using the stationary densities as initial densities results in a sufficiently low cost. Also, it provides a good

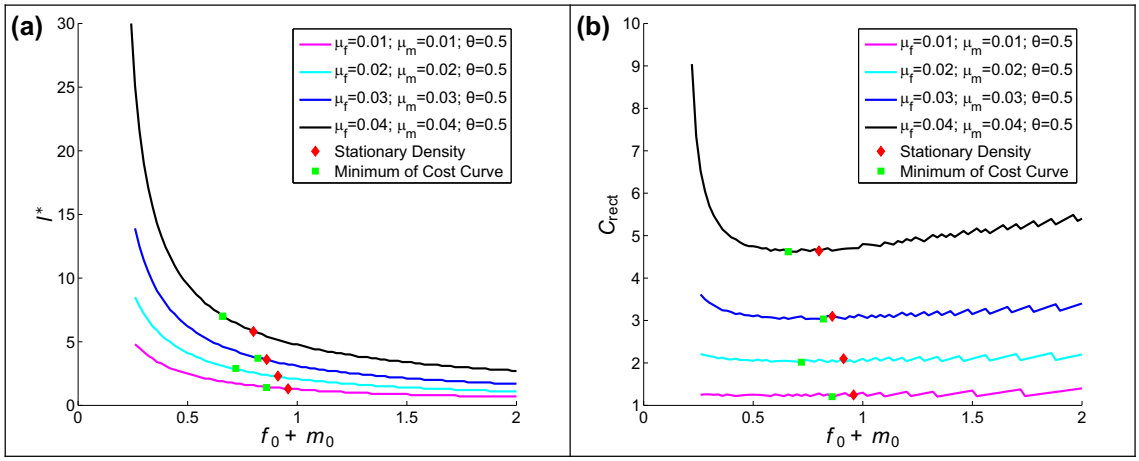

Figure 3.3: Scaling of (a) critical cluster length at introduction, and (b) cost, with respect to initial population density. The total density shown on the x axis is divided equally between males and females. The point markers on the curves show the stable stationary densities of local dynamics. For both figures, $D_f = D_m = 1.0$.

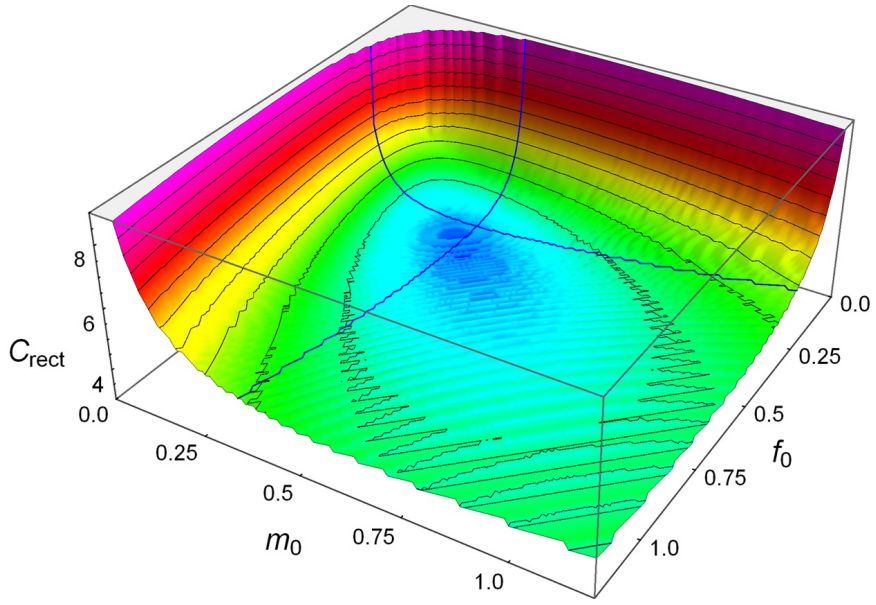

Figure 3.4: Cost landscape with respect to male and female densities. The blue cross marks the stable stationary density values; $\mu_f = 0.04$, $\mu_m = 0.03$, $\theta = 0.5$, $D_f = D_m = 1.0$.

choice of initial density based on model parameters, because the stationary densities depend on local dynamics, which in turn depend on model parameters. The formulas for the stationary densities are provided in Section 2.2.1.

We expand on the cost-minimizing property of stationary densities; we use them as initial densities throughout the rest of our study. We continue with the analysis of the critical cluster size's dependence on model parameters. In particular, we focus on the value of sex ratio at birth (θ), because both population stability and equilibrium density depend on this single parameter. For population stability and persistence, there is a range of permissible sex ratios, determined by the mortality rates (see Section 2.2.1). Within this range, we defined the optimal sex ratio θ^* as the value maximizing the equilibrium population density:

$$\theta^* = \frac{1}{1 + \sqrt{\mu_m/\mu_f}}. \tag{3.17}$$

Note, that when the mortality rates are equal, the optimal sex ratio is 0.5, which is the parametrization, by definition, in the symmetric single-sex model.

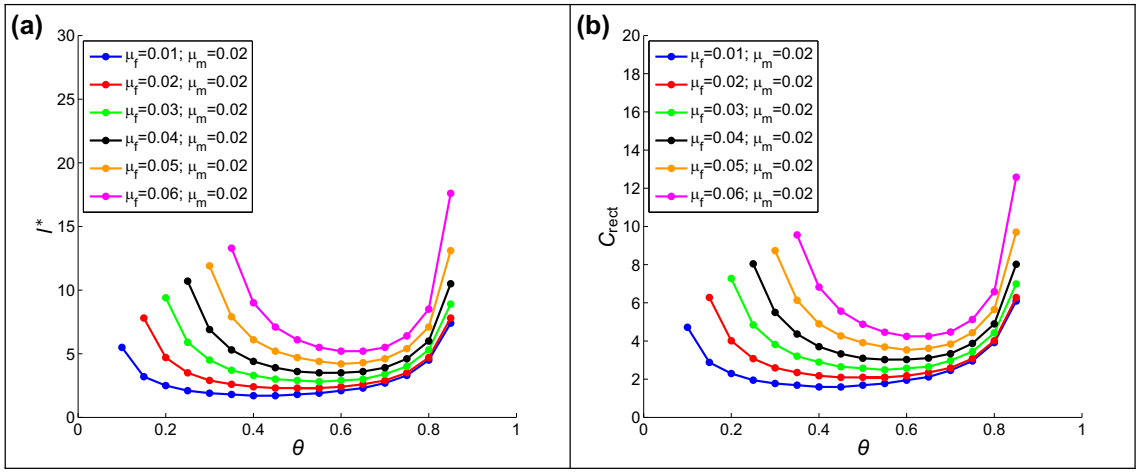

Figure 3.5: Scaling of (a) critical cluster size and (b) cost, with respect to sex ratio, at different mortality rate combinations. $D_f = D_m = 1.0$. **Rectangular initial populations were used with stationary population densities.**

We find that the smallest cluster size assuring restoration corresponds closely to the optimal sex ratio, and that any small deviation from the optimal value causes a small increase in the critical cluster size [Fig. 3.5(a)]. However, since the equilibrium population densities (serving as initial densities) decrease at suboptimal sex ratios (by definition; see Fig. 2.1), the combined effect on the cost is non-trivial. As we see on Fig. 3.5(b), restoration cost is minimized at approximately the same sex ratios minimizing the critical cluster size, indicating that the cluster size is more sensitive to biased sex ratios than to equilibrium densities. We also find that strongly biased sex ratios approaching the boundary of the stability range cause both cluster size and restoration cost to diverge.

To complete the relationship between the sex ratio minimizing restoration cost ($\widehat{\theta}$) and the sex ratio that locally maximizes total equilibrium population density (θ^*), we compare the two quantities numerically. For the latter we have an analytical expression [Eq. (3.17)]. But our cost-minimizing sex ratios have only limited precision, since for each value of θ we employed the bisection method to determine the critical cluster size, which, in turn, determines the cost. Therefore we define a computational error bound on $\widehat{\theta}$ as the range of θ values that give critical cluster sizes within the error range of the minimum point's cluster size found by binary search.

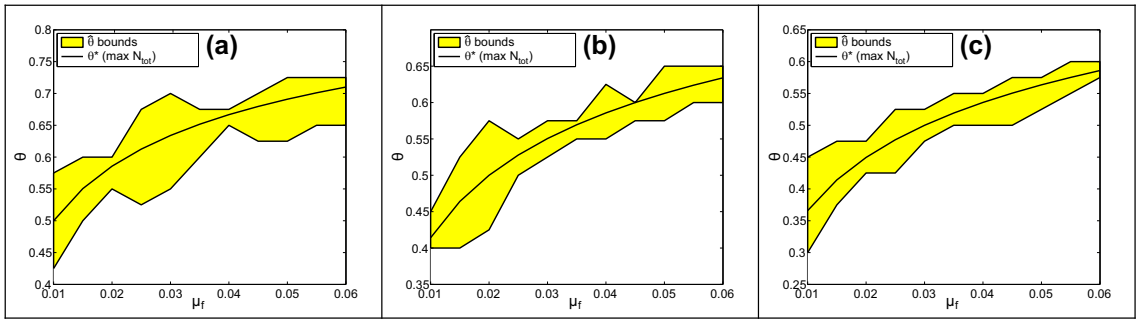

Figure 3.6: Comparison of density-maximizing and cost-minimizing sex ratios. Density-maximizing sex ratio θ^* [Eq. (3.17)] and numerical bounds of cost-minimizing sex ratios $\widehat{\theta}$ are calculated using rectangular population distributions with stationary initial densities, $D_f = D_m = 1.0$, (a) $\mu_m = 0.01$, (b) $\mu_m = 0.02$, (c) $\mu_m = 0.03$.

Figure 3.6 offers comparison of the density-maximizing and cost-minimizing sex ratios. We can conclude that the sex ratio maximizing equilibrium density is identical to the sex ratio minimizing restoration cost, up to computational error.

3.3.3 Different Cluster Sizes for Males and Females

To this point, our results reflect the assumption that individuals of each sex are introduced across the same extent of habitat. We now relax this constraint; that is, we permit $l_f^* \neq l_m^*$, and ask whether the cost of restoration can be reduced by introducing individuals into sex-specific lengths of habitat. We denote the ratio of costs obtained by unequal and equal cluster sizes as:

$$c_{\text{rel}} = \frac{C_{\text{rect}}[l_f^* \neq l_m^*]}{C_{\text{rect}}[l_f^* = l_m^*]} \tag{3.18}$$

The calculation of the sex-specific critical cluster sizes (l_f^* and l_m^*) requires bivariate optimization, implemented in the following way. First, we calculate the critical cluster size for males at a fixed cluster size for females using bisection method, resulting in a cost with respect to the given female cluster size. Then, we use gradient descent on this cost function to minimize it with respect to female cluster size. Note, that during the gradient descent we always change the female cluster size by one

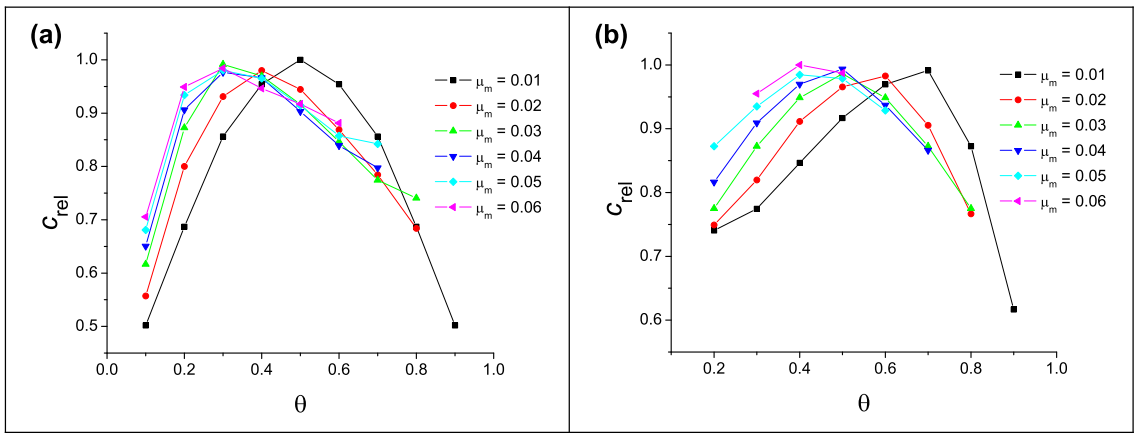

Figure 3.7: Minimum cost found by allowing different initial cluster sizes for males and females, relative to the case where cluster lengths are equal. Common parameters: $D_f = D_m = 1.0$. Individual parameters: (a) $\mu_f = 0.01$, (b) $\mu_f = 0.03$.

unit of spatial grid resolution, and we keep moving toward the negative gradient even if the local derivative is zero, because a small slope discretized on a grid can result in zero local gradients before reaching the actual minimum point. Note also, that we strongly rely on the convexity of the cost function with respect to male and female cluster sizes.

Figure 3.7 presents our results. We find that optimality of the sex ratio plays a central role in reducing the cost of restoration. In particular, when the sex ratio equals the optimal (density-maximizing) value determined by mortality rates, the minimum cost is achieved with equally sized clusters. No further cost reduction can be achieved by allowing different cluster sizes for males and females. However, the relative advantage of sex-specific cluster size increases as the sex ratio deviates from the optimal value. If we consider that the absolute cost value diverges for strongly biased sex ratios [see Fig. 3.5(b)] we conclude that in such cases the savings achieved by adjusting the initial cluster sizes of each sex could be substantial.

3.4 Simulated Annealing

Finally, we relax all constraints on the shape of the initial population distribution; we optimize the spatially discretized shape for lowest cost under the constraint

of successful restoration. Discretization is essential because it allows us to express the cost as an n-dimensional function, instead of a functional, where n is the size of the spatial grid. We use the same grid for shape discretization and numerical PDE integration, for practical reasons. Given the discretization, we use *simulated annealing* [181] for optimizing the shape function. This is essentially a Monte Carlo simulation, where random changes of the initial spatial distribution (the shape function) are accepted or rejected according to a specified acceptance probability function, such that the visited cost states have a Boltzmann-distribution characterized by a temperature-like parameter. As this parameter is lowered, the expected value of the cost is also lowered, eventually leading to the globally optimal, minimum cost state. The specific steps of simulated annealing are provided in Appendix A.3.

In order to determine whether the constraint of successful restoration is satisfied we must numerically integrate the model at every Monte Carlo step. To reduce computational time, we accelerated the PDE integration with GPGPU computation using CUDA [182, 183], which locates the global equilibrium of the PDE in a fraction of a second, giving a total time for the simulated annealing on the order of a few hours.

At this stage our only assumption is that the cost-minimizing distributions have a finite support, and we carry out the minimization accordingly. However, the support of the function is allowed to grow or shrink by random shape changes during simulated annealing; see Appendix A.3 for details.

By analyzing a series of minimum cost distributions obtained with simulated annealing, we observe the following. First, the distributions indeed have a finite support. Although we initialize them as such, the width of the optimal population distribution tends to become smaller, rather than larger, during simulated annealing. Typical final, optimized shapes are shown in Fig. 3.8. It is also remarkable that the edges of the distributions go to zero very sharply; this property develops without any influence inherent to the procedure.

Generally, the final result is an "arch"-shaped distribution, with similar dimensions for females and males. Note that as the sex ratio diverges from its optimal value, we observe a change in the sizes of the two initial population distributions

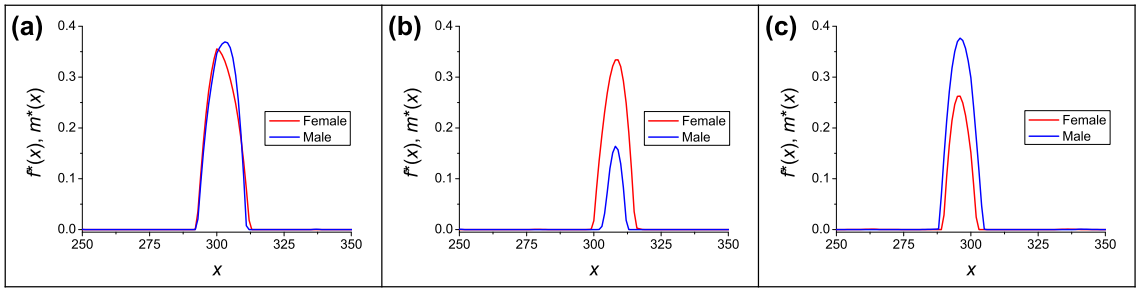

Figure 3.8: Shapes of initial density distributions that minimize cost, found by simulated annealing. The x axis shows discretization grid coordinates; $dx = 0.1$ length unit per grid point. Common parameters: $D_f = D_m = 1.0$. (a) $\theta = 0.5$, $\mu_f = 0.04$, $\mu_m = 0.04$, (b) $\theta = 0.1$, $\mu_f = 0.01$, $\mu_m = 0.02$, (c) $\theta = 0.7$, $\mu_f = 0.04$, $\mu_m = 0.02$.

and in the height of the peaks. These changes in spatial distributions occur roughly in proportion to the system's positive stationary densities [Fig. 3.8(b,c)]. Interestingly, the height of the peaks always falls between the stationary densities and the Allee threshold. Note that this shape provides the maximal rate of population growth possible during the first moments of the simulation, hence it combats the diffusion-amplified Allee effect most efficiently. Costs corresponding to the optimized distributions are denoted C_{SA} when we compare results with other methods.

3.5 Discussion and Conclusions

We examined three approaches to minimizing the cost of a species' restoration; the approaches differ in both ecological premises and mathematical methods. We considered the aperiodic, spatially inhomogeneous solution to the single-sex dynamics [Fig. 3.1(b)], critical cluster sizes of the rectangular initial setup [Fig. 3.5(a)], and simulated annealing of sex-specific initial distributions [Fig. 3.8].

Summary comparisons of the minimum restoration cost achieved by the different methods for the symmetric single sex model appear in Fig. 3.9. The aperiodic stationary solution to the dynamics gives significantly larger cost than the other approaches. Considering the shape of these distributions, they would likely prove difficult to implement in application. Restricting attention to the other two approaches,

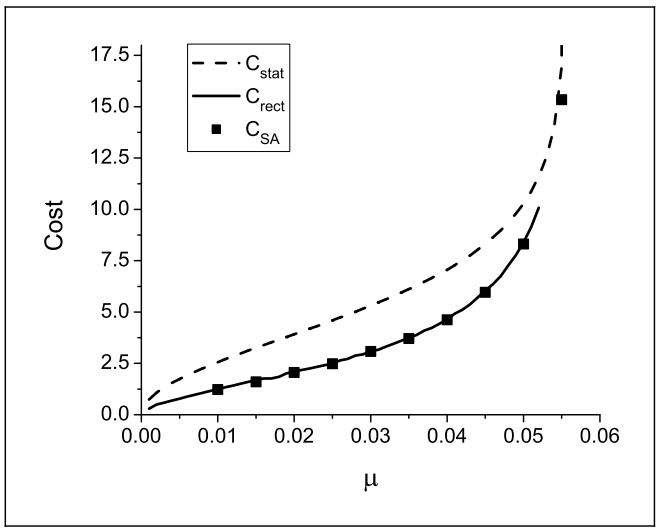

Figure 3.9: Comparison of minimum cost values in the symmetric model. Cost values found by integrating the aperiodic stationary density (C_{stat}), by using rectangular shape with stable stationary densities (C_{rect}), and with simulated annealing (C_{SA}) are presented. $D = 1.0$**.**

it is remarkable that simple rectangular distributions and the results of simulated annealing yield essentially identical costs.

Figure 3.10 compares minimum costs for each critical-cluster analysis (i.e., a single cluster size and sex-specific cluster sizes) and costs incurred under simulated annealing. The critical cluster methods assume uniform initial population density within cluster bounds; simulated annealing lets initial densities depend on spatial location. The minimum cost varies little among methods as long as the sex ratio at birth does not deviate too much from the optimal value (here, $\theta^* = 0.5$). As sex-ratio bias increases, optimal sex-specific initial cluster sizes can lower the minimum cost of restoration. Simulated annealing reduces restoration cost even further, but this advantage becomes significant only at strongly biased (and biologically rare) sex ratios, and implementing such spatial distributions in application could prove difficult, negating any cost advantage. The same qualitative conclusions hold when we fix the sex ratio and increase the difference between the sexes' respective mortality rates, because the mortality bias can also increase the difference between the optimal and any fixed sex ratio.

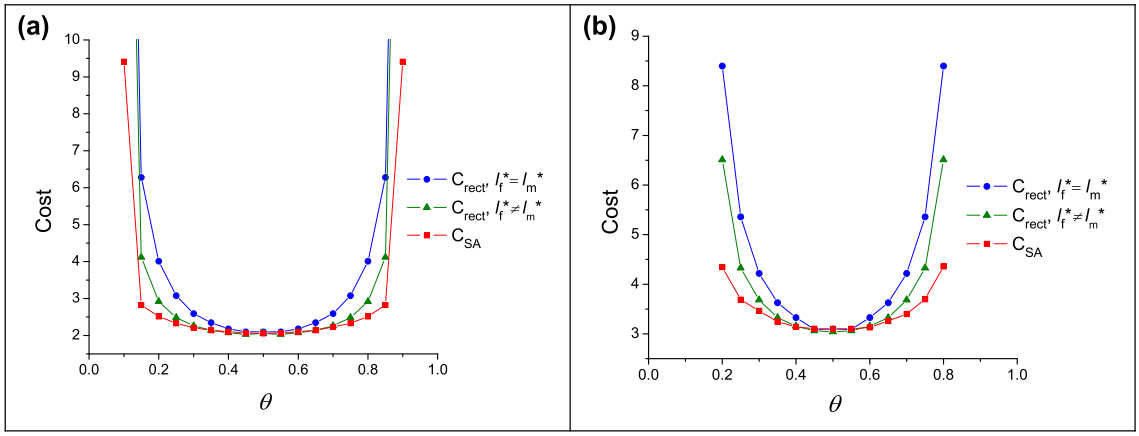

Figure 3.10: Comparison of minimum cost values in the two-sex model. Comparison includes cost values found by using rectangular shape with equal and unequal male and female cluster sizes, and by simulated annealing. $D_f = D_m = 1.0$**, (a)** $\mu_f = \mu_m = 0.02$**, (b)** $\mu_f = \mu_m = 0.03$**.**

Our model assumes deterministic dynamics, which does not account for extinction due to demographic stochasticity in populations near an extinction threshold [161, 184]. This effect can be exaggerated when a population's spatial dispersion leaves dynamically independent clusters near critical size [1, 185].

We assume diffusive dispersal. Many plants, and some animals, disperse only locally, i.e., the probability of long-distance dispersal is much lower than diffusion assumes [186–188]. Dispersal limitation becomes important when the number of discrete individuals is small [189], since random internal fluctuations can induce population extinction. Given discreteness and stochasticity, neither of which has a role in our cost-minimizing model, lattice-based results show that expected growth from rarity demands greater propagation, relative to mortality, as mean dispersal distance decreases [2, 190]. We also assume that no explicit interspecific interactions affect the population during restoration. Species occupying the community to be restored may facilitate restoration; for example, trees may attract birds that disperse seeds of other tree species [191]. Alternatively, resident species may resist the introduced species biotically [138, 192]. Interspecific interactions will often affect the likelihood of restoration success, as well as the cost. Consequences of these

interactions can sometimes be expressed abstractly through the introduced species' positive equilibrium density; in other cases, successful restoration may demand quantification of these interactions.

We assume an Allee effect arises from interaction of self-regulation with a birth rate that depends on the density of each sex. In the context of restoration, a two-sex dynamics may be essential to predicting spatial-expansion rate if dispersal differs between sexes [47]. We model mating encounters via mass-action, which should be reasonable for animals maintaining individual home ranges, or for dioecious plants with random mating. Alternative "marriage functions" [44] apply to certain species, particularly for polygynous or polyandrous mating systems.

We modeled a single species' restoration only. Habitat restoration may attempt to manage particular multi-species interactions, or may seek to promote growth of many threatened species [30]. Our cost function ignores feedback of a species' restoration on other biotic processes, or on economic stake-holders who incur post-restoration costs [193].

Our results suggest some considerations for species restoration. First, if a species disperses rapidly, individuals should be introduced concurrently, rather than serially. The initial population will increase only if the density exceeds any Allee threshold, and continues to do so as individuals disperse. Intuitively, the number/density of individuals introduced should increase with their dispersal rate.

Second, restoration cost declines little by introducing a species at a density below the (estimated) carrying capacity, unless the species disperses very slowly. Third, a rectangular spatial distribution adds little or no proportional cost over the ogive profile assumed in our simulated annealing method, as long as the sex ratio is close to optimal. Spatial uniformity will likely prove more practical for most animals. Given a uniform density close to the positive, stable equilibrium, restoration should focus on an initial population whose expanse exceeds the critical-cluster size, which (again) increases with dispersal rate.

Finally, adjusting frequencies of the sexes in an initial population may decrease the cost of successful restoration. Of course, if one sex always limits population growth, an excess of that sex promotes restoration. If population growth depends on

the density of each sex, introducing the sexes at different densities, or with different cluster sizes, may prove advantageous. Sex ratio at birth may be unbiased, but mortality rates may differ between sexes, particularly during dispersal. Adjusting the sex ratio at introduction to match frequencies at positive equilibrium densities should promote successful restoration, and reduce its cost.

CHAPTER 4
NUMERICAL TECHNIQUES FOR MODELING AND ANALYZING COMPLEX NETWORKS

We have developed a number of numerical methods and software tools to facilitate our research of dominating sets in scale-free networks, which we present here as a prerequisite before starting our study. These methods include techniques for efficiently storing heterogeneous networks (including scale-free networks) in memory, adapting universal network construction methods for building model scale-free networks with given parameters, and implementing efficient search algorithms to find approximations to the minimum dominating set (MDS).

4.1 Hybrid Storage of Heterogeneous Networks

The most commonly used data structures for representing the connectivity information of a network (graph) in computer memory are the adjacency list, and the adjacency matrix. In an adjacency list, as the name suggests, there is a list associated with each node that contains all nodes adjacent to that given node. For sparse networks, this is a very compact and efficient data structure, and thus it is used most commonly in applications. The adjacency matrix, on the other hand, is an $N \times N$ binary matrix (where N is the size of the network), where the values 1 and 0 at (i, j) indicate the presence or absence of an edge between nodes i and j, respectively. This data structure is more suitable for particularly dense networks.

Both data structures support directed networks as well as undirected ones. For an undirected network, every link is listed twice in an adjacency list, once for

Portions of this chapter previously appeared as: F. Molnár Jr., S. Sreenivasan, B. K. Szymanski, and G. Korniss, "Minimum Dominating Sets in Scale-Free Network Ensembles," Sci. Rep. **3**, 1736 (2013).

Portions of this chapter previously appeared as: F. Molnár Jr., N. Derzsy, É. Czabarka, L. Székely, B. K. Szymanski, and G. Korniss, "Dominating Scale-Free Networks Using Generalized Probabilistic Methods," Sci. Rep. **4,** 6308 (2014).

Portions of this chapter to appear as: F. Molnár Jr., N. Derzsy, B. K. Szymanski, and G. Korniss, "Building Damage-Resilient Dominating Sets in Complex Networks against Random and Targeted Attacks," (under review).

each node that the link connects. Similarly, the adjacency matrix is symmetric. For directed networks, directed links (arcs) are only listed on the adjacency list of the source node (tail), listing the adjacent (head) node where the link points toward (however, in some applications, storing links in the opposite direction may be more practical); and the adjacency matrix is no longer symmetric, having a user-specified convention on which of the row and column indices represent source and target nodes (heads and tails) of the links.

Additional data, such as weights on nodes and links, can also be incorporated into adjacency lists and adjacency matrices. In the former, for example, additional information may be allocated for each node on the adjacency list; and for the latter, the matrix may be real-valued, where any non-zero value indicates not just the presence of an edge, but also its associated weight. However, more complex data, either associated to nodes or links, are more suitably stored in a separate list or dictionary.

Unfortunately, neither the adjacency list nor the adjacency matrix are a good fit for storing heterogeneous networks, such as scale-free networks. In particular, due to the heavy tail of the degree distribution (which is more pronounced when the degree exponent is low), there are always a significant number of high-degree nodes that require very long adjacency lists to store connectivity information. For these nodes, an adjacency matrix would be more suitable. However, most nodes have low degrees, and for them the adjacency list would be optimal. To be specific, the exact threshold where the data structure should change from adjacency list to adjacency matrix is when the node degree exceeds aN/b, where a is the number of bits needed to store one element of the adjacency matrix, and b is the number of bits needed to store an element in the adjacency list. Assuming that the adjacency matrix elements can be stored as single bits, and nodes in the adjacency list are identified by 32-bit integers, the threshold becomes $N/32$.

Our hybrid network storage technique is based on this principle. For each node, the adjacent nodes are stored in a list when there are less than $N/32$ of them, and stored in an adjacency bitvector (a single row of an adjacency matrix) when their number exceeds $N/32$. The switch between storage formats is carried out

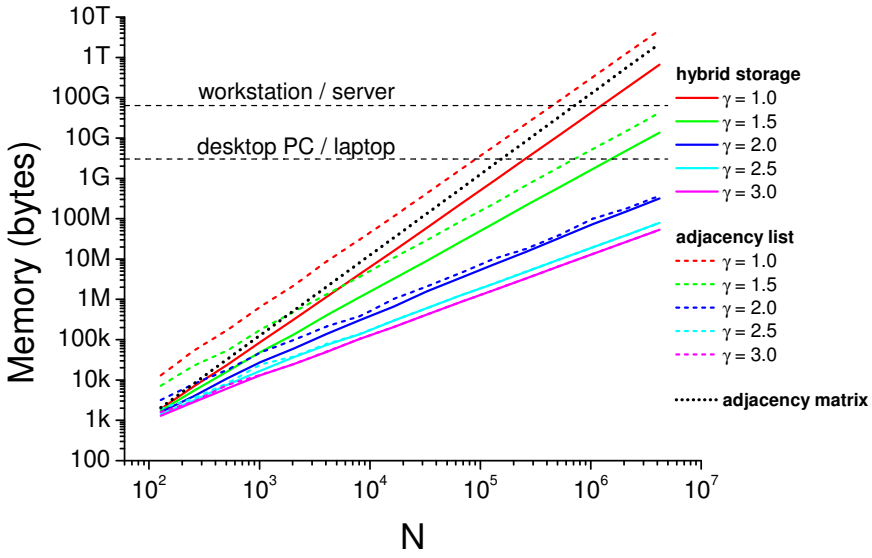

Figure 4.1: Comparison of memory requirements for storing scale-free networks using various data structures, with different power-law degree exponents. Typical amounts of available memory in personal computers and servers (as of 2014) are also shown for comparison.

automatically as links are added to or removed from the network. When needed, additional node data is stored in an array indexed by node IDs (32-bit integers from 0 to $N-1$) and additional edge data is stored in a hashed dictionary where the key is a 64-bit link ID calculated as the binary concatenation of the two node IDs connected by the link (for undirected links, the higher 32 bits always contain the smaller node ID). Note, a node ID is fixed for the lifetime of each node; IDs of deleted nodes are stored in a list and reused when new nodes are added to the network.

The computational overhead caused by the storage format switch is minimal. Since the number of steps needed to carry out the switch is exactly the same as the number of links that was added previously to a given node, the time complexity of adding one edge remains $O(1)$ in amortized time. For rapidly changing networks, where links are deleted as often as they are added, a switch back from adjacency bitvector to list may be omitted for higher performance, or carried out only if the available memory is running out, much like the garbage collection principle used in many object-oriented programming languages.

Memory requirements of storing scale-free networks using various data structures are compared in Fig. 4.1. It is clearly shown that for dense scale-free networks (with low degree exponents), our hybrid storage method uses an order of magnitude less memory than adjacency lists, and in general, our method is much more efficient than the adjacency matrix. From another perspective, the maximum size of networks that we can store and analyze on a computer with a fixed amount of memory is increased by an order of magnitude for dense scale-free networks.

4.2 Scale-Free Network Sample Generation

The most commonly cited methods for generating scale-free networks are variants of the preferential attachment model [65, 194, 195]. The main drawback of these methods is that they do not provide sufficient control over all network parameters. In order to have full control we must use universal methods that can build networks from any given degree sequence.

The first step, regardless of the specific construction model, is to obtain a degree sequence with the desired parameters. For this, we first create the (discrete) power-law degree distribution (with prescribed cutoffs), and calculate the cumulative distribution function (CDF). Then, using inverse sampling of the CDF, we generate the degree sequence of the network. We find that using a discrete distribution, rather than sampling degrees from a continuous distribution with rounding, results in much better accuracy of achieving the desired power-law exponent, for scale-free networks.

Once we have a degree sequence, it is passed on to one of three construction models as input. Each model (and thus the type of the resulting network) is identified by a four-letter abbreviation as follows:

- CONF; abbreviation for the configuration model [196, 197],
- HHMC; a Markov chain Monte Carlo method [198] initialized with Havel-Hakimi construction [199, 200],
- DKTB; a sequential construction algorithm capable of generating all possible realizations of a given degree sequence, named after the authors [201]. The algorithm is built on the underlying theorems proven in [202].

Based on the maximum degree cutoff, we also define two subtypes of these networks. Either the maximum cutoff is unrestricted, therefore $k_{\max} = N - 1$, or we introduce a structural cutoff, $k_{\max} = \sqrt{N}$, to create uncorrelated scale-free networks [203, 204] (see Chapter 5 for further details). In the latter case, having the cutoff is indicated in the network type name as cCONF, cDKTB, and cHHMC. Besides the different cutoff, the construction of these networks is identical to their unrestricted cutoff versions, illustrated in Fig. 4.2, and detailed below.

In CONF networks, we first assign as many edge stubs (half-links) to each node as prescribed by the degree of the node in the degree sequence. Then, we randomly pick any two edge stubs (uniformly among all stubs) and connect them, forming a link. We continue picking and connecting stubs until there are no more stubs left. The result of this procedure is a multigraph; we reduce multiple links to single links, and remove self-loops, to obtain a simple graph.

In HHMC networks, we first build a simple graph deterministically from the degree sequence using the Havel-Hakimi algorithm. Although this fundamental graph-theoretical method is designed to prove the graphicality of a degree sequence (that is, whether a simple graph exists with the given degree sequence), it does so by constructing the graph itself, thus we can use it as a starting point of a Markov chain to obtain a random network sample [198]. In the Markov chain we obtain the next realization of the degree sequence by carrying out a *double edge swap*, according to the following procedure. We select two edges randomly (uniformly among all edges): $(uv) \in E$ and $(xy) \in E$ with distinct nodes ($u \neq x$, $u \neq y$, $v \neq x$, $v \neq y$). These edges would be deleted and replaced by new ones; either (ux) and (vy), or (uy) and (vx) would become the new pair of edges. Considering that some of these edges may already exist, making that swap configuration invalid, we have three possibilities: (1) Both edge swap configurations are possible, then we pick one configuration randomly; (2) only one configuration is possible, then we pick that one; (3) no swap is possible, making the swap invalid. We proceed further only if we have a valid swap, otherwise, we continue with a new edge swap attempt. After a sufficiently large number of swaps we obtain a random (uniform, unbiased) sample from all possible realizations of the degree sequence, because the Markov chain is irreducible [205], aperiodic, and

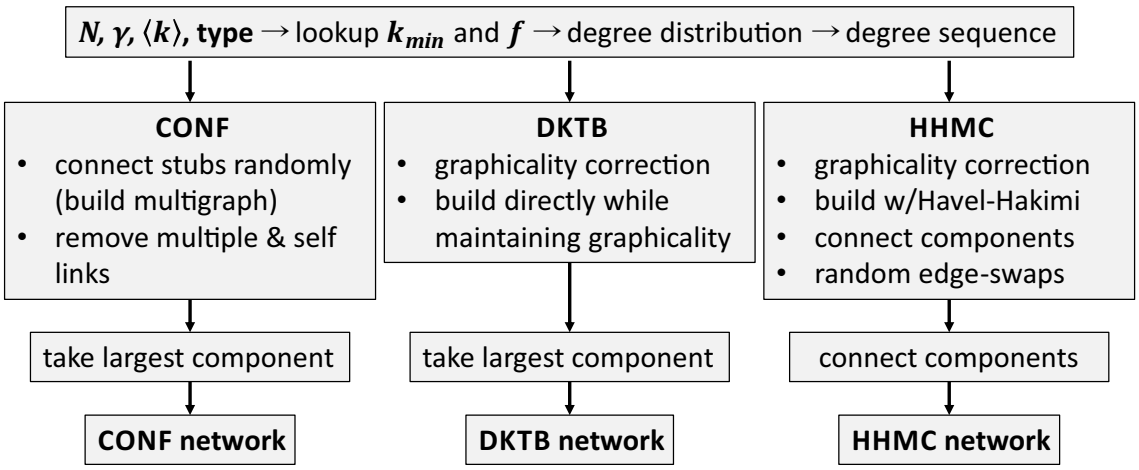

Figure 4.2: Flow diagram of scale-free network sample generation.

symmetric [198]. Although a theoretical proof is yet to be found, many experiments suggest that the Markov chain has short, $O(E)$ mixing time [206–208]. We found that attempting edge swaps four times the number of edges in the network is sufficient to obtain a random (unbiased) network sample, see Appendix B.1.

In DKTB networks, we follow a complex construction procedure, detailed in [201]. The main idea is to evaluate all possible choices before connecting the next link of each node, considering that after forming the new link, the remaining degree sequence (the unconnected edge stubs) must remain graphical. The procedure also tracks what the probability was to make each choice of link formation, thus it provides a statistical weighting factor at the end of the construction that tells us how likely it was to obtain that particular sample of the given degree sequence. However, calculation of this factor is difficult in practice, due to numerical overflows, caused by the extremely large number of possible choices, even with a network size of a few thousand nodes. Therefore, we ignore this weighting factor.

The final step of network generation is to ensure that we have a single connected component. In principle, all three network generation methods can produce networks with multiple components. However, due to the relatively high average degree we use ($8 < \langle k \rangle < 16$), the probability of having such networks is extremely low, as we never observed disconnected components in CONF and DKTB networks. Regardless, we always take the largest component of the resulting network (and remove the rest of nodes and edges), to ensure that we have a single component with 100% certainty.

In case of HHMC networks, the Havel-Hakimi algorithm can easily generate multiple components. The most common scenario is one large component, and numerous small fragments made of a few nodes. These can be deterministically connected into a single component without altering the degree sequence by performing well-picked double edge swaps: we must swap edges belonging to different components, provided that at least one of the swapped edges are not bridges (edges, whose removal would disconnect a component). We used Tarjan's algorithm [209] to find bridges that we needed to avoid.

In the original edge-mixing algorithm (in [198]) all edge swaps that would disconnect the network are ignored. In practice, testing this condition at every edge swap attempt is very expensive. Instead, we only check for multiple components after the edge mixing process has finished, and if we find multiple components, we connect them as we did before the random mixing. However, we never observed disconnected components after the random mixing, this step is only executed to ensure a single component with 100% certainty, just like taking the largest component after CONF and DKTB methods does.

HHMC and DKTB methods are "exact methods" in the sense that they do not alter the given degree sequence while constructing the network (unlike CONF, where the removal of multiple links alters the degree sequence). Therefore, we must supply these methods with graphical degree sequences, i.e., sequences for which it is assured that a simple graph realizing that sequence exists. To ensure graphicality, we devised a *graphicality correction* method (see algorithmic details in Appendix B.2), based on the Havel-Hakimi algorithm. The goal of the original algorithm is to test the graphicality of a degree sequence (while also building a possible network realization to prove graphicality). It reports failure when some stubs of a node cannot be connected to other nodes. Instead of reporting failure, we simply remove these stubs from the degree sequence, making the remaining sequence graphical. This correction precedes the network construction step for HHMC and DKTB networks, but it is not needed for CONF networks, because every degree sequence is graphical for a multigraph. We effectively remove any non-graphicality of the degree sequences when we remove multiple links and self loops to create a simple graph.

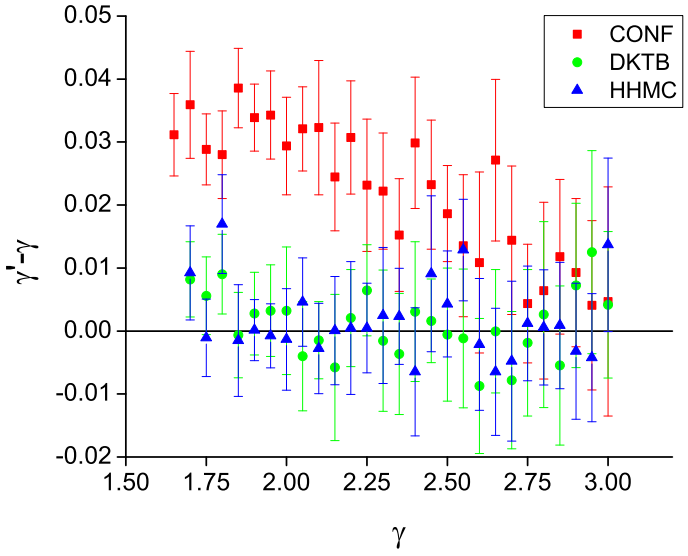

Figure 4.3: Accuracy of power-law exponents in CONF, DKTB and HHMC type networks, $N = 5000$, $\langle k \rangle = 14$. The horizontal axis shows the desired value, while the vertical axis shows the absolute difference between achieved and desired values, averaged over 20 samples.

To check whether the desired power-law exponents were achieved in the generated scale-free networks, we used the maximum likelihood estimator (MLE) [210] to measure actual values, denoted by γ'. It has been proven that MLE has superior performance compared to linear regression of the degree distribution on double-logarithmic plots [211, 212]. We measured γ' on the $(k_{\min} + 1, k_{\max})$ interval of the degree distribution. Figure 4.3 demonstrates the accuracy of power-law exponents in our generated networks. Although we can see a systematic error in the γ' values of CONF networks (possibly due to pruning multiple links), its value is very small, less than 3% deviation from the desired γ value, and it can be ignored.

4.3 Greedy Algorithm for Finding Approximate MDS

Since finding the MDS is NP-hard, we approximate the exact solution by using a sequential greedy algorithm [213]. Starting with an empty set $\mathcal{D}$, at each step the algorithm adds that node to $\mathcal{D}$ which yields the largest increase in the number of

dominated nodes in the network. When there are multiple candidate nodes yielding the maximal increase in domination, the algorithm chooses one randomly (uniformly among candidates). These steps continue until all nodes are dominated and then the algorithm terminates with $\mathcal{D}$ storing the approximated MDS. The greedy algorithm yields a $1 + \log N$ approximation to the size of MDS [87], and it has been also shown in the same work that finding a sublogarithmic approximation is NP-hard.

In order to implement the greedy procedure in $O(E)$ steps, we need to use appropriate data structures. First, we define the *dominating potential* of a node as the number of nodes that would become dominated if that node were selected at a given step. The greedy selection is based on this quantity; a node with the highest dominating potential is selected at every step. Initially, the dominating potential is set to $1 + d(i)$, where $d(i)$ is the degree of node i, because every node can dominate all its neighbors and itself. We use bucket sort to order nodes by their dominating potential in $O(N)$ time (which we can do because dominating potentials are integers in $[0, N+1]$ range), where we use a hashed list for each bucket. This allows lookup by value (by node ID), node addition and removal, and random node selection from any bucket to be performed in $O(1)$ amortized time, therefore it also allows to maintain the order of nodes in $O(1)$ time after any single change in a node's dominating potential. At every step of the algorithm, a node is selected into the dominating set, and the dominating potential of the node itself, its neighbors, and second neighbors are adjusted accordingly. Initially there are $\sum_{i:\text{nodes}} 1 + d(i) = 2E + N = O(E)$ dominating potentials. In every step, the affected nodes' potentials are reduced by one, except the selected node's potential, which reduces to zero instantly. There are at most $2E + N$ reductions of potentials, each completed in $O(1)$ amortized time, therefore the algorithm runs in $O(E)$ amortized time.

4.4 Controlling the Average Degree

There is an exact relationship between four parameters of the degree distribution: the power-law exponent γ, the minimum degree cutoff $k_{\min}$, the maximum

degree cutoff $k_{\max}$, and the average degree $\langle k \rangle$, namely:

$$\langle k \rangle = \int_{k_{\min}}^{k_{\max}} C k^{1-\gamma} \mathrm{d}k, \tag{4.1}$$

where C is the normalization for the degree distribution. Therefore, in order to control the average degree, we must adjust one of the other three parameters. The degree exponent is fixed as input, and although we have chosen to use either $k_{\max} = N-1$ or $k_{\max} = \sqrt{N}$ in our network subtypes, changing the maximum degree cutoff arbitrarily would, in principle, change the very essence of a fat-tail distribution. Thus, we are left with only one possible choice: the minimum degree cutoff, $k_{\min}$.

The problem with using $k_{\min}$ to control $\langle k \rangle$ is that changing $k_{\min}$ even by 1 would change $\langle k \rangle$ significantly (note, $k_{\min}$ is an integer, since we are using discrete degree distributions). Therefore, in order to have a very fine control over $\langle k \rangle$, we also remove a certain fraction f of the lowest degrees from the degree distribution, $f \in [0, 1)$. The exact formula for the degree distribution is:

$$p_K(k) = \begin{cases} C(1-f)k^{-\gamma} & \text{if } k = k_{\min} \\ Ck^{-\gamma} & \text{if } k_{\min} < k \leq k_{\max} \end{cases} \tag{4.2}$$

$$C = \left[(1-f)k_{\min}^{-\gamma} + \zeta(\gamma, 1+k_{\min}) - \zeta(\gamma, 1+k_{\max})\right]^{-1} \tag{4.3}$$

In principle, we can calculate the exact average degree using this distribution:

$$\langle k \rangle = \frac{(f-1)k_{\min} + k_{\min}^{\gamma}\left(\zeta(\gamma-1, k_{\max}+1) - \zeta(\gamma-1, k_{\min}+1)\right)}{f - 1 + k_{\min}^{\gamma}\left(\zeta(\gamma, k_{\max}+1) - \zeta(\gamma, k_{\min}+1)\right)}, \tag{4.4}$$

where $\zeta(s, q)$ is the generalized Riemann zeta function:

$$\zeta(s, q) = \sum_{n=0}^{\infty} \frac{1}{(q+n)^s}. \tag{4.5}$$

However, this formula has only a limited usefulness. The actual average degree in our networks can be different, due to graphicality corrections, pruning of multiple links and self-loops, and removal of small components during network construction. In order to calibrate which $k_{\min}$ and f values result in which average degrees, we use

high-accuracy empirical lookup tables. Essentially, we follow the principle of trial and error: we blindly loop over all possible $k_{\min}$ and f values in a given range, generate network samples, and record the achieved average degrees. This is illustrated in Figure 4.4. Afterward, if we want to generate a network with a given $\langle k \rangle$, we look for the appropriate $k_{\min}$ and f values that result in the wanted $\langle k \rangle$ value. We also need to loop over many possible network sizes N, and a wide range of γ values as well. The lookup table is therefore a multidimensional grid of known expected average degrees. For network parameters between grid points we perform a numerical sampling of this grid by linear interpolation. The resolution of the lookup table is given in Table 4.1. Note, that for each N, γ, $k_{\min}$ and f values we generate multiple network samples, and we record the average of average degrees among samples. The number of samples is given in Table 4.2.

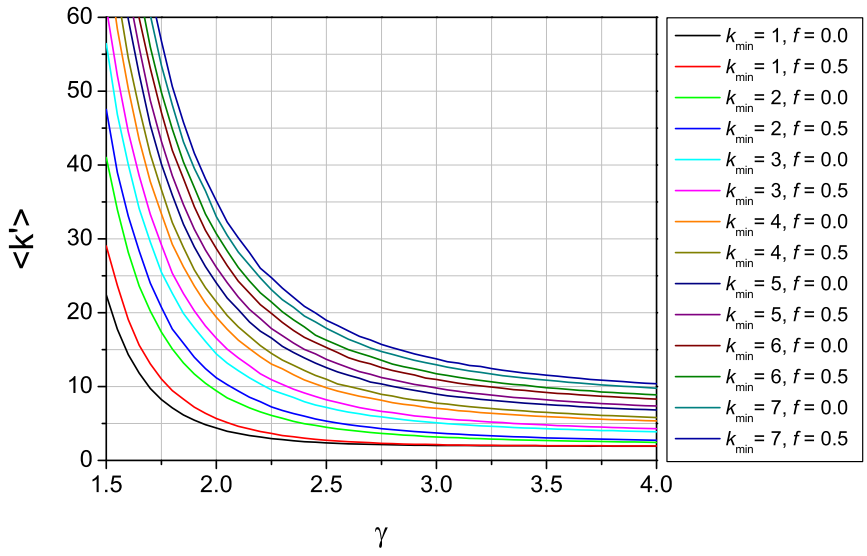

Figure 4.4: A cross-section of the empirical lookup table, showing the expectation of achieved average degrees vs. power-law exponent parameters at various minimum degree cutoffs and removed fraction of smallest degree nodes in CONF networks with $k_{\max} = N - 1$) and $N = 5000$.

Table 4.1: Parameter range and resolution of the average degree lookup tables created for all network subtypes.

Parameter	Lower bound	Upper bound	Step
$k_{\min}$	1	8	1
f	0.0	1.0	0.25
γ	1.5	4.0	0.05
N	100 1000 10000	1000 10000 40000	100 1000 5000

Table 4.2: Number of network samples generated for the average degree lookup table at each γ, $k_{\min}$ and f values. Bounds of N are inclusive.

Lower bound of N	Upper bound of N	Sample Count
100	1000	100
2000	10000	50
15000	30000	25
35000	40000	10

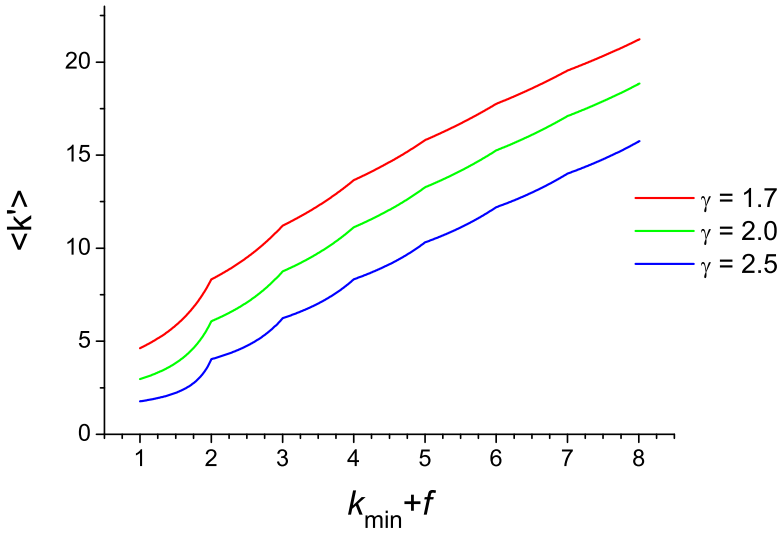

Figure 4.5: Relationship between control parameters $k_{\min}$ and f, and the achieved average degree, calculated from Eq. (4.4) for $\sqrt{N}$ cutoff networks with $N = 5000$. Here, $k_{\min}$ and f are combined into a single parameter to show continuous curves.

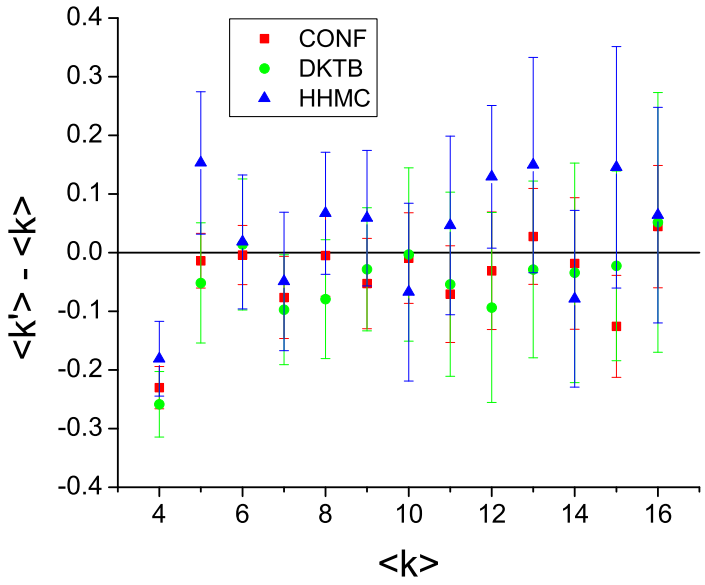

Figure 4.6: Accuracy of average degrees in CONF, DKTB and HHMC type networks, $N = 5000$, $\gamma = 2.5$. The horizontal axis shows the desired value, while the vertical axis shows the absolute difference between achieved and desired values, averaged over 50 samples.

It is important to note that the relationship between the control parameters ($k_{\min}$ and f) and the average degree is non-trivial, as shown in Figure 4.5. For clarity, we show analytically obtained values for a perfect degree distribution according to Eq. (4.3). We can find similar behavior for actual realized networks, but the effect is more difficult to observe due to noise from sampling errors and graphicality corrections.

We also checked the accuracy of the achieved average degrees, denoted by $\langle k' \rangle$, comparing them to the desired values. Figure 4.6 demonstrates the accuracy of average degrees in our generated networks.

4.5 Measuring and Controlling Assortativity

Assortativity describes mixing patterns in networks [118, 119], which is essentially a description of how likely it is to find edges between nodes of similar type. In particular, nodes are commonly classified by their degree, thus we have assortative

mixing (assortative networks) if nodes with similar degrees are more likely connected to each other, and disassortative mixing (disassortative networks), if nodes with significantly different degrees tend to be connected.

Assortativity is usually quantified as a correlation between the degrees of two nodes. The classical *assortativity coefficient*, introduced by Newman [118], is simply the Pearson correlation coefficient of the degrees of nodes connected by edges:

$$r = \frac{M^{-1}\sum_i j_i k_i - [M^{-1}\sum_i \frac{1}{2}(j_i + k_i)]^2}{M^{-1}\sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1}\sum_i \frac{1}{2}(j_i + k_i)]^2}, \tag{4.6}$$

where M denotes the number of edges; j_i and k_i are the degrees of the nodes at the end of the i-th edge. However, it has been shown recently [214] that the assortativity coefficient lacks accuracy and it is dependent on network size. Following the recommendation of [214], we measure network assortativity by Spearman's ρ [215, 216], a rank-correlation metric, defined as:

$$\rho_n^{rank} = \frac{\sum_{i=1}^{n}(r_i^X - (n+1)/2)(r_i^Y - (n+1)/2)}{\sqrt{\sum_{i=1}^{n}(r_i^X - (n+1)/2)^2 \sum_i^n (r_i^Y - (n+1)/2)^2}}, \tag{4.7}$$

where n is the network size, r_i^X and r_i^Y are the ranks assigned to X and Y that are the degrees of the two nodes found at the end of edge i. Spearman's ρ is independent of network size, allowing direct comparison of assortativity in networks of different sizes, and it can reveal strong dependencies more efficiently.

Control over network assortativity can be achieved by rewiring the edges of the network with a series of random, but selective double-edge swaps [198], similarly to the technique used in HHMC network construction. A double-edge swap does not change the degree sequence of a network, but it can potentially change its assortativity.

Without any bias, a double-edge swap can either increase or decrease a network's assortativity, with approximately equal probabilities, leading to an uncorrelated network in the steady state. The basis of our method is to bias the acceptance of otherwise random and valid edge swaps, such that the introduced bias shifts

assortativity toward a desired value. We introduce this bias in the following way:

$$\Pr(\text{accept}) = \begin{cases} a & \text{if } a > 0 \text{ and the swap is assortative,} \\ -a & \text{if } a < 0 \text{ and the swap is disassortative,} \\ 1 - |a| & \text{else,} \end{cases} \tag{4.8}$$

where $a \in [-1, 1]$ is a predetermined control parameter that controls the acceptance ratio of assortative and disassortative swaps; a swap is classified as assortative or disassortative if it increases or decreases the network's total assortativity, respectively. We obtain the most disassortative network when $a = -1$, the most assortative network when $a = 1$, and $a = 0$ corresponds to the case without bias, leading to uncorrelated networks.

While we generally measure assortativity using Spearman's ρ, we can track the intermediate changes of assortativity during the edge-mixing process using the assortativity coefficient [118, 119], because as long as one measure increases, so does the other (although the exact relationship is quite complex). The advantage of tracking changes with the assortativity coefficient is that we can classify a swap only by computing the change in assortativity caused by the two edges involved in the swap, since other edges and nodes in the network are not affected. Tracking the changes in Spearman's ρ would require to track the changes in degree ranks, an $O(\log N)$ calculation. In contrast, if we subtract the assortativity coefficients before and after the proposed swap, we obtain a simple formula: $d_u d_x + d_v d_y \geq d_u d_v + d_x d_y$, if the assortativity increases after the swap (and the opposite is true if assortativity is decreased, i.e. the swap is disassortative), where d_x, d_y, d_u, d_z are the degrees of the corresponding nodes, and the new edges are (ux) and (vy). Similarly for the other swap configuration, where the new edges are (uy) and (vx), the swap is assortative if $d_u d_y + d_v d_x \geq d_u d_v + d_x d_y$, and disassortative otherwise. Therefore, we can classify a swap in $O(1)$ time.

Since the acceptance probability effectively lowers the number of accepted swaps compared to the unbiased case, we must increase the number of swap attempts to ensure that the network has reached a well-mixed state. Unlike HHMC construction,

where we attempted edge-swaps four times the number of edges, here we attempt ten times the number of edges in the network.

Using our guided edge-mixing we can reach a wide range of Spearman's ρ values for any given network; however, determining the correct a control parameter for a desired ρ is nontrivial, see Fig. 4.7. Due to the random nature of the mixing procedure the resulting value of ρ is a random variable with unknown (most likely Gaussian) distribution, but it is clear that the mean of ρ monotonically increases as a increases. Therefore, we carry out a randomized bisection search to find the needed a for a desired ρ. The initial bounds are $a_{\min} = -1$ and $a_{\max} = 1$. The search is controlled by statistics of the resulting ρ samples: for any particular a value we run 24 edge-swap cycles (each with swap attempts ten times the number of edges) and record the resulting mean $\mu_\rho(a)$ and the confidence interval $c_\rho^{50\%}(a)$ with 50% two-sided confidence. These statistics of ρ are computed at the upper and lower bounds of a, and in the middle of the range, $a_{\text{mid}} = (a_{\min} + a_{\max})/2$. The new range becomes the lower half of the current range, if $\mu_\rho(a_{\min}) < \rho_{\text{wanted}} < \mu_\rho(a_{\text{mid}})$; or it becomes the upper half, if $\mu_\rho(a_{\text{mid}}) < \rho_{\text{wanted}} < \mu_\rho(a_{\max})$. We keep halfing the ranges until the confidence intervals of the upper and lower bounds overlap; the middle point of the last range gives the needed a value. The sample size and the confidence interval have been selected such that they provide two decimal digit accuracy for ρ. The overall search procedure is somewhat time consuming, but collecting statistics can be parallelized, and for a given set of network parameters a has to be computed only once.

The accuracy of this method is illustrated in Fig. 4.8. Larger sample sizes imply higher accuracy, however it also makes the search process slower. On the other hand, higher confidence level results in wider error bars for the achieved ρ, producing less accurate final results, but faster search. It is important to note that each search terminates when the bisection algorithm finds the upper and lower bounds with overlapping confidence intervals. Thus the confidence intervals do not refer to the accuracy of the bisection, instead they are used as a statistical tool for providing a well-defined terminating condition for the bisection algorithm. Figures 4.9 and 4.10 provide additional information about the achieved network assortativity using our method.

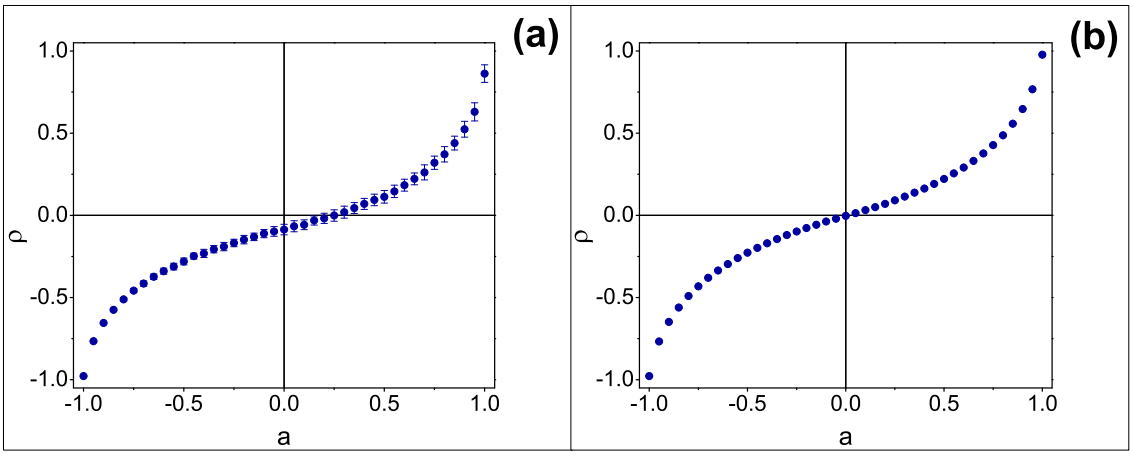

Figure 4.7: Relationship between the assortativity control parameter a and the achieved Spearman's ρ values, for (a) CONF and (b) cCONF networks. Parameters: $N = 5000, \gamma = 2.5$, $\langle k \rangle = 14$. Data is averaged over 100 network samples. Error bars indicate the sample standard deviation.

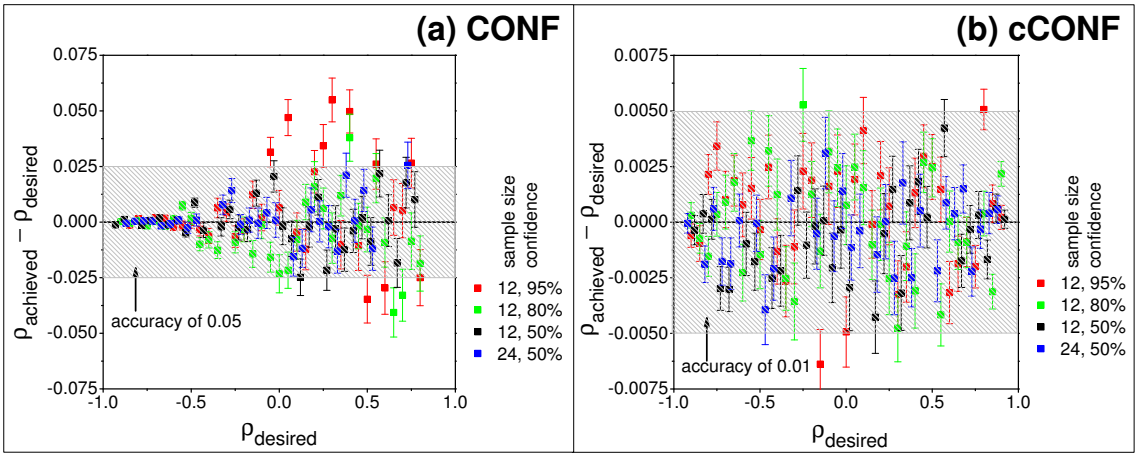

Figure 4.8: Controlling Spearman's ρ with randomized bisection method, using various sample sizes and confidence intervals of ρ used internally at each bisection step. The error bars show the actual confidence of the finally achieved ρ values after completing the bisection, for 100 samples with 95% confidence level. Network parameters are $N = 2000$, $\langle k \rangle = 14$ and $\gamma = 2.5$.

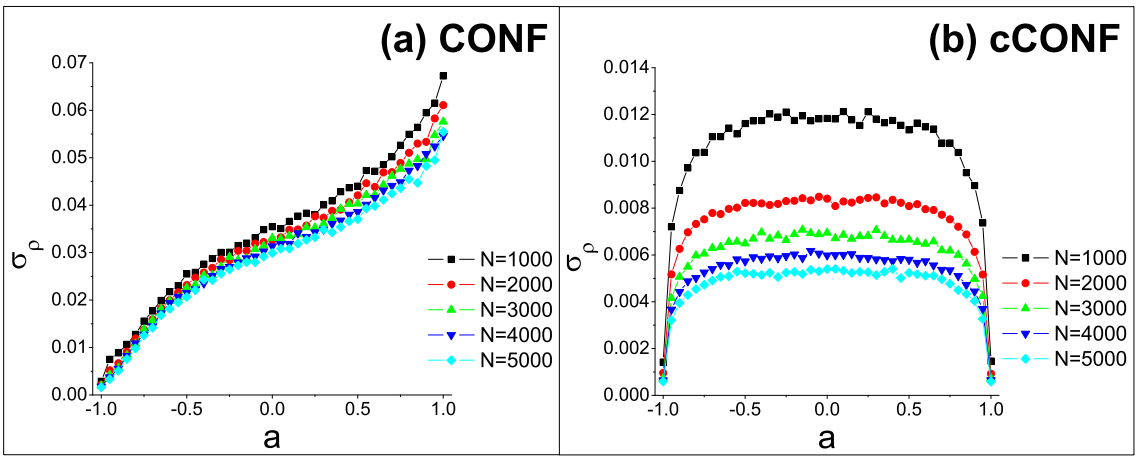

Figure 4.9: Standard deviation of Spearman's ρ achieved with controlled edge-mixing, across many network realizations. Each data point corresponds to a statistic of 2000 samples for each network size; $\langle k \rangle = 14$, $\gamma = 2.5$, measuring ρ only once for each sample.

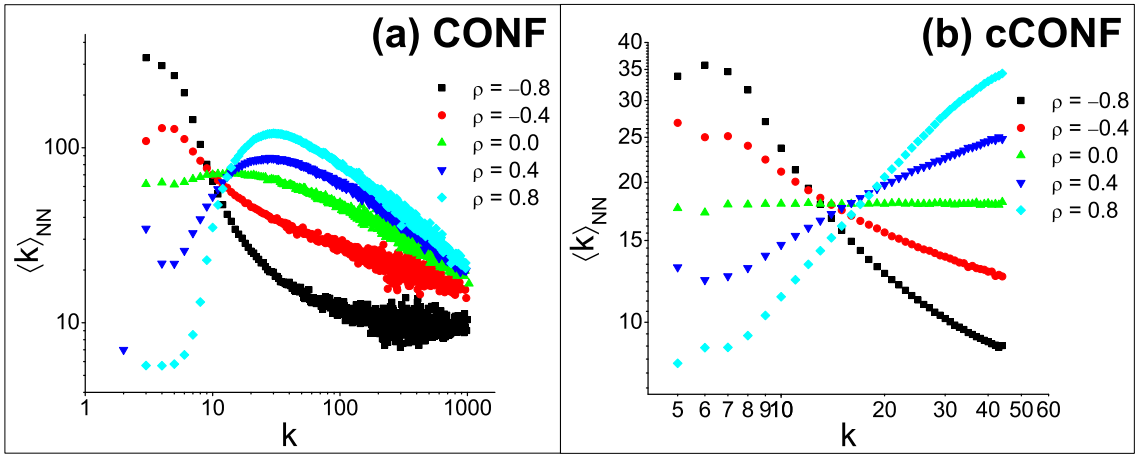

Figure 4.10: Average degree of nearest neighbor vs. node degree after mixing edges for various desired ρ values. Each data point is an averaged value over 100 network samples of size $N = 2000$, $\gamma = 2.5$ and $\langle k \rangle = 14$.

CHAPTER 5
SCALING OF MINIMUM DOMINATING SETS IN SCALE-FREE NETWORKS

The ability to efficiently control, track, monitor, or detect the behavior nodes is a central issue arising in complex networked systems. Assuming that nodes can control or influence their nearest neighbors most efficiently, the solution often involves finding a dominating set of a network. By definition, a dominating set of network (graph) $\mathcal{G}$ with node set V is a subset of nodes $S \subseteq V$ such that every node not in S is adjacent to at least one node in S; a minimum dominating set (MDS) is a smallest cardinality dominating set. Example problems in whose solution the MDS (or some variant of it) has been shown to play a part include optimal sensor placement for disease outbreak detection [217], controllability of networks [7, 73, 82], social influence propagation [10, 11], observability of power-grid stations [6], and finding high-impact optimal subsets in protein interaction networks [23].

In particular, we focus on the properties of the MDS in scale-free networks that are characterized by a power-law degree distribution ($P(k) \sim k^{-\gamma}$). These networks constitute a class of stylized networks which bear strong resemblance to several real-world networks including social, infrastructural and biological networks. Typically, values of the power-law exponent γ lie in the range $2 < \gamma < 3$ [218, 219], although there are few examples of networks with $\gamma < 2$ value; for example the co-authorship network in high-energy physics [220], and some email networks [221]. Our first goal is to understand the scaling behavior of the size of MDS against network parameters, in order to provide theoretical basis of the expected cost of implementing network control using dominating sets.

The mathematical literature focusing on bounds of dominating sets is vast [222]. In most prior works (with the exception of [92] to be discussed below), the MDS has not been studied systematically for scale-free networks over a significantly varying

Portions of this chapter previously appeared as: F. Molnár Jr., S. Sreenivasan, B. K. Szymanski, and G. Korniss, "Minimum Dominating Sets in Scale-Free Network Ensembles," Sci. Rep. **3,** 1736 (2013).

range of γ. For example, Cooper *et al.* [223] studied the behavior of MDS size on the special class of scale-free networks generated by preferential attachment [65] (corresponding to γ=3), and found that minimum dominating sets as well as minimum h-dominating sets (where every node needs to be dominated at least h times) have sizes that are bounded above and below by functions linear in N, where N is the number of nodes in the network. Other studies have focused on MDS sizes for random regular graphs and Erdős-Rényi (ER) [224] graphs. Zito [225] studied the size of the minimum independent dominating set on r-regular random graphs (with $3 \leq r \leq 7$) and showed that the size of this set (and therefore the MDS) is upper bounded by a linear function of N. Recently, Bíró *et al.* [226] improved the pre-factor of the $O(N)$ bound of the size of the MDS in r-regular graphs based on a greedy algorithm [88–90, 227]. Wieland *et al.* [228] derived general bounds for *dense* ER graphs (with fixed edge probability), showing that the MDS size scales as $\log N$ (with no direct applicability to sparse graphs with fixed average degree).

A recent study by Nacher *et al.* [92] has focused on the behavior of the MDS size on model scale-free networks with varying degree exponent, as well as empirical networks. The authors employed the Havel-Hakimi algorithm [199, 200] with random (Monte-Carlo) edge swaps [198] (HHMC) for constructing synthetic networks, and they used a binary integer programming method to obtain the MDS. They reported that the MDS size decreases as γ is lowered, making heterogeneous networks very easy to control. However, our study of a variety of scale-free network families suggests a more complicated picture. In particular, we find that details of the network generation process, and the choice of the maximum degree cutoff, bear an enormous influence on the dependence of MDS size on γ, even when the average degree $\langle k \rangle$ is kept fixed. The latter constraint is motivated by the need of comparing networks (from the MDS perspective) with the same amount of "resources," i.e., fixed average edge-cost per node, or equivalently, fixed average degree in unweighted networks. Naturally, for γ<2, and fixed the average degree, there is only a finite (but large) parameter range in terms of γ and N with realizable networks. Nevertheless, it is interesting to investigate how easy (or hard) it is to dominate networks from various ensembles in this regime, motivated by the existence of such real-world

networks [229, 230]. On the other hand, when keeping the minimum degree fixed in this regime, the number of edges increase faster than the number of nodes, and those networks are becoming inherently easy to dominate. For $\gamma>2$, keeping the average or the minimum degree fixed are equivalent constraints in the large network size limit.

In practice, one could imagine a scenario where rather than dominating all nodes of the network, it is sufficient to dominate some (large) fraction of nodes. This reduces to the problem of finding a *partial* minimum dominating set (pMDS) [231] which is the smallest subset of nodes (and possibly a subset of the full MDS) such that at least some given fraction of the nodes are either in the set or adjacent to a node in this set. We investigate the scaling of both MDS and pMDS with respect to the network size N.

5.1 Scaling with Network Size

5.1.1 Empirical Results

We start with a simple but thorough empirical analysis of MDS size using our scale-free network generating methods (CONF, DKTB, and HHMC, see Section 4.2), and the greedy MDS approximation algorithm (detailed in Section 4.3). We run a sweep of network size N at various power-law exponent γ values, generating hundreds of network realizations for each, and averaging the MDS size among them for each parameter combination.

Analysis of MDS via ensemble averages is justified by our observation that the MDS size obtained by the greedy algorithm for any single network follows a Gaussian distribution, with at least an order of magnitude smaller standard deviation than the standard deviation of MDS size among multiple network samples, which also follows a Gaussian (see Appendix C.1 for details). Therefore, for any given network, we find it sufficient to run the greedy algorithm five times to obtain a reliable estimate of the average MDS size; then, these are averaged over all network realizations (of the same type) to obtain an estimate of the mean greedily approximated MDS size.

We use two possible subclasses of each network class, according to the maximum degree cutoff $k_{\max}$. Either there is no explicit cutoff, having $k_{\max} = N-1$ (where N is the network size), or we use a structural cutoff $k_{\max} = \sqrt{N}$, resulting in uncorrelated

scale-free networks [203, 204]. When we have the $\sqrt{N}$ cutoff, we indicate it in the name of network type as cCONF, cDKTB, or cHHMC, where c stands for *cutoff*.

As indicated in the results and in figures, the average degree of each individual network is kept fixed at a predetermined value throughout all samples of each dataset. Details on controlling the average degree are included in Section 4.4.

Figures 5.1(a)–(c) show the MDS size for networks without any explicit upper cutoff on degree. For CONF networks, the MDS size scales linearly with N for all γ values. In striking contrast, DKTB networks and HHMC networks show a marked transition in the scaling behavior at $\gamma_c \approx 1.9$, which we call *domination transition*. For $\gamma > \gamma_c$, MDS size scales linearly with N, whereas for $\gamma < \gamma_c$, the MDS size appears to lose its dependence on N in the asymptotic limit. Figures 5.1(a)–(c) focus on a subset of all considered γ values, which range from $\gamma = 1.6$ to 3.0, to clearly show the scaling transition for DKTB and HHMC networks.

Figures 5.1(d)–(f) show in contrast that with the structural degree cutoff, ($k_{\max} = \sqrt{N}$), for all network classes the MDS size scales linearly with N irrespective of the γ value. Note also, that in this case the results for all three network types are quantitatively identical.

5.1.2 Analytic Lower Bound on MDS Size

To better understand the domination transition, we can derive a lower bound for the MDS size. For this, we consider a "best case scenario" for dominating the network, that is, under the best circumstances (detailed below), we count how many nodes are necessary at least to dominate all nodes in the network. In our calculation we use the continuous probability density function $f_K(k)$ equivalent to the discrete degree distribution, because using integrals instead of discrete sums simplifies calculation, and the distinction disappears in the infinite network size limit.

For the "best case" domination we assume that neighborhoods of all high degree nodes are disjoint sets (not overlapping), therefore each node with degree k dominates $k + 1$ nodes (itself, and its neighbors). Having no overlaps means that each dominator can be maximally efficient in covering (i.e., dominating) the largest possible fraction of the network. To cover all nodes, we start picking dominators

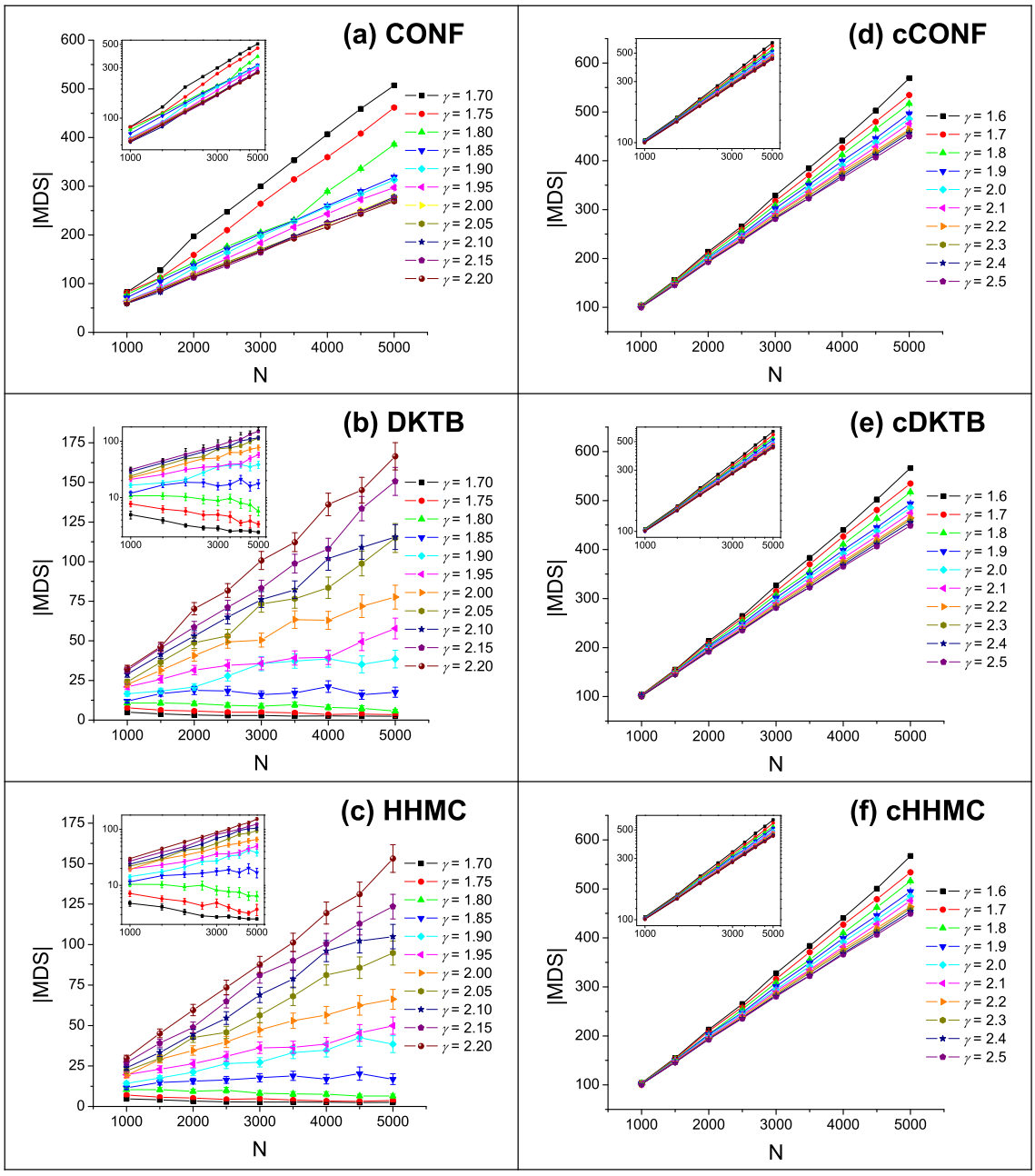

Figure 5.1: The size of MDS scaling with N, $\langle k \rangle = 14$, for all network types, averaged over 400 network realizations with 5 greedy searches for each at every data point. The figure insets show the same data on log-log scales. Error bars show 95% confidence for all data points (however, they may be very small and hidden by the larger symbols).

in degree-ranked order (starting with the largest degree), and keep selecting them until their theoretically maximal coverage includes all nodes of the network. The number of nodes we have to select this way is the lower bound, because in reality, node neighborhoods do overlap, leading to less efficient covering, thus requiring more dominators to cover the entire network.

In order to find the lower bound, we need to find the degree threshold first, above which all nodes must be selected in the theoretically best case to dominate all nodes. This threshold is denoted by k^* and expressed as:

$$k^* := \max\left\{k' : \int_{k'}^{k_{\max}} N(k+1) f_K(k)\,\mathrm{d}k \geq N\right\}. \tag{5.1}$$

For any given degree threshold k', the total number of nodes we expect to find, given a power-law degree distribution, is:

$$l(k') = \int_{k'}^{k_{\max}} N f_K(k)\,\mathrm{d}k. \tag{5.2}$$

Therefore, $l(k^*)$ gives the lower bound for the size of MDS. Note, that these formulae can be used with any degree distribution, and k^* can always be found numerically. Figure 5.2 shows the $l(k^*)$ bounds for power-law distributions as a function of N with $\langle k\rangle = 10$.

There are multiple consequences of the lower bound's scaling. For $k_{\max} = N-1$ networks, the possibility of an $O(N)$-to-$O(1)$ transition of MDS size is supported by $l(k^*)$ [Fig. 5.2]: it exhibits an $O(1)$ behavior for $\gamma < 2$, while it progresses to a linear scaling with N for $\gamma > 2$ [Fig. 5.2(a)], similar to the results of DKTB and HHMC networks. For networks with $k_{\max} = \sqrt{N}$, $l(k^*) \sim \sqrt{N}$ when $\gamma < 2$ and $l(k^*) \sim N$ when $\gamma > 2$. Note, however, that the convergence to the asymptotic behavior is extremely slow for $2 < \gamma < 3$ [see insets of Figs. 5.2(a) and (b)]. Thus, for the case of structural cutoff, the lower bound indicates that the MDS size can never become $O(1)$ and we cannot expect a sharp scaling transition. Derivation of the asymptotic behavior of $l(k^*)$ is included in Appendix C.2.

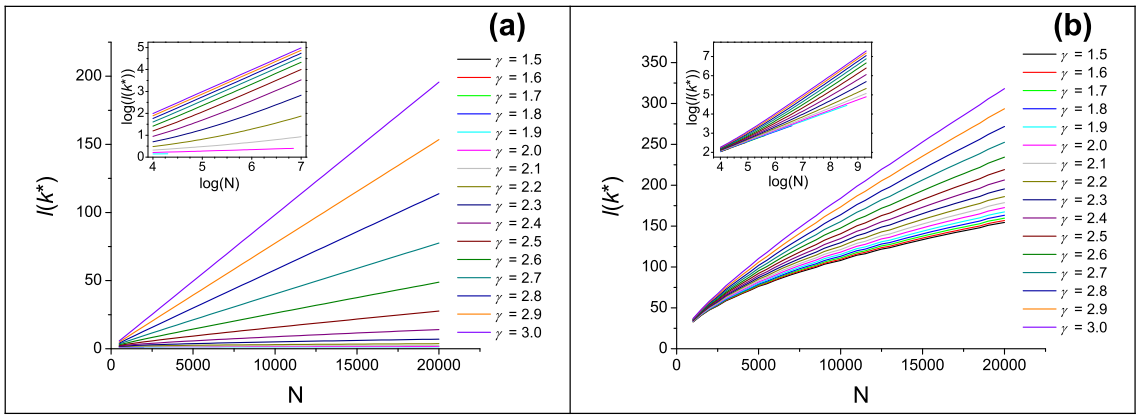

Figure 5.2: The scaling of the calculated lower bound of MDS size in power-law distributions, for various power-law exponents, $\langle k \rangle = 10$. (a) $k_{\max} = N - 1$, (b) $k_{\max} = \sqrt{N}$. Figure insets show $l(k^*)$ bounds on log-log scales. See Appendix C.2 for details.

5.1.3 Scaling of Partial Dominating Sets

Next, we study the scaling behavior of the partial MDS size with N as we vary the value of the required dominated fraction z. We present results for the DKTB class of networks in Fig. 5.3; our findings are qualitatively similar for CONF and HHMC network classes, and networks with $k_{\max} = \sqrt{N}$, as shown in Appendix C.3.

Below a certain value of $z = \widehat{k}_{\max}/N$, where $\widehat{k}_{\max}$ is the highest realized degree in the network, the pMDS trivially contains only the node with highest degree. Apart from these trivial cases, for $z > \widehat{k}_{\max}/N$, the size of the pMDS exhibits the same scaling as the full MDS in the different γ regimes (Fig. C.3). In other words, DKTB and HHMC networks display a transition in the scaling behavior of pMDS size from linear dependence on N to virtually no dependence on N at $\gamma \approx 1.9$, while CONF networks always show a linear scaling of pMDS size with N.

For a baseline-comparison to the partial MDS size obtained by greedy algorithm, we also study the expected number of nodes needed to dominate a given fraction of the network using *random node selection*, giving a partial *random* dominating set (pRDS). We run the random search five times on each realization to obtain an expected RDS size. Figure 5.4 compares pRDS with pMDS, showing that a simple random node selection gives approximately an order of magnitude larger dominating

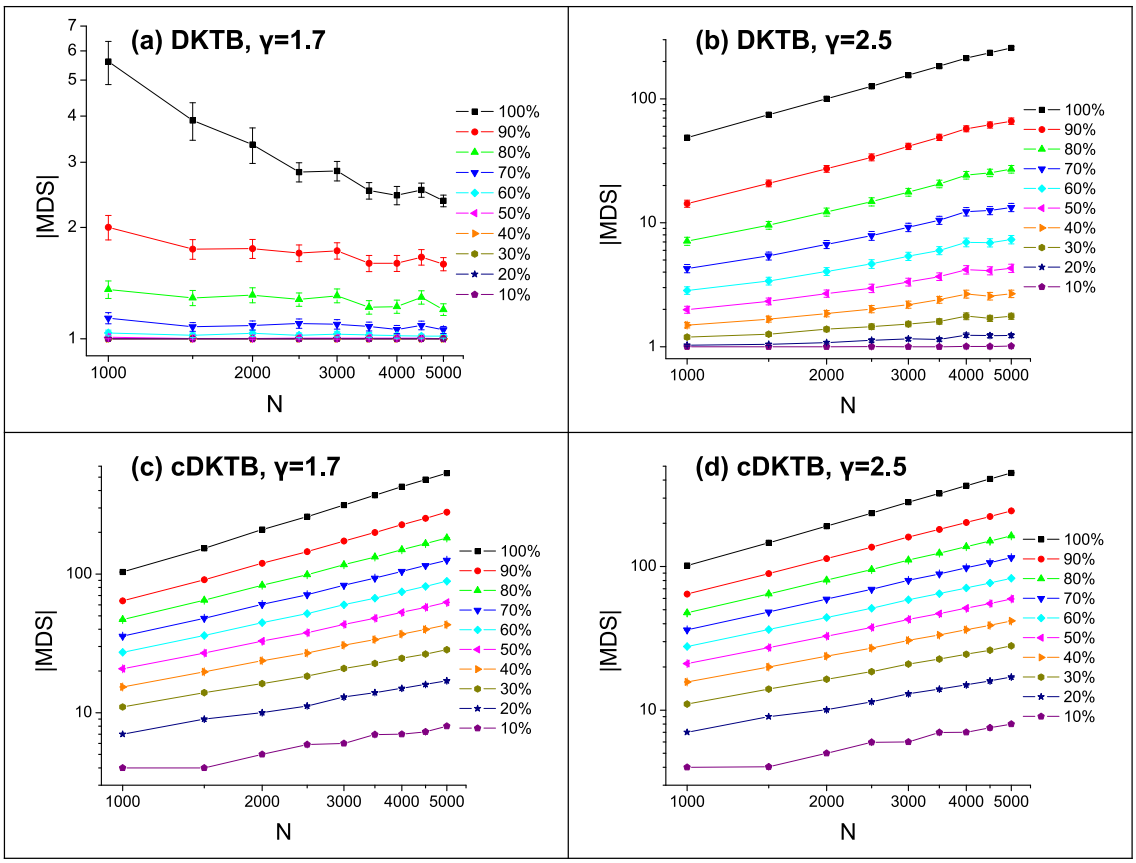

Figure 5.3: The size of partial MDS scaling with N, $\langle k\rangle = 14$, averaged over 400 samples, (a) DKTB, $\gamma = 1.7$, (b) DKTB, $\gamma = 2.5$, (c) cDKTB, $\gamma = 1.7$, (d) cDKTB, $\gamma = 2.5$. The dominated fraction of nodes is expressed as percentile of the network size.

set than the greedy method. Note also, that in order to reach full domination using random node selection, we would need to include almost all nodes of the network in the dominating set. Further, for reference, in Fig. 5.4 we also show the known upper bound, obtained for *optimized random* selection of the dominating set (oRDS) [88] but also applicable to the greedy algorithm [88, 89, 222], for a graph with minimum degree $k_{\min}$: $|oRDS| \leq N[1 - k_{\min}(1 + k_{\min})^{-1-1/k_{\min}}]$. Note, that in our network construction schemes with *fixed* average degree, $k_{\min} = k_{\min}(\gamma, \langle k\rangle, N)$, hence the small jumps in the above bound when plotted as a function of N for fixed γ and $\langle k\rangle$. The properties of random dominating sets are further studied in Chapter 6.

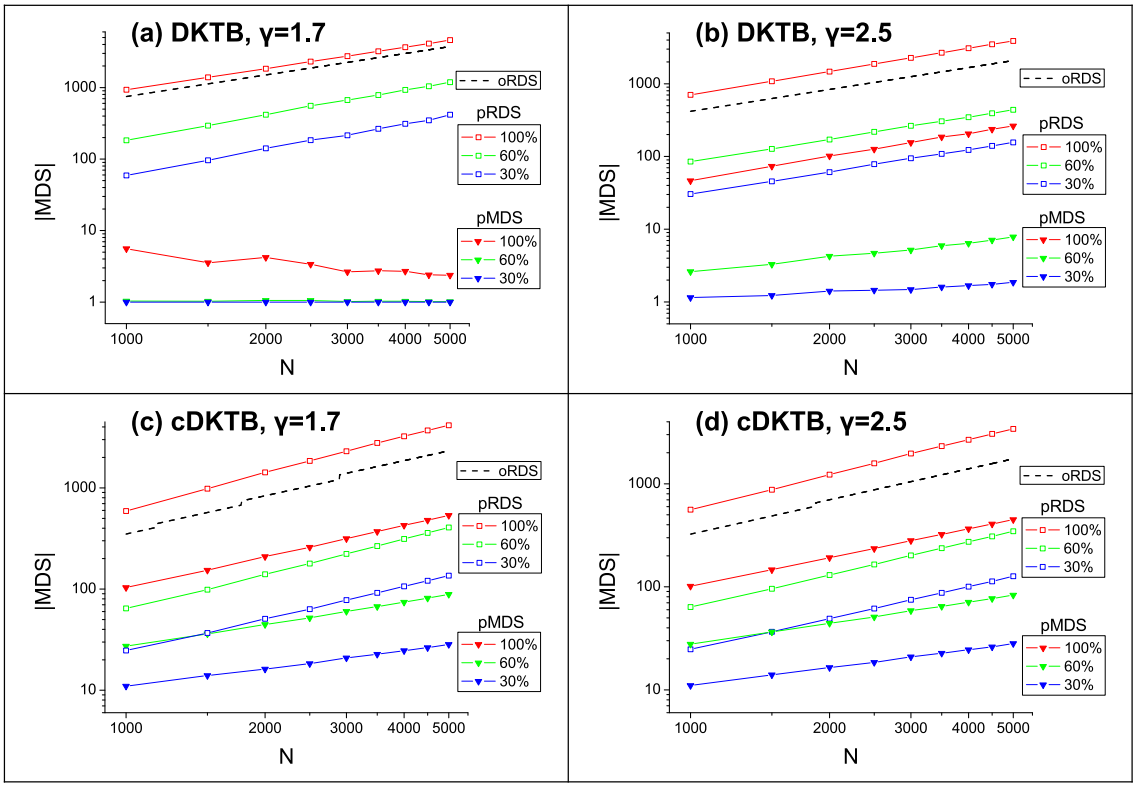

Figure 5.4: Comparison of partial MDS and partial RDS scaling with N for DKTB (a,b) and cDKTB (c,d) networks (without and with structural cutoff, respectively); $\langle k \rangle = 14$, averaged over 400 samples. The dominated fraction of nodes is expressed as a percentage of the network size. For reference, we also show the upper bound (dashed lines) for an optimized random dominating set (oRDS) [88, 222] (see text).

5.2 Scaling with Power-Law Degree Exponent

To measure the dependence of MDS size on γ, we find the MDS for a fixed network size of $N = 5000$ nodes. Results for various $\langle k \rangle$ values are presented in Fig. 5.5(a) for networks with no structural cutoff, and in Fig. 5.5(b) for networks with a structural cutoff.

We find a surprising trend in several cases. Perhaps, most intriguing is the trend seen in the case of CONF networks where the MDS size appears to have a non-monotonic dependence on γ. Traversing increasing γ values on a coarse scale, the MDS size starts out large at low γ, reaches a global minimum in the range $1.9 < \gamma < 2.3$, and then grows again as γ increases. However, generating network

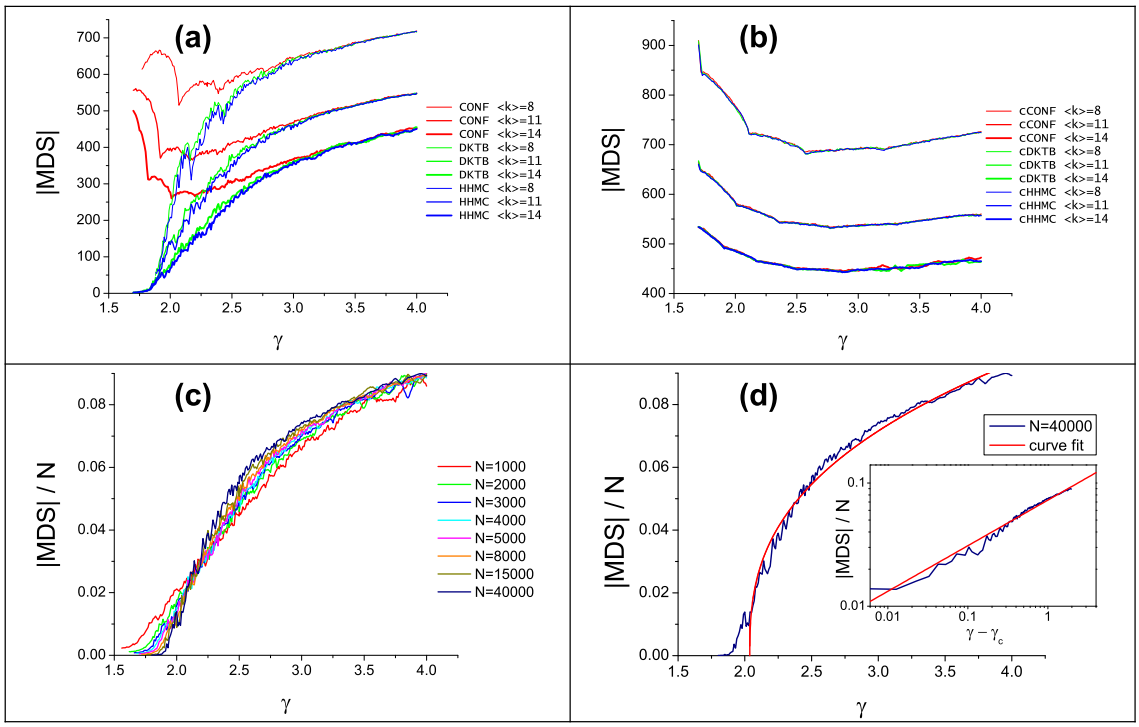

Figure 5.5: The size of MDS as a function of γ for various network types and average degrees, $N = 5000$, averaged over 200 network realizations with 5 greedy searches for each at every data point. (a) for networks with no degree cutoff; (b) for networks with the structural cutoff. (c) shows the scaled MDS size vs. γ for HHMC networks with $\langle k \rangle = 14$ for various system sizes. (d) Scaled MDS size for the largest network and the best-fit power-law (solid red curve) in the vicinity of (and above) the transition point, $|MDS|/N \propto (\gamma - \gamma_c)^{\beta}$ with $\beta \approx 0.37$. Inset: same data as in (d) after shifting the horizontal axis and on log-log scales.

samples with a finer resolution of γ values ($\Delta\gamma = 0.01$, reaching the resolution of error between desired and achieved γ values), we also notice the existence of kinks in addition to the large scale non-monotonicity.

By probing the dependence of MDS size on γ for DKTB and HHMC networks at fine resolution similar to one used for CONF, we find only minor traces of kinks, but they are within the error margin of MDS size. On a coarse scale, we find quantitatively similar results for both network classes. The MDS size curve is flat at very low values of γ, and then increases steadily beyond $\gamma \approx 1.8$. When $\gamma > 3$,

the MDS size of all three network types converge to the same value, indicating that beyond this point the structure of these networks are very similar [Fig. 5.5(a)].

The dependence of MDS size on γ is strikingly different for networks with a structural cutoff [Fig. 5.5(b)]. In this case, all three network classes show identical results for given network parameters. For increasing γ values the size of MDS first decreases, then reaches its minimum at approximately $2.5 < \gamma < 3.0$, and increases again when $\gamma > 3$. Since all three classes display almost indistinguishable results, we can infer that these networks are structurally very similar. Kinks like those seen in CONF networks are also observed here, but with a much smaller amplitude.

In the vicinity of (and above) the transition point, we also found that for sufficiently large DKTB and HHMC networks [Fig. 5.5(c)], the scaled MDS size can be reasonably well approximated with a power-law,

$$\frac{|MDS|}{N} = \text{const.} \cdot (\gamma - \gamma_c)^{\beta} \tag{5.3}$$

with $\beta \approx 0.37$ [Fig. 5.5(d)].

5.3 Discussion and Conclusions

As demonstrated clearly by the results, the specific method used for generating a network ensemble has a profound influence on the MDS size. This suggests that there are distinctive features in the structures of networks generated by the different classes. From the details of the generation methods, it might appear that DKTB and HHMC classes produce networks that are similar in structure, and this is certainly corroborated by our results. However, their distinction from networks in the CONF class seems to disappear when a structural cutoff is introduced in the degree distribution. Although we cannot rigorously demonstrate that particular structural features are responsible for the observed scaling behavior, we conjecture on the origin of the distinct behaviors.

It should be noted that Del Genio *et al.* [232] have shown the non-existence of realizable graphs with a power-law degree sequence with $0 \leq \gamma \leq 2$. However, as they point out, their arguments refer to the situation where the prescribed degree sequence

has to be perfectly satisfied. In our methods, networks with $1 \leq \gamma \leq 2$ are realized by removing some edge stubs from the degree sequence that cause non-graphicality. For CONF networks, pruning of multiple links and self loops perform this task, while for DKTB and HHMC networks, a Havel-Hakimi-based graphicality correction algorithm is applied (see Section 4.2 and Appendix B.2). It is notable, that we can choose appropriate parameters such that even after pruning, or graphicality-preserving stub removal, we have a network whose degree distribution approximately fits the desired power-law. As a result of these procedures, our networks in this range of γ are not exact realizations of perfect power-law degree sequences, and therefore do not contradict the fundamental results of [232].

When the structural cutoff is not imposed on the degree sequence, the non-graphicality below $\gamma = 2$ plays an important role in the scaling transition of the MDS size with N. When $\gamma < 2$, there are too many edge stubs in the prescribed degree sequence, and some of them have to be removed to resolve non-graphicality. Different network construction methods solve this problem differently. With respect to MDS scaling behavior, the key difference is in the treatment of the highest degrees. In case of CONF, the formation of multiple links is allowed during the stub connection process, and the duplicate links are pruned later. Since the realized multiple links predominantly connect stubs belonging to high-degree nodes with each other [203, 233], the large degrees of the hubs are effectively wasted in connections that do not improve their potential to dominate. Furthermore, as a consequence, the interconnections of low degree nodes become more dominant, forming a relatively sparse web, which necessitates the inclusion of many nodes in the dominating set, preventing it to become $O(1)$. However, in case of the Havel-Hakimi-based graphicality correction (used in HHMC and DKTB methods) the stubs of the highest degree nodes are connected first, ensuring that these nodes are present in the network as hubs. The MDS scaling transition can therefore be explained by the scaling of the maximum realized degree (also known as the natural cutoff of the degree distribution), $\widehat{k}_{\max} \sim N^{\frac{1}{\gamma-1}}$ [203, 234]. When $\gamma < 2$, the MDS size becomes $O(1)$ because the largest degree, and potentially the second and third largest degrees become $O(N)$, and the network is dominated by these nodes. In essence, we find that the domination

transition is directly related to the underlying graphicality transition [232]: the same underlying structural properties which are responsible for the graphicality transition in the infinite network-size limit allow for the $O(N)$-to-$O(1)$ domination transition for large but finite DKTB and HHMC networks. In other words, those finite DKTB and HHMC network realizations which happen to exist for $\gamma < 2$ can be dominated by an $O(1)$ MDS.

The small difference between our numerically observed value of the domination transition at around $\gamma \approx 1.9$ and the $\gamma = 2$ value of the graphicality transition [232] might lie in finite-size effects and in the $\log(N)$ accuracy of the greedy algorithm with respect to the size of the true MDS [Figs. 5.5(c) and (d)].

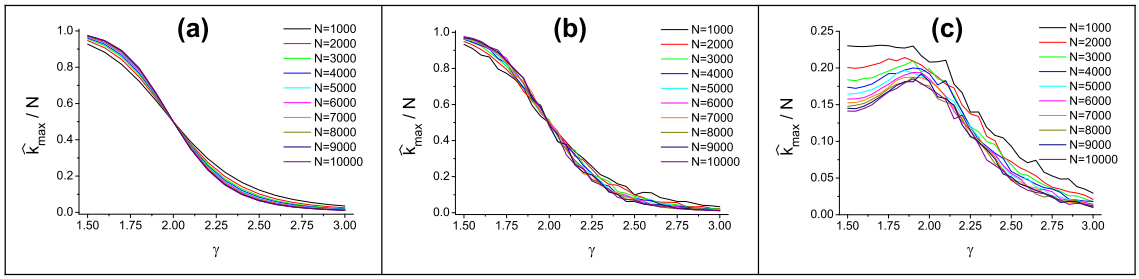

Figure 5.6: Scaling of maximum realized degree $\widehat{k}_{\max}$ with power-law exponent γ, for various network sizes. (a) theoretical expected value from power-law distribution, (b) degree sequence with graphicality correction (HHMC and DKTB networks), (c) degree sequence after pruning multiple links (CONF networks).

The different treatment of largest degrees in different network classes can be illustrated by plotting $\widehat{k}_{\max}$ against γ, see Fig. 5.6. Note, that for the theoretical value we need to derive and evaluate the exact formula from the degree distribution, see Appendix C.4. Further, the markedly different structure of CONF networks compared to HHMC and DKTB networks in the absence of a structural cutoff also shows up in the network visualizations in Fig. 5.7.

In contrast, with the structural cutoff, networks generated using the three different methods appear to share similar structure as can be seen in Fig. 5.8, from the similar scaling of MDS size with N, and the dependence of MDS size on γ. The restrictive $k_{\max} = \sqrt{N}$ cutoff precludes the scaling of MDS size from becoming $O(1)$, as shown by the $l(k^*)$ lower bound in Fig. 5.2(b).

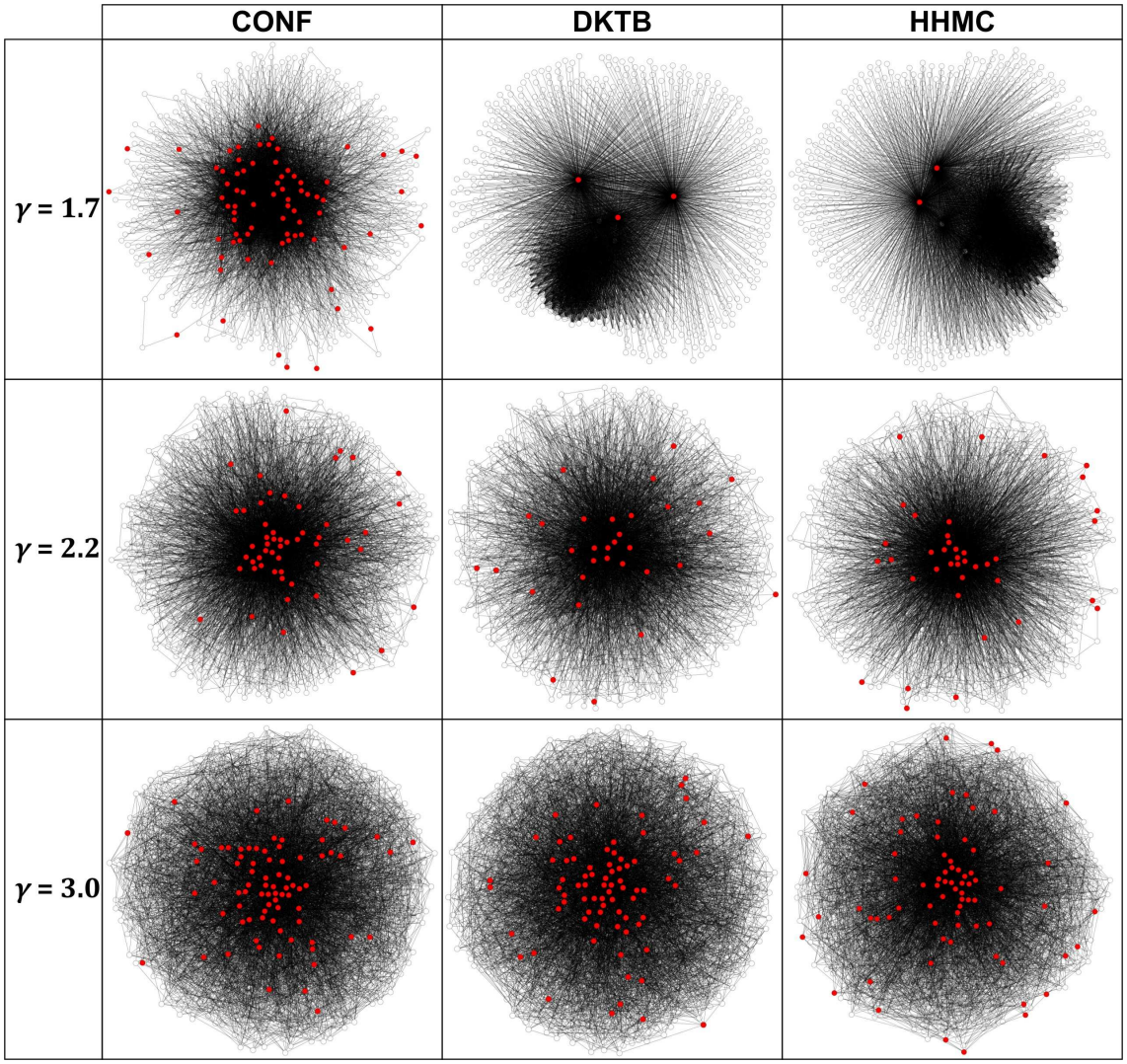

Figure 5.7: Visualization of typical scale-free networks of each type with $k_{\max} = N-1$ at three different power-law exponent values, embedded by the SFDP layout engine of Graphviz visualization software [235]. In all networks, $N = 1000$ and $\langle k \rangle = 14$; the colored nodes belong to the MDS.

The non-monotonic behavior of the size of the MDS with γ (with the exception of the DKTB and HHMC constructions with no maximum degree cutoff) are in part the consequence of the stringent constraint of resources for domination (fixed average degree, i.e., fixed number of edges for fixed N): for small and decreasing values of γ, while maintaining a fixed average degree for a given network size N, the minimum degree decreases, and there are $O(N)$ number of such nodes. However, in

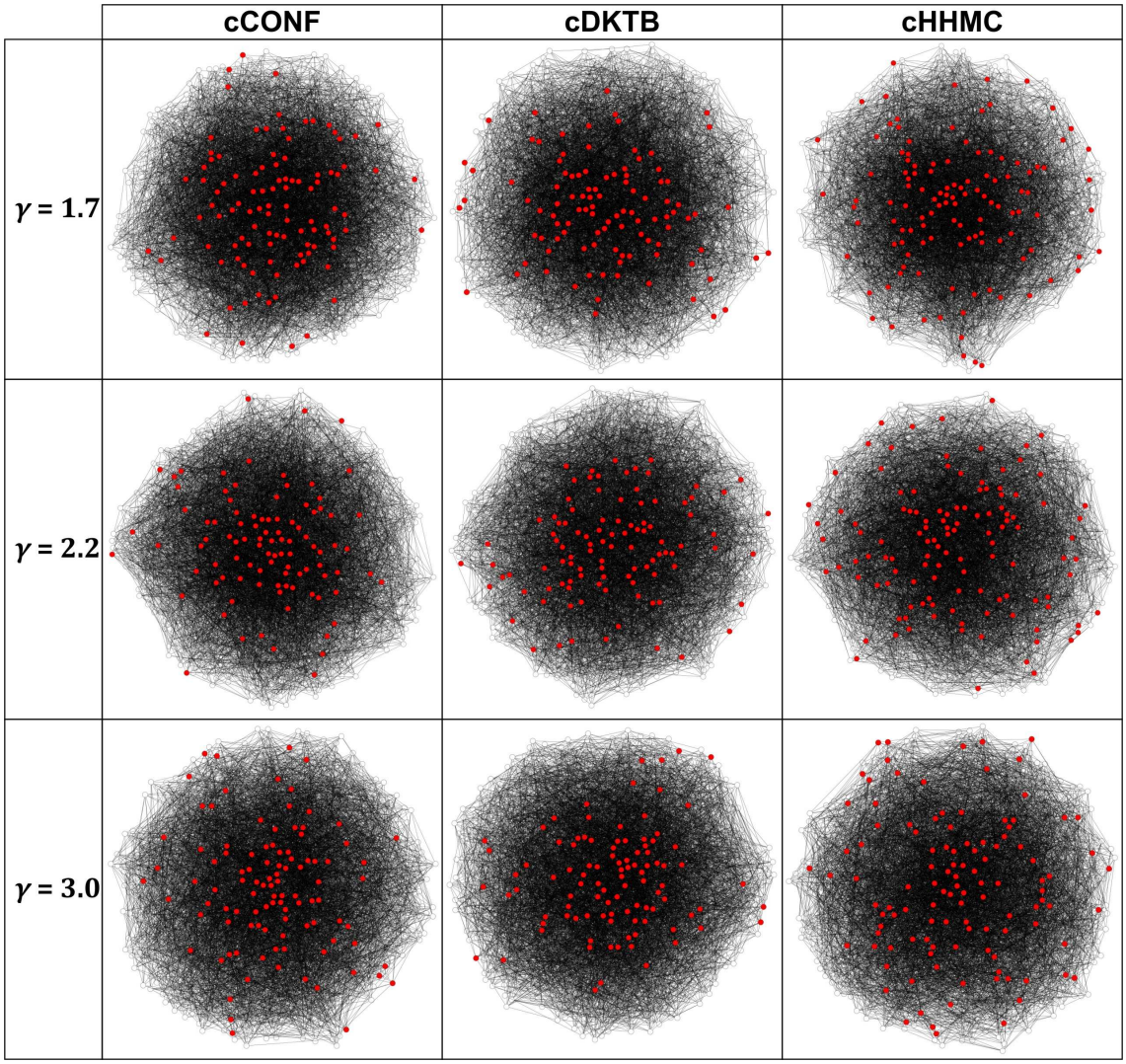

Figure 5.8: Visualization of typical scale-free networks of each type with $k_{\max} = \sqrt{N}$ at three different power-law exponent values, embedded by the SFDP layout engine of Graphviz visualization software [235]. In all networks, $N = 1000$ and $\langle k \rangle = 14$; the colored nodes belong to the MDS.

the DKTB and HHMC networks the largest hub can have $O(N)$ links, and it has the potential alone to connect to (and dominate) the nodes with the lowest degree, hence the monotonic behavior with γ (and the transition to $O(1)$ domination) for these networks [Fig. 5.5(a)].

Kinks seen in the curves of MDS size when plotted against γ are the result of controlling the average degree with very high precision. Smooth change of the control

parameters introduces gradual changes in the network structure, however, the average degree does not change smoothly (although it is monotonic; see Fig. 4.5 in Section 4.4). Conversely, when we probe a range of γ values, we need a smooth control over the average degree, requiring non-smooth changes in control parameters and hence in the network structure. Therefore, we can expect that any structure-dependent quantity, like the MDS size, will show kinks with respect to γ.

The results reported by Nacher *et al.* [92] suggest that for a given $\langle k \rangle$, decreasing γ results in a monotonic lowering of the MDS size. However, they only studied the HHMC method of network generation with a variable cutoff. By introducing a well-defined structural cutoff, and in addition studying two other classes of networks, we show that precise details of the network construction have a strong impact on the trend in MDS size as γ is varied.

In summary, we have shown through extensive numerical experiments, that the size of the minimum dominating set approximated by a greedy algorithm undergoes a transition in its scaling with respect to N only for particular methods of network construction in the absence of a structural cutoff. For the configuration model construction, or the other construction methods with a structural cutoff, no such transition is observed. However, intriguingly, in the presence of a structural cutoff, the MDS size increases as γ is lowered below 2. Concerning real-world scale-free networks, we conclude that it is not sufficient to have $\gamma < 2$ to have an easily dominated (easily controllable) network; intricate details in the wiring of the network must also be taken into consideration; in particular, the maximum realized node degrees.

CHAPTER 6
DOMINATING SCALE-FREE NETWORKS USING GENERALIZED PROBABILISTIC METHODS

In this chapter, we consider the problem of finding dominating sets with the additional factor of local connectivity information availability (or lack thereof) that affects the cost of selecting dominators. Most existing dominating set search methods require full knowledge of network structure and connectivity patterns (i.e., adjacency matrix, or equivalent adjacency information). This information may not be available in large networks (over tens of millions of nodes and edges), and obtaining this information involves additional expenses (in addition to actually running the dominating set search method) that can ultimately lead to overall suboptimal costs. Moreover, sophisticated search methods tend to have polynomial computational time complexity with high orders in the number of nodes or edges, therefore their applicability to large real networks is questionable. Our goal is to develop dominating set selection strategies that satisfy the cost-efficiency demands in terms of required connectivity information, computational complexity, and the size of the resulting dominating set. We focus on scale-free networks, as most large networks with potential applications of MDS (e.g., social, infrastructural, and communication networks) fall into this category. Without full adjacency information, we must select nodes based solely on their individual (local) properties, such as the node degree, and potentially a limited amount of global network information, such as the number of nodes and edges, average degree, and power-law degree exponent.

6.1 Probabilistic (Random) Dominating Sets (RDS)

The results of Alon and Spencer [88] provide a graph-theoretical approach to find an upper bound for the MDS size, and as part of it they propose a probabilistic

Portions of this chapter previously appeared as: F. Molnár Jr., N. Derzsy, É. Czabarka, L. Székely, B. K. Szymanski, and G. Korniss, "Dominating Scale-Free Networks Using Generalized Probabilistic Methods," Sci. Rep. **4,** 6308 (2014).

method for selecting dominator nodes. While their approach is theoretical, we can carry out their method, numerically, to obtain a probabilistic dominating set, and study its properties in scale-free networks.

Finding a probabilistic (random) dominating set (RDS) in a graph has the following steps. First, we visit each node, and add it to an initially empty set X, with probability p (a parameter chosen arbitrarily, $p \in [0, 1]$), independently of other nodes. Then, the remaining nodes that are not in X nor adjacent to any node in X are placed in set Y. The dominating set is obtained by $X \cup Y$. Alon and Spencer showed [88] that the expected size of this set is

$$|RDS| = |X| + |Y| \leq Np + N(1-p)^{k_{\min}+1} , \tag{6.1}$$

where $k_{\min}$ is the minimum degree and N is the number of nodes in the graph. By differentiation of this formula with respect to p we can find the optimal p value that minimizes $|RDS|$ (the corresponding dominating set is denoted by oRDS; o stands for optimal), which is then further bounded from above [88]:

$$|MDS| \leq |oRDS| \leq N[1 - k_{\min}(1 + k_{\min})^{-1-1/k_{\min}}] . \tag{6.2}$$

Our numerical results on scale-free network samples in comparison with the analytical values are shown in Fig. 6.1 for a wide range of parameters ($2 \leq \gamma \leq 4$; $4 \leq \langle k \rangle \leq 16$). We find that our numerically obtained RDS size is significantly lower than the analytical one, for optimal p values. However, when $p \gtrsim 0.5$ the size of the RDS found numerically closely approaches the analytical curve. The difference we see can be understood easily: The upper bound in Eq. (6.1) assumes the worst case, that all nodes not dominated by the X set are nodes of the smallest degree. Thus, the difference between their bound and the numerical results shows the relative number of nodes whose neighbors are higher than minimum degree, yet not dominated. In scale-free networks, we indeed expect to find a significant number of low-degree nodes with high-degree neighbors (especially in disassortative networks), explaining our observations.

An alternative strategy to random node selection (without adjacency infor-

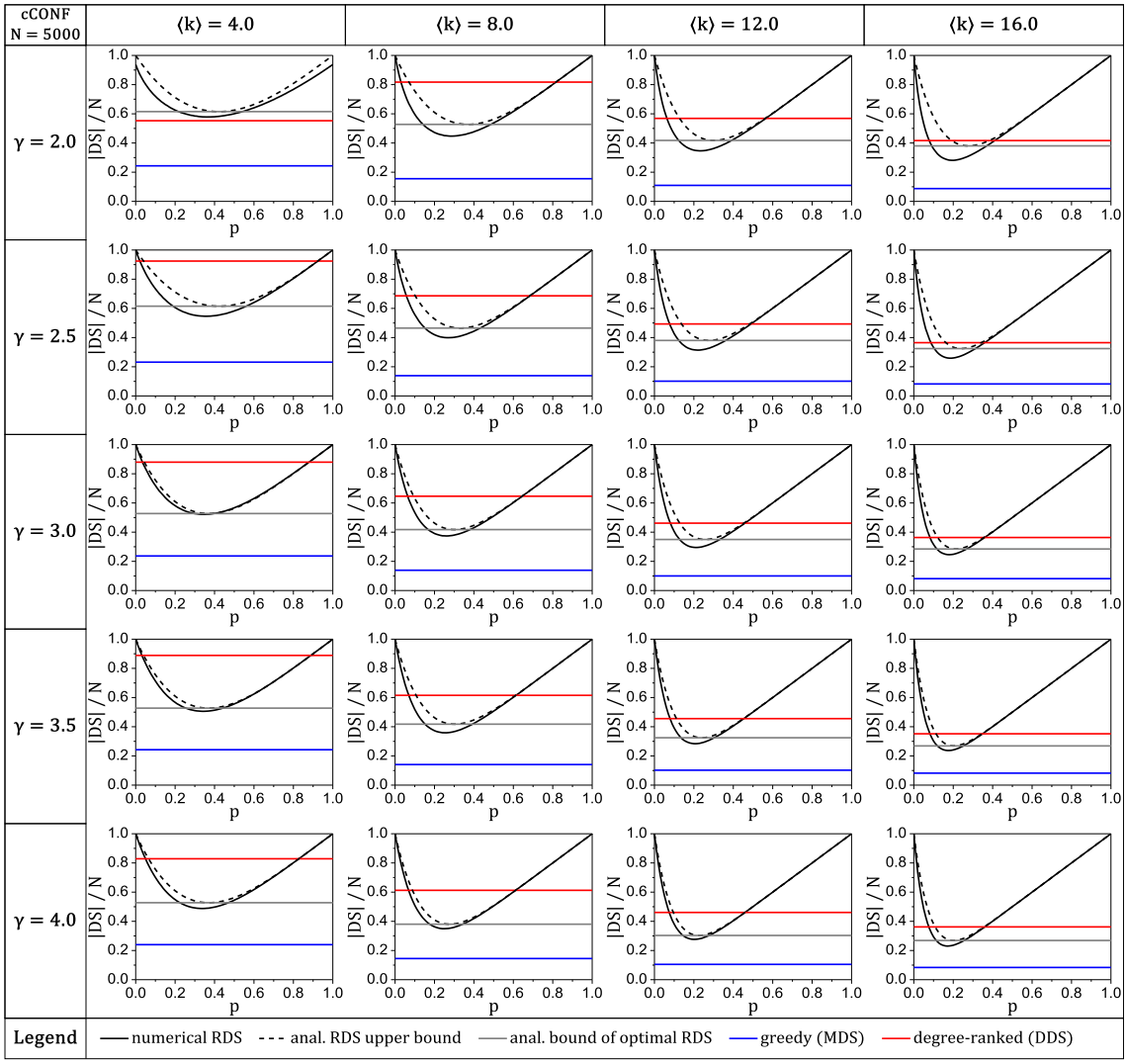

Figure 6.1: Comparison of the numerically computed probabilistic (random) dominating set to its analytical estimate, its analytical upper bound, the greedy minimum dominating set, and the degree-ranked dominating set, for various network parameters in cCONF networks. $N = 5000$**, averaged over** 100 **network samples.**

mation) is to pick nodes in degree-ranked order, until the selection dominates all nodes. This is called a degree-ranked dominating set (DDS). Although one may expect DDS to be a better strategy, simply because high degree nodes can dominate more neighbors, our results surprisingly show the opposite. As we compare the size of RDS to DDS in Fig. 6.1, we find that DDS is outperformed by RDS for optimally chosen p values, for every combination of network parameters.

Figure 6.1 also shows the MDS approximation provided by the greedy method, for comparison. Since the greedy method uses full adjacency information, it is not surprising that it provides much smaller dominating sets than RDS or DDS.

6.2 Degree-Dependent Random Dominating Sets

In order to improve the results of RDS we have to consider that complex networks are heterogeneous, and it would be beneficial to exploit this characteristic in the probabilistic node selection strategy. Although we have seen that using *only* high degree nodes for dominators (DDS) is a bad strategy, introducing a bias toward high degree nodes while keeping the random nature of node selection could prove beneficial. Therefore, we propose a novel degree-dependent probability function for selecting nodes that are placed in set X:

$$p_i = \min\left\{1, p\left(\frac{k_i}{k_{\max}}\right)^{\beta}\right\}, \tag{6.3}$$

where k_i is the degree of node i, $k_{\max}$ is the maximum degree in the network, and p and β are parameters. Note that we no longer require p to be a probability but rather a prefactor that can have any positive value. Similarly to the case of degree-independent selection probability, set Y contains nodes that are not dominated by X, and the ultimate result, RDS is obtained by $X \cup Y$. Note, that when $p > 1$, we can have $p_i = 1$, in which case node i is surely selected.

Figure 6.2 compares RDS with degree-dependent and degree-independent node selections for a wide range of β values (note, $\beta = 0$ is identical to the degree-independent case). In agreement with our expectations, our results clearly show that degree-dependent node selection provides a much smaller dominating set than the simple degree-independent selection, and thus it also outperforms DDS by far. We can also observe that as the β parameter is increased the smallest possible RDS size decreases, and it approaches the greedily approximated MDS size. Notice however, that for finding the smallest possible RDS the value of p has to increase as well.

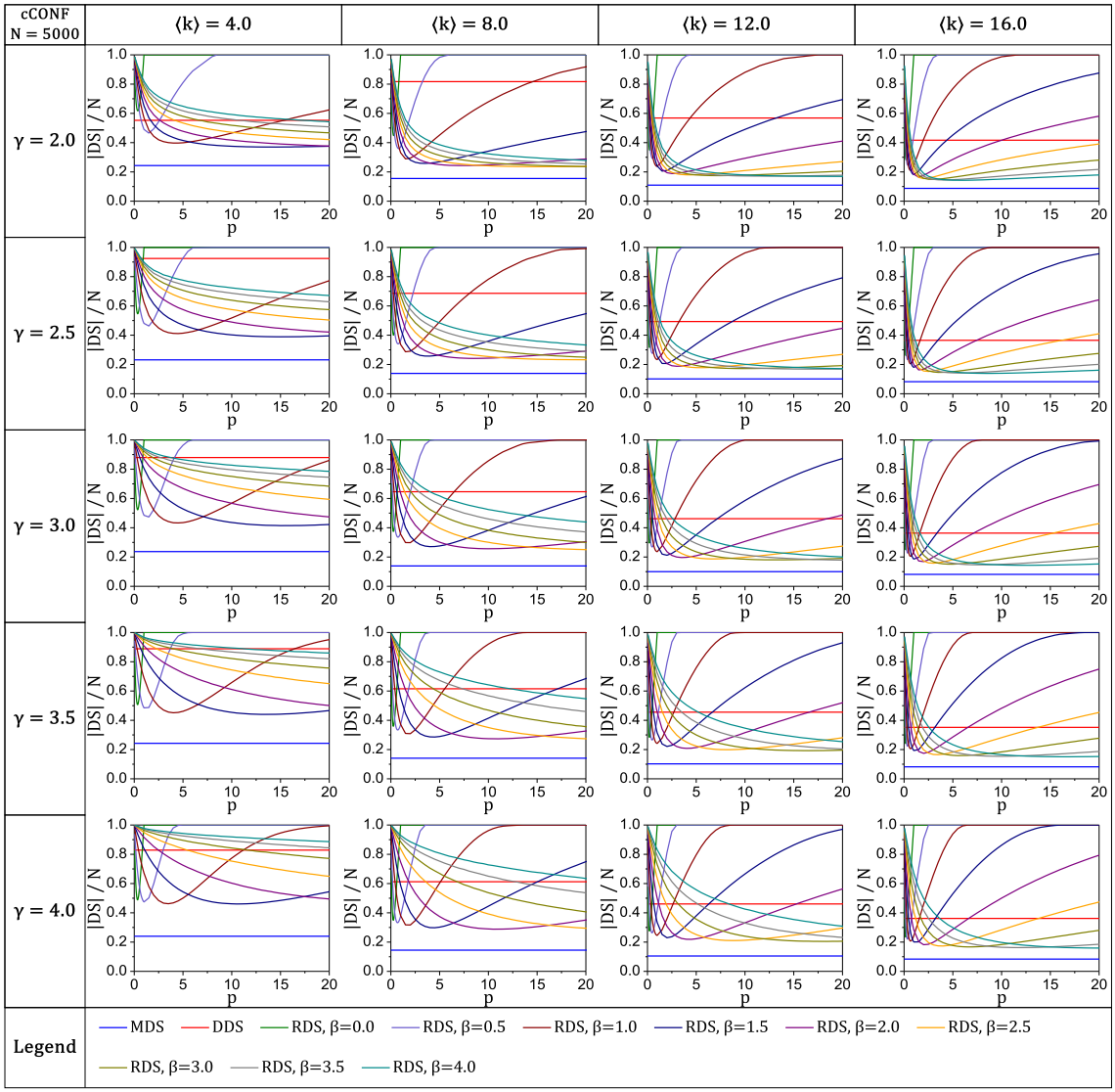

Figure 6.2: Size of random dominating sets (RDS) as a function of p prefactor in the degree-dependent node selection probability [Eq. (6.3)]. Data is averaged over 200 network samples and 20 repetitions of dominating set searches for each sample, in cCONF networks with $N = 5000$.

6.3 Cutoff Dominating Sets (CDS)

Since the smallest RDS size obtained seems to become lower for ever increasing β values, we expect to find the minimum with $\beta \to \infty$. Notice, that in this case all nodes with degree $k_i > k_{\max} p^{-1/\beta}$ are selected with probability 1 and nodes with smaller degrees are selected with probability 0. Thus, we have a degree threshold, $\kappa \equiv k_{\max} p^{-1/\beta}$ that now deterministically decides whether nodes will be added to set

X or not. We can use this κ to reparametrize the node selection probability in RDS as well: Eq. (6.3) now becomes

$$p_i = \min\left\{1, \left(\frac{k}{\kappa}\right)^{\beta}\right\} . \tag{6.4}$$

This form shows even more explicitly that the $\beta \to \infty$ case transforms the probabilistic selection into a deterministic one based on the κ degree cutoff. Therefore, we call the final result a *cutoff dominating set* (CDS).

Figure 6.3 shows CDS in comparison with RDS for various β values. We can see that CDS indeed provides the smallest dominating set size among probabilistic methods, and when κ is optimal (i.e., it minimizes the size of CDS) the size of CDS almost reaches the greedy MDS approximation. Considering how much simpler CDS is compared to the greedy approximation, this result is quite remarkable.

In order to further verify the performance of CDS, we calculate it on real-world network samples and compare it to RDS, as well as greedy MDS approximation and DDS. We use scale-free networks from the Stanford large network dataset collection [230], namely a snapshot of the peer-to-peer Gnutella network, and the web graph of the University of Notre Dame (domain nd.edu). Figures 6.4 and 6.5 show these results. In both cases, we see the same behavior of CDS as in synthetic networks: CDS reaches the smallest possible size of all probabilistic dominating sets, approaching the greedy MDS approximation for optimal κ values.

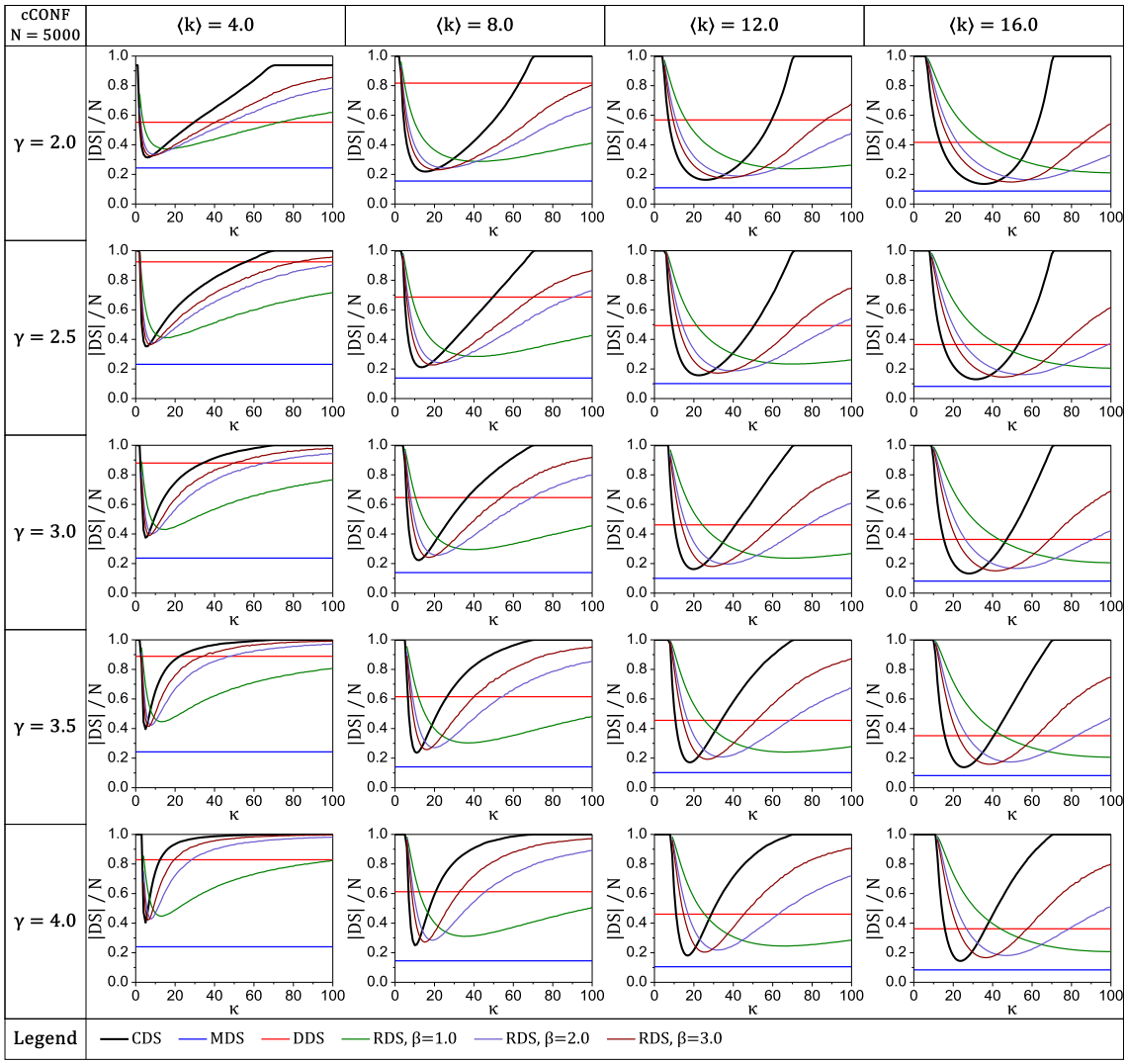

Figure 6.3: Comparison of the cutoff dominating set (CDS) to the greedy minimum dominating set (MDS), the degree-ranked dominating set (DDS), and degree-dependent random dominating sets (RDS), for various network parameters in cCONF networks. $N = 5000$, averaged over 100 network samples.

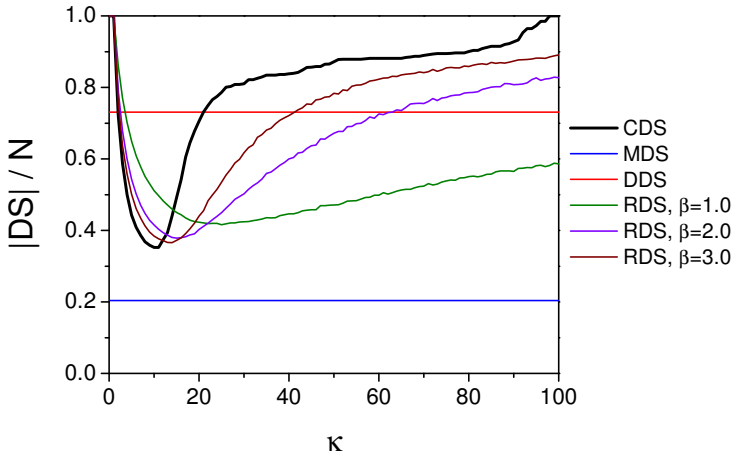

Figure 6.4: **Comparison of the cutoff dominating set (CDS) to the greedy minimum dominating set (MDS), the degree-ranked dominating set (DDS), and degree-dependent random dominating sets (RDS), in p2p-Gnutella08 [230] network.**

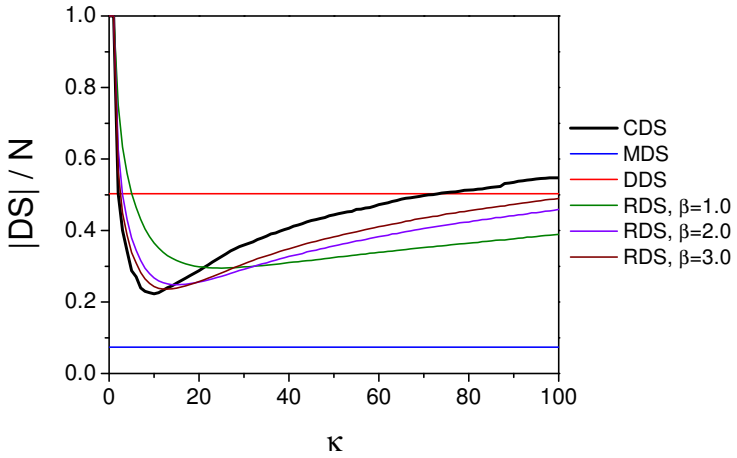

Figure 6.5: **Comparison of the cutoff dominating set (CDS) to the greedy minimum dominating set (MDS), the degree-ranked dominating set (DDS), and degree-dependent random dominating sets (RDS), in web-NotreDame [230] network.**

6.4 Analytical Estimates of RDS and CDS

Since both RDS and CDS require only the degree of each node to decide whether to place that node in the X set, we can estimate the size of RDS and CDS in the infinite network size limit using continuous degree distributions. In general, we can estimate the size of any probabilistic dominating set in a network with any degree distribution and degree correlations as follows:

$$\frac{\langle DS \rangle}{N} = \int_{k_{\min}}^{k_{\max}} X(k)P(k)\mathrm{d}k + \\ + \int_{k_{\min}}^{k_{\max}} (1 - X(k))P(k) \left[\int_{k_{\min}}^{k_{\max}} (1 - X(k'))P(k'|k)\mathrm{d}k' \right]^k \mathrm{d}k, \tag{6.5}$$

where $P(k)$ is the degree distribution on the domain of $[k_{\min}, k_{\max}]$, $X(k)$ is the probability of selecting a node with degree k into set X, $P(k'|k)$ is the degree distribution of the neighbors of a node with degree k. The first integral calculates the expectation of $|X|/N$, while the rest is the expectation of $|Y|/N$. The latter is obtained by counting the nodes that are not in X (the first part), but only those that also have no neighbors in X (the expression in square brackets).

We can plug in the properly normalized power-law degree distribution in $P(k)$. Further, for uncorrelated networks we have $P(k'|k) = k'P(k')/\langle k \rangle$. For RDS with uniform node selection probability we have $X(k) = p$, resulting in:

$$\frac{\langle RDS \rangle}{N} = p + \\ + \frac{(1-p)(1-\gamma)}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}} \left[k_{\min}^{1-\gamma} E_\gamma(-k_{\min} \log(1-p)) - k_{\max}^{1-\gamma} E_\gamma(-k_{\max} \log(1-p)) \right] \tag{6.6}$$

For RDS with degree-dependent probability we have $X(k) = \min(1, (k/\kappa)^\beta)$, resulting:

$$\frac{\langle RDS \rangle}{N} = \frac{k_{\max}^{1-\gamma} - \kappa^{1-\gamma} + (1-\gamma)\left[y_1 + \kappa^{-\beta}(x + y_2) \right]}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}}, \tag{6.7}$$

with

$$x = \frac{\kappa^{1+\beta-\gamma} - k_{\min}^{1+\beta-\gamma}}{1+\beta-\gamma} \tag{6.8}$$

$$y_1 = k_{\min}^{1-\gamma} E_\gamma(-k_{\min}\log a) - \kappa^{1-\gamma} E_\gamma(-\kappa \log a) \tag{6.9}$$

$$y_2 = \kappa^{1+\beta-\gamma} E_{\gamma-\beta}(-\kappa \log a) - k_{\min}^{1+\beta-\gamma} E_{\gamma-\beta}(-k_{\min}\log a) \tag{6.10}$$

$$a = \frac{\kappa^{2-\gamma} - k_{\min}^{2-\gamma}}{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}} - \frac{(2-\gamma)\kappa^{-\beta}}{2+\beta-\gamma}\left(\frac{\kappa^{2+\beta-\gamma} - k_{\min}^{2+\beta-\gamma}}{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}}\right). \tag{6.11}$$

Finally, for CDS we have $X(k) = \Theta(k-\kappa)$, where Θ is the Heaviside step function that returns 1 for positive arguments and 0 otherwise, yielding:

$$\frac{\langle CDS\rangle}{N} = \frac{k_{\max}^{1-\gamma} - \kappa^{1-\gamma} + (1-\gamma)[k_{\min}^{1-\gamma} E_\gamma(-k_{\min}\log b) - \kappa^{1-\gamma} E_\gamma(-\kappa \log b)]}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}}, \tag{6.12}$$

with

$$b = \frac{\kappa^{2-\gamma} - k_{\min}^{2-\gamma}}{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}}. \tag{6.13}$$

Note, that in all the above formulas, $E_n(z)$ denotes the exponential integral function, $E_n(z) = \int_1^\infty e^{-zt} t^{-n} \mathrm{d}t$. The detailed derivation of the analytical estimates can be found in Appendix D.1.

Figure 6.6 shows the accuracy of our analytical estimates in comparison with the numerical results of RDS and CDS. Further results on scale-free networks with different $\langle k\rangle$ and γ values are provided in Appendix D.1.4, showing that as $\langle k\rangle$ increases, the accuracy of the analytical estimates improves. For CDS and degree-independent RDS the estimates are very close to the numerically obtained values, even with a small $\langle k\rangle$. The estimates for degree-dependent RDS are slightly less accurate, but still sufficient to provide a useful approximation of the expected dominating set size. Therefore, we can easily calculate a very accurate expected size of these dominating sets in uncorrelated scale-free networks, based on nothing beyond basic network parameters.

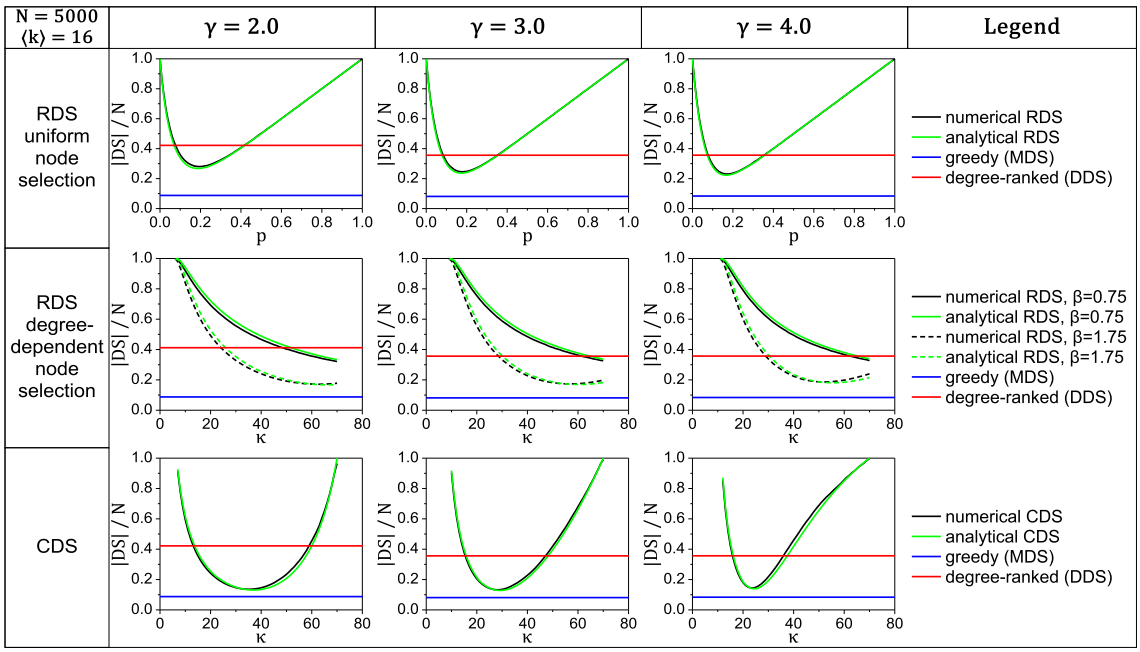

Figure 6.6: Comparison of analytical estimates and numerically computed sizes of RDS and CDS in uncorrelated (cCONF) scale-free networks. For numerical results, data is averaged over 200 network samples. Parameters: $N = 5000$ and $\langle k \rangle = 16$.

6.5 Effects of Network Assortativity

Using our edge-mixing method to control the assortativity of a network (see Section 4.5), we have compared the sizes of dominating sets as a function of assortativity, measured by Spearman's ρ. Figure 6.7 shows our results for a synthetic network and a real social network, while the same comparison for different network parameters is provided in Appendix D.2 for artificial networks, and in Appendix D.3 for real networks.

As expected, the size of most dominating sets increase with higher assortativity, except for RDS with degree-independent selection probability. The most dramatic size increase is observed in DDS, which indicates that this method can only be considered viable in real-world applications for highly disassortative networks. Also, as the assortativity increases, CDS becomes larger than the simple RDS at a certain point, indicating that favoring high-degree nodes as dominators is not an effective strategy when the network is highly assortative. While the MDS size obtained by greedy search also increases with increasing assortativity, it shows the smallest

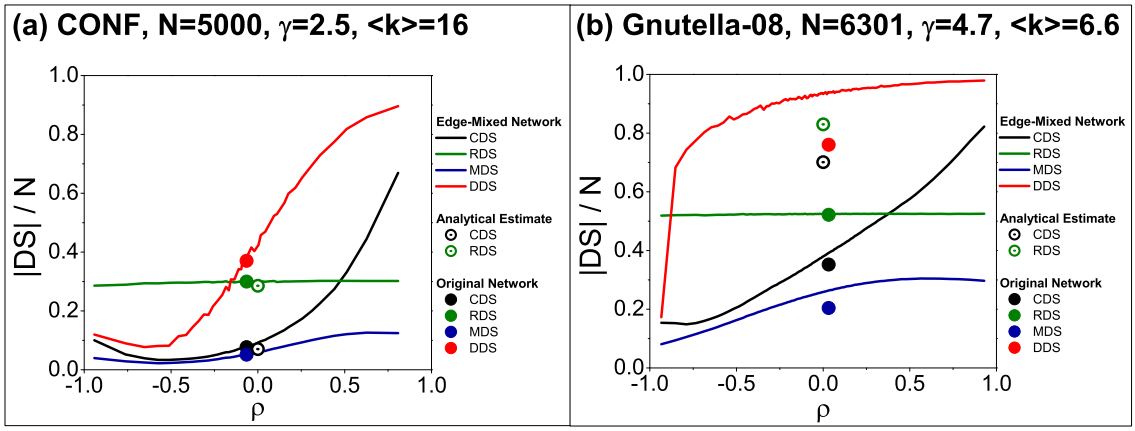

Figure 6.7: Dominating set sizes as a function of Spearman's ρ assortativity measure in (a) a synthetic network and (b) a real-world network (Gnutella08 [230]). Networks with assortativity values different from the original network are obtained by guided edge-mixing with double-edge swaps.

increase, thus the advantage of greedy search over other methods is more pronounced.

We also analyze the effects of assortativity on the *optimal* κ degree threshold value that minimizes the size of CDS. Figure 6.8 provides a complete dependence map of the optimal κ with respect to two vital network parameters: power-law degree exponent γ, and assortativity, measured by Spearman's ρ. Regardless of γ and ρ, we can see that κ is roughly proportional to the network's average degree. Also, we observe that for any particular network assortativity (and ρ value), $\kappa \sim e^{-\gamma}$. However, it is intriguing that for a fixed γ value, κ has a maximum approximately at $\rho = -0.4$.

6.6 Discussion and Conclusions

It is remarkable that RDS (either dependent or independent of degree) with optimally chosen p parameter can always provide a smaller dominating set than a simple degree-ranked node selection. While the latter may be favored for its simplicity and plausibility to be effective in heterogeneous networks, our results show that it is not the case; the usefulness of degree-ranked dominating sets beyond theoretical studies is very limited.

When we calculate a CDS, there is an algorithmic optimization we can use to

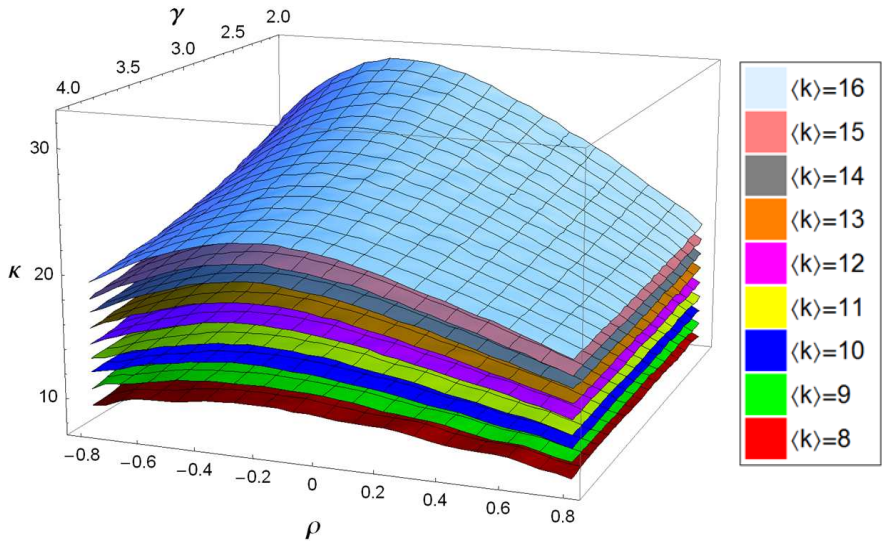

Figure 6.8: Optimal κ degree cutoff values (that minimize the size of CDS) as a function of degree exponent γ and Spearman's ρ. Each layer represents different average degrees, in cCONF networks with $N = 2000$. Data is averaged over 200 network realizations. Data grid resolution: $\Delta\gamma = 0.05$, $\Delta\rho = 0.05$.

find the CDS size for *all* possible κ values, including the optimal one that minimizes CDS size, in the same time complexity as finding CDS for only one κ value. First, we sort nodes into degree classes in $O(N)$ time using counting sort (or bucket-sort). The linear time complexity comes from the fact that both the number of nodes and the range of their degree values are $O(N)$. Then, we loop over all degree classes in decreasing order of degree, and for each degree class we add all nodes to set X (and remove them and their neighbors from set Y at the same time). This way, we can check the value of $|X| + |Y|$ after finishing each degree class, which is exactly the size of CDS with κ equal to the current class degree. We can either output the size of CDS at the current degree, or simply record which CDS size at which κ was the smallest. Since we process each node exactly the same way as in RDS (except for the specific order in which they are processed), we have the same $O(E)$ time complexity, and it is not increased by the $O(N)$ time needed to sort the nodes.

Clearly, the performance of CDS depends on the optimal selection of κ. When full adjacency information is available, we can use our algorithmic optimization to find the optimal κ (along with the CDS size for all κ values). Note, that since

the algorithm only uses local connectivity information, a distributed version can be easily designed to speed up processing large networks. However, when adjacency information is not available, we must estimate κ. Given basic (estimated or sampled) information about the degree distribution, we can use analytical estimates for the CDS size (see Section 6.4); differentiation with respect to κ leads to the optimal value. Further, the optimal κ has little dependence on particular network parameters (power-law exponent and assortativity), as shown in Fig. 6.8. Therefore, as long as the average degree is accurately known, κ can be chosen correctly.

The lower accuracy of our analytical estimates for RDS and CDS observed at low $\langle k \rangle$ and γ values can be easily explained: It is an artifact of our average degree control method (Section 4.4), which controls $\langle k \rangle$ by adjusting $k_{\min}$, and removing a certain fraction of smallest degree nodes. The latter becomes significant when $k_{\min} \to 1$ (for low $k_{\min}$), because it causes a slight deviation from a perfect power-law degree distribution. In order to use the analytical formulas (which are very sensitive to $k_{\min}$), we have to estimate a fractional $k_{\min}$, as if it were a cutoff of a continuous and otherwise perfectly satisfied power-law distribution. In reality, we deviate from power-law, leading to the inaccurate estimates. However, as $\langle k \rangle$ increases, $k_{\min}$ also increases, and the relative deviation from a perfect power-law decreases, hence the increased accuracy. The implication for real networks is that we can expect similarly less accurate estimates if the degree distribution deviates from power-law.

While the analytical estimates for RDS and CDS are highly accurate, they are only applicable to uncorrelated scale-free networks. However, the base formula [Eq. (6.5)] can be used for any network (not only scale-free), if the degree distribution and degree correlations can be expressed (or approximated) by some formula. Without analytical expressions, one can still calculate the base formula numerically, using observed (sampled) estimates of the degree distribution and degree correlations, assuming that collecting these estimates requires less time than actually running the RDS or CDS algorithms, or if full adjacency information is not available.

We can also understand CDS as a method that bridges the degree-ranked and greedy methods. When selecting the very first nodes of the dominating sets, both greedy and degree-ranked methods start by selecting the highest degree nodes. Later,

they diverge; the degree-ranked selection continues with the high-degree nodes, while greedy specifically seeks out nodes that increase domination maximally, typically smaller degree nodes. The degree-ranked selection eventually becomes very inefficient only because of the presence of low degree nodes connected only to each other (hard to reach). Thus, degree-ranked selection is efficient at first, but there is a point at which the method should be abandoned and instead look for nodes that are still not dominated, and target them specifically. This is exactly what CDS does: it is essentially a degree-ranked selection until κ is reached (set X), and then the remaining undominated part is simply added as dominators (set Y).

Our numerical study of dominating set sizes with respect to assortativity reveals a general tendency that the dominating set becomes larger as assortativity increases. We can understand this easily. In case of a disassortative network, high degree nodes connect mostly to low degree nodes, therefore we can expect small dominating sets, due to efficient domination via high-degree nodes. In fact, when $\gamma < 2$, scale-free networks may become so disassortative that star subgraphs form and the size of MDS becomes $O(1)$ [236]. On the other hand, hubs are less effective in dominating assortative networks, since most of their connections are used to connect to other high degree nodes. Therefore, the impact of assortativity on each dominating set selection method depends on how much the method relies on high-degree nodes as dominators. This is why the degree-ranked selection shows the worst performance on highly assortative networks, followed by the degree-dependent RDS (and its limiting case, the CDS), which also favors high-degree nodes. Since technological scale-free networks tend to be disassortative, and although social networks tend to be assortative, extreme assortativity is rare, we can safely conclude that CDS is a viable alternative of greedy selection for most scale-free networks.

In summary, we explored probabilistic dominating set selection strategies in scale-free networks with respect to various network properties. We found that as a particular limiting case of degree-dependent random node selection, a deterministic cutoff dominating set (CDS) provides the smallest dominating set among probabilistic methods, and is widely applicable to heterogeneous networks. Even if full adjacency information is not available, the size of CDS (and RDS) can be accurately predicted using our analytical estimates.

CHAPTER 7
BUILDING DAMAGE-RESILIENT DOMINATING SETS IN COMPLEX NETWORKS

Understanding the effects of network damage is essential for designing resilient and long-lasting networks. If we implement network control using dominating sets, then the connectivity of the surviving network structures and the fraction of the remaining set of nodes still dominated following failures or attacks are both essential for sustainable network operations and carrying out network functions. While the former (structural integrity) has been studied in great detail over the past two decades [106–112], the latter (domination stability) has not received any attention.

We assume that the network damage is relatively small, and although the network may become fragmented due to the loss of nodes, we assume it remains functional. In such cases efficient domination over the network is still important and desirable, just as it is in undamaged networks. However, considering that most dominating set search methods aim for the smallest possible set size (and corresponding cost) in a fixed topology network, even a small damage could severely disrupt the complete domination "coverage." Our goal is to understand how fragile dominating sets are, how to improve them, and ultimately to provide new methods for selecting dominating sets with adjustable balance between resilience and cost.

In order to quantify the resilience of a dominating set against node removal, we define *domination stability* as the fraction of the remaining network that is still dominated after some nodes have been removed from the network (and thus from the dominating set):

$$s(f) := \frac{|\bigcup_{j\in \mathrm{DS}} \mathrm{N}^+(j)|}{N(1-f)}, \tag{7.1}$$

where DS is a dominating set of the original (undamaged) network, f is the fraction of nodes removed from the network, and $\mathrm{N}^+(j)$ is the closed neighborhood of node j

Portions of this chapter to appear as: F. Molnár Jr., N. Derzsy, B. K. Szymanski, and G. Korniss, "Building Damage-Resilient Dominating Sets in Complex Networks against Random and Targeted Attacks," (under review).

that still exists in the remaining network. We measure stability by simulating network damage, i.e., by actually removing nodes from the network and calculating the remaining dominated fraction.

Domination stability does not only depend on the fraction of removed nodes, but also on the order in which nodes have been removed from the network. Similarly to many studies in the literature we consider two damage scenarios: random and targeted node removals. The random node removal strategy models network damage produced by natural causes or errors, while the targeted node removal method reflects the impact of intentional, targeted attacks on a network. In the random damage scenario nodes are removed with equal probability, in random order. In case of targeted attacks, the nodes are removed in degree-ranked order, with highest degrees being removed first. We indicate which strategy we consider in the subscript of stability: s_{rand} denotes the stability against random damage, and s_{deg} corresponds to the stability against degree-ranked removal.

7.1 Stability of Various Fixed Dominating Sets

We start our analysis by measuring the stability of three different dominating sets, that we use for baseline comparison with our new methods. These are the following:

- minimum dominating set (MDS) [88, 222, 236], where nodes are selected by a sequential greedy search algorithm in order to approximate the actual (NP-hard) smallest dominating set (see Section 4.3),

- cutoff dominating set (CDS) [237], where all nodes above a degree threshold are selected into set X, and the nodes not dominated by any nodes in set X are selected into set Y. The dominating set is then given by $X \cup Y$. The degree threshold is selected such that it minimizes the size of the resulting dominating set (see Section 6.3),

- degree-ranked dominating set (DDS), where we select all nodes in decreasing order of degree (with random tie-breaking) as dominators until the selected set dominates the entire network.

Our first choice is MDS, due to its importance in cost-efficient control of complex networks [236]. The other methods we have chosen are potentially useful when finding the MDS (using either a greedy algorithm or by solving the binary integer programming equivalent) is impractical, e.g., when the adjacency information of the network is incomplete, or the network is too large to run optimal MDS search algorithms. In these cases heuristic algorithms, such as CDS or DDS can find suboptimal (not the smallest possible), yet small enough dominating sets that are still useful for practical applications. In particular, the excess nodes selected by these methods may help to increase domination stability.

Figure 7.1 shows the stability of MDS, CDS and DDS against the fraction of removed nodes in the entire remaining network [Fig. 7.1(a), (b)] and in the remaining giant component [Fig. 7.1(c), (d)]. It is clear that the degree-ranked node removal reduces the dominated fraction much faster than the random node removal, because high-degree nodes are more likely to be dominator nodes than low degree nodes. The giant component itself also breaks down much faster, as shown in the insets of Fig. 7.1(c) and (d). However, as long as a giant component exists, it has higher domination stability than the entire network, in both scenarios. The slight increase of stability at high damage rates is a side effect caused by removal of nodes that had lost domination by earlier removals. When the network damage is high, it becomes more likely that these nodes are deleted, causing the dominated fraction of the remaining network to increase. At this point, however, the network is almost completely destroyed and domination stability becomes meaningless.

The stability curves show much more disturbed shapes in degree-ranked removal than random removal, due to the differences in the degree structure of each dominating set. In MDS, there is no preference toward any particular node degree during selection of dominators (besides the natural effect of the greedy selection, where the high-degree nodes provide a larger increase in the number of dominated nodes, hence they are more likely selected), which means that removal of high-degree nodes has a smooth (albeit strong) impact on stability. In CDS, we can see a fast initial drop as we remove the very high degree nodes that were specifically selected for dominators (in set X), then continuing at a more gentle slope as the dominators from the Y

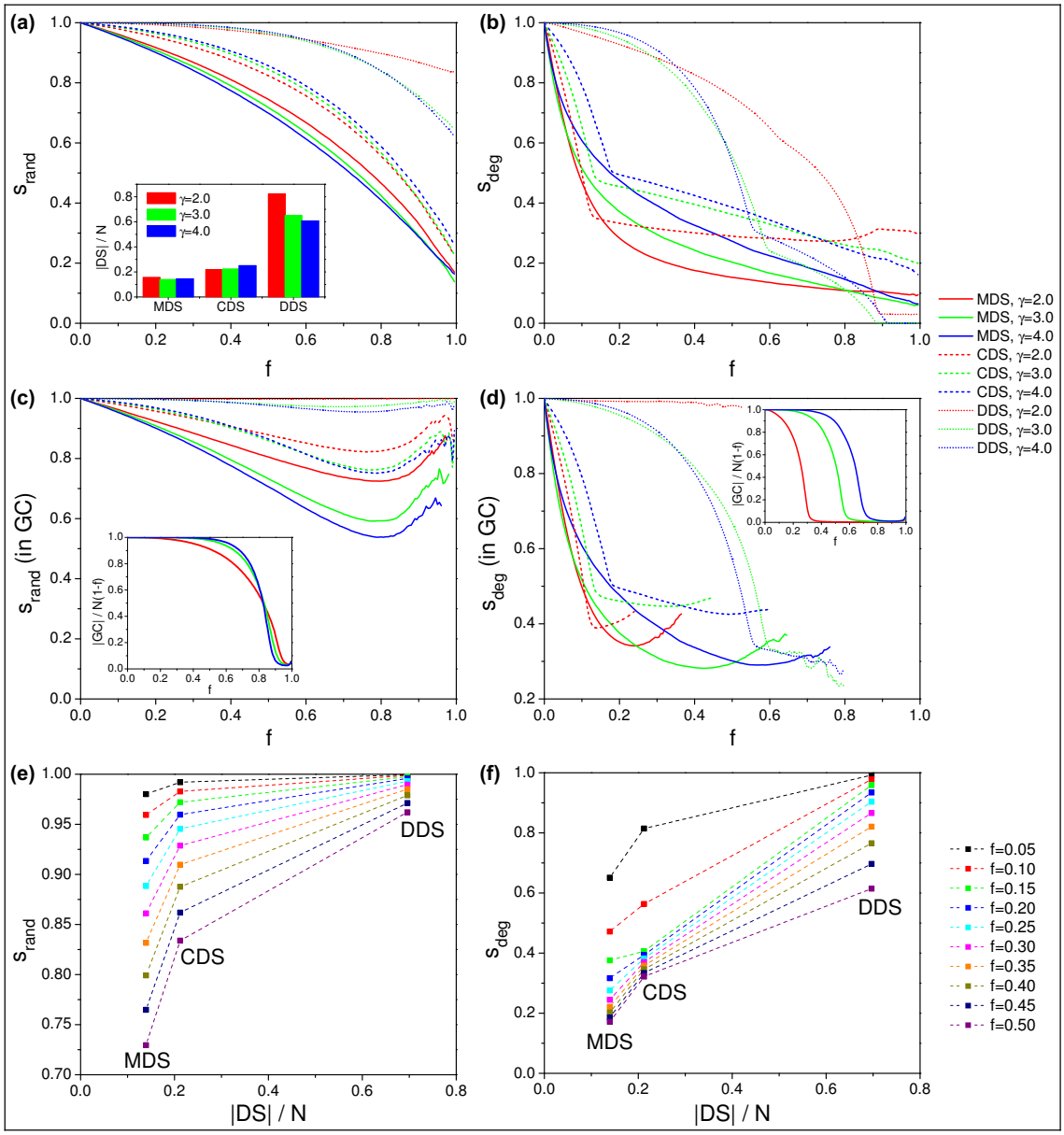

Figure 7.1: Stability of various dominating sets against random and degree-ranked node removal. The first column of subfigures shows random node removal, the second shows degree ranked node removal. (a) and (b) show stability in the entire network, while (c) and (d) show stability within the remaining giant component. Inset in (a) shows the corresponding sizes of dominating sets; insets in (c) and (d) show the size of the corresponding giant component. (e) and (f) show a correlation between set size and stability, at $\gamma = 2.5$. All plots show synthetic scale-free networks, $N = 5000$, $\langle k \rangle = 8$, averaged over 200 network samples.

set are removed, since any node that was not dominated by X, regardless of degree, may be in set Y. On one hand the Y set may seem wasteful in its construction, but with the right degree threshold the size of the CDS is actually very close to the MDS [237], and the excess nodes provide a fair increase in stability. DDS is the simplest but most inefficient method for finding a dominating set because it selects *all* nodes starting from the highest degrees until all nodes are dominated. However, the resulting redundancy of dominators in the network is providing the highest stability of all three methods.

We can also observe the general tendency that a larger dominating set provides higher stability. At any given fraction of removed nodes, there is a positive correlation between stability and the size of the original dominating set, in both random [Fig. 7.1(a)] and degree-ranked [Fig. 7.1(b)] node removals. We clearly illustrate this correlation in Fig. 7.1(e) and (f), where we show stability as a function of the dominating set size, at various damage levels. This means that the MDS, which is the smallest (most cost-efficient) dominating set, is also the most vulnerable, to both random damage and targeted attacks.

Note, that Fig. 7.1 only shows the stability for a certain network type with given degree exponent and uncorrelated networks (where Spearman's $\rho = 0$). We have included the analysis of stability at different values of these parameters in Appendix E.3.

The main conclusion we can draw is that the extra amount of dominating nodes selected by heuristic methods CDS and DDS, compared to the smaller and more optimal MDS, can effectively increase the stability of domination. However, all three methods are "fixed" in the sense that they give only a single possible dominating set size (and corresponding stability) for a given network.

7.2 Flexible-Redundancy Dominating Set (frDS)

In order to overcome the limitations of fixed methods, we must analyze in detail how domination is lost when the network is damaged. First, we realize that loss of domination occurs locally at each node: those nodes that lose all dominators will reduce the domination stability of the network. Therefore, stability can be expressed

locally, as the domination *redundancy* of each node. This quantity simply counts how many dominator nodes are within the closed neighborhood of a given node. A large dominating set can successfully increase domination stability, if the extra nodes are distributed in a way that they increase domination redundancy on many nodes. This seems to occur naturally for CDS and DDS, however we cannot guarantee that redundancy was increased in the most optimal way (relative to MDS), nor can we control the number of selected nodes.

We introduce the flexible-redundancy dominating set (frDS) to solve these problems. We explicitly set an average domination redundancy in the network, denoted by r, that must be guaranteed by frDS, while aiming for minimum set size. Note, that $r = 1$ is equivalent to the minimum dominating set (MDS), and when r is an integer, the frDS is identical to the *h-dominating set* (with $h = r$) studied by Cooper *et al.* [223]. Finding an frDS is most likely NP-hard, since it is also NP-hard to find an MDS [87] or an h-dominating set [238], but we can use a modified greedy algorithm to find an approximation.

The steps of finding an frDS are as follows. First, we assign a domination redundancy requirement, $r(i)$ for each node i as an integer value indicating at least how many dominators node i must have in the dominating set. Given the desired average (non-integer) r value for the entire network, we assign the nearest integer values $\lfloor r \rfloor$ and $\lceil r \rceil$ to each node randomly, such that the network average will be r (the probability of assigning $\lceil r \rceil$ is $r - \lfloor r \rfloor$, which is analogous to a biased coin toss). For the greedy selection we define a dominating potential $p(i)$ as the number of nodes in the closed neighborhood of i that have not yet reached their domination requirement, and therefore selecting node i can help them advance toward their goal. (Note, by definition, the potential of an already selected node is zero.) At each greedy step we select one node with maximum dominating potential (with random tie-breaking), until the requirements of all nodes have been fulfilled. Note, that since dominating potential is an integer number between 0 and N, nodes can be sorted according to their potential in $O(N)$ steps, and it is possible to maintain sortedness after changing the potential of a node in $O(1)$ step (see Appendix E.1 for further details and pseudocode). This results in the same computational time complexity as

for the greedy MDS approximation, $O(E)$. Also note, that if $r > N$, then the node requirements can never be satisfied, in which case the greedy selection naturally falls back to selecting nodes in degree-ranked order, because at every step every neighbor of a node may be advanced toward its goal.

7.3 Flexible-Cost Dominating Set (fcDS)

When we aim for a desired dominating set size (fixed cost level, i.e., limited budget), we can, in principle, aim for the necessary redundancy level in frDS to achieve that desired cost. However, we can further improve stability by considering the expected attack pattern on the network (if the information is available), and optimize the selected dominating set accordingly. For example, if the attack is expected at high-degree nodes, we should avoid selecting many of those nodes as dominators, despite their ability to cover large fractions of the network.

We can optimize our choice of dominators by including the probability of losing each node into the calculation of local stability, which we aim to maximize. First, we assign a strength value $s(i) \in (0, 1)$ to each node i, which represents the a-priori estimated probability for not losing that node after the attack (i.e., the anticipated attack pattern). Then, we calculate the current domination stability of node i as follows:

$$\text{stability}(\text{DS}, i) = \begin{cases} 0 & \text{if } \text{DS} \cap N^+(i) = \varnothing \\ 1 - \prod_{j \in \text{DS} \cap N^+(i)} (1 - s(j)) & \text{else,} \end{cases} \tag{7.2}$$

which is the probability that node i will remain dominated (not lose all dominators), assuming nodes will be deleted independently; DS denotes the currently selected dominating set. For selecting the next dominator, we choose one that increases the total stability of the network maximally. The total potential increase of stability can be calculated for each node as follows:

$$\text{potential}(i) = \sum_{j \in N^+(i)} \text{stability}(\text{DS} \cup \{\text{i}\}, j) - \text{stability}(\text{DS}, j) \tag{7.3}$$

$$= \sum_{j \in N^+(i)} (1 - \text{stability}(\text{DS}, i)) \cdot s(i). \tag{7.4}$$

Therefore, we always select a node with maximum potential (with random tie-breaking). Note, that unlike in frDS, the potential here is a non-integer value, thus we can only use comparative sorting to order nodes by potential, which needs $O(N \log N)$ steps. In addition, after selecting each dominator, the stability values have to be recomputed in the selected node's closed neighborhood, and the potentials up to the node's second neighborhood. This involves $O(d^2)$ nodes, where d is the average degree. Thus, maintaining sortedness of nodes by their potential requires $O(d^2 \log N)$ steps after selecting each dominator.

In order to compare stability of fcDS with frDS and other dominating sets, we calculate the "a-priori" node strength values as follows: $s(i) = 0.5$ for random node removal, and $s(i) = 1 - d(i)/N$ for degree-ranked node removal. Here, we assume the size of the anticipated damage is unknown, thus strength values are expressing relative probabilities only. The strength value for a random damage is arbitrary, as long as it is uniform among the nodes, and it is inversely proportional to node degree in a degree-ranked attack. Further details of fcDS and pseudocode are included in Appendix E.2.

7.4 Stability Comparison of Dominating Sets

We seek to answer two main questions in our analysis. First, we want to see how much stability we can achieve by selecting various sizes of dominating sets (in other words, how does the stability scale with larger invested cost of domination). Second, we want to know how much more efficient our methods are compared to the fixed dominating sets, that is, given the same size of dominating set as MDS, CDS, or DDS, how much higher stability can our methods provide.

Figures 7.2 and 7.3 show domination stability achieved by frDS and fcDS as a function of redundancy and dominating set size, respectively. Stability achieved by the fixed methods (MDS, CDS, DDS) are also shown at their corresponding cost values for comparison. The general shape of the curves in both figures are similar, since the dominating set size is roughly proportional to redundancy (see Fig. 7.2 inset and Appendix E.5). In case of random damage, the stability rapidly increases with cost, until the size of MDS is reached, then the curve saturates. There

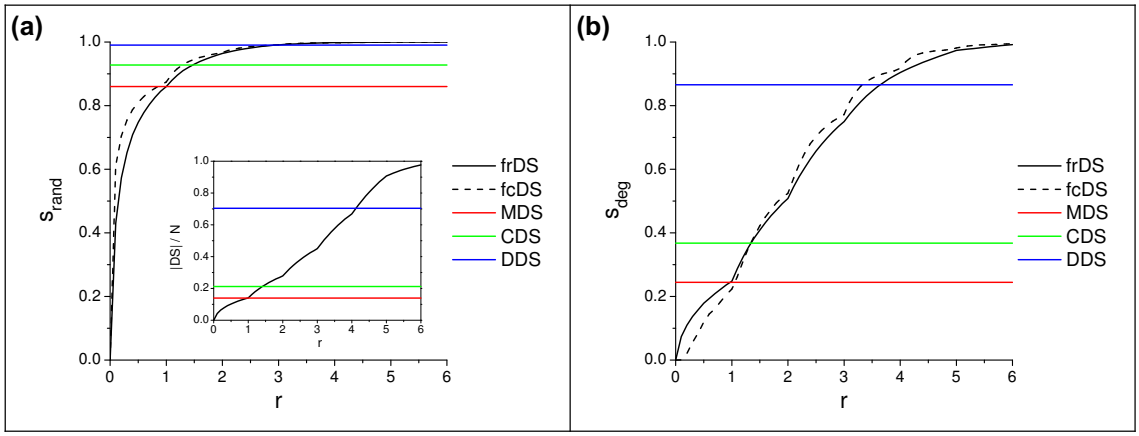

Figure 7.2: Domination stability of frDS and fcDS as a function of domination redundancy. (a) shows random node removal, (b) shows degree-ranked node removal. The inset shows the sizes of the corresponding dominating sets. The size of fcDS is set to match frDS at any given r value. Synthetic networks, $N = 5000$, $\langle k \rangle = 8$, $\gamma = 2.5$, $f = 0.3$, averaged over 200 network samples.

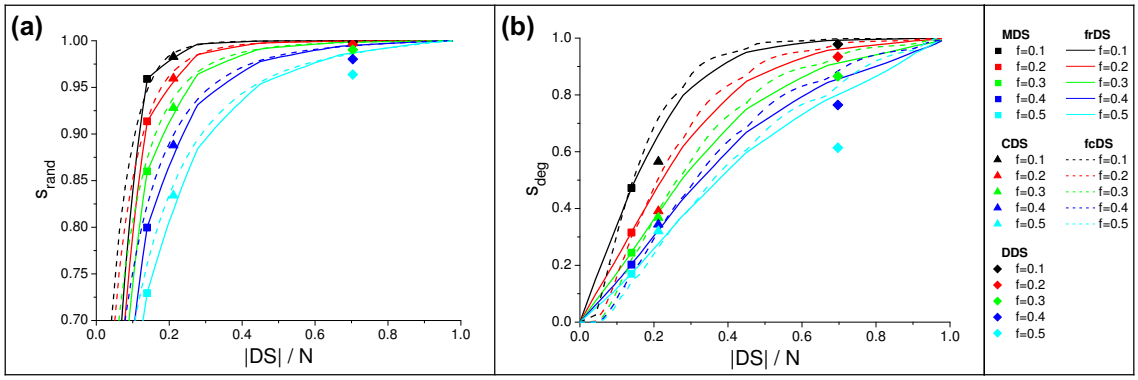

Figure 7.3: Domination stability of frDS and fcDS as a function of dominating set size (cost) for various network damage fractions. Stabilities of MDS, CDS, and DDS are presented at their corresponding cost values. Subfigure (a) shows random node removal, (b) shows degree-ranked node removal, for synthetic networks, $N = 5000$, $\langle k \rangle = 8$, $\gamma = 2.5$, averaged over 200 network samples.

is little advantage in selecting a dominating set larger than approximately twice the size of MDS, because stability is already very close to 1, even at large damage values. However, in case of degree-ranked damage, there is a steady increase in stability as more nodes are selected as dominators. In both cases, fcDS provides

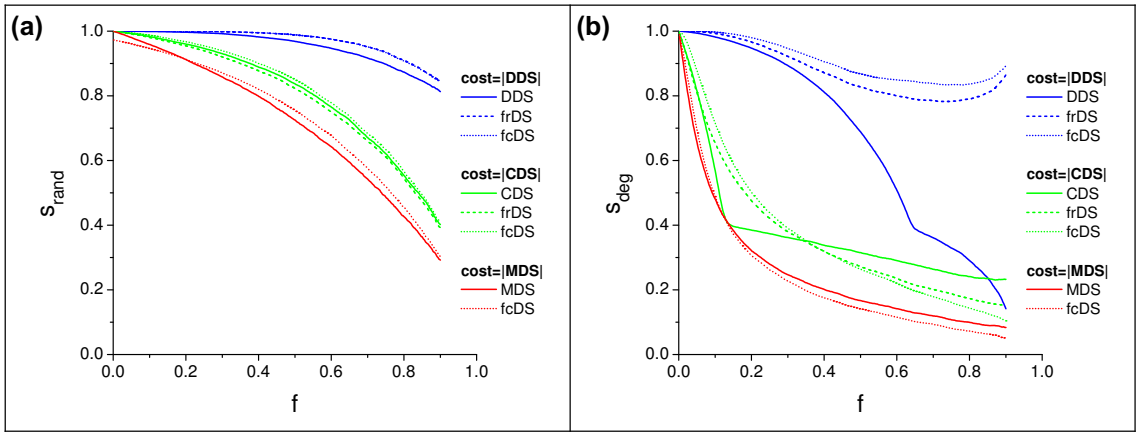

Figure 7.4: Comparison of domination stability at fixed cost levels, as a function of network damage fraction. Stability of frDS and fcDS are plotted at cost values identical to MDS, CDS and DDS. Subfigure (a) shows random node removal, (b) shows degree-ranked node removal, for synthetic networks, $N = 5000$, $\langle k \rangle = 8$, $\gamma = 2.5$, averaged over 200 network samples.

somewhat higher stability than frDS at moderate damage levels, but frDS is more stable at small damage levels. These observations hold across a wide range of network parameters, see Appendix E.6. It is also clear that both frDS and fcDS can provide great flexibility in adjusting the size of the dominating set and stability.

The stability of frDS and fcDS at cost levels identical to MDS, CDS, and DDS are presented in Fig. 7.4. Our results show that frDS provides stability values very similar to the fixed methods (in case of MDS, it is identical by definition, thus it is not shown), while fcDS shows a minor improvement in stability. On the other hand, both frDS and fcDS show significant improvement over the fixed methods against degree-ranked attacks, at low network damage fractions. MDS and CDS show a tipping point in damage, where these methods become slightly more effective than frDS or fcDS, but the difference is minimal, and it occurs only at moderate to high network damage ($f \gtrsim 0.3$).

7.5 Stability in Real Networks

We analyze stability of frDS and fcDS, as well as other dominating sets, in several real complex networks, listed in Table 7.1. These include an internet peer-to-peer network (p2p-Gnutella08) [230], the power transmission network of continental

Europe (ENTSO-E powergrid) [239, 240], and one brain graph extracted from MRI data (KKI21-KKI2009-26) [24, 241]. Note, that we only use the giant component of these networks. A brief analysis of the degree distributions of these networks is provided in Appendix E.7.

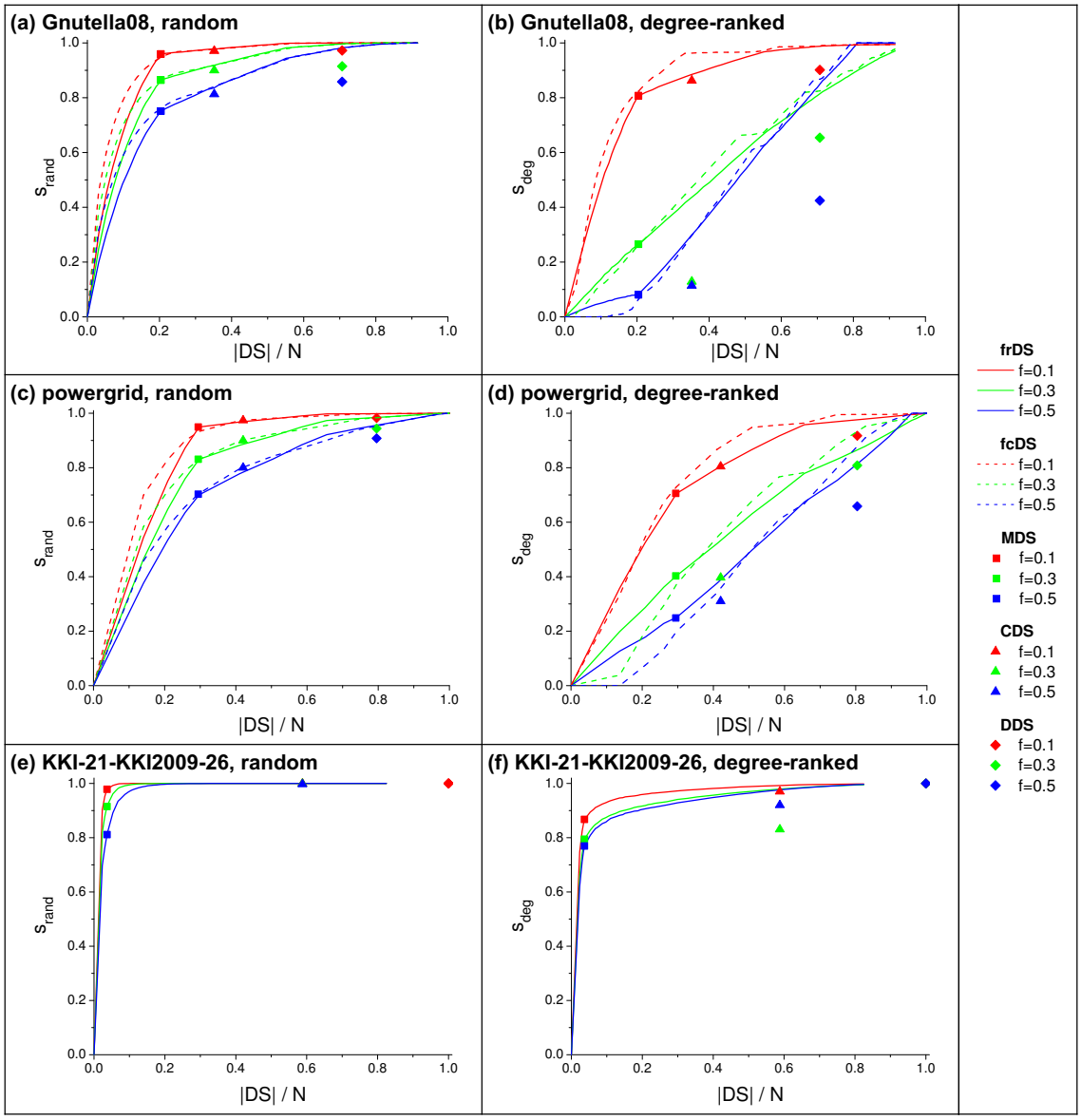

Figure 7.5: Stability of frDS, fcDS, and other dominating sets in real networks, for various damage fractions. Data is averaged over 20 independent runs of node removal. See Table 7.1 for network parameters.

Table 7.1: Parameters of real networks used in our analysis. The data refers exclusively to the giant component.

Name	Source	N	$k_{\min}$	$k_{\max}$	$\langle k \rangle$	Spearman's ρ
Gnutella08	[230]	6299	1	97	6.60	0.03
powergrid	[239, 240]	1494	1	13	2.89	−0.18
KKI-21-KKI2009-26	[24, 241]	813479	1	5171	171.9	0.57

The brain graph we analyze here (KKI-21-KKI2009-19) is one of 200 graphs available from [241]. These graphs have peculiar structural properties, and are very similar to each other. In particular, all brain graphs are very dense: $\langle k \rangle \approx 150$; they are all very assortative: $\rho \approx 0.6$; and they have very similar degree distributions (see Appendix E.8 for details). It is also interesting that the size of MDS is very small, only 3-4% network size, while the size of CDS and DDS is very large, around 60% and 100% of network size, respectively. We attempt to separate the effects of density and assortativity in order to identify their impact on domination stability.

Figure 7.5 shows domination stability as a function of dominating set size for the real network samples. In general, we see that stability of frDS and fcDS matches the stability of MDS, and exceeds the stability of CDS and DDS, at identical set sizes. In case of Gnutella08 and the powergrid, the stability curves saturate slowly, and the curve shapes are not as smooth as for synthetic scale-free networks, due to having more disturbed (non-scale-free) degree distributions. However, the brain graph shows very high domination stability against both random and targeted attacks. In all cases, the relative advantage of frDS and fcDS over CDS and DDS (i.e., cost-efficiency) remains as high as in synthetic scale-free networks.

We can observe the effects of assortativity separately from other structural properties by artificially changing the network's assortativity, using a biased edge-mixing method (see in [237] and Section 4.5), which rewires the edges in the graph, while keeping the degree sequence unchanged. Using this method we present a brief analysis of dominating set size vs. assortativity in Appendix E.9. In general, we see the expected behavior that dominating sets tend to become larger in more assortative networks [237]. Note, that the size of DDS in the brain graph being 100% of nodes

regardless of assortativity is the result of a particular topological feature; there are a small number of leaves (degree 1 nodes) connected to degree 2 nodes, thus DDS has to select al nodes down to degree 2 (essentially all nodes) to dominate these off-hanging leaves — a feature left unchanged by edge-mixing.

Figure 7.6 presents the effects of assortativity on domination stability. We see an unexpected behavior: as assortativity increases, domination stability decreases against random damage, but increases against an attack on high-degree nodes. We can understand this behavior by considering the effects of assortativity on dominator node degrees. In disassortative networks dominators are mostly high-degree hubs, while in assortative networks dominators have a full range of degrees. Thus, when the network is disassortative and the damage is random, it is less likely to remove high-degree hubs and more likely to remove low degree nodes, the latter rarely being a dominator, leading to increased stability. On the other hand, the result is reversed when high-degree nodes are targeted, in which case we are more likely removing dominators, leading to decreased stability.

Finally, we can conjecture that the outstandingly high domination stability in brain graphs can be attributed to both their high average degree and high assortativity. High average degree results in a highly redundant dominating set (regardless of method) which resists random damage successfully, while high assortativity guarantees that an attack targeted at high degrees leaves the network with plenty of lower-degree dominators.

7.6 Partial Flexible-Redundancy Dominating Sets

There are two possible ways to achieve a certain desired cost (dominating set size) with frDS. Either we aim for the lowest r value that provides the desired cost, or we may choose a larger r value, and use only a fraction of the larger dominating set it provides. In the latter case we would select nodes in the same order as the greedy algorithm picked them. In other words, we can either select a full frDS with small r or a partial frDS with the same size but larger r. Figure 7.7 shows the comparison of these two cases (see Appendix E.4 for analysis over a wide range of network parameters). The contour curves of fixed stability values are monotonically

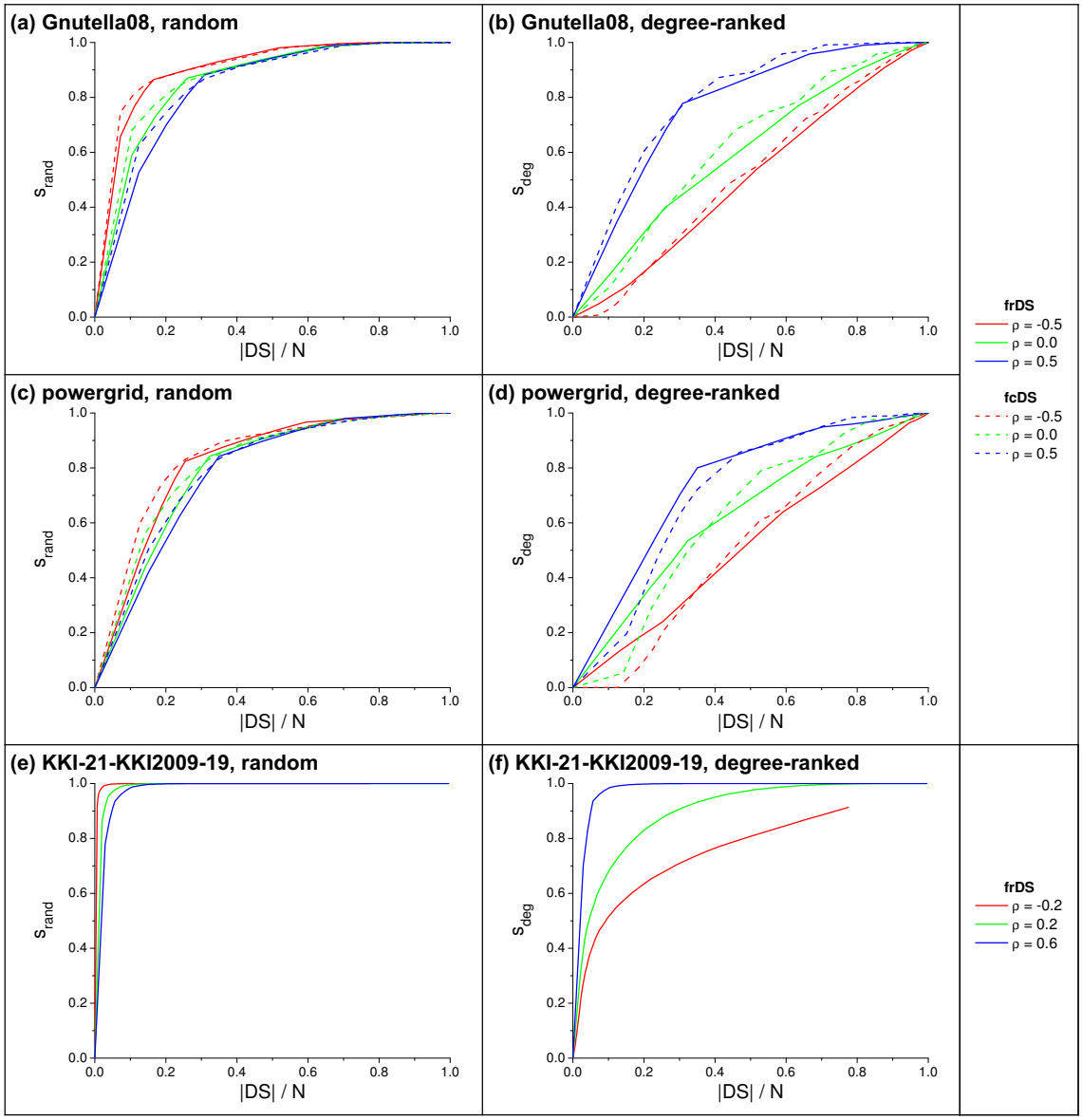

Figure 7.6: Stability of frDS and fcDS in edge-mixed real networks against random and degree-ranked attacks, for various assortativity levels: (a,b) Gnutella peer-to-peer network; (c,d) ENTSO-E powergrid; (e,f) Brain (MRI) network. Network damage fraction $f = 0.3$. For (a-d) data is averaged over 50 independent runs edge mixing and node removal; (e,f) is from a single run. See Table 7.1 for parameters of the original networks.

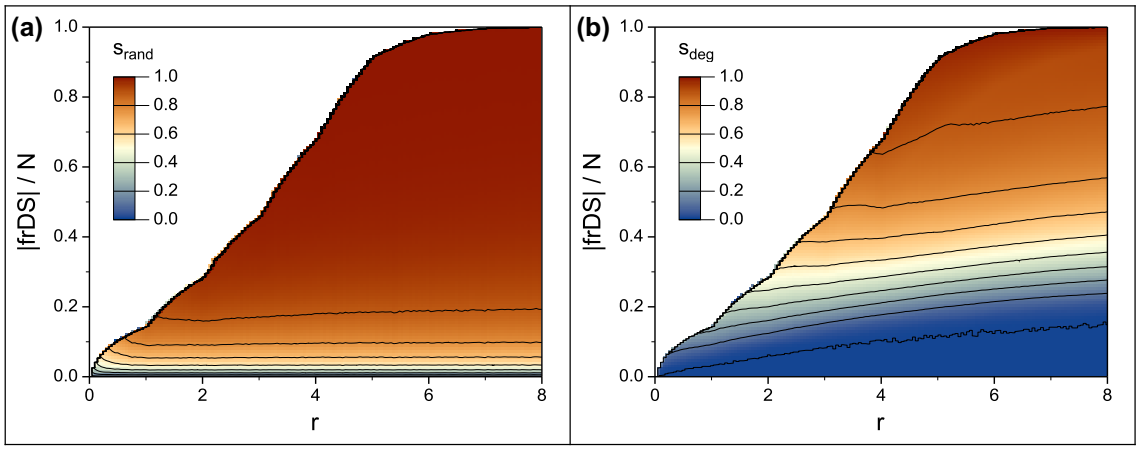

Figure 7.7: Domination stability of partial frDS as a function of domination redundancy and dominating set size. The plotted area is bounded by the size of the full frDS at any given r. Subfigure (a) shows random node removal, (b) shows degree-ranked node removal, for synthetic networks, $N = 5000$, $\langle k \rangle = 8$, $\gamma = 2.5$, $f = 0.3$, averaged over 50 network samples.

increasing for larger r values, indicating that the cost for a certain stability level increases if we use partial frDS with higher r values. This also means that using full frDS with the smallest possible r value provides the highest possible stability.

In order to find the needed r value for a desired cost we must look at the relationship between r and the size of the resulting dominating set (see Fig. 7.2(a) inset, and Appendix E.5). The frDS size curve has a complex shape, but it is always monotonically increasing. Therefore, we can use a bisection method for finding the desired r value. Without any assumptions (other than monotonicity) about the size of frDS we must calculate the full frDS for every tested r, each taking $O(E)$ time, leading to $O(E \log N)$ time complexity for the entire procedure.

It is also interesting to note that the cost of stability increases slightly for smaller r values when $r < 1$, in case of a random damage [in Fig. 7.6(a)]. In this case even the full frDS is providing only a partial dominating set (dominating only a fraction of nodes in the undamaged network). This indicates that r should never be smaller than 1; if a smaller cost is needed than the one provided by frDS with $r = 1$ (which is the MDS by definition), then a partial MDS (given by the greedy MDS algorithm) is a more optimal solution.

7.7 Effects of Incorrectly Estimated Damage in fcDS

For practical applications of fcDS, it is necessary to understand how stability is affected, when the network damage is estimated incorrectly. We can check this effect for a degree-ranked attack by using the following sigmoid strength function for a node with degree k:

$$s(k) = \frac{1}{1 + e^{\alpha(k-\kappa(\alpha,f))}}. \tag{7.5}$$

There are two control parameters for the anticipation. The slope parameter $\alpha \in (-\infty, \infty)$ describes the attack distribution: it expresses whether low degrees ($\alpha < 0$) or high degrees ($\alpha > 0$) are targeted, and how sharp the difference is between targeted and non-targeted node strengths; parameter f is the anticipated damage fraction. The $\kappa(\alpha, f)$ function gives the threshold for the sigmoid, such that the expected number of lost nodes equals the anticipated damage, $\sum_k (1 - s(k))p(k) = f$ (where $p(k)$ is the degree distribution). Note, that $\alpha = \infty$ gives a sharp cutoff selecting all nodes above κ, corresponding to the actual attack; $0 < \alpha \lesssim 5$ corresponds to an uncertain transition point but correct anticipation; $a \approx 0$ corresponds to a random

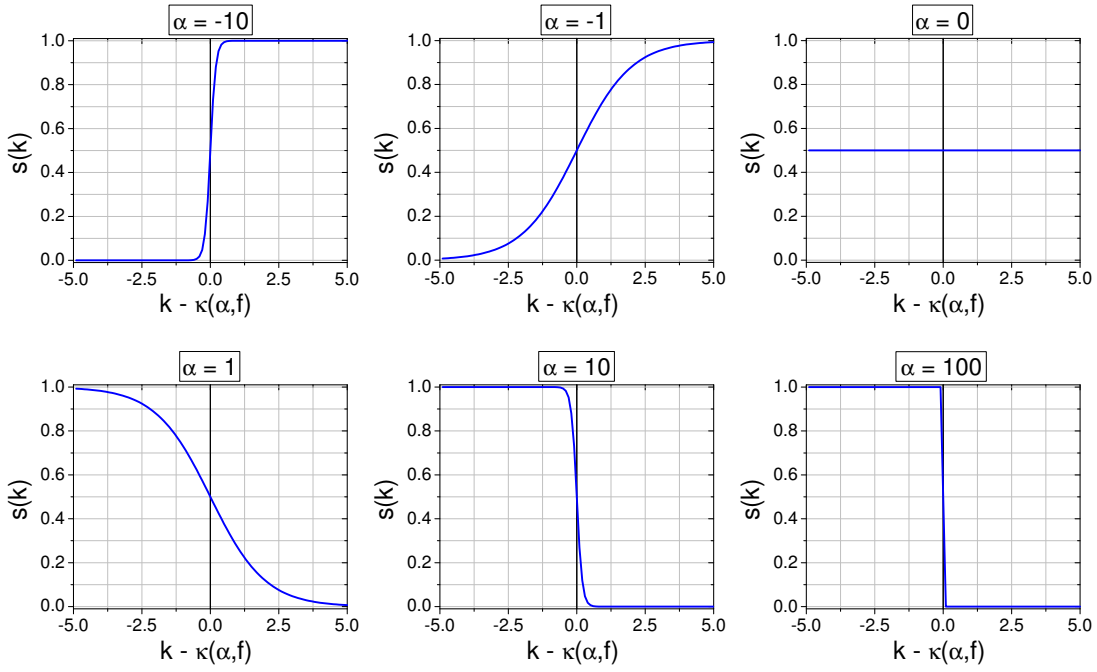

Figure 7.8: Node strength functions determined by the control parameters of fcDS anticipation accuracy (α and f), see Eq. (7.5).

guess; $-5 \lesssim \alpha < 0$ corresponds to an incorrect anticipation (i.e., anticipating attack on low degree nodes, when the attack occurs at high-degree nodes); and $\alpha \ll -5$ is the complete opposite of the actual attack.

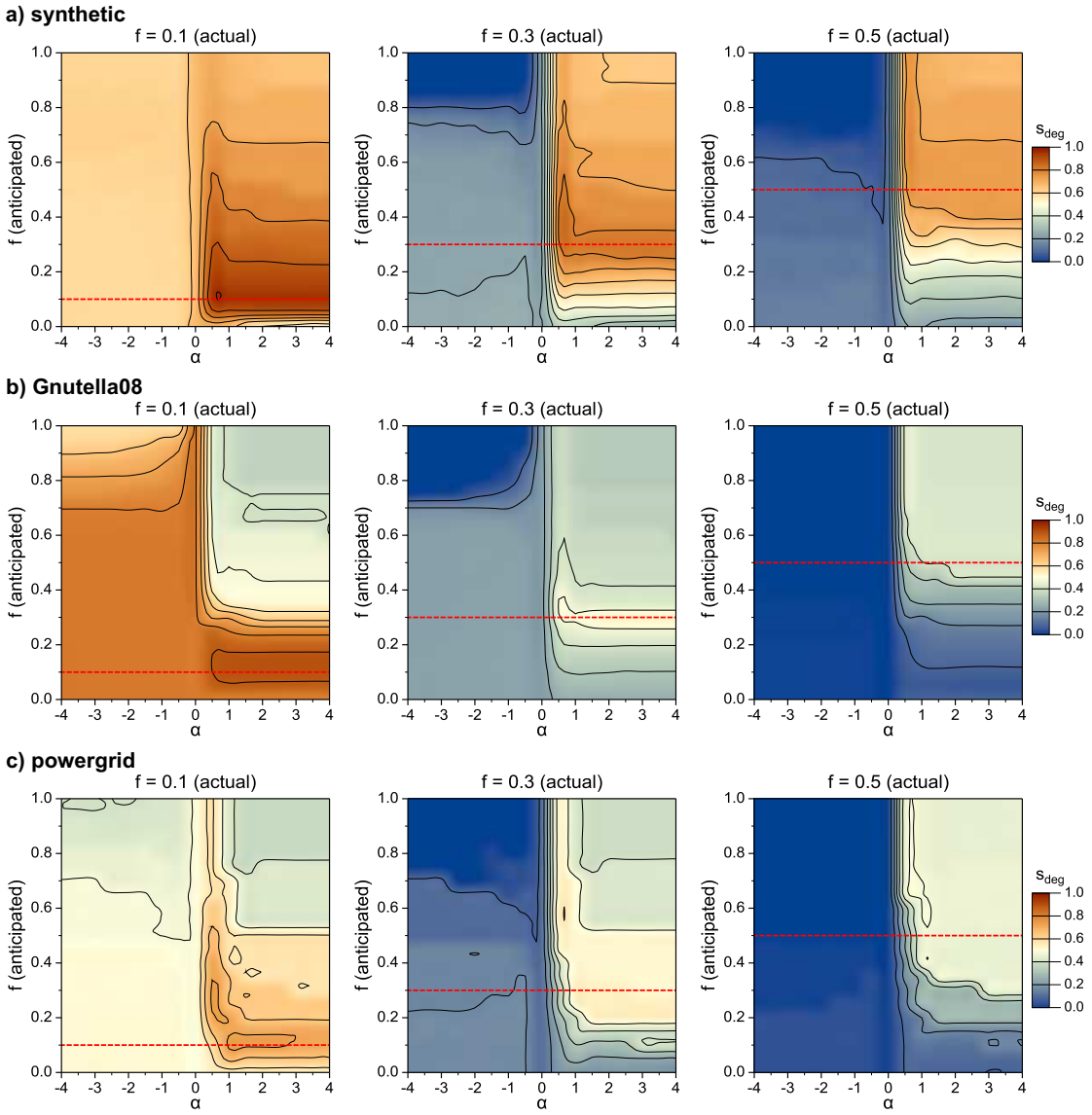

Figure 7.9: Stability of fcDS against degree-ranked node removal as a function of damage anticipation accuracy. (a) shows a synthetic network with $N = 5000$, $\langle k \rangle = 8$, $\gamma = 2.5$, (b) and (c) show real networks. The actual damage fraction is indicated above the plots and marked by red dashed lines; the actual degree distribution of the damage corresponds to $\alpha \geq 4$ values.

Figure 7.9 shows the landscape of stability as a function of the control parameters. As expected, we obtain the highest stability when the attacked degrees and the size of the attack are correctly estimated. For small damage fractions ($f = 0.1$) we lose stability mostly for overestimating the size of the attack, while for moderate ($f = 0.3$) and large ($f = 0.5$) damages we lose stability for incorrectly anticipating which degrees are targeted.

7.8 Discussion and Conclusions

We must clarify the distinction between the prescribed domination redundancy and the actual achieved domination redundancy in a network, when using frDS. The former is the one denoted by the r parameter, while the latter (i.e., the actual number of dominators in the closed neighborhood of a node) can be easily calculated for any given dominating set (not just frDS), and its average always exceeds the prescribed value. For example, even an MDS could have an actual average redundancy of 2.5 in certain networks, although most nodes would have only one dominator. However, an frDS with $r = 2.5$ would guarantee not only that the actual redundancy is at least 2.5, but also that no nodes will have less than 2 dominators.

The usage of frDS against degree-ranked or any other targeted attacks seems counter-intuitive, since in frDS, we aim for an overall increased redundancy that is most effective against random damage. However, the greedy algorithm has no preference toward selecting low-degree or high-degree dominators when trying to fulfill domination requirements, and in general, we observe empirically that the selected dominators have a large variability in degrees. This indicates that dominators of a given node may have significantly different degrees, which helps to keep the node dominated even if high degree nodes are targeted by an attack.

In the calculation of node stability in fcDS we assumed that nodes are deleted independently. In a realistic scenario, an attack may have between-node correlations, especially, in spatial graphs (e.g., clustered attack on a power grid). Taking this into count would add more complexity to the calculations, which we postpone for future work. However, it is important to emphasize that even without correlations, the fcDS algorithm can use arbitrary node strength values, irrespective of node degrees,

therefore its applicability goes much beyond our studied scenario of a degree-ranked attack.

Currently, the time complexity of fcDS is $O(d^2 \log N)$ for selecting each dominator node, which makes it prohibitive for very large graphs. In order to speed up the algorithm, the only obstacle we need to overcome is maintaining the sortedness of nodes by their potentials efficiently, which takes $O(\log N)$ steps after each change with comparative sorting. In principle, the potentials could be discretized and assigned to bins (the same optimization we use in frDS), which would lead to $O(E)$ complexity, as long as the bin count remains $O(N)$. However, the effects of such discretization on the dominating set and its stability is unclear, and it would require a thorough analysis to test the method's viability.

We can easily explain that fcDS has a slightly lower stability than frDS at low damage fractions, which we can observe in all graphs, by looking at the effects of incorrect attack anticipation. When the actual damage is very small, we overestimate the damage with our degree-dependent strength formula ($s(i) = 1 - d(i)/N$), because we assign nonzero probabilities to losing nodes with medium to low degrees. In reality, these nodes will not be deleted in a small targeted attack, thus the overestimated damage causes fcDS to lose stability, dropping slightly below the levels of frDS. This also underlines the need to correctly estimate the size as well as the distribution of the expected attack to achieve the most optimal domination stability.

Finally, we can provide a simple guide for selecting one of our two methods for practical applications. For large networks, or with no information about a potential attack, frDS is a good choice for providing a dominating set with decent stability against any form of damage (mostly against random damage originating from natural sources), with a short computational time. However, if there is a fixed budget for dominators, or detailed (and reliable) information is available about potential attacks, then fcDS can be used to optimize the selected dominating set for the highest stability.

CHAPTER 8
APPLICATION OF DOMINATING SETS IN THE THRESHOLD MODEL: A PROOF OF CONCEPT

Social influencing and spreading processes in social networks is an active field of research in network science and in social sciences [242]. In particular, dynamics of information spreading is studied to describe how a small initial seed of information can spread to the entire network [243, 244]; propagation models are studied in epidemiology (spread of diseases) [245], and economics (word of mouth information spreading [246], viral marketing and influencing [247, 248]); opinion dynamics are studied in the context of social agreement formation [242]. Spreading processes on complex networks are conceptually very similar to the spread of populations in an environment; we can draw analogy between the critical initiators of a spreading process in a complex network, and the critical cluster phenomenon in the diffusive spatial spread of populations. The goals are the same in both topics: finding a critical subset of the system that can initiate global change, and its relationship to the underlying dynamics.

The purpose of this short chapter is showing that dominating sets can play a significant role in social influencing. We propose that using nodes of the minimum dominating set (MDS) are very effective in initiating global opinion cascades in social networks.

We will use one of the most basic models of opinion adoption dynamics in our study: the linear *threshold model* [8, 124–127]. In this model each individual (node) has one of two states; it is either active, or inactive. Initially, all nodes are inactive, except for a small set of active nodes, called initiators. The rules of dynamics are very simple: An inactive node becomes active, if the fraction of its neighbors that are already active exceeds a predefined threshold (hence the name "threshold model"). Once activated, a node cannot return to inactive state. As time progresses, the number of active nodes increases, until no more change is possible. Spread S is defined as the subset of nodes that eventually become activated (either

being initially active or activated during the spreading process). We focus on the spread size $s = |S|/N$; and in particular, we are interested in *global cascades*, i.e., the eventual activation of all nodes in the network. The smallest fraction of nodes that must be chosen as initiators for reaching a global cascade is called the *critical fraction of initiators*.

Several strategies have been proposed for initiator node selection, in order to minimize the critical fraction of initiators. In particular, selecting nodes in degree-ranked order (selecting highest degree nodes first), have been shown to be effective [8]. We assert whether the MDS can provide a better alternative. A priori, the MDS could be effective for initiating adoption cascades, because we understand it as a means to reach nodes efficiently via nearest neighbors, which is how opinion (adoption) spreads in the threshold model.

It is notable, that *influence maximization* is a closely related problem to opinion cascades [249–251]. It is essentially a combinatorical optimization problem that focuses on the optimal placement of initiators to maximize the spread size. This is a hard problem, computationally [250], greedy approximations and node selection heuristics have been proposed as alternatives [251], much like for minimizing the critical initiator fraction. However, in influence maximization, it is implicitly assumed that the spread will not become global, therefore an optimal solution is more dependent on "early" spread dynamics (a few nodes causing a large spread), while for global cascades the optimal solution may include initiators that cause only a small early spread, as long as the total initiator set results in a global cascade. Therefore, the best strategies for the two problems could be different.

8.1 Critical Fraction of Initiators

In a recent work, Singh *et al.* [8] studied the critical fraction of initiators in Erdős-Rényi (ER) random networks, and in a real social network. We start with a similar analysis here, on synthetic scale-free networks. Similarly to [8], first we use uniform threshold values for all nodes, denoted by ϕ.

Figure 8.1 shows the cascade size s as a function of initiator fraction p, when initiators are randomly selected. It is clearly visible that similarly to earlier results [8],

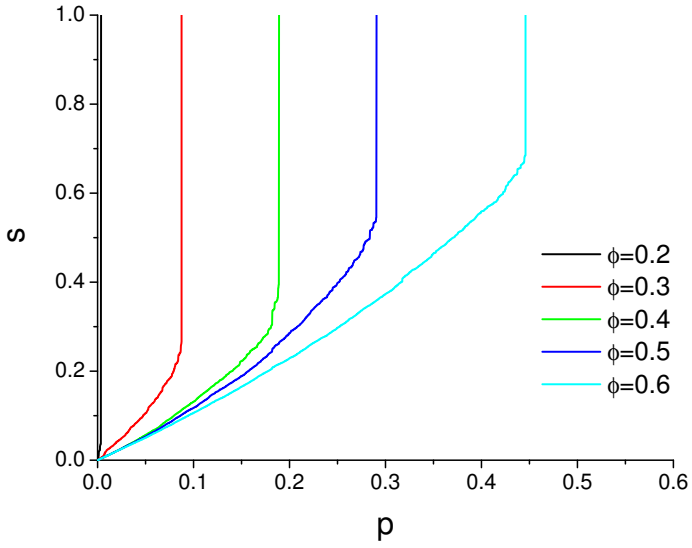

Figure 8.1: Cascade size vs. fraction of initiators, at various levels of uniform adoption thresholds, in a synthetic scale-free network (cCONF), with $N = 5000$, $\gamma = 2.5$, $\langle k \rangle = 8$.

global activation cascades occur in scale-free networks, when p reaches the critical initiator fraction, p_c.

In order to compare initiator selection strategies, we must be able to measure p_c numerically. Based on Fig. 8.1, we would simply need to record the smallest p that results in the activation on all nodes. However, in principle, certain network topologies could show only a sudden increase of spread when p_c is reached, but not a complete cascade; a small fraction of nodes may remain inactive (until more initiators are selected and these nodes are also activated, eventually). We can correctly measure p_c in these cases, as well as in case of complete cascades, by redefining and measuring p_c as the location of the maximal derivative of the $s(p)$ curve.

8.2 Comparison of Initiator Selection Strategies

Before we can compare strategies, we must consider the implementation details of our proposed MDS strategy. First, we need to establish an order in which nodes are selected as initiators. Since we use a sequential greedy algorithm to find an MDS, we can use the same ordering as the greedy algorithm picked the nodes. If an alternative method provides the MDS nodes without ordering, then using them in

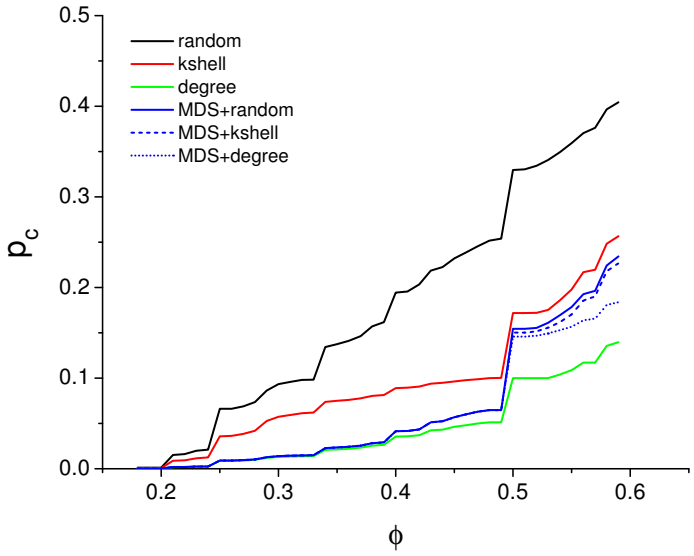

Figure 8.2: Critical fraction of initiators as a function of uniform adoption thresholds, for different initiator selection strategies. Synthetic scale-free networks (cCONF), with $N = 5000$, $\gamma = 2.5$, $\langle k \rangle = 8$, averaged over 20 samples.

degree-ranked order would be a viable alternative. Second, we realize that since the MDS is only a subset of nodes, it may not provide sufficient amount of initiators to reach a global cascade, especially when the adoption threshold ϕ is high. In this case we can fall back to other strategies for selecting the rest of initiators, until p_c is found. The fallback strategy is indicated after a "+" sign, e.g., "MDS+degree" denotes initiator selection by MDS nodes first, then the rest of nodes in degree-ranked order.

Figure 8.2 compares previously studied strategies (degree-ranked, k-shell-ranked and random node selections) [8] to our MDS strategy. Note, that the sudden jumps in the curve originates from the discreteness of node degrees, and not from small sample size. For example, in case of a degree 3 node, $\phi = 0.55$ is equivalent to $\phi = 0.65$, because in both cases, at least two neighbors must be activated first, to activate the given node. Quantitative changes in threshold behavior only occur at multiples of the reciprocal of the node degree, e.g., when $\phi = {}^1/_3$ or ${}^2/_3$, for a degree 3 node. Since low degree nodes are the most common in scale-free networks, these "threshold jumps" add up and become visible in p_c vs. ϕ curves. We have confirmed this observation by repeating the calculation for Fig. 8.2 and averaging over 2000

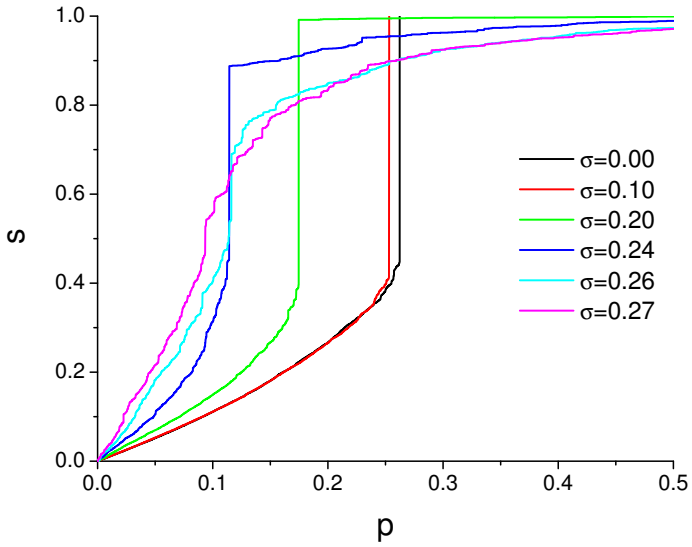

Figure 8.3: Cascade size vs. fraction of initiators. Adoption thresholds follow a truncated normal distribution, shown for various standard deviation σ values. Note, the truncated normal distribution converges to standard uniform distribution at $\sigma \approx 0.279$. Parameters: synthetic scale-free network (cCONF), $N = 10000$, $\gamma = 2.5$, $\langle k \rangle = 10$, mean threshold $\phi = 0.5$.

runs (each with a different network sample); the result was essentially identical.

We can avoid the discrete threshold jumps, and ease the comparison of strategies, by applying a normally distributed adoption threshold, $\phi \sim N(\mu, \sigma)$ (realized independently for each node), where we set the mean μ to match the uniform ϕ threshold that we used earlier. It is important to note that large standard deviation of the threshold distribution has a profound effect on the presence or absence of global cascades, see Fig. 8.3. However, we use small standard deviation ($\sigma = 0.04$), therefore we preserve the cascades. The resulting smoothness of the critical fractions allows for easier comparison of strategies, as shown in Fig. 8.4. Note, that the normal distribution must be truncated to $(0, 1)$ interval.

Using normally distributed thresholds, we compare initiator selection strategies on a wide range of network parameters, in Fig 8.5 for networks with $\langle k \rangle = 4$, and in Fig. 8.6 for networks with $\langle k \rangle = 8$. We can observe that for high γ power-law exponents, and for assortative networks, the MDS provides smaller, thus more

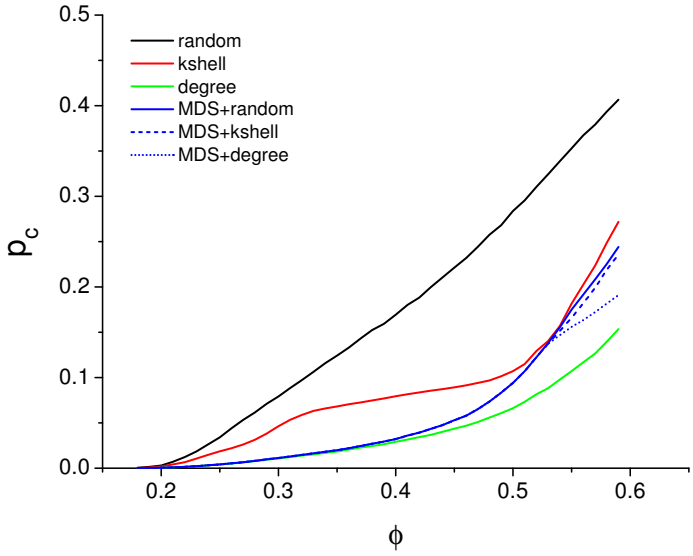

Figure 8.4: Critical fraction of initiators as a function of the mean of normally distributed adoption thresholds, for different initiator selection strategies. Standard deviation $\sigma = 0.04$. Synthetic scale-free networks (cCONF), with $N = 5000$, $\gamma = 2.5$, $\langle k \rangle = 8$, averaged over 20 samples.

efficient set of initiators than other strategies. Note, that these network parameters typically correspond to real social networks. In addition, the performance of MDS follows the degree-ranked selection (best among the other methods) very closely for other network parameters.

While we lack a theoretical reasoning for the effectiveness of MDS, we can conjecture that degree-based strategies have poor performance in assortative networks, because they are wasting initiators on high degree nodes that have few connections to low degree areas of the network, while the MDS is optimized for reaching every node equally. This is essentially the same explanation that we have for the increased size of degree-based dominating sets in assortative networks.

To summarize, we have shown that using MDS nodes as initiators in spreading dynamics is an efficient strategy, and it underlines the importance and applicability of dominating sets in complex networks.

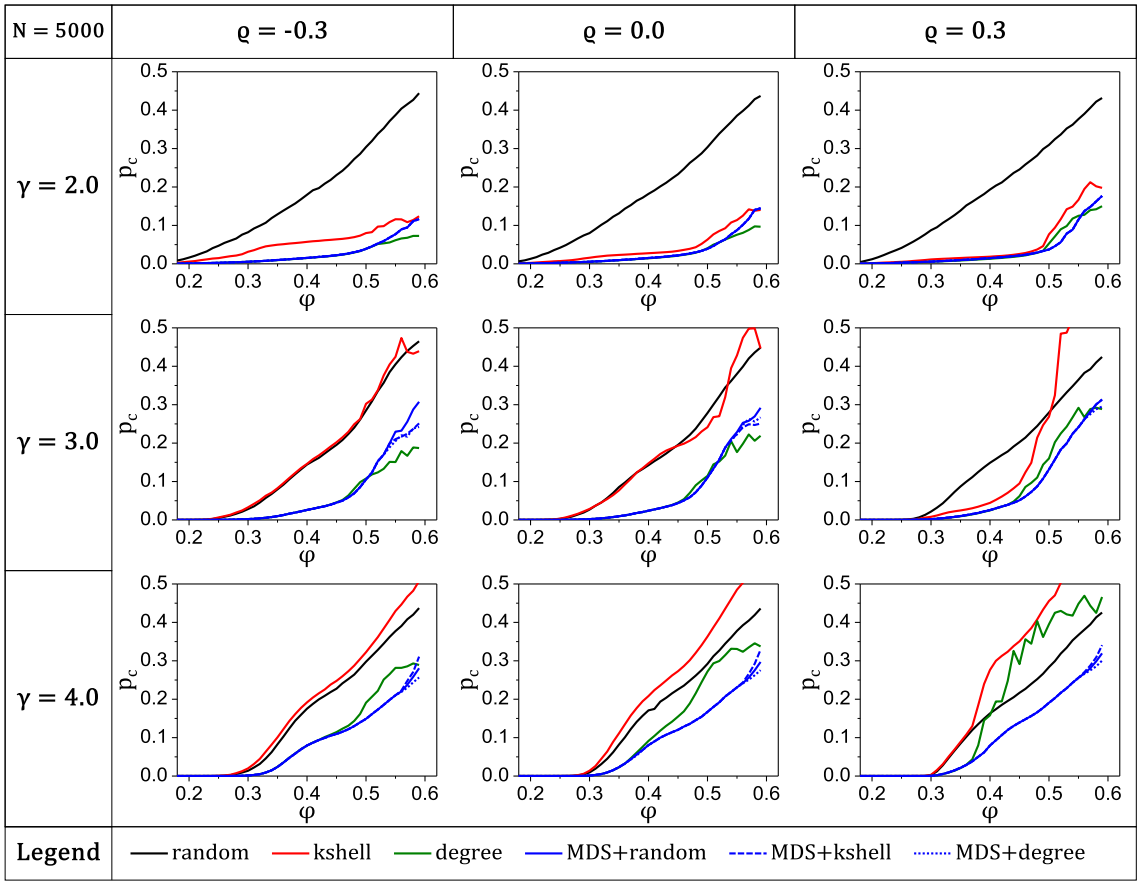

Figure 8.5: Critical fraction of initiators as a function of the mean of normally distributed adoption thresholds, over a wide range of power-law degree exponents (γ) and assortativities (Spearman's ρ). Standard deviation $\sigma = 0.04$. Synthetic scale-free networks (cCONF), with $N = 5000$, $\langle k \rangle = 4$, averaged over 20 samples.

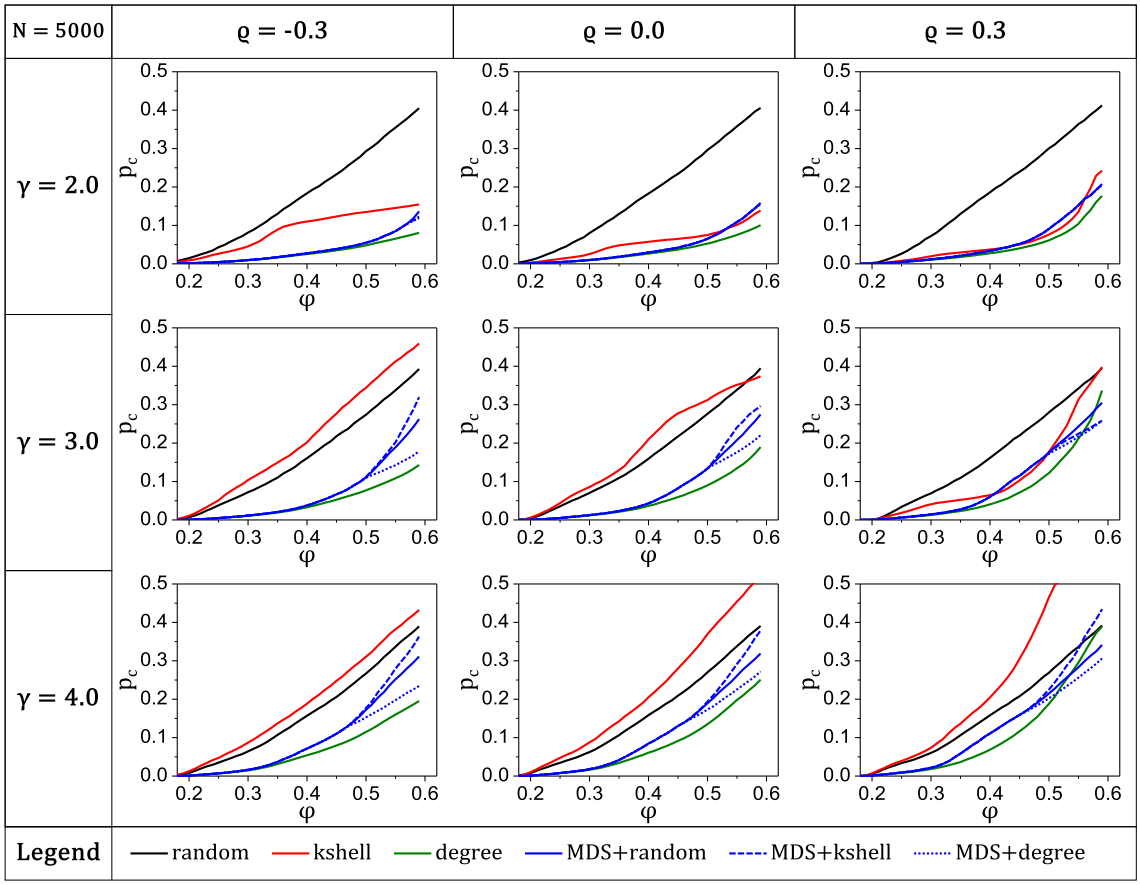

Figure 8.6: Critical fraction of initiators as a function of the mean of normally distributed adoption thresholds, over a wide range of power-law degree exponents (γ) and assortativities (Spearman's ρ). Standard deviation $\sigma = 0.04$. Synthetic scale-free networks (cCONF), with $N = 5000$, $\langle k \rangle = 8$, averaged over 20 samples.

CHAPTER 9
SUMMARY

We explored two distinct topics in complex systems analysis, around the common notion of finding the critical components of the system that can initiate global change, and provide efficient means of control. In the context of population dynamics, we studied the role of genetic sex ratio traits in competition dynamics and critical cluster behavior in spatial spreading. In the context of networks, we conducted a thorough analysis of the properties and applications of minimum dominating sets for observing, controlling, and influencing complex networks.

In the first part of this dissertation, we analyzed a theoretical population dynamics model that describes potential interactions between genetic sex-ratio traits, controlled by only one gender, and culturally transmitted sex-specific mortality traits. We found that there is a fundamental limit for sex ratio bias that can be sustained by counter-balancing mortalities. We also found that the coexistence of multiple mortality traits (in multiple coexisting groups) can successfully avert extinction, even if each single group could not sustain itself alone.

Our study of spatial effects and diffusive spreading in the same model revealed the presence of a critical cluster size, below which an initial population patch cannot achieve positive growth, which is closely related to the critical radius phenomenon in statistical physics. We also found that if the habitat is already occupied by a stable resident group, which can be exploited by an invader, then any small patch of invader can advance from rarity and exclude the resident. When a group invades an empty environment, which is desirable in ecological restoration, the cost of a successful spread becomes an important economical question. We introduced a novel application of simulated annealing for minimizing the total number of individuals in the initial patch, and thus minimizing the cost of successful restoration.

In future research, our theoretical models could be extended in two distinct directions. One possibility is specializing the model for a specific species, based on empirical data, in order to test the predicting capability of the underlying theoretical

principles. However, sex ratio dynamics and evolution could be futher studied in theoretical directions as well, by extending the basic assumptions of sex ratios and their interactions with other traits of organisms.

It would also be interesting to test our findings of ecological restoration in a real environment. However, one would have to consider the specific dynamics of growth of the given species being studied, and the implementation of non-uniform spatial distributions would have to be developed. Nonetheless, the theoretical approach of cost minimization is generally applicable to reaction-diffusion systems, which could have broader applications, e.g., in materials science and microstructure engineering.

In the second part of this dissertation, we conducted an extensive numerical study of minimum dominating sets (MDS) in model scale-free networks, using a greedy algorithm for finding approximate MDS solutions. This work also involved the development and implementation of several scale-free network construction models, and techniques for controlling the average degree and assortativity. Our work revealed a domination transition of MDS size from linear to $O(1)$ with respect to network size, as the degree exponent becomes less than 2, which leads to easily dominated networks, under certain structural conditions on the maximum realized degrees. Our contribution is not only finding this transition, but also providing a graph-theoretical explanation, and uncovering a relationship to the graphicality of power-law degree sequences.

We have explored the possibilities of using probabilistic node selection methods for selecting dominating sets, based on the method of Alon and Spencer [88]. Our work revealed that using a degree-dependent node selection probability leads to more efficient dominating sets. We also found a special limit of selection probabilities, where the node selection becomes deterministic, and the dominating set size is minimized, approaching the quality of state-of-the-art MDS approximations. Since this method only uses local information, it can be distributed and parallelized. In essence, it provides an alternative tool for finding small dominating sets, thereby contributing to their real-world applicability. Further, the special nature of this method provides an important theoretical insight of MDS approximation methods, as it forms a bridge between greedy and degree-ranked selection of dominating nodes.

We evaluated the damage resilience of dominating sets, i.e., the post-damage dominated fraction, in scale-free networks, and found that resilience can be improved by adding a few carefully selected nodes to the dominating set. This simple approach was effective against both random and targeted removal of nodes from the network. Based on this idea, we have developed two new methods that can select dominating sets with a flexible balance between resilience and set size (cost): one for providing a specific resilience level, another for maximizing resilience at a given cost level while considering the expected attack pattern.

As a proof of concept, we completed a short analysis of the linear threshold model, and found that using MDS nodes as the initiators of influence propagation can maximize opinion spread, outperforming previous methods, such as selection by degree or k-shell value, for assortative networks. This result holds even if the opinion thresholds are not uniform. In addition, we found that the standard deviation of the threshold distribution has a profound influence on the presence or absence of a global opinion cascade.

Our work revealed, on several occasions, that mixing patterns play a crucial role in network domination. Technological networks tend to be dissortative, and easily dominated, while social networks are assortative, and harder to dominate. Understanding the quantitative (theoretical) relationship between assortativity and MDS size, in the future, would be significant for designing domination strategies for real networks.

While the analysis of network damage can be viewed as a first step, domination, control, and influencing of dynamically changing network topologies remain unexplored. Interactions of the two dynamics (changing topology and changing dynamics on top of it) could provide new insights and depper understading of spatiotemporal pattern formation in complex networked systems.

REFERENCES

[1] G. Korniss and T. Caraco, "Spatial Dynamics of Invasion: The Geometry of Introduced Species," J. Theor. Biol. **233,** 137–150 (2005).

[2] L. O'Malley, J. Basham, J. A. Yasi, G. Korniss, A. Allstadt, and T. Caraco, "Invasive advance of an advantageous mutation: nucleation theory," Theor. Popul. Biol. **70,** 464–478 (2006).

[3] A. Allstadt, T. Caraco, and G. Korniss, "Ecological invasion: spatial clustering and the critical radius," Evol. Ecol. Res. **9,** 375–394 (2007).

[4] G. Gallavotti, *Statistical Mechanics: A Short Treatise* (Springer, Berlin, Heidelberg, New York, 1999).

[5] P.-J. Wan, K. M. Alzoubi, and O. Frieder, "Distributed Construction of Connected Dominating Set in Wireless Ad Hoc Networks," Mobile Netw. Appl. **9,** 141–149 (2004).

[6] Y. Yang, J. Wang, and A. E. Motter, "Network Observability Transitions," Phys. Rev. Lett. **109,** 258701 (2012).

[7] J. C. Nacher and T. Akutsu, "Analysis on Critical Nodes in Controlling Complex Networks Using Dominating Sets," in *2013 International Conference on Signal-Image Technology & Internet-Based Systems* (IEEE, New York, NY, 2013) pp. 649–654.

[8] P. Singh, S. Sreenivasan, B. K. Szymanski, and G. Korniss, "Threshold-limited spreading in social networks with multiple initiators," Sci. Rep. **3,** 2330 (2013).

[9] S. Eubank, H. Guclu, V. Anil Kumar, M. Marathe, A. Srinivasan, Z. Toroczkai, and N. Wang, "Modelling disease outbreaks in realistic urban social networks," Nature **429,** 180–184 (2004).

[10] L. Kelleher and M. Cozzens, "Dominating Sets in Social Network Graphs," Math. Social Sci. **16,** 267–279 (1988).

[11] F. Wang, H. Du, E. Camacho, K. Xu, W. Lee, Y. Shi, and S. Shan, "On positive influence dominating sets in social networks," Theoret. Comput. Sci. **412,** 265–269 (2011).

[12] P. Bak and M. Paczuski, "Complexity, contingency, and criticality," Proc. Natl. Acad. Sci. USA **92,** 6689–6696 (1995).

[13] N. Boccara, *Modeling Complex Systems*, 2nd ed. (Springer, New York, NY, 2010).

[14] T. Vicsek and A. Zafeiris, "Collective motion," Phys. Rep. **517,** 71–140 (2012).

[15] J. D. Murray, *Mathematical Biology*, Vol. 2. (Springer, New York, NY, 2003).

[16] I. R. Epstein and K. Showalter, "Nonlinear chemical dynamics: oscillations, patterns, and chaos," J. Phys. Chem. **100,** 13132–13147 (1996).

[17] B. Drossel and F. Schwabl, "Self-Organized Critical Forest-Fire Model," Phys. Rev. Lett. **69,** 1629–1632 (1992).

[18] S. H. White, A. M. del Rey, and G. R. Sánchez, "Modeling epidemics using cellular automata," Appl. Math. Comput. **186,** 193–202 (2007).

[19] L. Arnold, "Hasselmann's Program Revisited: The Analysis of Stochasticity in Deterministic Climate Models," in *Stochastic Climate Models*, edited by P. Imkeller and J.-S. von Storch (Birkhäuser, Basel, 2001), pp. 141–157.

[20] A. Liebscher, C. Proppe, C. Redenbach, and D. Schwarzer, "Stochastic multiscale modeling of metal foams," Procedia IUTAM **6,** 87–96 (2013).

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

[22] M. E. J. Newman, "The structure and function of complex networks," SIAM Rev. **45,** 167–256 (2003).

[23] S. Wuchty, "Controllability in protein interaction networks," Proc. Natl. Acad. Sci. USA **111,** 7156–7160 (2014).

[24] W. G. Roncal *et al.*, "MIGRAINE: MRI Graph Reliability Analysis and Inference for Connectomics," in *Proceedings of Global Conference on Signal and Information Processing (GlobalSIP 2013)* (IEEE, 2013) pp. 313–316.

[25] R. H. Kurvers, J. Krause, D. P. Croft, A. D. Wilson, and M. Wolf, "The evolutionary and ecological consequences of animal social networks: emerging issues," Trends Ecol. Evol. **29,** 326–335 (2014).

[26] M. O. Jackson, *Social and Economic Networks* (Princeton University Press, Princeton, NJ, 2008).

[27] R. W. Griffiths, D. W. Schloesser, J. H. Leach, and W. P. Kovalak, "Distribution and Dispersal of the Zebra Mussel (Dreissena polymorpha) in the Great Lakes Region," Can. J. Fish. Aquat. Sci. **48,** 1381–1388 (1991).

[28] R. J. Blaustein, "Kudzu's invasion into Southern United States life and culture," in *The Great Reshuffling: Human Dimensions of Invasive Alien Species*, edited by J. A. McNeely (IUCN—The World Conservation Union, Cambridge, UK, 2001) pp. 55–62.

[29] D. Pimentel, "Environmental consequences and economic costs of alien species," in *Invasive Plants: Ecological and Agricultural Aspects*, edited by P. Inderjit (Birkhäuser, Basel, 2005) pp. 269–276.

[30] D. B. Lindenmayer, A. D. Manning, P. L. Smith, H. P. Possingham, J. Fischer, I. Oliver, and M. A. McCarthy, "The focal-species approach and landscape restoration: a critique," Conserv. Biol. **16,** 338–345 (2002).

[31] R. J. Hall, "Restoration ecology," in *Encyclopedia of Theoretical Ecology*, edited by A. Hastings and L. J. Gross (Univeristy of California Press, Berkeley, CA, 2012) pp. 629–632.

[32] K. D. Holl KD and R. B. Howarth, "Paying for restoration," Restor. Ecol. **8,** 260–267 (2000).

[33] K. Shea and H. P. Possingham, "Optimal release strategies for biological control agents: an application of stochastic dynamic programming to population management," J. Appl. Ecol. **37,** 77–86 (2000).

[34] New York State Department of Environmental Conservation, "Canada Lynx," http://www.dec.ny.gov/animals/6980.html (Accessed November 01, 2012).

[35] M. R. Evans *et al.* "Do simple models lead to generality in ecology?," Trends Ecol. Evol. **28,** 578–583 (2013).

[36] E. A. Codling and A. J. Dumbrell, "Mathematical and theoretical ecology: linking models with ecological processes," Interface Focus **2,** 144–149 (2012).

[37] L. J. Jackson, A. S. Trebitz, and K. L. Cottingham, "An Introduction to the Practice of Ecological Modeling," BioScience **50,** 694–706 (2000).

[38] J. Molofsky and J. D. Bever, "A new kind of ecology?," BioScience **54,** 440–446 (2004).

[39] D. Grünbaum, "The logic of ecological patchiness," Interface Focus **2,** 150–155 (2012).

[40] J. Osenbaugh, "Idealized, inaccurate but successful: A pragmatic approach to evaluating models in theoretical ecology," Biol. Philos. **20,** 231–255 (2005).

[41] R. F. Shaw and J. D. Mohler, "The selective significance of the sex ratio," Am. Nat. **87,** 337–342 (1953).

[42] I. Eshel, "Selection on sex-ratio and the evolution of sex-determination," Heredity **34,** 351–361 (1975).

[43] S. Karlin and S. Lessard, *Theoretical Studies on Sex Ratio Evolution* (Princeton University Press, Princeton, NJ, 1986).

[44] H. Caswell and D. E. Weeks, "Two-sex models: chasos, extinction, and other dynamic consequences of sex," Am. Nat. **128,** 707–735 (1986).

[45] A. C. Ashih and W. G. Wilson, "Two-sex population dynamics in space: effects of gestation time on persistence," Theor. Popul. Biol. **60,** 93–106 (2001).

[46] S. Engen, R. Lande, and B.-E. Sæther, "Demographic stochasticity and Allee effects in populations with two sexes," Ecology **84,** 2378–2386 (2003).

[47] T. E. X. Miller, A. K. Shaw, B. D. Inouye, and M. G. Neubert, "Sex-biased dispersal and the speed of two-sex invasions," Am. Nat. **177,** 549–561 (2011).

[48] W. D. Hamilton, "Extraordinary sex ratios," Science **156,** 477–488 (1967).

[49] R. A. Fisher, *The Genetical Theory of Natural Selection* (Clarendon Press, Oxford, UK, 1930).

[50] J. Kumm, K. N. Laland, and M. W. Feldman, "Gene-culture coevolution and sex ratios: the effects of infanticide, sex-selective abortion, sex selection, and sex-biased parental investment on the evolution of sex ratios," Theor. Popul. Biol. **46,** 249–278 (1994).

[51] M. Dickemann, "Demographic consequences of infanticide in man," Ann. Rev. Ecol. Syst. **6,** 107–137 (1975).

[52] G. Hausfater and S. B. Hrdy, *Infanticide: Comparative and Evolutionary Perspectives* (Aldine Transaction, Hawthorne, NY, 1984).

[53] K. N. Laland, J. Kumm, and Feldman, "Gene-Culture Coevolutionary Theory: A Test Case," Curr. Anthropol. **36,** 131–156 (1995).

[54] M. D. Gupta, "Explaining Asia's 'Missing Women': A New Look at the Data," Popul. Dev. Rev. **31,** 529–535 (2005).

[55] J. M. Diamond, "Learned specializations of birds," Nature **330,** 16–17 (1987).

[56] E. V. Lonsdorf, L. E. Eberly, and A. E. Pusey, "Sex differences in learning in chimpanzees," Nature **428,** 715–716 (2004).

[57] M. Krützen, J. Mann, M. R. Heithaus, R. C. Connor, L. Bejder, and W. R. Sherwin, "Cultural transmission of tool use in bottlenose dolphins," Proc. Natl. Acad. Sci. USA **102,** 8939–8943 (2005).

[58] T. Slagsvold and K. L. Wiebe, "Social learning in birds and its role in shaping a foraging niche," Philos. Trans. R. Soc. B **366,** 969–977 (2011).

[59] P. A. Rikvold, H. Tomita, S. Miyashita, and S. W. Sides, "Metastable lifetimes in a kinetic Ising model: dependence on field and system size," Phys. Rev. E **49,** 5080–5090 (1994).

[60] R. A. Ramos, P. A. Rikvold, and M. A. Novotny, "Test of the Kolmogorov-Johnson-Mehl-Avrami picture of meta-stable decay in a model with microscopic dynamics," Phys. Rev. B **59,** 9053–9069 (1999).

[61] E. Machado, G. M. Buendía, and P. A. Rikvold, "Decay of metastable phases in a model for the catalytic oxidation of CO," Phys. Rev. E **71,** 031603 (2005).

[62] A. Gandhi, S. Levin, and S. Ország, "Nucleation and relaxation from meta-stability in spatial ecological models," J. Theor. Biol. **200,** 121–146 (1999).

[63] D. J. Watts and S. H. Strogatz, "Collective dynamics of 'small-world' networks," Nature **393,** 440–442 (1998).

[64] S. H. Strogatz, "Exploring Complex Networks," Nature **410,** 268–276 (2001).

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

[66] A.-L. Barabási, R. Albert, and H. Jeong, "Scale-free characteristics of random networks: the topology of the world wide web," Physica A **281,** 69–77 (2000).

[67] A. Barabasi and E. Bonabeau, "Scale-Free Networks," Sci. Am. **288,** 50–59 (2003).

[68] R. Cohen and S. Havlin, "Scale-free networks are ultrasmall," Phys. Rev. Lett. **90,** 058701 (2003).

[69] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach* (Cambridge University Press, Cambridge, UK, 2004).

[70] B. Bollobás, *Random Graphs*, 2nd Ed. (Cambridge University Press, Cambridge, UK, 2001).

[71] R. Cohen and S. Havlin, *Complex Networks: Structure, Robustness and Function* (Cambridge University Press, Cambridge, UK, 2010).

[72] Y.-Y. Liu, J.-J. Slotine, and A.-L. Barabási, "Controllability of complex networks," Nature **473,** 167–173 (2011).

[73] N. J. Cowan, E. J. Chastain, D. A. Vilhena, J. S. Freudenberg, and C. T. Bergstrom, "Nodal Dynamics, Not Degree Distributions, Determine the Structural Controllability of Complex Networks," PLoS One **7,** e38398 (2012).

[74] T. Nepusz and T. Vicsek, "Controlling edge dynamics in complex networks," Nat. Phys. **8,** 568–573 (2012).

[75] T. Jia, Y.-Y. Liu, E. Csóka, M. Pósfai, J.-J. Slotine, and A.-L. Barabási, "Emergence of bimodality in controlling complex networks," Nat. Commun. **4,** 2002 (2013).

[76] T. Jia and A.-L. Barabási, "Control Capacity and a Random Sampling Method in Exploring Controllability of Complex Networks," Sci. Rep. **3,** 2354 (2013).

[77] B. Wang *et al.*, "Maintain the structural controllability under malicious attacks on directed networks," Europhys. Lett. **101,** 58003 (2013).

[78] Z. Yuan, C. Zhao, Z. Di, W. X. Wang, and Y. C. Lai, "Exact controllability of complex networks," Nat. Commun. **4,** 2447 (2013).

[79] S. P. Cornelius, W. L. Kath, and A. E. Motter, "Realistic control of network dynamics," Nat. Commun. **4,** 1942 (2013).

[80] J. Sun and A. E. Motter, "Controllability transition and nonlocality in network control," Phys. Rev. Lett. **110,** 208701 (2013).

[81] J. Ruths and D. Ruths, "Control Profiles of Complex Networks," Science **343,** 1373–1376 (2014).

[82] J. C. Nacher and T. Akutsu, "Structural controllability of unidirectional bipartite networks," Sci. Rep. **3,** 1647 (2013).

[83] T. Milenković, V. Memišević, A. Bonato, and N. Pržulj, "Dominating Biological Networks," PLoS One **6,** e23016 (2011).

[84] P. Echenique, J. Gómez-Gardeñes, Y. Moreno, and A. Vázquez, "Distance-d covering problems in scale-free networks with degree correlations," Phys. Rev. E **71,** 035102(R) (2005).

[85] P. Gogas, T. Papadimitriou, and M.-A. Matthaiou, "A Novel Banking Supervision Method Using the Minimum Dominating Set," in *Network Models in Economics and Finance*, edited by V. A. Kalyagin, P. M. Pardalos, and T. M. Rassias (Springer, Switzerland, 2014) pp. 277–286.

[86] C. Yang, X. Li, and J. Wu, "Dominating-Set-Based Searching in Peer-to-Peer Networks," in *Grid and Cooperative Computing*, edited by M. Li, X.-H. Sun, Q. Deng, and J. Ni (Springer, Berlin, Heidelberg, 2004) pp. 332–339.

[87] R. Raz and S. Safra, "A Sub-Constant Error-Probability Low-Degree Test, and a Sub-Constant Error-Probability PCP Characterization of NP," in *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing* (ACM, New York, NY, 1997) pp. 475–484.

[88] N. Alon and J. H. Spencer, *The Probabilsitic Method*, 2nd Ed. (Willey, New York, NY, 2000).

[89] N. Alon, "Transversal numbers of uniform hypergraphs," Graphs Combin. **6,** 1–4 (1990).

[90] V. I. Arnautov, "Estimation of the exterior stability number of a graph by means of the minimal degrees of the vertices," (in Russian), Prikl. Mat. i Prog. **11,** 3–8, 126 (1974).

[91] A. H. Land and A. G. Doig, "An automatic method of solving discrete programming problems," Econometrica **28,** 497–520 (1960).

[92] J. C. Nacher and T. Akutsu, "Dominating scale-free networks with variable scaling exponent: heterogeneous networks are not difficult to control," New J. Phys. **14,** 073005 (2012).

[93] A Potluri and A. Singh, "Two Hybrid Meta-heuristic Approaches for Minimum Dominating Set Problem," in *Swarm, Evolutionary, and Memetic Computing*, edited by B. K. Panigrahi, P. N. Suganthan, S. Das, and S. C. Satapathy (Springer, Berlin, Heidelberg, 2011) pp. 97–104.

[94] A. R. Hedar and R. Ismail, "Hybrid Genetic Algorithm for Minimum Dominating Set Problem," in *Computational Science and Its Applications – ICCSA 2010*, edited by D. Taniar, O. Gervasi, B. Murgante, E. Pardede, and B. O. Apduhan (Springer, Berlin, Heidelberg, 2010) pp. 457–467.

[95] J. Blum, M. Ding, A. Thaeler, and X. Cheng, "Connected Dominating Set in Sensor Networks and MANETs," in *Handbook of Combinatorial Optimization*, edited by D.-Z. Du and P. M. Pardalos (Springer US, New York, NY, 2005) pp. 329–369.

[96] L. Ruan, H. Du, X. Jia, W. Wu, Y. Li, and K.-I. Ko, "A greedy approximation for minimum connected dominating sets," Theoret. Comput. Sci. **329,** 325–330 (2004).

[97] Q. Chen, W. T. Fan, and M. Zhang, "Distributed heuristic approximation algorithm for minimum connected dominating set," Comput. Engin. **35,** 92–94 (2009).

[98] L. Simonetti, A. S. da Cunha, and A. Lucena, "The Minimum Connected Dominating Set Problem: Formulation, Valid Inequalities and a

Branch-and-Cut Algorithm," in *Network Optimization*, edited by J. Pahl, T. Reiners, and S. Voß (Springer, Berlin, Heidelberg, 2011) pp. 162–169.

[99] S. Gaspers, M. Liedloff, "A branch-and-reduce algorithm for finding a minimum independent dominating set in graphs," in *Graph-Theoretic Concepts in Computer Science*, edited by F. V. Fomin (Springer, Berlin, Heidelberg, 2006) pp. 78–89.

[100] N. Bourgeois, F. Della Croce, B. Escoffier, and V. Th. Paschos, "Fast algorithms for min independent dominating set," Discrete Appl. Math. **161,** 558–572 (2013).

[101] W. Goddard, and M. A. Henning, "Independent domination in graphs: A survey and recent results," Discrete Math. **313,** 839–854 (2013).

[102] F. V. Fomin, F. Grandoni, A. V. Pyatkin, and A. A. Stepanov, "Combinatorial Bounds via Measure and Conquer: Bounding Minimal Dominating Sets and Applications," ACM Trans. Algorithms **5,** 9 (2008).

[103] F. V. Fomin, F. Grandoni, and D. Kratsch, "A Measure & Conquer Approach for the Analysis of Exact Algorithms," J. ACM **56,** 25 (2009).

[104] J. M. M. van Rooij, J. Nederlof, and T. C. van Dijk, "Inclusion/Exclusion Meets Measure and Conquer: Exact Algorithms for Counting Dominating Sets," in *Algorithms—ESA 2009*, edited by A. Fiat and P. Sanders (Springer, Berlin, Heidelberg, 2009) pp. 554–565.

[105] J. M. M. van Rooij and H. L. Bodlaender, "Exact algorithms for dominating set," Discrete Appl. Math. **159,** 2147–2164 (2011).

[106] R. Albert, H. Jeong, and A.-L. Barabási, "Error and attack tolerance of complex networks," Nature **406,** 378–382 (2000).

[107] J. Duch and A. Arenas, "Effect of random failures on traffic in complex networks," in *Noise and Stochastics in Complex Systems and Finance*, edited by J. Kertész, S. Bornholdt, and R. N. Mantega (SPIE, Bellingham, WA, 2007).

[108] L. K. Gallos, R. Cohen, P. Argyrakis, A. Bunde, and S. Havlin, "Stability and topology of scale-free networks under attack and defense strategies," Phys. Rev. Lett. **94,** 188701 (2005).

[109] P. Holme and B. J. Kim, "Attack vulnerability of complex networks," Phys. Rev. E **65,** 056109 (2002).

[110] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, "Resilience of the Internet to Random Breakdowns," Phys. Rev. Lett. **85,** 4626–4628 (2000).

[111] R. Cohen, K. Erez, D. ben-Avraham, and S. Havlin, "Breakdown of the Internet under Intentional Attack," Phys. Rev. Lett. **86,** 3682–3685 (2001).

[112] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, "Network Robustness and Fragility: Percolation on Random Graphs," Phys. Rev. Lett. **85,** 5468–5471 (2000).

[113] T. Tanizawa, "Structural robustness and transport efficiency of complex networks with degree correlation," arXiv:1209.4897 [physics.soc-ph] (Accessed May 05, 2014).

[114] G. Paul, T. Tanizawa, S. Havlin, and H. E. Stanley, "Optimization of robustness of complex networks," Eur. Phys. J. B **38,** 187–191 (2004).

[115] T. Tanizawa, G. Paul, R. Cohen, S. Havlin, and H. E. Stanley, "Optimization of network robustness to waves of targeted and random attacks," Phys. Rev. E **71,** 047101 (2005).

[116] Y. Hayashi and T. Miyazaki, "Emergent rewirings for cascades on correlated networks," arXiv:cond-mat/0503615 [cond-mat.dis-nn] (Accessed May 09, 2014).

[117] A. Asztalos, S. Sreenivasan, B. K. Szymanski, and G. Korniss, "Cascading failures in spatialy-embedded random networks," PLoS One **9,** e84563 (2014).

[118] M. E. J. Newman, "Assortative mixing in networks," Phys. Rev. Lett. **89,** 208701 (2002).

[119] M. E. J. Newman, "Mixing patterns in networks," Phys. Rev. E **67,** 026126 (2003).

[120] P. F. Lazarsfeld and R. K. Merton, *Freedom and Control in Modern Society* (Van Nostrand, New York, NY, 1954), Vol. 18, pp. 18–66.

[121] V. Eguíluz and K. Klemm, "Epidemic Threshold in Structured Scale-Free Networks," Phys. Rev. Lett. **89,** 108701 (2002).

[122] S. Fortunato, M. Boguñá, A. Flammini, and F. Menczer, "On Local Estimations of PageRank: A Mean Field Approach," Internet Math. **4,** 245–266 (2007).

[123] L. Li, D. Alderson, J. Doyle, and W. Willinger, "Towards a theory of scale-free graphs: Definition, properties, and implications," Internet Math. **2,** 431–523 (2005).

[124] M. Granovetter, "Threshold models of collective behavior," Am. J. Sociol. **83,** 1420 (1978).

[125] B. Latané and T. L'Herrou, "Spatial clustering in the conformity game: dynamic social impact electronic groups," J. Pers. Soc. Psychol. **70,** 1218–1230 (1996).

[126] D. J. Watts, "A simple model of global cascades on random networks," Proc. Natl. Acad. Sci. USA **99,** 5766 (2002).

[127] D. J. Watts and P. S. Dodds, "Influentials, networks, and public opinion formation," J. Consum. Res. **34,** 441 (2007).

[128] E. L. Charnov, *The Theory of Sex Allocation* (Princeton University Press, Princeton, NJ, 1982).

[129] J. Jaenike "Sex chromosome meiotic drive," Ann. Rev. Ecol. Syst. **32,** 25–49 (2001).

[130] K. Tainaka, T. Hayashi, and J. Yoshimura, "Sustainable sex ratio in lattice populations," Europhys. Lett. **74,** 554–559 (2006).

[131] S. A. West, *Sex Allocation* (Princeton University Press, Princeton, NJ, 2009).

[132] L. Aviles, P. Abbot, and A. D. Cutter, "Population ecology, nonlinear dynamics, and social evolution. I. Associations among nonrelatives," Am. Nat. **159,** 115–127 (2002).

[133] D. S. Boukal and L. Berec, "Single-species models of the Allee effect: extinction boundaries, sex ratios and mate encounters," J. Theor. Biol. **218,** 375–394 (2002).

[134] L. Berec, E. Angulo, and F. Courchamp, "Multiple Allee effects and population management," Trends Ecol. Evol. **22,** 185–191 (2006).

[135] L. Berec, D. S. Boukal, and M. Berec, "Linking the Allee effect, sexual reproduction, and temperature-dependent sex determination via spatial dynamics," Am. Nat. **157,** 217–230 (2001).

[136] L. L. Cavalli-Sforza and M. W. Feldman, *Cultural Transmission and Evolution: A Quantitative Approach* (Princeton University Press, Princeton, NJ, 1981).

[137] P. Amarasekare, "Competitive coexistence in spatially structured environments: a synthesis," Ecol. Lett. **6,** 1109–1122 (2003).

[138] A. Allstadt, T. Caraco, and G. Korniss, "Preemptive spatial competition under a reproduction-mortality constraint," J. Theor. Biol. **258,** 537–549 (2009).

[139] B. M. Going, J. Hillerislambers, and J. M. Levine, "Abiotic and biotic resistance to grass invasion in serpentine annual plant communities," Oecologia **159,** 839–847 (2009).

[140] E. Sober, *The Nature of Selection: Evolutionary Theory in Philosophical Focus* (MIT Press, Cambridge, MA, 1984)

[141] I. M. Hastings, "Manifestations of sexual selection may depend on the genetic basis of sex determination," Proc. R. Soc. B **258,** 83–87 (1994).

[142] B. M. Appleby, S. J. Petty, J. K. Blakey, P. Rainey, and D. MacDonald, "Does variation of sex ratio enhance reproductive success of offspring in tawny owls (*Strix aluco*)?," Proc. R. Soc. B **264,** 1111–1116 (1997)

[143] J. Komdeur, D. Daan, J. Tinbergen, and C. Mateman, "Extreme adaptive mofification in sex ratio of the Seychelles warbler's eggs," Nature **385,** 522–525 (1997)

[144] C. Dijkstra, S. Daan, and J. B. Buker, "Adaptive seasonal variation in the sex ratio of kestrel broods," Funct. Ecol. **4,** 143–147 (1990).

[145] A. C. Chandley and H. J. Cooke, "Human male fertility—Y-linked genes and spermatogenesis," Hum. Mol. Genet. **3,** 1449–1452 (1994).

[146] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipies in C: The Art of Scientific Computing*, 2nd ed. (Cambridge University Press, New York, NY, 1992).

[147] A. Gandhi, S. Levin, and S. Orszag, "Nucleation and relaxation from meta-stability in spatial ecological models," J. Theor. Biol. **200,** 121–146 (1999).

[148] B. Oborny, G. Meszéna, and G. Szabó, "Dynamics of populations on the verge of extinction," Oikos **109,** 291–296 (2005)

[149] W. E. Schiesser, *The Numerical Method of Lines: Integration of Partial Differential Equations* (Academic Press, San Diego, CA, 1991).

[150] A. J. McKane and T. J. Newman, "Stochastic models in population biology and their deterministic analogues," Phys. Rev. E **70,** 041902 (2004).

[151] G. Korniss, B. Schmittmann, and R. K. P. Zia, "Novel phase transitions in biased diffusion of two species," Europhys. Lett. **32,** 49–54 (1995).

[152] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam, 1981).

[153] H. Hinrichsen, "Non-equilibrium critical phenomena and phase transitions into absorbing states," Adv. Phys. **49,** 815–958 (2000).

[154] M. A. Lewis and P. Kareiva, "Allee dynamics and the spread of invading organisms," Theor. Popul. Biol. **43,** 141–158 (1993).

[155] S. A. Frank, "Hierarchical selection theory and sex ratios I. General solutions for structured populations," Theor. Popul. Biol. **29,** 312–342 (1986).

[156] M. K. Uyenoyama and B. O. Bengtsson, "Towards a genetic theory for the evolution of the sex ratio. III. Parental and sibling control of brood investment ratio under partial sib-mating," Theor. Popul. Biol. **22,** 43–68 (1982).

[157] T. Nitta, K.-I. Tainaka, Y. Sakisaka, B. Saito, T. Togashi, and J. Yoshimura, "Lattice population and optimality of sex ratio: effect of sterile male," in *Cellular Automata*, edited by H. Umeo, S. Morishita, K. Nishinari, T. Komatsuzaki, and S. Bandini, (Springer, Berlin, Heidelberg, 2008) pp. 368–373.

[158] T. E. Harris, "Contact interaction on a lattice," Ann. Probab. **2,** 969–988 (1974)

[159] T. M. Liggett, *Stochastic Interacting Systems: Contact, Voter, and Exclusion Processes* (Springer, New York, NY, 1999).

[160] J. Marro and R. Dickman, *Nonequilibrium Phase Transition in Lattice Models* (Cambridge University Press, Cambridge, UK, 1999).

[161] R. Lande, "Demographic stochasticity and Allee effect on a scale with isotropic noise," Oikos **83,** 353–358 (1998).

[162] A. O. Shelton, "The ecological and evolutionary drivers of female-biased sex ratios: two-sex models of perennial seagrasses," Am. Nat. **175,** 302–315 (2010).

[163] F. Molnár Jr., C. Caragine, T. Caraco, and G. Korniss, "Restoration Ecology: Two-Sex Dynamics and Cost Minimization," PLoS One **8,** e77332 (2013).

[164] E. E. Holmes, M. A. Lewis, J. E. Banks, and R. R. Veit, "Partial differential equations in ecology: spatial interactions and population dynamics," Ecology **75,** 17–29 (1994).

[165] T. Schmickl and I. Karsai, "The interplay of sex ratio, male success and density-independent mortality affects population dynamics," Ecol. Model. **221,** 1089–1097 (2010).

[166] F. Molnár Jr., T. Caraco, and G. Korniss, "Extraordinary sex ratios: cultural effects on ecological consequences," PLoS One **7,** e43364 (2012)

[167] T. H. Keitt, M. A. Lewis, and R. D. Holt, "Allee effects, invasion pinning, and species' borders," Am. Nat. **157,** 203–216 (2001).

[168] F. Courchamp, T. Clutton-Brick, and B. Grenfell, "Inverse density dependence and the Allee effect," Trends Ecol. Evol. **14,** 405–410 (1999).

[169] T. Caraco, G. W. Uetz, R. G. Gillespie, and L.-A. Giraldeau, "Resource consumption variance within and among individuals: on coloniality in spiders," Ecology **76,** 196–205 (1995).

[170] P. A. Stephens and W. J. Sutherland, "Consequences of the Allee effect for behaviour, ecology and conservation," Trends Ecol. Evol. **14,** 401–405 (1999).

[171] G. Korniss, "Structure factors and their distributions in driven two-species models," Phys. Rev. E **56,** 4072–4084 (1997).

[172] C. Escudero, J. Buceta, F. J. de la Rubia, and K. Lindenberg, "Extinction in population dynamics," Phys. Rev. E **69,** 021908 (2004).

[173] S. Pigolotti, R. Benzi, P. Perlekar, M. H. Jensen, F. Toschi, and D. R. Nelson, "Growth, competition and cooperation in spatial population genetics," Theor. Popul. Biol. **84,** 72–86 (2013).

[174] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry and the Natural Sciences* (Springer, Berlin, Heidelberg, 1985) p. 442.

[175] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (Elsevier Science Ltd., Amsterdam, 1981) p. 434.

[176] B. Schmittmann and R. K. P. Zia, *Statistical Mechanics of Driven Diffusive Systems,* Vol. 17, Phase Transitions and Critical Phenomena (Academic Press, New York, NY, 1995) p. 220.

[177] M. Doi, "Stochastic theory of diffusion-controlled reaction," J. Phys. A **9,** 1479–1495 (1976).

[178] U. C. Täuber, M. Howard, and B. P. Vollmayr-Lee, "Applications of field-theoretic renormalization group methods to reaction-diffusion problems," J. Phys. A **38,** R79–R131 (2005).

[179] U. C. Täuber, "Population oscillations in spatial stochastic Lotka-Volterra models: a field-theoretic perturbational analysis," J. Phys. A **45,** 405002 (2012).

[180] S. V. Petrovskii and B.-L. Li, *Exactly Solvable Models of Biological Invasion* (Chapman & Hall/CRC, Boca Raton, FL, 2006) pp. 102–116.

[181] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, "Optimization by simulated annealing," Science **220,** 671–680 (1983).

[182] NVIDIA Corp., *CUDA C Programming Guide*, http://docs.nvidia.com/cuda/cuda-c-programming-guide/ (Accessed June 01, 2013).

[183] F. Molnár Jr., F. Izsák, R. Mészáros, and I. Lagzi, "Simulation of reaction-diffusion processes in three dimensions using CUDA," Chemom. Intell. Lab. Syst. **108,** 76–85 (2011).

[184] B. Dennis, "Allee effects: population growth, critical density, and chance of extinction," Nat. Resour. Model. **3,** 481–538 (1989).

[185] T. Caraco, M. C. Duryea, S. Glavanakov, W. Maniatty, and B. K. Szymanski, "Host spatial heterogeneity and the spread of vector-borne infection," Theor. Popul. Biol. **59,** 185–206 (2001).

[186] S. P. Ellner, A. Sasaki, A. Y. Haraguchi, and H. Matsuda, "Speed of invasion in lattice population models: pair-edge approximation," J. Math. Biol. **36,** 469–484 (1998).

[187] L. O'Malley, B. Kozma, G. Korniss, Z. Rácz, and T. Caraco, "Fisher waves and front roughening in a two-species invasion model with preemptive competition," Phys. Rev. E **74,** 041116 (2006).

[188] L. O'Malley, G. Korniss, and T. Caraco, "Ecological invasion, roughened fronts, and a competitor's extreme advance: integrating stochastic spatial-growth models," Bull. Math. Biol. **71,** 1160–1188 (2009).

[189] E. Moro, "Emergence of pulled fronts in fermionic microscopic particle models," Phys. Rev. E **68,** 025102(R) (2003).

[190] T. Caraco, M. Duryea, G. Gardner, W. Maniatty, and B. K. Szymanski, "Host spatial heterogeneity and extinction of an SIS epidemic," J. Theor. Biol. **192,** 351–361 (1998).

[191] G. R. Robinson and S. N. Handel, "Directing spatial patterns of recruitment during an experimental urban woodland reclamation," Ecol. Appl. **10,** 174–188 (2000).

[192] J. M. Levine, P. B. Adler, and S. G. Yelenik, "A meta-analysis of biotic resistance to exotic plant invasions," Ecol. Lett. **7,** 975–989 (2004).

[193] M. Buckley and B. M. Haddad, "Socially Strategic Ecological Restoration: A Game-Theoretic Analysis Shortened: Socially Stratgeic Restoration," Environ. Manage. **38,** 48–61 (2006).

[194] S. N. Dorogovtsev, J. F. F. Mendes, and A. N. Samukhin, "Exact solution of the Barabási-Albert model," Phys. Rev. Lett. **85,** 4633 (2000).

[195] P. L. Krapivski, S. Redner, and F. Leyvraz, "Connectivity of growing random networks," Phys. Rev. Lett. **85,** 4629 (2000).

[196] M. Molloy and B. Reed, "A Critical Point for Random Graphs with a Given Degree Sequence," Random Structures Algorithms **6,** 161–180 (1995).

[197] T. Britton, M. Deijfen, and A. Martin-Löf, "Generating simple random graphs with prescribed degree distribution," J. Stat. Phys. **124,** 1377–1397 (2005).

[198] F. Viger and M. Latapy, "Efficient and simple generation of random simple connected graphs with prescribed degree sequence," in *Computing and Combinatorics*, edited by L. Wang (Springer, Berlin, Heidelberg, 2005) pp. 440–449.

[199] V. Havel, "A remark on the existence of finite graphs," (in Chech), Casopis Pest. Mat. **80,** 477–480 (1955).

[200] S. L. Hakimi, "On the realizability of integers as the degrees of the vertices of a linear graph I.," J. Soc. Ind. Appl. Math. **10,** 496–506 (1962).

[201] C. I. del Genio, H. Kim, Z. Toroczkai, and K. E. Bassler, "Efficient and Exact Sampling of Simple Graphs with Given Arbitrary Degree Sequence," PLoS One **5,** e10012 (2010).

[202] H. Kim, Z. Toroczkai, P. L. Erdős, I. Miklós, and L. A. Székely, "Degree-based graph construction," J. Phys. A **42,** 392001 (2009).

[203] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, "Cut-offs and finite size effects in scale-free networks," Eur. Phys. J. B **38,** 205–509 (2004).

[204] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, "Generation of uncorrelated random scale-free networks," Phys. Rev. E **71,** 027103, (2005).

[205] R. Taylor, "Switchings Constrained to 2-Connectivity in Simple Graphs," SIAM J. Algebra Discr. **3,** 114–121 (1982).

[206] J. Ray, A. Pinar, and C. Seshadhri, "Are we there yet? When to stop a markov chain while generating random graphs?," in *9th Workshop on Algorithms and Models for the Web Graph* (Springer, New York, NY, 2012) pp. 155–164.

[207] L. Tabourier, C. Roth, and J. Cointet, "Generating constrained random graphs using multiple edge switches," ACM J. Exp. Algorithmics **16,** 205–217 (2011).

[208] I. Stanton and A. Pinar, "Constructing and sampling graphs with a prescribed joint degree distribution," ACM J. Exp. Algorithmics **17,** Article No. 3.5 (2012).

[209] R. E. Tarjan, "A note on finding the bridges of a graph," Inform. Process. Lett. **2,** 160–161 (1974).

[210] H. Bauke, "Parameter estimation for power-law distributions by maximum likelihood methods," Eur. Phys. J. B **58,** 167–173 (2007).

[211] A. Clauset, C. R. Shalizi, and M. E. J. Newman, "Power-law distributions in empirical data," SIAM Rev. **51,** 661–703 (2009).

[212] M. L. Goldstein, S. A. Morris, and G. G. Yen, "Problems with fitting to the power-law distribution," Eur. Phys. J. B **41,** 255–258 (2004).

[213] A. K. Parekh, "Analysis of a greedy heuristic for finding small dominating sets in graphs," Inform. Process. Lett. **39,** 237–240 (1991).

[214] N. Litvak and R. van der Hofstad, "Uncovering disassortativity in large scale-free networks," Phys. Rev. E **87,** 022801 (2013).

[215] C. Spearman, "The Proof and Measurement of Association between Two Things," Am. J. Psychol. **15,** 72–101 (1904).

[216] C. B. Borkowf, "Computing the nonnull asymptotic variance and the asymptotic relative efficiency of Spearman's rank correlation," Comput. Statist. Data Anal. **39,** 271–286 (2002).

[217] S. Eubank, V. S. Anil Kumar, M. V. Marathe, A. Srinivasan, and N. Wang, "Structural and algorithmic aspects of massive social networks," in *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms* (Society for Industrial and Applied Mathematics, Philadelphia, PA, 2004) pp. 718–727.

[218] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener, "Graph structure in the web," Comput. Netw. **33,** 309–320 (2000).

[219] H. Jeong, S. Mason, A.-L. Barabási, and Z. N. Oltvai, "Lethality and centrality in protein networks," Nature **411,** 41–42 (2001).

[220] M. E. J. Newman, "Scientific collaboration networks: I. Network construction and fundamental results," Phys. Rev. E **64,** 016131 (2001).

[221] H. Ebel, S. Mielsch, and S. Bornholdt, "Scale-free topology of e-mail networks," Phys. Rev. E **66,** 035103(R) (2002).

[222] T. W. Haynes, S. T. Hedetniemi, and P. J. Slater, *Fundamentals of Domination in Graphs* (Marcel Dekker, New York, NY, 1998).

[223] C. Cooper, R. Klasing, and M. Zito, "Lower bounds and algorithms for dominating sets in web graphs," Internet Math. **2,** 275–300 (2005).

[224] P. Erdős and A. Rényi, "On the evolution of random graphs," Publ. Math. Inst. Hung. Acad. Sci. **5,** 17–61 (1960).

[225] M. Zito, "Greedy Algorithms for Minimisation Problems in Random Regular Graphs," in *Algorithms—ESA 2001*, edited by F. M. auf der Heide (Springer, Berlin, Heidelberg, 2001) pp. 525–536.

[226] Cs. Bíró, É. Czabarka, P. Dankelmann, and L. Székely, "Remarks on the domination number of graphs," Bull. Inst. Combin. Appl. **64,** 73–82 (2012).

[227] W. E. Clark, B. Shekhtman, S. Suen, and D. C. Fisher, "Upper bounds for the domination number of a graph," Congr. Numer. **132,** 99–123 (1998).

[228] B. Wieland and A. P. Godbole, "On the Domination Number of a Random Graph," Electron. J. Combin. **8,** R37 (2001).

[229] H. Seyed-Allaei, G. Bianconi, and M. Marsili, "Scale-free networks with an exponent less than two," Phys. Rev. E **73,** 046113 (2006).

[230] Stanford Network Analysis Project (SNAP), http://snap.stanford.edu/data (Accessed August 02, 2014).

[231] J. Kneis, D. Molle, P. Rossmanith, "Partial vs. Complete Domination: t-Dominating Set," in *SOFSEM 2007: Theory and Practice of Computer Science*, edited by J. van Leeuwen, G. F. Italiano, W. van der Hoek, C. Meinel, H. Sack, and F. Plášil (Springer, Berlin, Heidelberg, 2007) pp. 367–376.

[232] C. I. Del Genio, T. Gross, and K. E. Bassler, "All Scale-Free Networks Are Sparse," Phys. Rev. Lett. **107,** 178701 (2011).

[233] "Weighted Configuration Model," AIP Conf. Proc. **776,** 101–107 (2005).

[234] S. N. Dorogovtsev and J. F. F. Mendes, "Evolution of Networks," Adv. Phys. **51,** 1079–1187 (2002).

[235] E. R. Gansner and S. C. North, "An open graph visualization system and its applications to software engineering", Software Pract. Exper. **30**, 1203–1233 (2000), http://www.graphviz.org (Accessed June 04, 2013).

[236] F. Molnár, Jr., S. Sreenivasan, B. K. Szymanski, and G. Korniss, "Minimum dominating sets in scale-free network ensembles," Sci. Rep. **3,** 1736 (2013).

[237] F. Molnár Jr., N. Derzsy, É. Czabarka, L. Székely, B. K. Szymanski, and G. Korniss, "Dominating Scale-Free Networks Using Generalized Probabilistic Methods," Sci. Rep. **4,** 6308 (2014).

[238] R. Klasing and C. Laforest, "Hardness results and approximation algorithms of k-tuple domination in graphs," Inform. Process. Lett. **89,** 75–83 (2004).

[239] N. Hutcheon and J. W. Bialek, "Updated and validated power flow model of the main continental European transmission network," in *Proceedings of the IEEE PowerTech 2013*, (IEEE, 2013) pp. 1–5.

[240] Continental European Transmission Network (2009 winter data), http://www.powerworld.com/bialek (Accessed August 01, 2014).

[241] Open Connectome Project, http://mrbrain.cs.jhu.edu/disa/download (Accessed July 01, 2014).

[242] C. Castellano, S. Fortunato, and V. Loreto, "Statistical physics of social dynamics," Rev. Mod. Phys. **81,** 591–646 (2009).

[243] J. Zoller and S. Montangero, "Probing models of information spreading in social networks," J. Phys. A **47,** 435102 (2014).

[244] C. Liu and Z.-K. Zhang, "Information spreading on dynamic social networks," Commun. Nonlinear Sci. Numer. Simul. **19,** 896–904 (2014).

[245] D. J. Daley and J. Gani, *Epidemic Modeling: An Introduction* (Cambridge University Press, Cambridge, UK, 2005).

[246] J. Kostka, Y. A. Oswald, and R. Wattenhofer, "Word of mouth: Rumor dissemination in social networks," in *Structural Information and Communication Complexity*, edited by A. A. Shvartsman and P. Felber (Springer, Berlin, Heidelberg, 2008) pp. 185–196.

[247] P. Domingos, "Mining social networks for viral marketing," IEEE Intell. Syst. **20,** 8082 (2005).

[248] Y. Hu, S. Havlin, and H. A. Makse, "Conditions for Viral Influence Spreading through Multiplex Correlated Social Networks," Phys. Rev. X **4,** 021031 (2014).

[249] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, NY, 2001) pp. 57–66.

[250] D. Kempe, J. Kleinberg, and E. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, New York, NY, 2003), pp. 137–146.

[251] W. Chen, Y. Yuan, and L. Zhang, "Scalable Influence Maximization in Social Networks under the Linear Threshold Model," in *Proceedings of the 2010 IEEE International Conference on Data Mining* (IEEE Computer Society, Washington, DC, 2010), pp. 88–97.

APPENDIX A
Population Dynamics

A.1 Separatrices of Stationary Solutions for Single-Sex Model

In Section 3.2 we have derived an analytical expression for the relationship between density and its spatial derivative for the stationary solutions:

$$v(u) = \pm\sqrt{\frac{3u^4 - 2u^3 + 6\mu u^2 + 12DE}{6D}}. \tag{A.1}$$

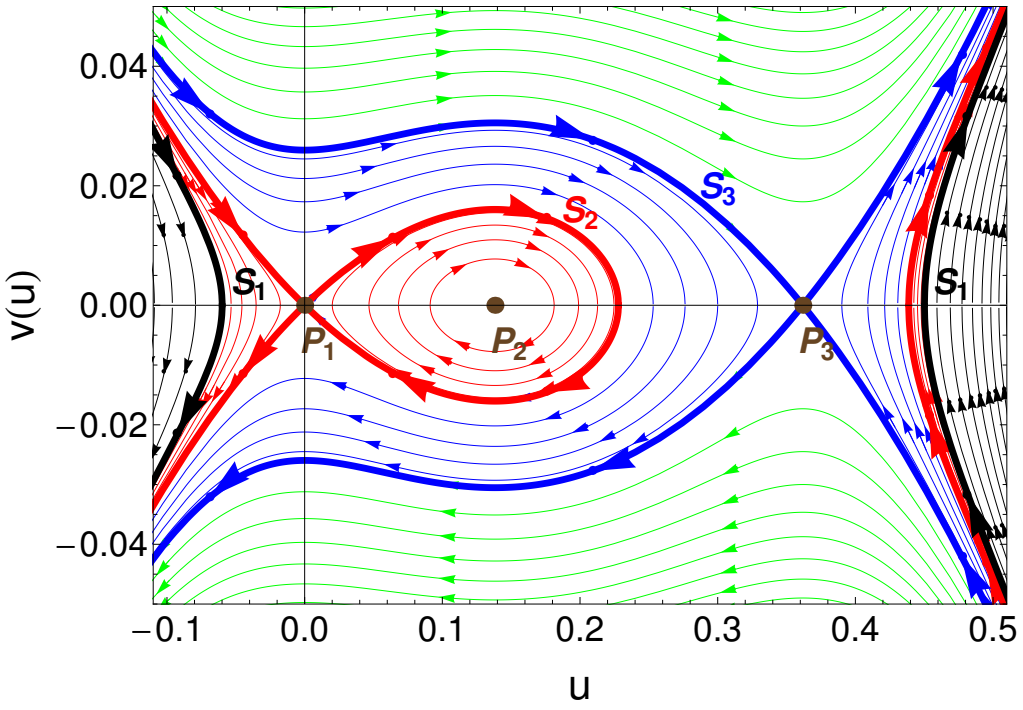

Figure A.1: Phase plot of dynamics in the single-sex model, described by Eq. (A.1). $D = 1.0$, $\mu = 0.05$. The dots indicate fixed points, the thick lines indicate separatrices. Different curves correspond to different E parameters, however, values of E were not chosen uniformly, for aesthetic reasons. This figure is identical to Fig. 3.1(a); it is repeated here for the reader's convenience.

Here we present further analysis to find the special values for E corresponding to separatrices, by analyzing local extremum points of $v(u)$. To find such points, we need only to study the fourth-order polynomial under the square root. Such a polynomial generally has three extrema. We utilize the phase plot [Fig. A.1] to identify minima and maxima.

First, notice that when we are outside separatrix S_3 (green curves), $v(u)$ has a minimum at P_1, maximum at P_2 and minimum at P_3. The minimum at P_3 disappears when S_3 is reached. Therefore, we substitute the formula for u_3^* (the fixed point at P_3) into the polynomial under the square root in $v(u)$ and equate to zero:

$$3(u_3^*)^4 - 2(u_3^*)^3 + 6\mu(u_3^*)^2 + 12DE = 0, \tag{A.2}$$

and we solve for E. We denote the result with the subscript of the corresponding separatrix:

$$E_3 = \frac{1}{24D}(u_3^*)^2(u_3^* - 6\mu) \tag{A.3}$$

Therefore, if $E = E_3$, we obtain separatrix S_3.

Following the same logic, we find that the local minimum of $v(u)$ at P_1 becomes zero, when we reach separatrix S_2. We now substitute u_1^* (the fixed point at P_1) into the polynomial. Note that $u_1^* = 0$, which gives

$$12DE = 0. \tag{A.4}$$

Therefore, we obtain separatrix S_2 when $E = E_2 = 0$. Further, in this case we can find the zero point values of $v(u)$ easily:

$$3u^4 - 2u^3 + 6\mu u^2 = 0. \tag{A.5}$$

This gives the following zero points:

$$u_1 = 0, \tag{A.6}$$

$$u_{2,3} = \frac{1}{3} \pm \sqrt{\frac{1}{9} - 2\mu}. \tag{A.7}$$

This gives the necessary condition $\mu < \frac{1}{18}$ for the existence of S_2 and the corresponding aperiodic solution. Note, that this is a stricter condition than the existence of nonzero fixed points.

Separatrix S_1 is found again using the same method. It corresponds to the case when the local maximum at P_2 becomes zero (and disappears). We substitute u_2^* into the polynomial to obtain:

$$3(u_2^*)^4 - 2(u_2^*)^3 + 6\mu(u_2^*)^2 + 12DE = 0 \tag{A.8}$$

We solve for E and denote the result with the subscript number of the separatrix:

$$E_1 = \frac{1}{48D}(u_2^* - 2\mu)(u_2^* - 6\mu). \tag{A.9}$$

Therefore, if $E = E_1$ we obtain separatrix S_1. Note that E_1 is negative. To summarize, the separatrices have corresponding E values that relate to each other as

$$E_1 < E_2 = 0 < E_3. \tag{A.10}$$

A.2 Analysis of Stationary Solutions of the Single-Sex Model

Stationary solutions outsite S_2 separatrix (blue, green, and black curves on Fig. A.1) have no physical or biological meaning, because they are unbounded, and extend to negative density ranges. Therefore we restrict our analysis to the periodic stationary solutions (red curves), and show that there is a minimum required spatial extent for their existence. In other words, we show that the period length cannot be smaller than a certain value.

In general, the length of period, L, specified by the value of E (note, $E_1 < E < E_2$ for periodic solutions) cannot be derived analytically, but we can find it numerically by integrating Eq. (A.1) for given model parameters, see Fig. A.2. As expected, L goes to infinity as E goes to E_2, that is, the solutions converge to S_2 separatrix, which corresponds to the aperiodic solution. It is also clearly visible that L is monotonically decreasing for smaller values of E. As E goes to E_1 the solutions converge to P_2 fixed point, which is a center. Since E cannot be smaller that E_1,

the period length corresponding to E_1, denoted by L^*, is the minimum length. To find L^* explicitly, we need to analyze the asymptotic behavior of solutions around P_2. First, we find the stability matrix and its eigenvalues.

$$f(u,v) = \frac{\partial u}{\partial x} = v \tag{A.11}$$

$$g(u,v) = \frac{\partial v}{\partial x} = \frac{-1}{D}\left(\frac{1}{2}u^2(1-2u) - \mu u\right) \tag{A.12}$$

$$J = \begin{pmatrix} \frac{\partial f}{\partial u} & \frac{\partial f}{\partial v} \\ \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ \frac{3u^2-u+\mu}{D} & 0 \end{pmatrix} \tag{A.13}$$

We obtain the following eigenvalues:

$$\lambda_{1,2} = \pm\sqrt{\frac{3u^2-u+\mu}{D}}. \tag{A.14}$$

We are interested in the imaginary eigenvalues corresponding to the center (P_2):

$$\lambda_{P2} = \lambda_{1,2}\left(u_2^*\right) = \pm i\sqrt{\frac{16\mu - 1 + \sqrt{1-16\mu}}{8D}}. \tag{A.15}$$

Asymptotically, as the periodic solutions approach P_2 they become circular orbits with an angular frequency given by the imaginary eigenvalue:

$$\omega = \sqrt{\frac{16\mu - 1 + \sqrt{1-16\mu}}{8D}}. \tag{A.16}$$

From this we obtain the length of period (wavelength) in the limit of P_2:

$$L^* = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{8D}{16\mu - 1 + \sqrt{1-16\mu}}}. \tag{A.17}$$

This is the minimum habitat size allowing periodic stationary solutions, corresponding to $E = E_1$.

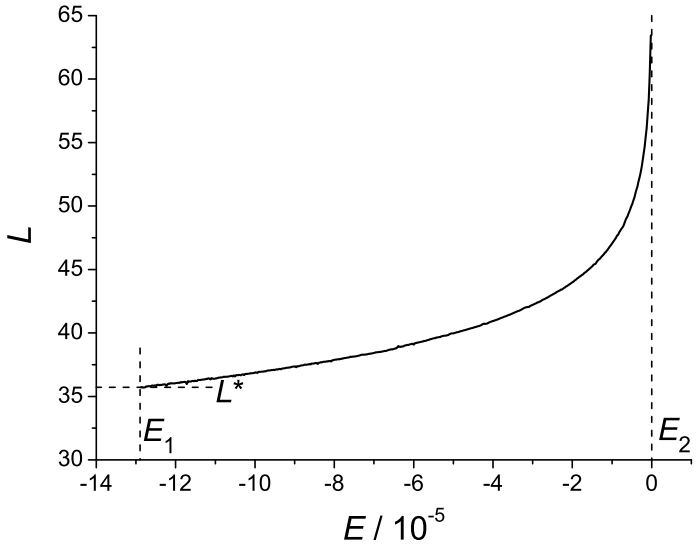

Figure A.2: Length of period in periodic stationary solutions, found by numerical integration of $v(u) \equiv \mathrm{d}u/\mathrm{d}x$. $D = 1.0$, $\mu = 0.05$. L^* is the minimum habitat size allowing periodic stationary solutions.

A.3 Simulated Annealing

To move beyond the homogeneous distribution of the initial population as we seek to minimize cost, we need a method that finds the minimum cost in the domain of infinitely many possible distribution functions. The key to solving this problem is to define the initial distribution in discretized form. We can determine the eventual persistence or extinction of an initial population only by running the simulation to global equilibrium, and this simulation requires a spatial discretization for the integration. We use the same grid to discretize the initial distribution function. Further, we can assume that whatever initial distribution gives the minimum cost, it will have a finite support, therefore we can restrict the spatial extent of the distribution to a region in the center of the simulated space. If this restricted spatial domain includes k grid points, we essentially reduce an infinitely detailed initial distribution to a k-dimensional function.

To find the minimum cost distribution, we use a global optimization method on the k-dimensional cost function, given the constraint that the population must

persist. We use simulated annealing [181], which has the ability to explore the multidimensional domain randomly without getting stuck in local minima. It slowly guides itself to the global minimum of the given function (in our case, the cost function) by lowering a control parameter, which slowly lowers the expected function values the random moves can take. Eventually, a global minimum is found.

Simulated annealing requires an initial guess for the minimum-cost shape, and we must define a transition function that can transform any shape to any other in some finite number of iterations. At every step of the minimization, the new shape is proposed by applying the transition function to the previous shape. It is then accepted or rejected by the following criterion:

$$Pr(\text{accepted}) = \begin{cases} 1 & \text{if } \Delta C < 0 \\ \exp\left[\frac{-\Delta C}{T}\right] & \text{if } \Delta C \geq 0, \end{cases} \tag{A.18}$$

where ΔC is the difference of cost between the newly proposed and previous functions, and T is a temperature-like control parameter that is lowered over time by a given "cooling schedule," so the probability of accepting changes that increase the cost function continuously declines. Eventually only changes that lower the cost are accepted, which finally moves the cost to a local minimum, but due to the gradual cooling and the stochastic nature of the procedure, this is also the global minimum with high probability.

Our minimization procedure starts with the following initial conditions:

$$z(x) := \max\left(0, H\left(1 - \frac{|(x - N/2)|}{W/2}\right)\right) \tag{A.19}$$

$$T(0) := 0.1 \tag{A.20}$$

Here, T is the temperature parameter and $z(x)$ denotes either male or female initial density distributions. The shape of $z(x)$ is an isosceles triangle standing on its shorter side of length W and having altitude H. These values are arbitrary, as long as the initial population they define generates a cost well above the sought minimum, so that they do not influence the minimization procedure. N denotes the width of simulated habitat; $W < N$. In our simulations, we used the following values:

$N = 600$, $W = 100$ (measured in discretized grid points), $H = 0.4$, which is above the Allee threshold in every case we studied.

To minimize restoration cost with simulated annealing, we iterate the following steps. First, we determine the current spatial distribution's finite support:

$$a := \{x : z(x) > \xi \wedge \forall i < x : z(i) \leq \xi\} \tag{A.21}$$

$$b := \{x : z(x) < \xi \wedge \forall i : a < i < b : z(i) > \xi\} \tag{A.22}$$

$$\widetilde{a} := a - (b - a)/4 \tag{A.23}$$

$$\widetilde{b} := b + (b - a)/4, \tag{A.24}$$

where ξ is a cutoff threshold set to 10^{-3}. We use a randomized Gaussian function as transition function to generate a new proposed shape. The mean of this function is chosen from $[\widetilde{a}, \widetilde{b}]$ interval. Therefore, we allow for increasing the width of the density distributions, should the simulated annealing take that direction. The bell-shaped curve is defined by the following parameters:

$$M := \widetilde{a} + \alpha(\widetilde{b} - \widetilde{a}) \tag{A.25}$$

$$A := 2\beta T \tag{A.26}$$

$$V := (\widetilde{b} - \widetilde{a})/8 + 5\gamma T(\widetilde{b} - \widetilde{a}) \tag{A.27}$$

$$S := \begin{cases} -1 & \text{if } \delta < 1/2 \\ 1 & \text{if } \delta \geq 1/2 \end{cases}, \tag{A.28}$$

where M, A, V and S represent mean, amplitude, variance and sign, respectively, and α, β, γ and δ are uniform random numbers in the range $[0, 1)$. Note, that amplitude and variance also depend on the current temperature, T, which helps the minimization process by making smaller changes as T is reduced. The transition function is defined as

$$g(x) := \frac{SA}{\sqrt{2\pi V}} \exp\left[\frac{-(x - M)^2}{2V}\right]. \tag{A.29}$$

We then add $g(x)$ to the current shape, resulting in a new proposed shape:

$$\widetilde{z}(x) := \max\left(0, z(x) + g(x)\right). \tag{A.30}$$

We calculate shape distributions separately (independently) for both males and females. After the new shapes have been generated, the cost of the proposal is evaluated, and then accepted or rejected according to equation (A.18). Finally, the temperature is lowered using a simple cooling schedule:

$$T(t+1) = 0.9999T(t). \tag{A.31}$$

The iteration of these steps starts at $T(0) = 0.1$ and continues until $T(t) < 10^{-4}$.

The constraint requiring successful restoration must be checked by running the simulation until convergence to a homogenous stationary state, persistence or extinction. If the newly proposed shapes result in extinction, the shapes are always rejected regardless of Eq. (A.18). Therefore, given that all previously accepted shapes resulted in survival, and given the convexity of the transition function (Gaussian), we know that when the newly proposed shapes increase the cost, then survival is guaranteed and there is no need to check it with a simulation. A test is only necessary when the cost is reduced. Still, this means that we need to run a numerical simulation for almost every second Monte Carlo step, which is computationally very intensive. To improve performance, we use GPGPU computation (using graphics processing units of video cards for general purpose computations) with CUDA [182]. Using GPUs can significantly improve the performance of PDE integration [183]. This technology lets us carry out each simulation within a fraction of a second, giving a total time for the simulated annealing in the order of a few hours.

APPENDIX B
Synthetic Scale-Free Network Generation

B.1 Mixing Time of HHMC Networks

In order to ensure that HHMC gives a uniform sample of all possible networks with the given degree sequence, we must study the mixing time of the Markov chain. For the lack of theoretical bounds on the exact mixing times of networks with general degree sequences (or power-law degree sequences), we can only empirically inspect the change of a particular metric of the network during mixing, and visually select a mixing time after which the observed value does not change significantly. In our study, the obvious quantity we observe is the MDS size. Figure B.1 shows the MDS size of a single network during mixing. Using multiple observations with various γ parameters, we find that repeating edge swap attempts four times the number of edges is sufficient to reach a properly mixed network.

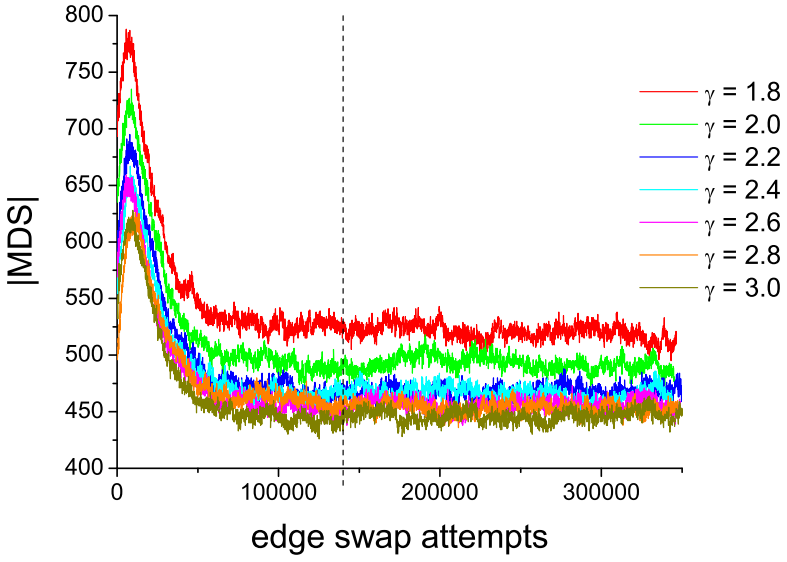

Figure B.1: Evolution of MDS size in HHMC networks during random edge-swaps. $N = 5000$, $\langle k \rangle = 14$. The vertical dashed line indicates where the HHMC method normally stops during network construction.

B.2 Graphicality Correction Algorithm

The algorithm is built on the idea that the Havel-Hakimi algorithm constructs a network during the testing of a degree sequence, and proceeds as far as possible, until the non-graphicality of the sequence is determined, or all edge stubs are connected. In case of a non-graphical sequence, we eventually arrive at a node with too many stubs that cannot be connected. We simply ignore these stubs, and keep the network that has already been constructed. Since there could be more nodes with stubs that could be connected among themselves, we continue the Havel-Hakimi procedure until all node stubs are connected or deleted. The exact steps are given in Alg. 1. Note, that it is not necessary to actually construct the network during this algorithm, it is sufficient to record the degrees of nodes that would be created. The result is a graphical degree sequence, which can be passed to any degree sequence sampling method.

Algorithm 1 Graphicality Correction

Require: $U = (u_1, u_2, ..., u_N)$: input degree sequence, $1 \leq u_i < N$
Ensure: S is graphical
  if there are odd number of odd degrees **then**
    $j :=$ ARGMAX(u_i)
    $u_j := u_j - 1$
  end if
  Sort U to nonincreasing order
  $V := (v_1, v_2, ..., v_N)$ ▷ new degree sequence
  Initialize V to zeros
  for $i := 1$ to N **do**
    if $u_i = 0$ **then**
      break ▷ no more degrees
    end if
    for $j := i + 1$ to $i + u_i$ **do**
      if $u_j > 0$ **then**
        $u_j := u_j - 1$
        $v_i := v_i + 1$
        $v_j := v_j + 1$
      else
        break ▷ non-graphical, skip remaining stubs of node i
      end if
    end for
    Sort U in range $[i + 1, N]$ to nonincreasing order and move elements of V simultaneously with U.
  end for
  $U := V$

APPENDIX C
Scaling of Minimum Dominating Sets

C.1 Probability Distribution of the Size of MDS

In order to justify the averaging of MDS sizes found in realizations, we present the distributions of greedy MDS sizes in each network realization [Fig. C.1(a)], and the distribution of averaged MDS sizes across multiple realizations [Fig. C.1(b)]. In both cases we find approximately Gaussian distributions. However, the width of the distribution of MDS sizes in a single realization is about an order of magnitude smaller than the width of distribution across realizations. Therefore, we can indeed study the MDS size scaling correctly by ensemble averages.

C.2 Asymptotic Scaling of the MDS Lower Bound

We can derive the asymptotic scaling of the lower bound for the MDS in power-law degree distributions with fixed average degree. While starting from slightly different assumptions, our derivation is very similar to that of Ref. [92]. We then explicitly apply the integral formulae to both kinds of $k_{\max}$ cutoffs we considered in our study.

Using continuous degree distributions, $P(k) \simeq Ck^{-\gamma}$, we may work with integrals instead of cumbersome discrete sums. First, we derive the normalization constant C for a power-law distribution with lower cutoff $k_{\min}$ and upper cutoff $k_{\max}$:

$$C = \left[\int_{k_{\min}}^{k_{\max}} k^{-\gamma}\,\mathrm{d}k\right]^{-1} = \frac{1-\gamma}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}} \tag{C.1}$$

The average degree $\langle k\rangle$ is found by:

$$\langle k\rangle = C\int_{k_{\min}}^{k_{\max}} k^{1-\gamma}\,\mathrm{d}k = \left(\frac{1-\gamma}{2-\gamma}\right)\frac{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}} \tag{C.2}$$

Assuming that $k_{\max} \to \infty$ as $N \to \infty$, and $\gamma > 2$, the formula above can be simplified

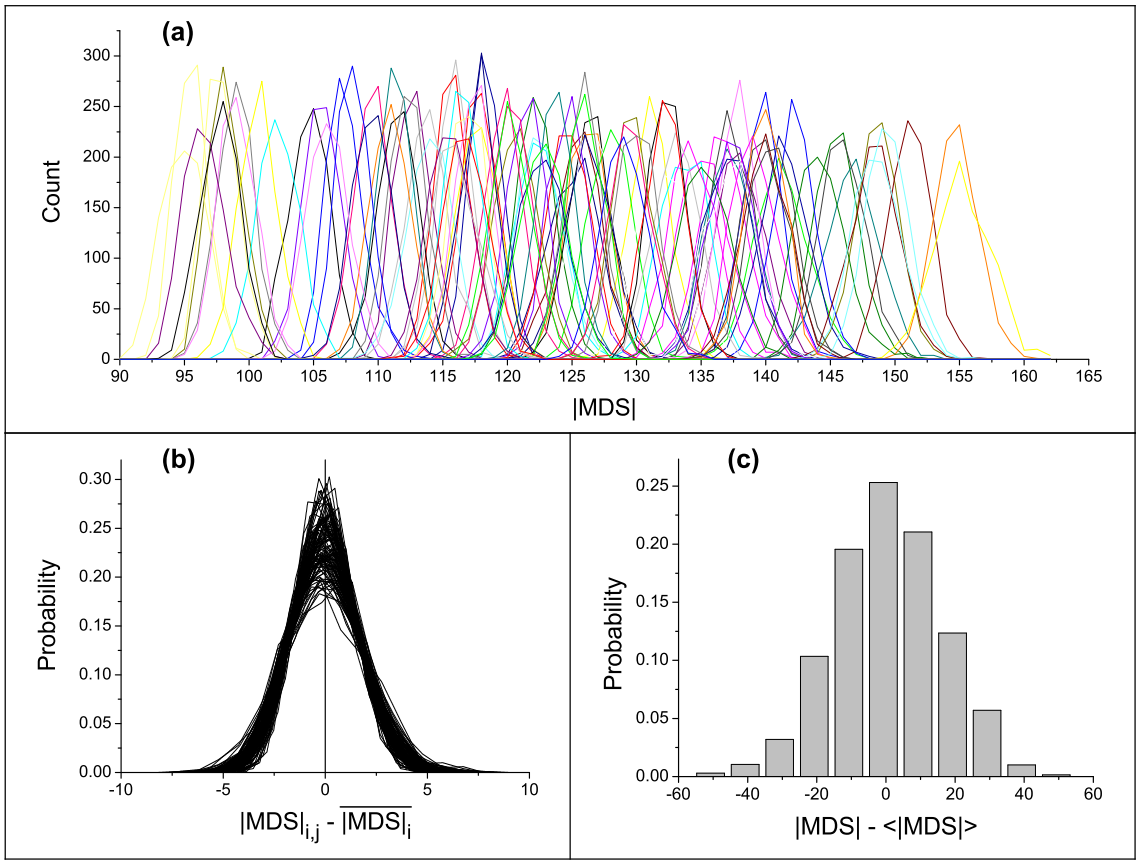

Figure C.1: (a) Distributions of the MDS size for 100 network realizations, found by greedy search executed 1000 times on each realization. One distribution curve is plotted for each realization. (b) Distribution of the MDS sizes across realizations, for 1000 network realizations. All realizations are CONF networks with $N = 2000$, $\gamma = 2.5$, $\langle k \rangle = 14$.

for large N:

$$\langle k \rangle \simeq \left(\frac{\gamma - 1}{\gamma - 2} \right) k_{\min} \tag{C.3}$$

Now we select all nodes above a certain degree threshold k'. The number of these nodes is

$$l(k') = NC \int_{k'}^{k_{\max}} k^{-\gamma} \, \mathrm{d}k = N \frac{k_{\max}^{1-\gamma} - k'^{1-\gamma}}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}} \tag{C.4}$$

Assuming that all nodes with degree $> k'$ have non-overlapping neighbors, these nodes dominate as many nodes as their degree. Asymptotically, we can ignore that these nodes also dominate themselves. Therefore, the number of nodes dominated

by the selected $l(k')$ nodes is

$$d(k') = NC \int_{k'}^{k_{\max}} k^{1-\gamma}\,\mathrm{d}k = N\left(\frac{1-\gamma}{2-\gamma}\right)\frac{k_{\max}^{2-\gamma} - k'^{2-\gamma}}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}} \tag{C.5}$$

We look for the special degree threshold $k' = k^*$, such that $d(k^*) = N$, which means the entire network is dominated. Then for $\gamma > 2$ and large N from Eq. (C.5) it follows that

$$k^* \simeq \left(\frac{\gamma-2}{\gamma-1}\right)^{\frac{1}{2-\gamma}} k_{\min}^{\frac{\gamma-1}{\gamma-2}}\,, \tag{C.6}$$

and combining it with Eq. (C.3) we find

$$k^* \simeq \left(\frac{\gamma-2}{\gamma-1}\right)\langle k\rangle^{\frac{\gamma-1}{\gamma-2}}\,. \tag{C.7}$$

Using this degree threshold, from Eq. (C.4), the lower bound of the size of MDS becomes

$$l(k^*) \simeq N\left(\frac{k^*}{k_{\min}}\right)^{1-\gamma} = N\langle k\rangle^{-\frac{\gamma-1}{\gamma-2}}\,. \tag{C.8}$$

Since we use a fixed average degree, the number of selected nodes is proportional to N, therefore the lower bound is $O(N)$. Further, this linear bound of the MDS size is asymptotically tight, because the natural upper bound of the dominating set for a graphs with no isolated vertices is $N/2$ [222] (either a maximal independent set or its complement is a dominating set).

Note, that the result above always holds when $k_{\max} \to \infty$ as $N \to \infty$, therefore it holds for networks with $k_{\max} = N - 1$ or $k_{\max} = \sqrt{N}$. However, in the latter case the convergence of the bound to its asymptotic limit is very slow. If $l(k^*)$ is evaluated on the range of network sizes we study (see Fig. C.2), we cannot diretly observe the $O(N)$ scaling of the bound. Numerical experiments show that the bound's $O(N)$ scaling can only be observed when $N > 10^8$ (see Fig. 5.2).

In the case when $\gamma < 2$ (note, we always assume $\gamma > 1$), the average degree cannot be fixed asymptotically, because it diverges as $k_{\max} \to \infty$. To derive an asymptotic bound, we need to assume that $k_{\min} = 1$. From the condition that all

nodes in the network must be dominated, $d(k^*) = N$, Eq. (C.5) yields

$$k^* \simeq \left[k_{\max}^{2-\gamma} - \left(\frac{2-\gamma}{\gamma-1}\right)\right]^{\frac{1}{2-\gamma}} \simeq k_{\max} \ . \tag{C.9}$$

Inserting this result into Eq. (C.4) would give $l(k^*) = 0$ in the asymptotic limit. However, for the correct interpretation of this result we must consider that a network actually has a discrete degree distribution. The convergence of k^* to $k_{\max}$, which is valid for the discrete case as well, means that the dominating set ony contains maximum degree nodes, therefore the lower bound of MDS size becomes $N/k_{\max}$. For $k_{\max} = N - 1$ networks this gives an $O(1)$ bound, and for $k_{\max} = \sqrt{N}$ networks this gives an $O(N^{1/2})$ bound. Finally, an observation worth noting: when $\gamma < 2$, $l(k^*)$ gives no better asymptotic lower bound for scale-free networks than the well-known formula $|\text{MDS}| \geq N/(k_{\max} + 1)$ [222].

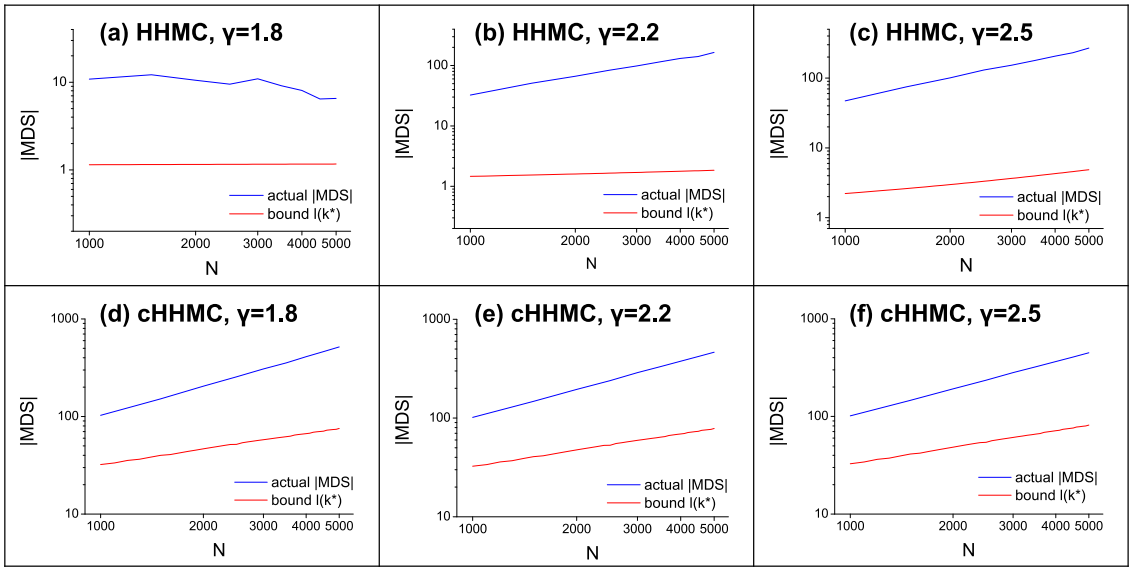

Figure C.2: Comparision of measured MDS size obtained by greedy algorithm and its corresponding $l(k^*)$ lower bound, in HHMC and cHHMC networks, $\langle k \rangle = 14$.

C.3 Partial MDS Scaling with Network Size

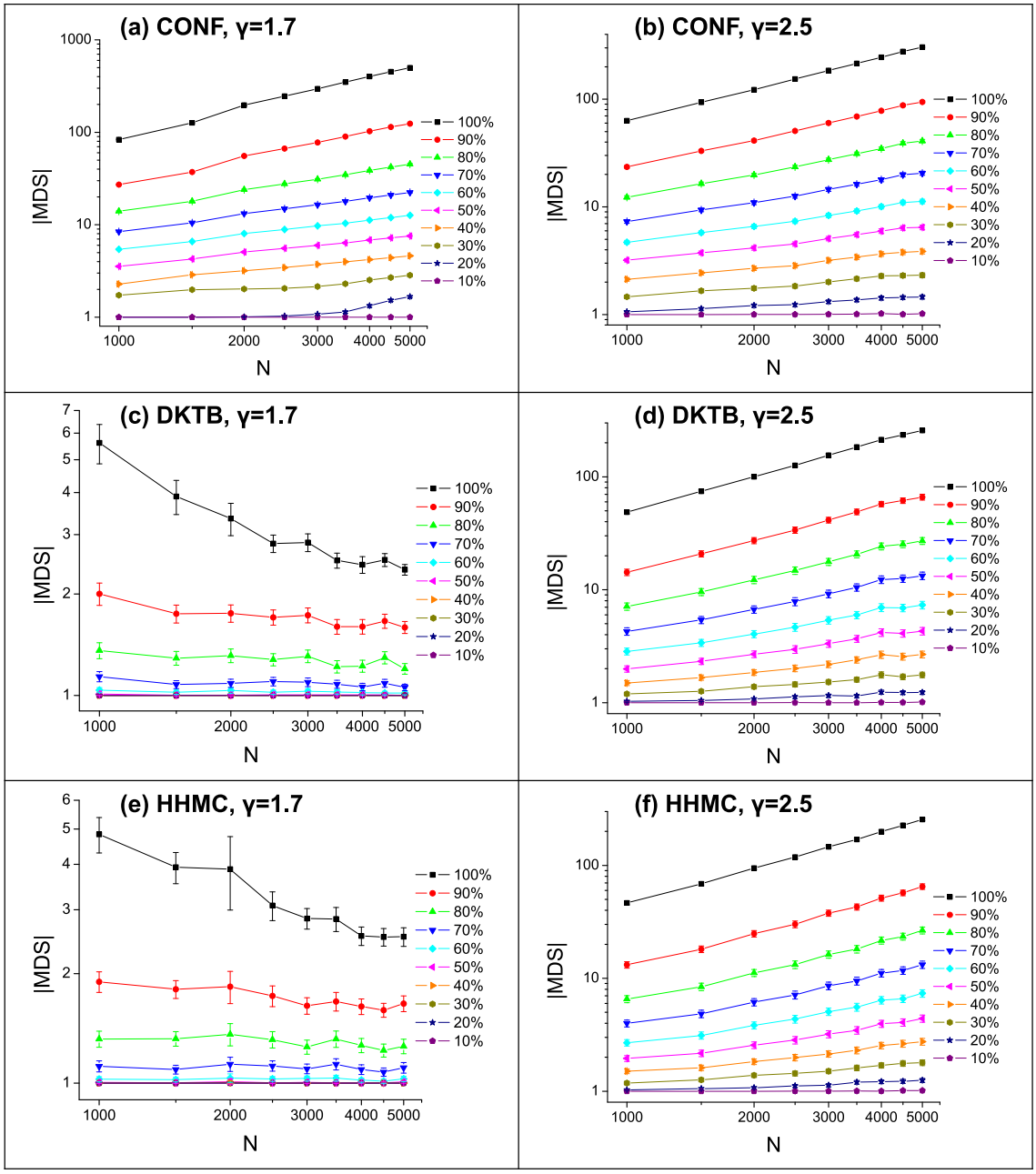

Figure C.3: The size of partial MDS scaling with N in regular cutoff ($k_{\max} = N - 1$) networks, $\langle k \rangle = 14$, averaged over 400 samples at every data point. The dominated fraction of nodes is expressed as percentage of N.

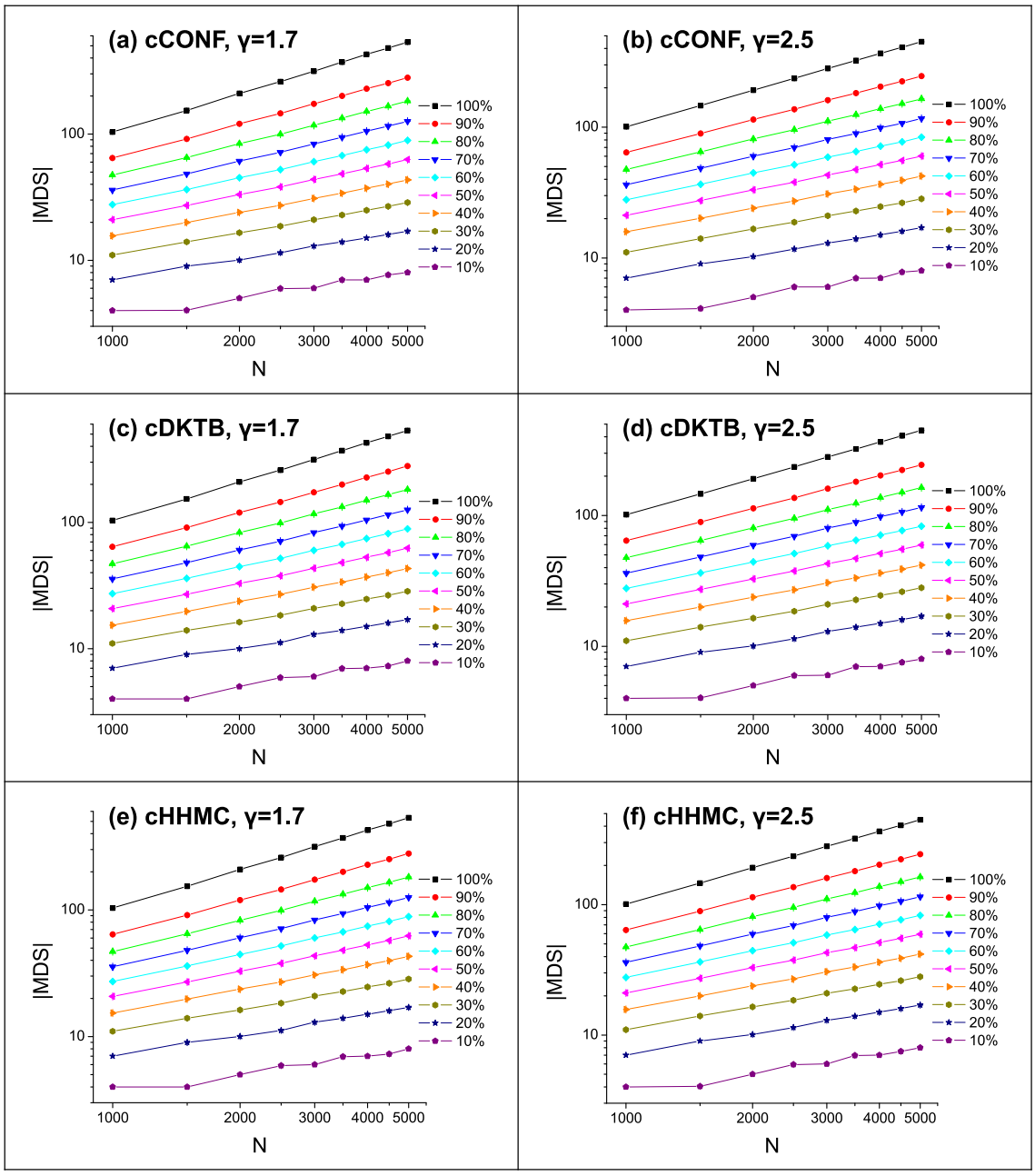

Figure C.4: The size of partial MDS scaling with N in networks with structural cutoff ($k_{\max} = \sqrt{N}$), $\langle k \rangle = 14$, averaged over 400 samples at every data point. The dominated fraction of nodes is expressed as percentage of N.

C.4 Maximum Realized Degree

The expected maximum degree of a network can be calculated analytically from the degree distribution. We start with a power-law PMF with cutoffs $k_{\min}$ and $k_{\max}$:

$$p_K(k) = k^{-\gamma} \left(\zeta\left(\gamma, k_{\min}\right) - \zeta\left(\gamma, k_{\max}+1\right)\right)^{-1}. \tag{C.10}$$

The probability that any given sample K from this distribution is below a certain k' bound is:

$$\Pr(K \leq k') = \sum_{k=k_{\min}}^{k'} p_K(k) = \frac{\zeta(\gamma, k_{\min}) - \zeta(\gamma, 1+k')}{\zeta(\gamma, k_{\min}) - \zeta(\gamma, 1+k_{\max})}, \tag{C.11}$$

and assuming all N node degrees are sampled independently, the probability that all degrees are below k' is:

$$\Pr(\text{all degrees} \leq k') = \left(\Pr(K \leq k')\right)^N, \tag{C.12}$$

which is by definition the cumulative distribution function (CDF) for the largest degree, $Z = \max\{K_1, K_2, ..., K_N\}$, with K_i being a sampled node degree. Then, the expected value of Z is:

$$\begin{aligned} \mathbb{E}Z &= \sum_{k'=k_{\min}}^{k_{\max}} \left[1 - \Pr(\text{all degrees} \leq k')\right] = \\ &= \sum_{k'=k_{\min}}^{k_{\max}} \left[1 - \left(\frac{\zeta(\gamma, k_{\min}) - \zeta(\gamma, 1+k')}{\zeta(\gamma, k_{\min}) - \zeta(\gamma, 1+k_{\max})}\right)^N\right]. \end{aligned} \tag{C.13}$$

By numerically evaluating this formula, we can find the expected largest degree in a network with any degree distribution. Figure 5.6(a) shows these values for a power-law distribution as a function of the power-law exponent.

APPENDIX D
Probabilistic Dominating Sets

D.1 Analytical Estimates of the Size of Probabilistic Dominating Sets

In the main text we introduced the analytical estimates of the RDS and CDS methods for uncorrelated scale-free networks. Here, we present the detailed derivations of these estimates. We start by deriving the expected sizes of sets X and Y that constitute the probabilistic dominating set. In general, we have the following:

$$\frac{|X|}{N} = \int_{k_{\min}}^{k_{\max}} X(k)P(k)\mathrm{d}k, \tag{D.1}$$

where $X(k)$ is the probability that a node with degree k is added to set X, and $P(k)$ is the degree distribution of the network, with lower bound $k_{\min}$ and upper bound $k_{\max}$. Using the same notation, we have for the Y set:

$$\begin{aligned} \frac{|Y|}{N} &= \int_{k_{\min}}^{k_{\max}} (1 - X(k))P(k)\,\Pr(\text{not dominated})\mathrm{d}k & \text{(D.2)} \\ &= \int_{k_{\min}}^{k_{\max}} (1 - X(k))P(k) \left[\int_{k_{\min}}^{k_{\max}} (1 - X(k'))P(k'|k)\mathrm{d}k'\right]^k \mathrm{d}k. & \text{(D.3)} \end{aligned}$$

Here, we count all nodes that are not in X (the first integral), but only those that are also not dominated by any node in X, which means that none of the k neighbors of any given node with degree k are in set X (the second integral inside the square brackets). $P(k'|k)$ is the conditional probability distribution of the degrees of neighbors for a given node with degree k. Since we obtain a probabilistic dominating set by $X \cup Y$, and these sets are disjoint, we have:

$$\begin{aligned} \frac{\langle DS \rangle}{N} = &\int_{k_{\min}}^{k_{\max}} X(k)P(k)\mathrm{d}k + \\ &+ \int_{k_{\min}}^{k_{\max}} (1 - X(k))P(k) \left[\int_{k_{\min}}^{k_{\max}} (1 - X(k'))P(k'|k)\mathrm{d}k'\right]^k \mathrm{d}k. \end{aligned} \tag{D.4}$$

This is a general formula for the size of any probabilistic dominating set in a network with arbitrary degree distribution and degree correlations.

To estimate the size of RDS and CDS in uncorrelated scale-free networks, first we need to find the formula for the properly normalized power-law degree distribution, $P(k) = Ck^{-\gamma}$. The normalization constant is found by:

$$C = \left[\int_{k_{\min}}^{k_{\max}} k^{-\gamma} \mathrm{d}k\right]^{-1} = \frac{1-\gamma}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}}. \tag{D.5}$$

Further, we have for uncorrelated networks:

$$P(k'|k) = \frac{k'P(k')}{\langle k \rangle}. \tag{D.6}$$

We can obtain $\langle k \rangle$ as follows:

$$\langle k \rangle = \int_{k_{\min}}^{k_{\max}} kP(k)\mathrm{d}k = C\int_{k_{\min}}^{k_{\max}} k^{1-\gamma}\mathrm{d}k = \left(\frac{1-\gamma}{2-\gamma}\right)\frac{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}}. \tag{D.7}$$

D.1.1 RDS with Uniform Node Selection

In case of RDS, with a uniform node selection probability, we have: $X(k) = p$, where $p \in (0, 1)$ is the probability parameter. Substituting into Eq. (D.4) yields:

$$\begin{aligned}
\frac{\langle RDS \rangle}{N} &= \int_{k_{\min}}^{k_{\max}} pP(k)\mathrm{d}k + \\
&\quad + \int_{k_{\min}}^{k_{\max}} (1-p)P(k)\left[\frac{1-p}{\langle k \rangle}\int_{k_{\min}}^{k_{\max}} k'P(k')\mathrm{d}k'\right]^k \mathrm{d}k && \text{(D.8)} \\
&= p + \int_{k_{\min}}^{k_{\max}} (1-p)^{k+1}P(k)\mathrm{d}k && \text{(D.9)} \\
&= p + \frac{1-\gamma}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}}\int_{k_{\min}}^{k_{\max}} (1-p)^{k+1}k^{-\gamma}\mathrm{d}k. && \text{(D.10)}
\end{aligned}$$

The integration above can be calculated as follows:

$$\int_{k_{\min}}^{k_{\max}} (1-p)^{k+1} k^{-\gamma} \mathrm{d}k =$$
$$(p-1)\left[\Gamma(1-\gamma, -k_{\max}\log(1-p)) - \Gamma(1-\gamma, -k_{\min}\log(1-p))\right]\left[-\log(1-p)\right]^{\gamma-1}, \tag{D.11}$$

where $\Gamma(a,z)$ is the incomplete gamma function, $\Gamma(a,z) = \int_z^\infty t^{a-1}e^{-t}\mathrm{d}t$. We can simplify Eq. (D.11) by using the following:

$$\Gamma(1-\gamma, z) = z^{1-\gamma} E_\gamma(z), \tag{D.12}$$

where $E_\gamma(z)$ is the exponential integral function, $E_\gamma(z) = \int_1^\infty e^{-zt}t^{-\gamma}dt$. Detailed explanation for using exponential integrals instead of gamma functions is included in Section D.1.5. Substituting into Eq. (D.11) yields:

$$\int_{k_{\min}}^{k_{\max}} (1-p)^{k+1} k^{-\gamma} \mathrm{d}k =$$
$$(p-1)\left[k_{\max}^{1-\gamma} E_\gamma(-k_{\max}\log(1-p)) - k_{\min}^{1-\gamma} E_\gamma(-k_{\min}\log(1-p))\right]. \tag{D.13}$$

Therefore, we have the estimated size of RDS as:

$$\frac{\langle RDS\rangle}{N} = p +$$
$$+ \frac{(1-\gamma)(1-p)}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}}\left[k_{\min}^{1-\gamma} E_\gamma(-k_{\min}\log(1-p)) - k_{\max}^{1-\gamma} E_\gamma(-k_{\max}\log(1-p))\right]. \tag{D.14}$$

D.1.2 RDS with Degree-Dependent Node Selection

In this case we have a degree-dependent formula for $X(k)$, making the derivation longer, since we cannot simplify the integrals. We use the reparametrized selection probability that is expressed as a function of the κ degree threshold, we have $X(k) = \min(1, (k/\kappa)^\beta)$, where κ is defined as $\kappa = k_{\max}p^{-1/\beta}$ in the main text. Thus

we get from Eq. (D.4):

$$\frac{\langle RDS\rangle}{N} = \int_{k_{\min}}^{k_{\max}} \min\left(1, (k/\kappa)^{\beta}\right) P(k)\mathrm{d}k + \\ + \int_{k_{\min}}^{k_{\max}} \left(1 - \min\left(1, (k/\kappa)^{\beta}\right)\right) P(k) \left[\int_{k_{\min}}^{k_{\max}} \left(1 - \min\left(1, (k'/\kappa)^{\beta}\right)\right) \frac{k'P(k')}{\langle k\rangle}\mathrm{d}k'\right]^{k} \mathrm{d}k. \tag{D.15}$$

For the first integral in Eq. (D.15), corresponding to $\langle X\rangle/N$, we have:

$$\int_{k_{\min}}^{k_{\max}} \min\left(1, (k/\kappa)^{\beta}\right) P(k)\mathrm{d}k = \int_{k_{\min}}^{\min(k_{\max},\kappa)} (k/\kappa)^{\beta} P(k)\mathrm{d}k + \int_{\min(k_{\max},\kappa)}^{k_{\max}} P(k)\mathrm{d}k, \tag{D.16}$$

where we can simplify the integration bounds, since always $\kappa < k_{\max}$:

$$\int_{k_{\min}}^{k_{\max}} \min\left(1, (k/\kappa)^{\beta}\right) P(k)\mathrm{d}k = \int_{k_{\min}}^{\kappa} (k/\kappa)^{\beta} P(k)\mathrm{d}k + \int_{\kappa}^{k_{\max}} P(k)\mathrm{d}k. \tag{D.17}$$

The first integral of Eq. (D.17) can be calculated as

$$\int_{k_{\min}}^{\kappa} (k/\kappa)^{\beta} P(k)\mathrm{d}k = \frac{(1-\gamma)\kappa^{-\beta}}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}} \int_{k_{\min}}^{\kappa} k^{\beta-\gamma}\mathrm{d}k = \frac{(1-\gamma)\kappa^{-\beta}\left(\kappa^{1+\beta-\gamma} - {k_{\min}}^{1+\beta-\gamma}\right)}{(1+\beta-\gamma)\left(k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}\right)}, \tag{D.18}$$

while for the second part of Eq. (D.17) we obtain:

$$\int_{\kappa}^{k_{\max}} P(k)\mathrm{d}k = \frac{k_{\max}^{1-\gamma} - \kappa^{1-\gamma}}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}}. \tag{D.19}$$

For the rest of Eq. (D.15) we first note that we can simplify the integration by changing the bounds in the following way:

$$\int_{k_{\min}}^{k_{\max}} \left(1 - \min\left(1, (k/\kappa)^{\beta}\right)\right) z(k)\mathrm{d}k = \int_{k_{\min}}^{\kappa} \left(1 - (k/\kappa)^{\beta}\right) z(k)\mathrm{d}k, \tag{D.20}$$

where $z(k)$ is some function of k. We use this simplification in Eq. (D.15) for the second integral and for the expression in the square brackets. First, the expression

in the square brackets of Eq. (D.15) is calculated as follows:

$$\begin{aligned}\int_{k_{\min}}^{k_{\max}} \left(1-\min\left(1,(k'/\kappa)^{\beta}\right)\right)\frac{k'P(k')}{\langle k\rangle}\mathrm{d}k' &= \int_{k_{\min}}^{\kappa}\left(1-(k/\kappa)^{\beta}\right)\frac{k'P(k')}{\langle k\rangle}\mathrm{d}k = \\ &= \frac{C}{\langle k\rangle}\left[\int_{k_{\min}}^{\kappa} k'^{1-\gamma}\mathrm{d}k' - \kappa^{-\beta}\int_{k_{\min}}^{\kappa} k'^{1+\beta-\gamma}\mathrm{d}k'\right] = \\ &= \frac{2-\gamma}{k_{\max}^{2-\gamma}-k_{\min}^{2-\gamma}}\left[\frac{\kappa^{2-\gamma}-k_{\min}^{2-\gamma}}{2-\gamma} - \kappa^{-\beta}\frac{\kappa^{2+\beta-\gamma}-k_{\min}^{2+\beta-\gamma}}{2+\beta-\gamma}\right] = \\ &= \frac{\kappa^{2-\gamma}-k_{\min}^{2-\gamma}}{k_{\max}^{2-\gamma}-k_{\min}^{2-\gamma}} - \frac{(2-\gamma)\kappa^{-\beta}}{2+\beta-\gamma}\left(\frac{\kappa^{2+\beta-\gamma}-k_{\min}^{2+\beta-\gamma}}{k_{\max}^{2-\gamma}-k_{\min}^{2-\gamma}}\right). \end{aligned} \tag{D.21}$$

We introduce a variable for this expression to simplify further calculations:

$$a := \frac{\kappa^{2-\gamma}-k_{\min}^{2-\gamma}}{k_{\max}^{2-\gamma}-k_{\min}^{2-\gamma}} - \frac{(2-\gamma)\kappa^{-\beta}}{2+\beta-\gamma}\left(\frac{\kappa^{2+\beta-\gamma}-k_{\min}^{2+\beta-\gamma}}{k_{\max}^{2-\gamma}-k_{\min}^{2-\gamma}}\right). \tag{D.22}$$

Then the second integral of Eq. (D.15), corresponding to $\langle Y\rangle/N$, can be calculated as:

$$\begin{aligned}\int_{k_{\min}}^{k_{\max}} \left(1-\min\left(1,(k/\kappa)^{\beta}\right)\right)P(k)a^k\mathrm{d}k &= \int_{k_{\min}}^{\kappa}\left(1-(k/\kappa)^{\beta}\right)P(k)a^k\mathrm{d}k = \\ &= C\left[\int_{k_{\min}}^{\kappa} k^{-\gamma}a^k\mathrm{d}k - \kappa^{-\beta}\int_{k_{\min}}^{\kappa} k^{\beta-\gamma}a^k\mathrm{d}k\right] = \\ &= \frac{C(-\log a)^{\gamma}}{\log a}\left[\Gamma(1-\gamma,-\kappa\log a)-\Gamma(1-\gamma,-k_{\min}\log a)\right] + \\ &+ \frac{C(-\log a)^{\gamma-\beta}\kappa^{-\beta}}{\log a}\left[\Gamma(1+\beta-\gamma,-k_{\min}\log a)-\Gamma(1+\beta-\gamma,-\kappa\log a)\right]. \end{aligned} \tag{D.23}$$

Replacing the incomplete gamma functions with exponential integrals we get:

$$\begin{aligned}\int_{k_{\min}}^{k_{\max}} \left(1-\min\left(1,(k/\kappa)^{\beta}\right)\right)P(k)a^k\mathrm{d}k &= \\ &= C\left[k_{\min}^{1-\gamma}E_{\gamma}(-k_{\min}\log a) - \kappa^{1-\gamma}E_{\gamma}(-\kappa\log a)\right] + \\ &+ C\kappa^{-\beta}\left[\kappa^{1+\beta-\gamma}E_{\gamma-\beta}(-\kappa\log a) - k_{\min}^{1+\beta-\gamma}E_{\gamma-\beta}(-k_{\min}\log a)\right]. \end{aligned} \tag{D.24}$$

Finally, we put all parts together and we obtain the estimated size of RDS as:

$$\frac{\langle RDS \rangle}{N} = \frac{k_{\max}^{1-\gamma} - \kappa^{1-\gamma} + (1-\gamma)\left[y_1 + \kappa^{-\beta}(x + y_2)\right]}{k_{\max}^{1-\gamma} - k_{\min}^{1-\gamma}}, \tag{D.25}$$

with

$$x = \frac{\kappa^{1+\beta-\gamma} - k_{\min}^{1+\beta-\gamma}}{1+\beta-\gamma} \tag{D.26}$$

$$y_1 = k_{\min}^{1-\gamma} E_\gamma(-k_{\min} \log a) - \kappa^{1-\gamma} E_\gamma(-\kappa \log a) \tag{D.27}$$

$$y_2 = \kappa^{1+\beta-\gamma} E_{\gamma-\beta}(-\kappa \log a) - k_{\min}^{1+\beta-\gamma} E_{\gamma-\beta}(-k_{\min} \log a) \tag{D.28}$$

$$a = \frac{\kappa^{2-\gamma} - k_{\min}^{2-\gamma}}{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}} - \frac{(2-\gamma)\kappa^{-\beta}}{2+\beta-\gamma}\left(\frac{\kappa^{2+\beta-\gamma} - k_{\min}^{2+\beta-\gamma}}{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}}\right). \tag{D.29}$$

D.1.3 CDS

In case of CDS, the node selection probability is degree-dependent, and can be expressed by a Heaviside step function: $X(k) = \Theta(k - \kappa)$, where κ is the degree threshold above which all nodes are selected. When substituting into Eq. (D.4), we can simply incorporate this by changing the integration bounds, thus we have:

$$\frac{\langle CDS \rangle}{N} = \int_{\kappa}^{k_{\max}} P(k)\mathrm{d}k + \int_{k_{\min}}^{\kappa} P(k)\left[\int_{k_{\min}}^{\kappa} \frac{k' P(k')}{\langle k \rangle}\mathrm{d}k'\right]^k \mathrm{d}k. \tag{D.30}$$

Substituting the expression of the average degree and the power-law degree distribution in the expression inside the sqare brackets in Eq. (D.30) gives:

$$\int_{k_{\min}}^{\kappa} \frac{k' P(k')}{\langle k \rangle} dk' = \frac{2-\gamma}{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}} \int_{k_{\min}}^{\kappa} k'^{1-\gamma} dk' = \frac{\kappa^{2-\gamma} - k_{\min}^{2-\gamma}}{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}}. \tag{D.31}$$

We denote this expression to simplify further calculations:

$$b := \frac{\kappa^{2-\gamma} - k_{\min}^{2-\gamma}}{k_{\max}^{2-\gamma} - k_{\min}^{2-\gamma}}, \tag{D.32}$$

thus we have for the second part of Eq. (D.30):

$$\int_{k_{\min}}^{\kappa} P(k)\left[\int_{k_{\min}}^{\kappa}\frac{k'P(k')}{\langle k\rangle}\mathrm{d}k'\right]^k \mathrm{d}k = \int_{k_{\min}}^{\kappa} P(k)b^k\mathrm{d}k = C\int_{k_{\min}}^{\kappa} k^{-\gamma}b^k\mathrm{d}k = \\ = \frac{1-\gamma}{k_{\max}^{1-\gamma}-k_{\min}^{1-\gamma}}\left[k_{\min}^{1-\gamma}E_\gamma(-k_{\min}\log b)-\kappa^{1-\gamma}E_\gamma(-\kappa\log b)\right]. \quad \text{(D.33)}$$

For the first integral in Eq. D.30 we have:

$$\int_{\kappa}^{k_{\max}} P(k)\mathrm{d}k = C\int_{\kappa}^{k_{\max}} k^{-\gamma}\mathrm{d}k = \frac{k_{\max}^{1-\gamma}-\kappa^{1-\gamma}}{k_{\max}^{1-\gamma}-k_{\min}^{1-\gamma}}, \quad \text{(D.34)}$$

therefore we have for the estimated size of CDS:

$$\frac{\langle CDS\rangle}{N} = \frac{k_{\max}^{1-\gamma}-\kappa^{1-\gamma}+(1-\gamma)[k_{\min}^{1-\gamma}E_\gamma(-k_{\min}\log b)-\kappa^{1-\gamma}E_\gamma(-\kappa\log b)]}{k_{\max}^{1-\gamma}-k_{\min}^{1-\gamma}}. \quad \text{(D.35)}$$

D.1.4 Comparison of Analytical and Numerical Dominating Set Sizes

By evaluating the analytical formulas we can find the expected size of RDS and CDS without the need to perform numerical calculations (the actual dominating set selection algorithms) on the network. Figures D.1, D.2, and D.3 show the accuracy of the analytical curves in comparison with the numerical ones.

The accuracy of our analytical estimates for RDS and CDS seem to be lower for low $\langle k\rangle$ and γ values. This inaccuracy is an artifact of our average degree control method, which controls $\langle k\rangle$ by adjusting $k_{\min}$, and removing a certain fraction of smallest degree nodes. The latter becomes significant when $k_{\min}\to 1$ (for low $\langle k\rangle$), because it causes a slight deviation from a perfect power-law degree distribution. In order to use the analytical formulas (which are very sensitive to $k_{\min}$), we have to estimate a fractional $k_{\min}$, as if it were a cutoff of a continuous and otherwise perfectly satisfied power-law distribution. In reality, we deviate from power-law, leading to inaccuracy. However, as $\langle k\rangle$ increases, $k_{\min}$ also increases, and the relative deviation from a perfect power-law decreases, hence the increased accuracy. The implication for real networks is that we can expect similarly less accurate estimates if the degree distribution deviates from power-law.

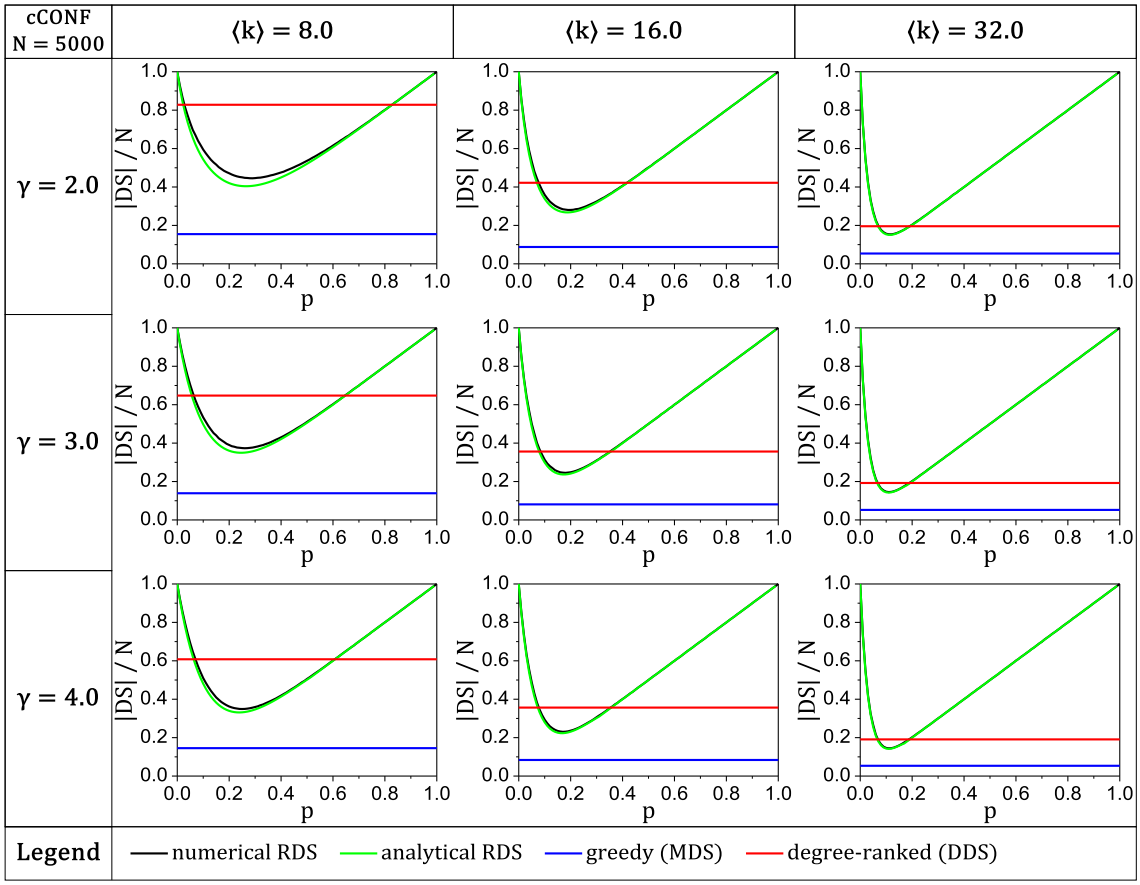

Figure D.1: Analytical estimates vs. numerical results of RDS with degree-independent node selection probability. The figures represent averaged data over 50 network samples of uncorrelated (cCONF) scale-free networks with $N = 5000$ and various $\langle k \rangle$ and γ values.

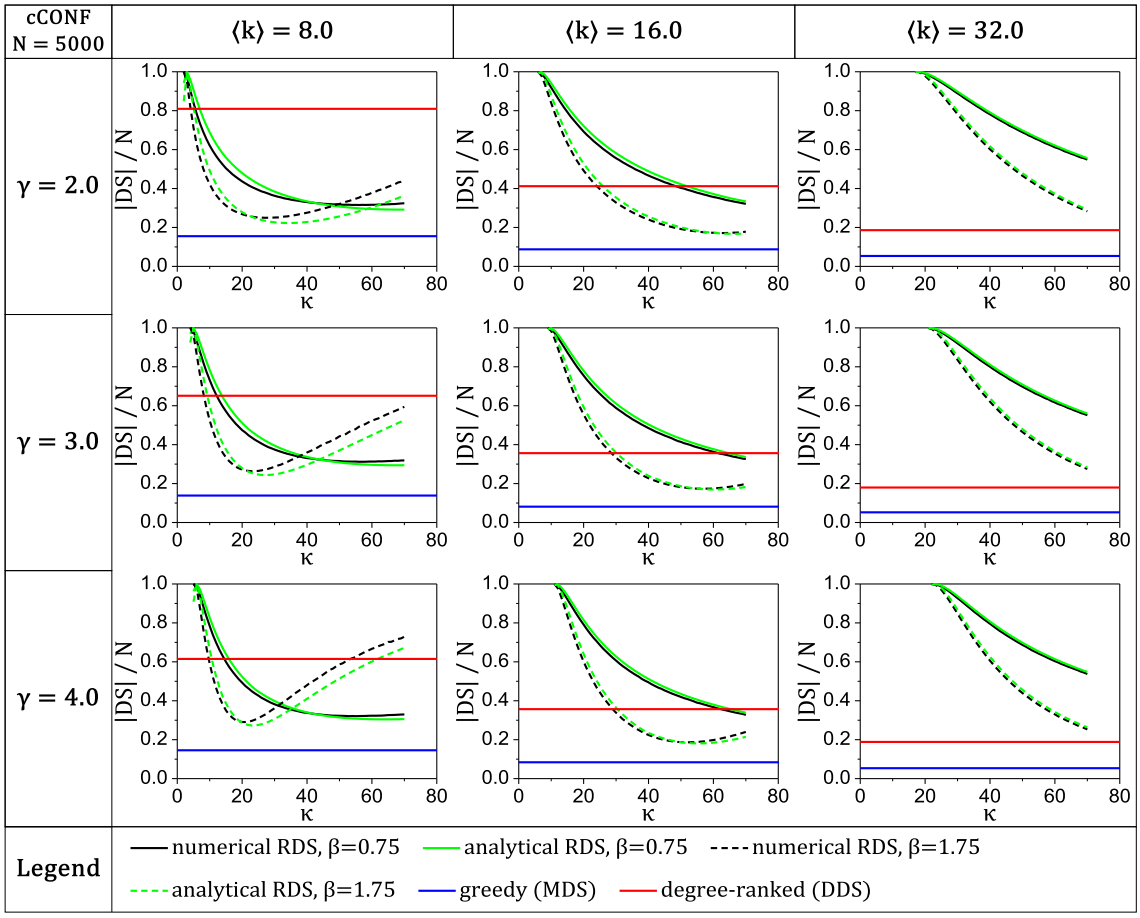

Figure D.2: Analytical estimates vs. numerical results of RDS with degree-dependent node selection probability. The figures represent averaged data over 50 network samples of uncorrelated (cCONF) scale-free networks with $N = 5000$ and various $\langle k \rangle$ and γ values.

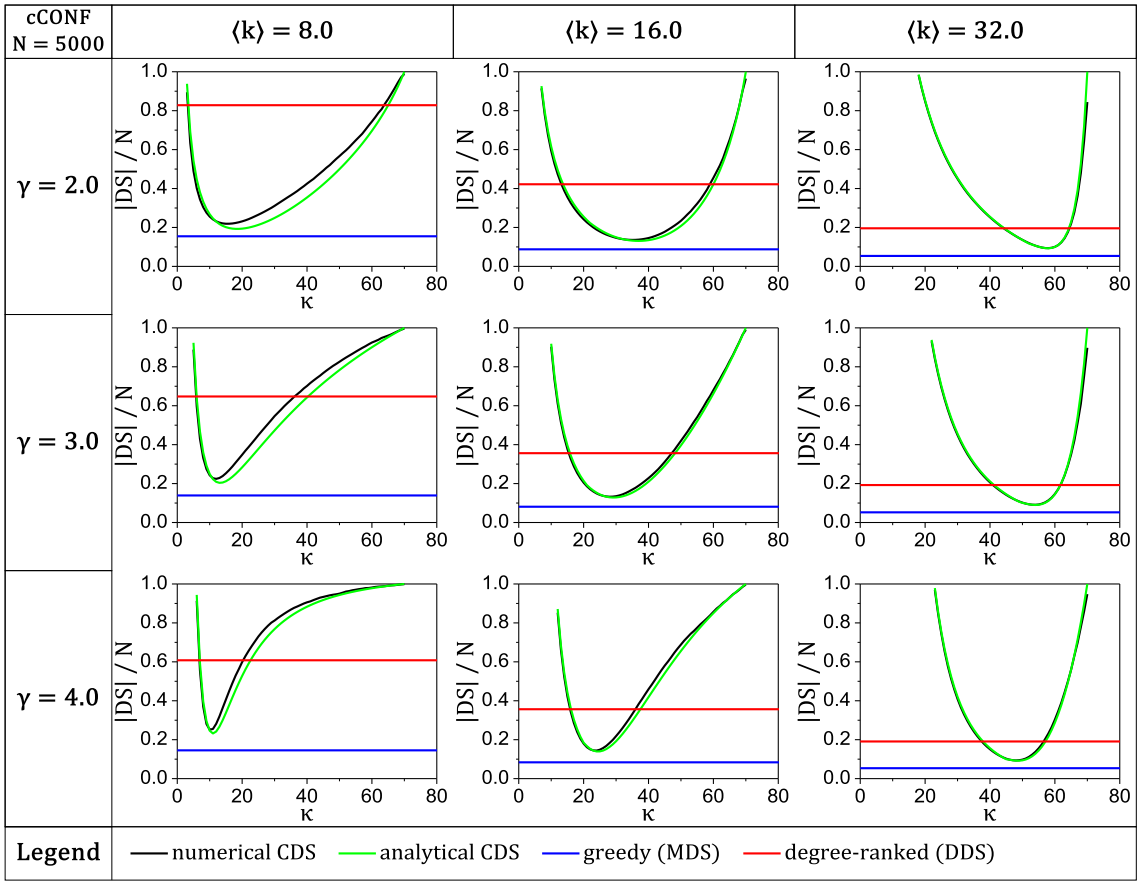

Figure D.3: Analytical estimates vs. numerical results of CDS. The figures represent averaged data over 50 network samples of uncorrelated (cCONF) scale-free networks with $N = 5000$ and various $\langle k \rangle$ and γ values. The curves start at nonzero κ values, because the smallest cutoff that can be applied is the minimum degree, $k_{\min}$. Similarly, the largest possible κ value is the maximum degree, $k_{\max}$.

D.1.5 Exponential Integral Function

The usage of exponential integrals instead of incomplete gamma functions in the estimates of RDS and CDS is justified not only by mathematical aesthetics, but also by numerical computational issues. Notice, in Eq. (D.11), that the first argument of the incomplete gamma function is negative. There are very few numerical libraries that can compute gamma functions with such arguments. On the other hand, an exponential integral function can be computed numerically as:

$$E_\gamma(z) = \int_1^\infty e^{-zt} t^{-\gamma} dt = z^{\gamma-1}\Gamma(1-\gamma) + \sum_{n=0}^{\infty} \frac{(-1)^n z^n}{n!(\gamma-1-n)} \tag{D.36}$$

Here, we only need to compute the regular gamma function with a negative argument, which is readily available in most numerical libraries. There are two issues, however, that we must consider. First, notice that for integer γ values $\Gamma(1-\gamma)$ evaluates to complex infinity. This can be avoided by simply adding a very small number to integer γ values. For example, if $\gamma = 3$, we use $\gamma = 3.01$ instead during the calculation, which has only a negligible effect on the final estimated RDS or CDS size. Second, notice that when the infinite sum is truncated, the result will always diverge for large enough arguments; it diverges to positive infinity if the sum is terminated at an odd n number, and it diverges to negative infinity when the sum is terminated at an even n number. However, the true exponential integral always converges to zero very quickly. We can use this to our advantage. We make sure the sum diverges to negative infinity by running the sum up to an even number. In this case, if our calculated exponential integral is a negative value, we can simply return zero instead, since we know the true value should never be negative. We just need to make sure that we have added sufficiently many terms, such that we have accurate calculated values up to several decimal digits, before the divergence occurs, and we truncate to zero. Considering both the accuracy requirement, and the limits of precision and representability of floating point numbers (especially for the $n!$ in the denominator), we use $n = 12$ as the summation limit. Figure D.4 below illustrates the achieved numerical accuracy of calculating the exponential integral.

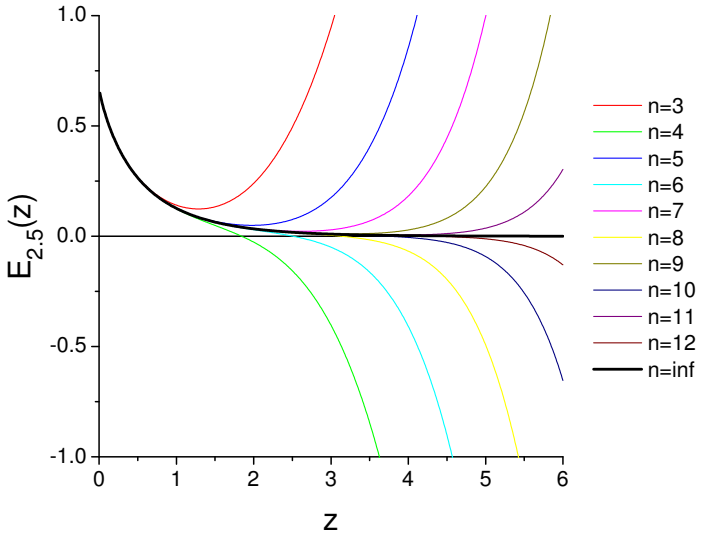

Figure D.4: Numerical estimates of the exponential integral function at various truncations of the infinite sum. For high enough and even n values we accept the calculated value of the exponential integral if it is positive, or return zero if negative, because the true exponential integral function converges to zero quickly.

D.2 Impact of Assortativity on RDS and CDS in Artificial Networks

We present here the comparison of dominating set sizes as a function of network assortativity, for CONF and cCONF networks, in Figs. D.5 and D.6, respectively. Various levels of assortativity are obtained by using our edge-mixing method with biased double edge-swaps, detailed in Section 4.5.

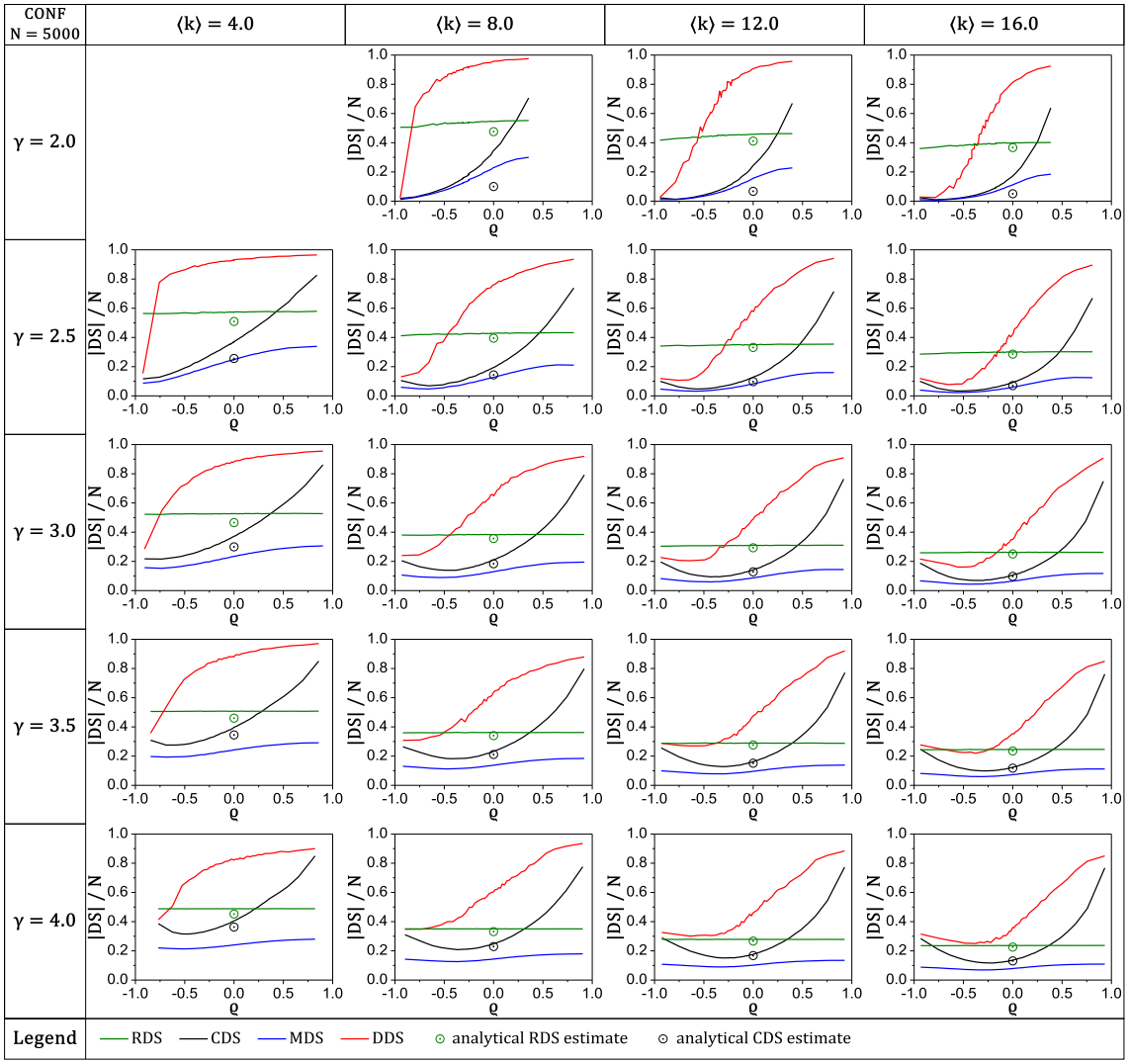

Figure D.5: Comparison of the sizes of dominating sets against assortativity, measured by Spearman's ρ. Each tile shows the following: probabilistic dominating set with uniform selection probability (RDS), cutoff dominating set (CDS), greedy minimum dominating set (MDS), degree-ranked dominating set (DDS). $N = 5000$, CONF networks, averaged over 100 samples at each data point. Analytical estimates of RDS and CDS are provided by Eq. (D.14) and Eq. (D.35), respectively. The missing tile has network parameters that cannot be realized.

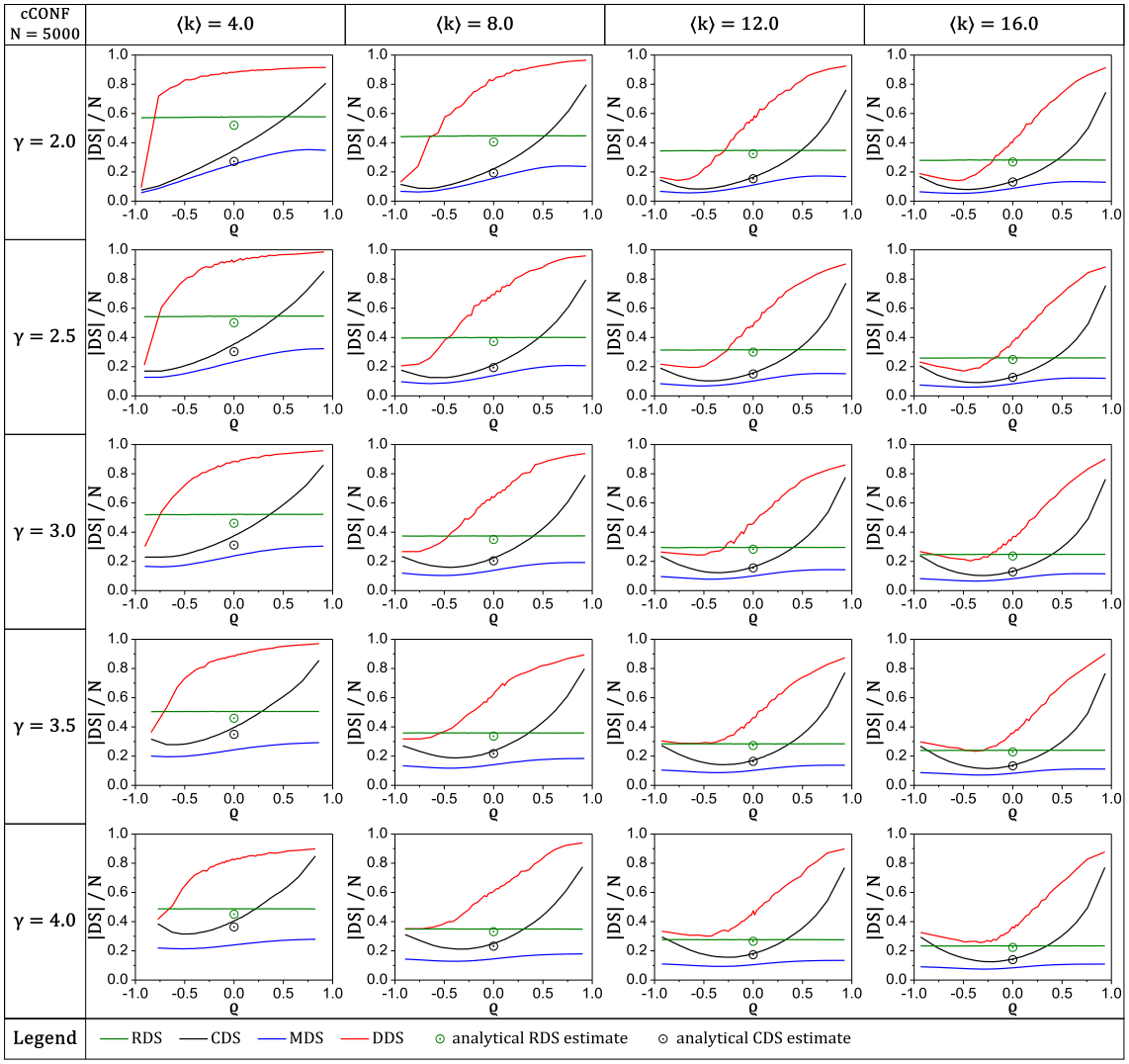

Figure D.6: Comparison of the sizes of dominating sets against assortativity, measured by Spearman's ρ. Each tile shows the following: probabilistic dominating set with uniform selection probability (RDS), cutoff dominating set (CDS), greedy minimum dominating set (MDS), degree-ranked dominating set (DDS). $N = 5000$, cCONF networks, averaged over 100 samples at each data point. Analytical estimates of RDS and CDS are provided by Eq. (D.14) and Eq. (D.35), respectively.

D.3 Impact of Assortativity on RDS and CDS in Real Networks

We analyzed the effect of changing assortativity on the dominating sets of real network samples. For each sample, we used the original network as a starting point, and used our assortativity control method (mixing edges by biased double edge-swaps) to achieve a certain desired assortativity level, measured by Spearman's ρ. We show the comparison of dominating set sizes for the following networks:

- Gnutella-08: A peer-to-peer network downloaded from the Stanford Large Network Dataset Collection [230]. It has directed edges that we have symmetrized to obtain an undirected network.
- Flickr: Social network of Flickr users, where edges represent either friendship or who-follows-whom relationship (symmetrized)
- Foursquare: Friendship network of Foursquare users.

The Flickr and Foursquare networks were collected at Rensselear Polytechnic Institute using BFS crawling. These networks are undirected and unweighted.

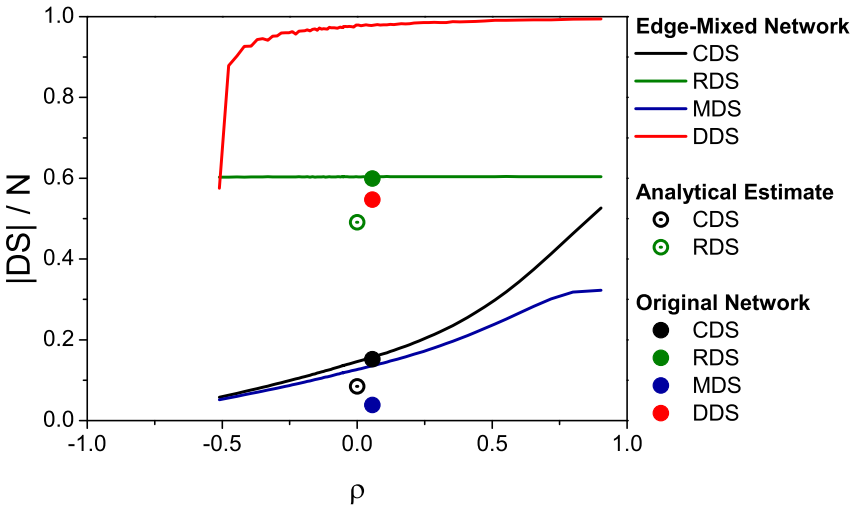

Figure D.7: Dominating sets as a function of Spearman's ρ assortativity measure in the Flickr network. Properties: $N = 248000$, $\gamma = 1.5$ **and** $\langle k \rangle = 32.9$.

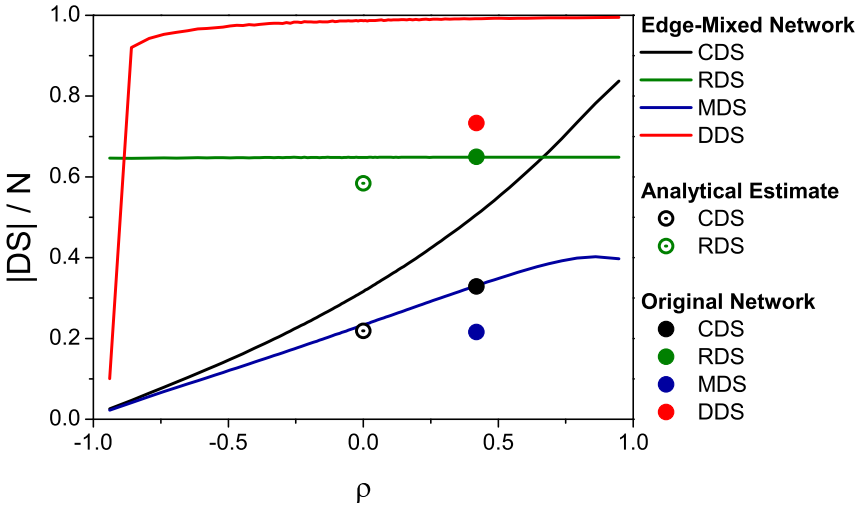

Figure D.8: Dominating sets as a function of Spearman's ρ assortativity measure in the Foursquare network. Properties: $N = 115201$, $\gamma = 1.9$ and $\langle k \rangle = 4.7$.

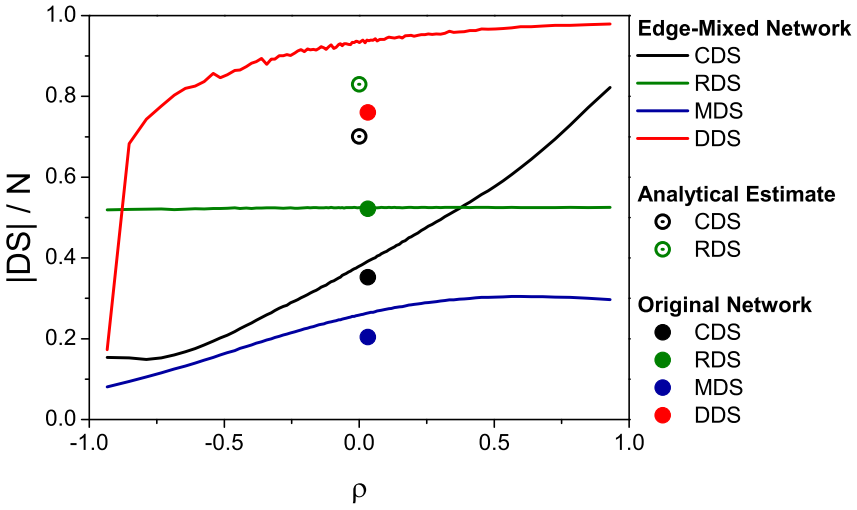

Figure D.9: Dominating sets as a function of Spearman's ρ assortativity measure in Gnutella network [230]. Properties: $N = 6301$, $\gamma = 4.7$ and $\langle k \rangle = 6.6$.

APPENDIX E
Damage-Resilient Dominating Sets

E.1 Finding a Flexible-Redundancy Dominating Set (frDS)

The algorithm for finding a flexible-redundancy dominating set (frDS) is based on greedy search. At each step we add one node to the dominating set, which helps the maximal number of nodes to advance toward their required domination goals. There are several variables that we must define and track for each node.

First, we define the *domination requirement* $r(i)$ as the number of required dominators for node i among its closed neighbors. This value is calculated and assigned randomly for each node before the search begins. The requirement is set to either $\lfloor r \rfloor$ or $\lceil r \rceil$ (where r is the global requirement for the entire network), the probability for the latter is exactly the fractional part of r (that is, $r - \lfloor r \rfloor$). Note, that $r(i)$ can be zero if $r < 1$, and it is also possible that $r > d(i) + 1$ (where $d(i)$ is the degree of node i), in which case all nodes in the closed neighborhood are required to be in the dominating set.

Second, we define $score(i)$ as the current number of dominators of node i at any given step. Initially, $score(i) = 0$ for every node, and it increases by one in the closed neighborhood of the selected node.

Finally, we track the dominating $potential(i)$ of node i, which counts how many nodes in the closed neighborhood of i have not yet reached their domination requirement. Specifically, $potential(i) = \sum_{j \in N^+(i)} I[score(j) < r(j)]$, where $N^+(i)$ is the closed neighborhood of i and $I[x]$ is an indicator function that returns 1 if x is true and 0 else. In other words, the potential is the number of nodes in the closed neighborhood that can be advanced toward their goal by selecting i as the next dominator. The greedy search is based on this quantity: at every step we select a node with maximum potential (with random tie-breaking among the candidates).

The key to implementing the algorithm with optimal time complexity is the use of an efficient data structure for maintaining a list of nodes sorted by their potentials. Note that the potential is an integer value between 0 and $N + 1$, therefore we can use

bucket-sort for initial sorting. We assign one bucket for each possible potential value, and we implement each bucket by a hashed set. This way we can add or remove a node from any bucket in $O(1)$ step, therefore we can perform the initial sorting in $O(N)$ and maintain sortedness in $O(1)$ step after any single change in a node's potential.

The time complexity of the algorithm can be found by analyzing the changes in scores and potentials of nodes. The initial calculation of potentials requires a loop over all nodes' all neighbors. Assuming we can enumerate the neighbors of node i in $d(i)$ steps, this calculation takes $\sum_{i \in V(G)} 1 + d(i) = 2E + N = O(E)$ steps. Then in the main loop one node is selected at every step, which increases the score of the selected node and its neighbors by one. In principle, the scores could increase until all nodes are selected (e.g., when $r > N$), therefore again all nodes' all neighbors are processed, taking $O(E)$ steps. However, during this procedure, there are additional steps for updating the node potentials. Some (usually all) nodes will reach their predefined requirement at one point or another, after which the dominating potentials change. We count these changes as follows. Initially, all nodes can increase all their neighbors' score toward their requirement (including the nodes themselves), therefore the initial sum of potentials is $\sum_{i \in V(G)} 1 + d(i) = 2E + N$, or less, if some nodes have zero requirement. The potential of a node can either be reduced by one if a neighbor reaches its requirement (and thus that neighbor can no longer be advanced to its goal by the current node), or it becomes zero by definition if the node is actually selected. At most, there are $2E + N = O(E)$ changes (reductions) of potentials, each computed in $O(1)$ time (maintaining sortedness of nodes after each change), therefore during the procedure there are at most $O(E)$ additional steps for updating node potentials. This means the entire algorithm runs in $O(E)$ steps. Note that in sparse networks, $O(E) = O(N)$.

Algorithm 2 Find an frDS

```
procedure FRDS(G: graph, r: domination redundancy)
    finished ← 0
    for all i ∈ V(G) do                    ▷ initialization of score, r, and potential
        score(i) ← 0
        potential(i) ← 0
        if Random(0, 1) < r − ⌊r⌋ then
            r(i) ← ⌈r⌉
        else
            r(i) ← ⌊r⌋
        end if
        if score(i) ≥ r(i) then
            finished ← finished + 1
        end if
    end for
    for all i ∈ V(G) do                    ▷ initial calculation of potentials
        for all j : (i, j) ∈ E(G) do
            if score(j) < r(j) then
                potential(i) ← potential(i) + 1
            end if
        end for
        if score(i) < goal(i) then
            potential(i) ← potential(i) + 1
        end if
    end for                                ▷ Alg. continues on the next page
```

```
    while finished < |V(G)| ∧ max(potential) > 0 do                ▷ main loop
        k := random node with maximum potential                    ▷ greedy step
        Add k to Dominating Set                                    ▷ construct the output
        score(k) ← score(k) + 1                                    ▷ count self-domination
        potential(k) ← 0                                           ▷ remove k from further consideration
        change ← score(k) = r(k)                                   ▷ requirement of k reached in this iteration?
        if change then
            finished ← finished + 1
        end if
        for all j : (j, k) ∈ E(G) do                               ▷ update neighbors of k
            if change then                                         ▷ neighbors cannot increase score(k) any more
                potential(j) ← max(0, potential(j) − 1)
            end if
            score(j) ← score(j) + 1                                ▷ k adds domination score to all its neighbors
            if score(j) = r(j) then                                ▷ requirement reached for the neighbor?
                finished ← finished + 1
                potential(j) ← max(0, potential(j) − 1)
                for all x : (x, j) ∈ E(G) do                       ▷ update second neighbors
                    if x ≠ k then                                  ▷ skip when second neighbor is k
                        potential(x) ← max(0, potential(x) − 1)
                    end if
                end for
            end if
        end for
    end while
end procedure
```

E.2 Finding a Flexible-Cost Dominating Set (fcDS)

The fcDS algorithm is also a form of greedy search, since it builds the dominating set by selecting one node at a time with maximum potential, similarly to the frDS algorithm. However, in this method the potential is calculated from the changes in probability of losing all dominators for the nodes in the neighborhood of the given node.

First, we define $strength(i)$ for each node i as an input ($0 < strength(i) < 1$), which defines the probability of not losing node i after the anticipated damage:

$$strength(i) := \Pr(i \text{ is not lost}). \tag{E.1}$$

We also keep a record of $instability(i)$ for each node i, which is defined as the probability of losing all dominators after the damage has occured:

$$instability(i) = \prod_{j \in \mathrm{DS} \cap N^+(i)} 1 - strength(j). \tag{E.2}$$

Initially, $instability(i) = 1.0$ for all i. The $potential(i)$ of node i, which is used in the greedy node selection, is calculated as the sum of the changes in instabilities over the closed neighborhood of node i, if i was selected:

$$\begin{aligned} potential(i) &= -\sum_{j \in N^+(i)} instability(j) \Pr(i \text{ is lost}) - instability(j) \\ &= -\sum_{j \in N^+(i)} instability(j) \left[\Pr(i \text{ is lost}) - 1\right] \\ &= \sum_{j \in N^+(i)} instability(j) \cdot strength(i). \end{aligned} \tag{E.3}$$

Note, that the negative sign is manually inserted to make the potential a positive value, for practical reasons. Without it, the change in instabilities would be negative, because by each node selection the stability always increases.

With the definition above, we select a node with maximum potential at each greedy step. After the node has been selected and added to the dominating set, the instabilities in the closed neighborhood, and the potentials for all nodes in the

second neighborhood of the selected node must be recalculated, and the nodes must be sorted again based on the new potentials. Since the potentials are non-integer values (and cannot be mapped to integers) we can only use comparative sorting, where it takes $O(\log N)$ steps to find the new place for each node in the list. With a simple approximation for sparse networks, a node in a network with average degree d will have $O(d^2)$ nodes in its second neighborhood, therefore the selection of each dominator involves $O(d^2 \log N)$ steps.

Algorithm 3 Find an fcDS

```
procedure FCDS(G: graph, strength: array, c: number of nodes to select)
    for all i ∈ V(G) do                          ▷ initialization of instability and potential
        instability(i) ← 1.0
        potential(i) ← (degree(i) + 1)(1 − strength(i))
    end for
    for a ← 1 ... c do                                   ▷ a simply counts the output
        k ← random node with maximum potential                    ▷ greedy selection
        Add k to Dominating Set                                   ▷ construct the output
        S ← ∅                          ▷ set of nodes whose potential must be updated
        instability(k) ← instability(k)(1 − strength(k))   ▷ update self instability
        for all j : (k, j) ∈ E(G) do
            instability(j) ← instability(j)(1 − strength(k)  ▷ update instability of
neighbors
            S ← S ∪ {j}                                   ▷ request potential update for j
            for all i : (j, i) ∈ E(G) do
                S ← S ∪ {i}          ▷ request potential update for second neighbors
            end for
        end for
        for all i ∈ S do                                          ▷ update potentials
            potential(i) ← 0
            if i ∉ Dominating Set then
                for all j : (i, j) ∈ E(G) do
                    potential(i) ← potential(i) + instability(j)strength(i)
                end for
                potential(i) ← potential(i) + instability(i)strength(i)
            end if
        end for
    end for
end procedure
```

E.3 Dependence of Domination Stability on Degree Exponent and Assortativity

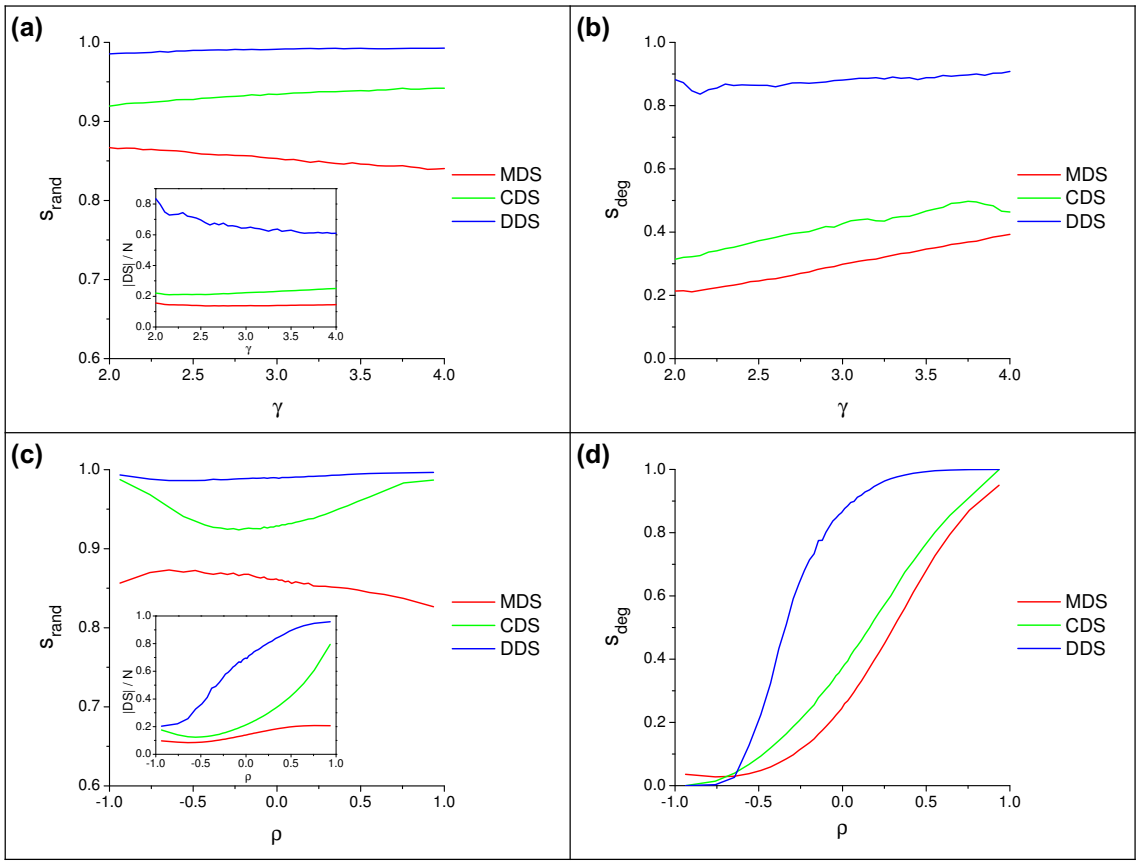

Figure E.1: Stability of dominating sets vs. power-law degree exponent γ and Spearman's ρ assortativity measure. (a) and (c) present random node removal, (b) and (d) show degree-ranked node removal. The insets illustrate the sizes of the corresponding dominating sets. In (a) and (b): $\rho = 0.0$; in (c) and (d): $\gamma = 2.5$. Common parameters: $N = 5000$, $\langle k \rangle = 8$, $f = 0.3$. Results are averaged over 200 network samples.

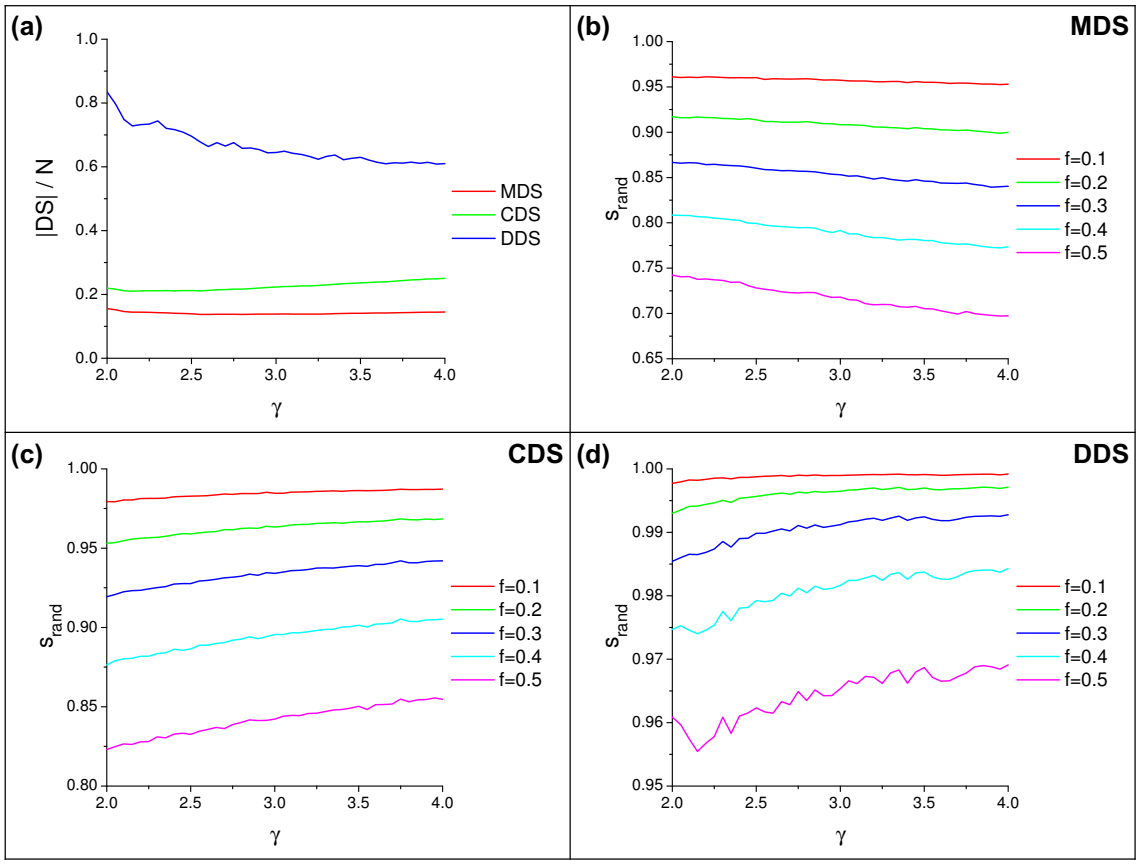

Figure E.2: Comparison of size and stability of dominating sets vs. power-law degree exponent, at various fractions of random node removal. Synthetic networks, $N = 5000$, $\langle k \rangle = 8$, $\rho = 0.0$. Results are averaged over 200 network samples.

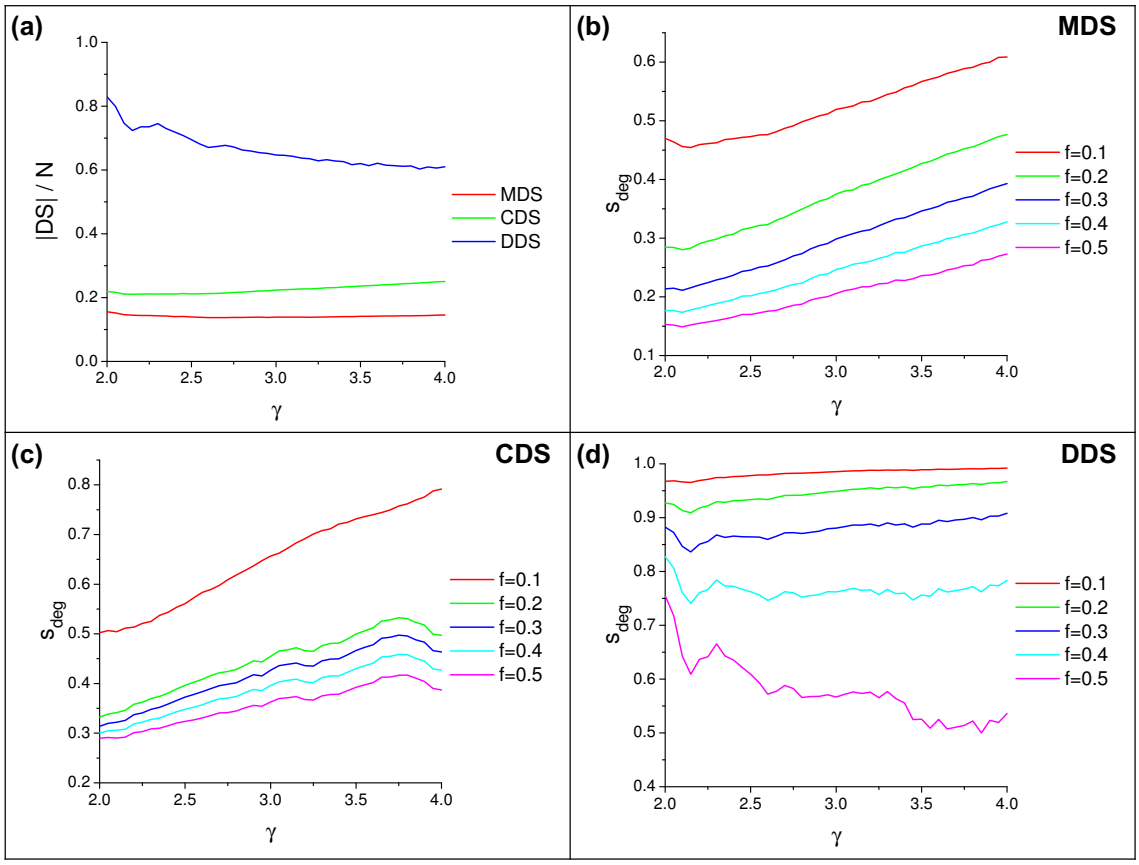

Figure E.3: Comparison of size and stability of dominating sets vs. power-law degree exponent at various fractions of degree ranked node removal. Synthetic networks, $N = 5000$, $\langle k \rangle = 8$, $\rho = 0.0$. Results are averaged over 200 network samples.

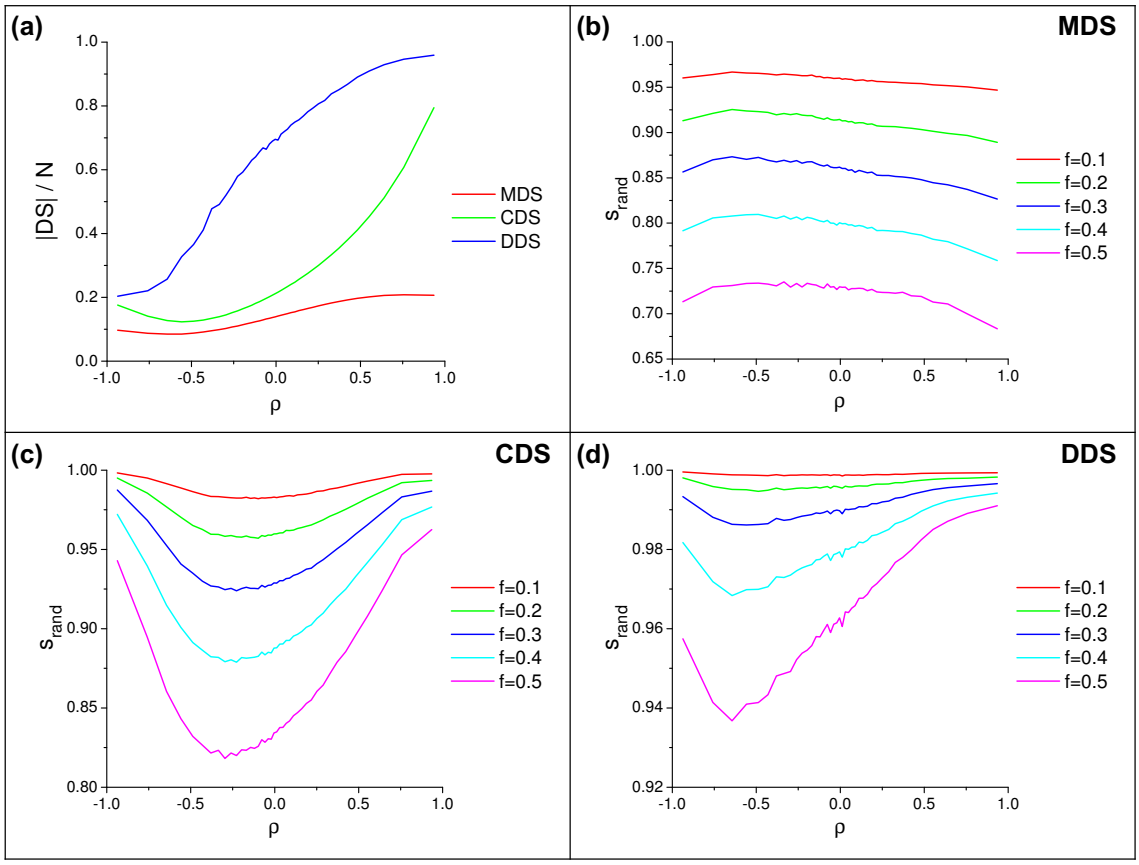

Figure E.4: Comparison of size and stability of dominating sets vs. assortativity at various fractions of random node removal. Synthetic networks, $N = 5000$, $\langle k \rangle = 8$, $\gamma = 2.5$. Results are averaged over 200 network samples.

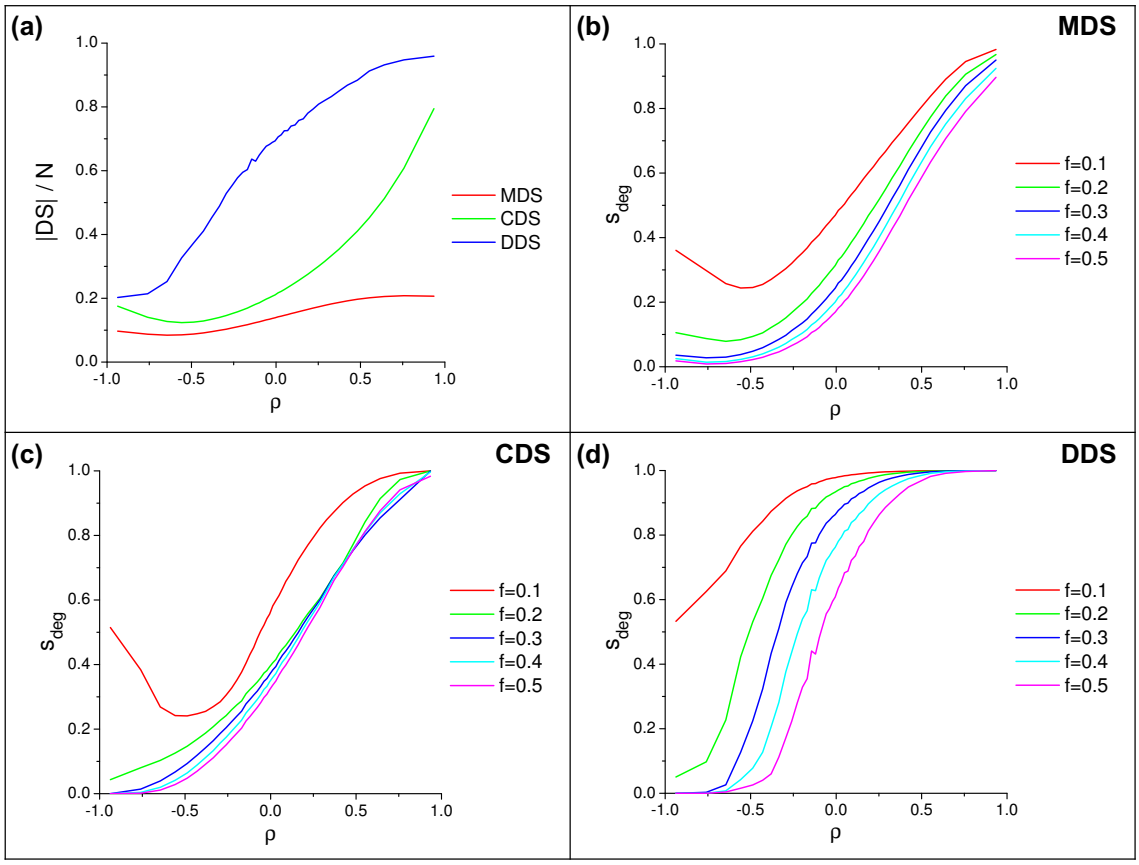

Figure E.5: Comparison of size and stability of dominating sets vs. assortativity at various fractions of degree-ranked node removal. Synthetic networks, $N = 5000$, $\langle k \rangle = 8$, $\gamma = 2.5$. Results are averaged over 200 network samples.

E.4 Stability of Partial Flexible-Redundancy Dominating Sets

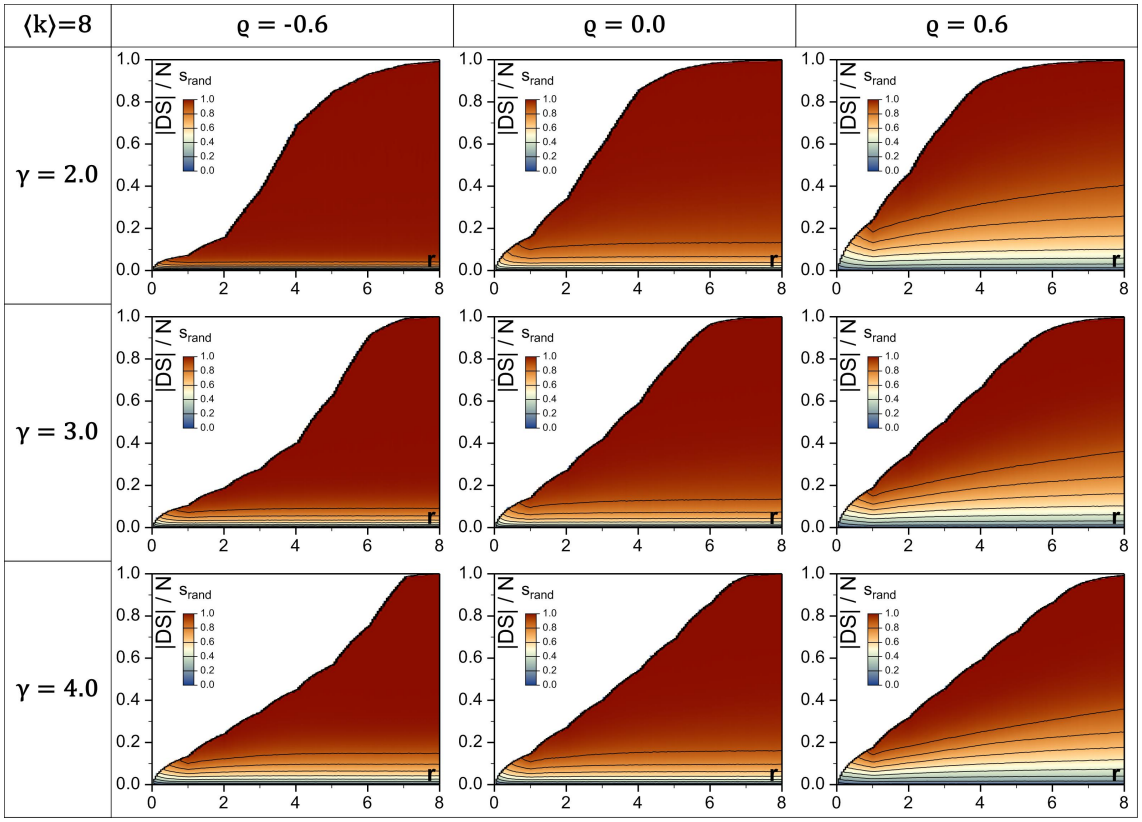

Figure E.6: Stability of partial frDS against random damage, as a function of redundancy level and dominating set size, at various power-law degree exponents and Spearman's ρ values. Synthetic networks, $N = 5000$, damage fraction $f = 0.1$.

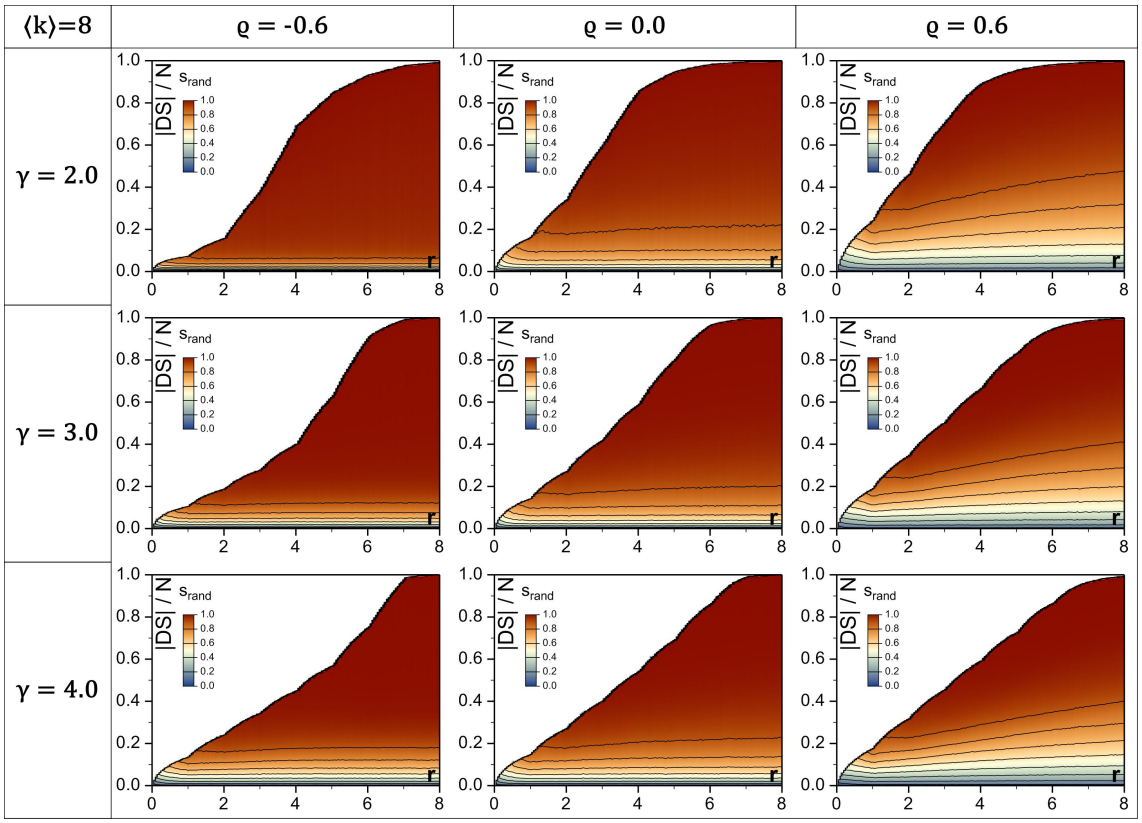

Figure E.7: Stability of partial frDS against random damage, as a function of redundancy level and dominating set size, at various power-law degree exponents and Spearman's ρ values. Synthetic networks, $N = 5000$, damage fraction $f = 0.3$.

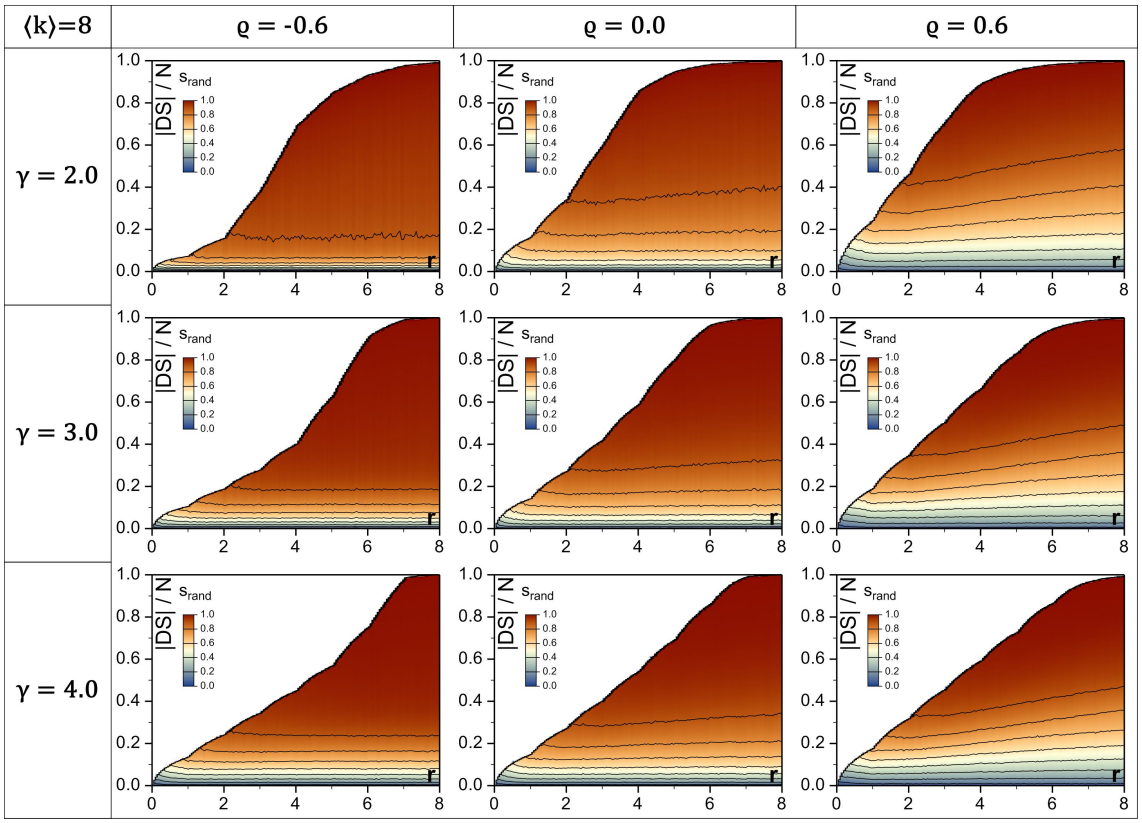

Figure E.8: Stability of partial frDS against random damage, as a function of redundancy level and dominating set size, at various power-law degree exponents and Spearman's ρ values. Synthetic networks, $N = 5000$, damage fraction $f = 0.5$.

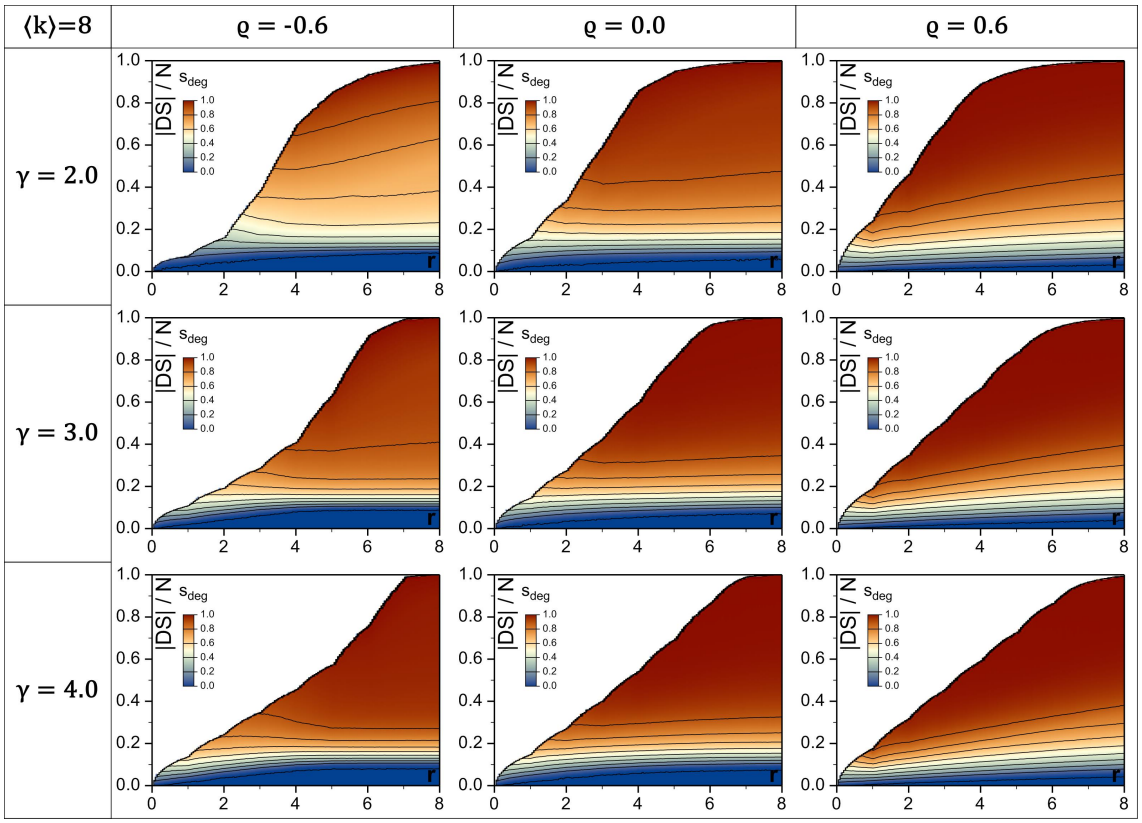

Figure E.9: Stability of partial frDS against targeted attack, as a function of redundancy level and dominating set size, at various power-law degree exponents and Spearman's ρ values. Synthetic networks, $N = 5000$, damage fraction $f = 0.1$.

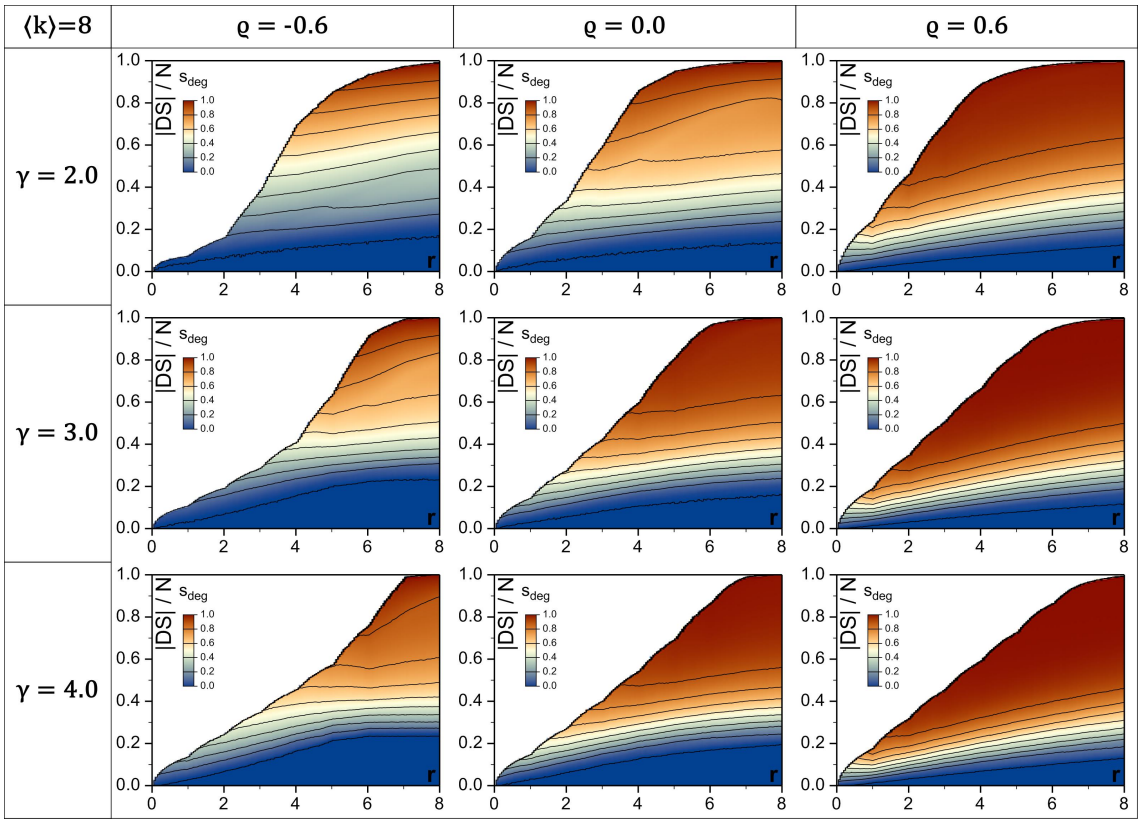

Figure E.10: Stability of partial frDS against targeted attack, as a function of redundancy level and dominating set size, at various power-law degree exponents and Spearman's ρ values. Synthetic networks, $N = 5000$, damage fraction $f = 0.3$.

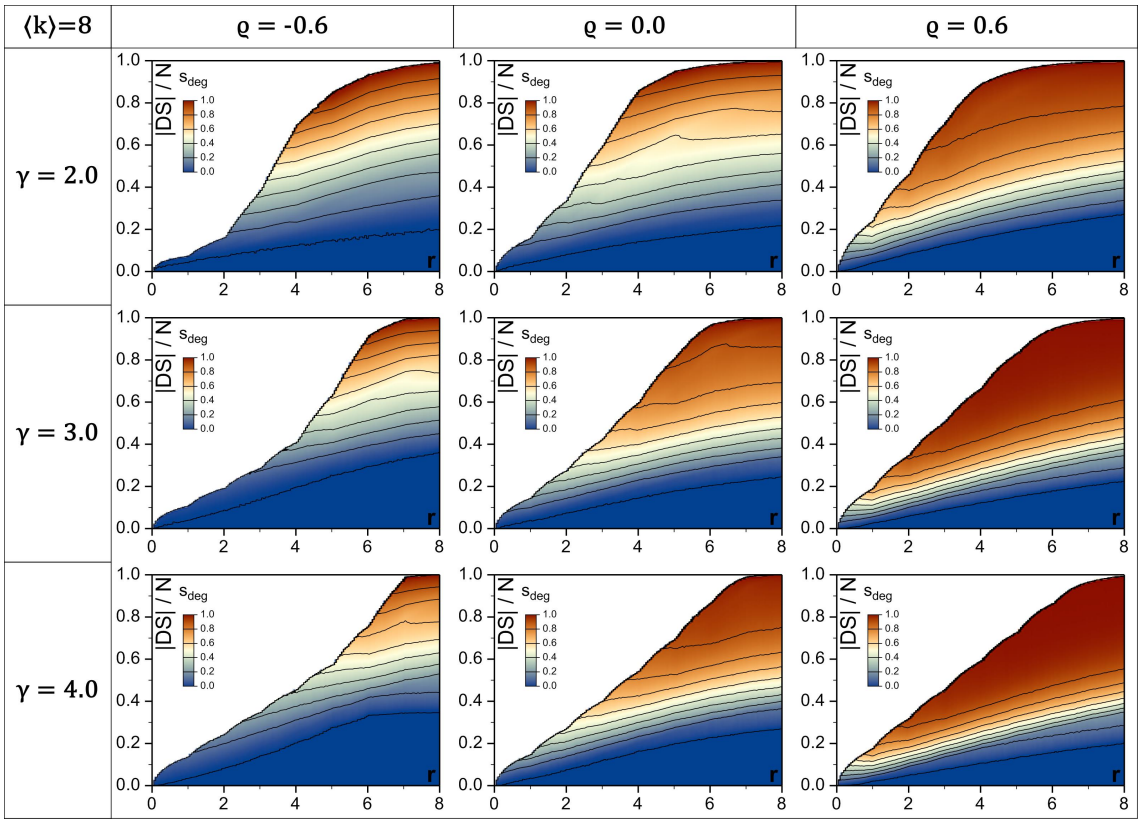

Figure E.11: Stability of partial frDS against targeted attack, as a function of redundancy level and dominating set size, at various power-law degree exponents and Spearman's ρ values. Synthetic networks, $N = 5000$, damage fraction $f = 0.5$.

E.5 Size of Flexible-Redundancy Dominating Sets

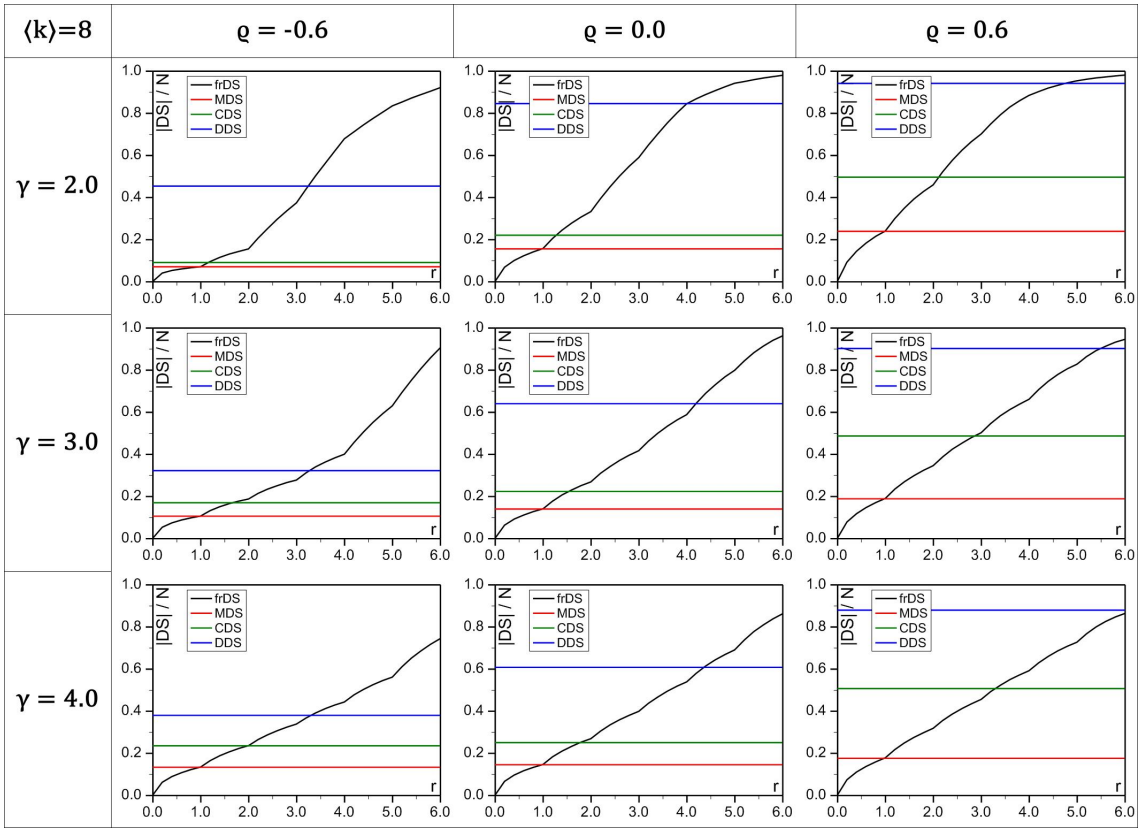

Figure E.12: Size of frDS as a function of domination redundancy, at various power-law degree exponents and Spearman's ρ values, in synthetic networks with $N = 5000$. The sizes of MDS, CDS, and DDS are shown for comparison.

E.6 Stability Comparison of frDS and fcDS against Dominating Set Size

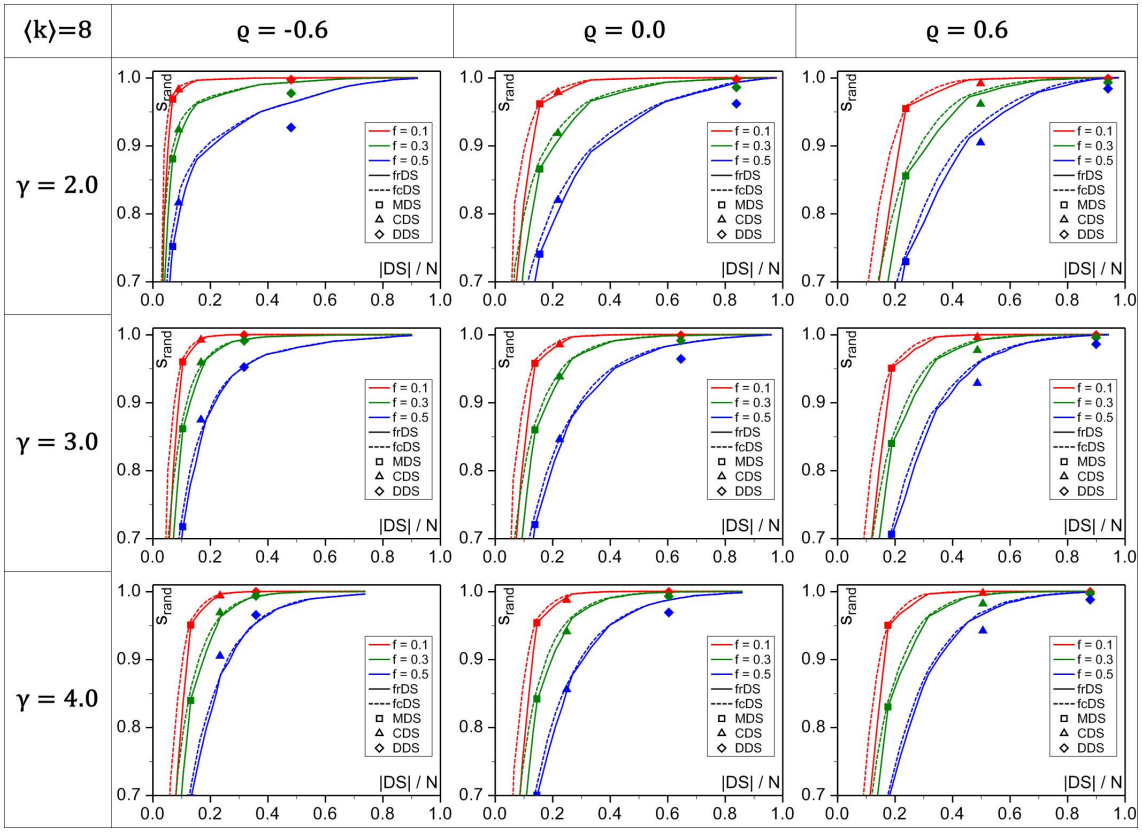

Figure E.13: Stability of frDS and fcDS against random damage, as a function of dominating set size (cost), at various power-law degree exponents and Spearman's ρ values, in synthetic networks with $N = 5000$. The stabilities of MDS, CDS, and DDS are shown for comparison at their corresponding set sizes. Black legend symbols refer to the shape only, colors refer to damage fractions.

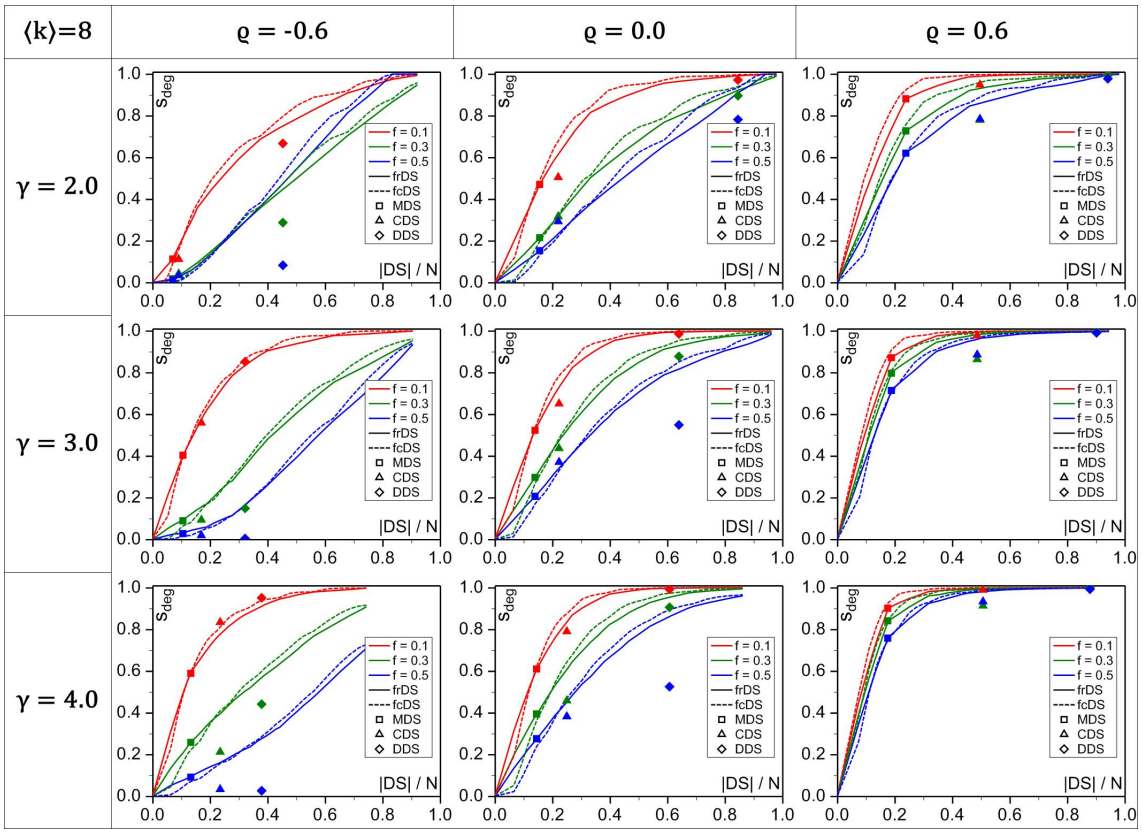

Figure E.14: Stability of frDS and fcDS against targeted attack, as a function of dominating set size (cost), at various power-law degree exponents and Spearman's ρ values, in synthetic networks with $N = 5000$. The stabilities of MDS, CDS, and DDS are shown for comparison at their corresponding set sizes. Black legend symbols refer to the shape only, colors refer to damage fractions.

E.7 Degree Distributions of Real Networks Used in Domination Stability Analysis

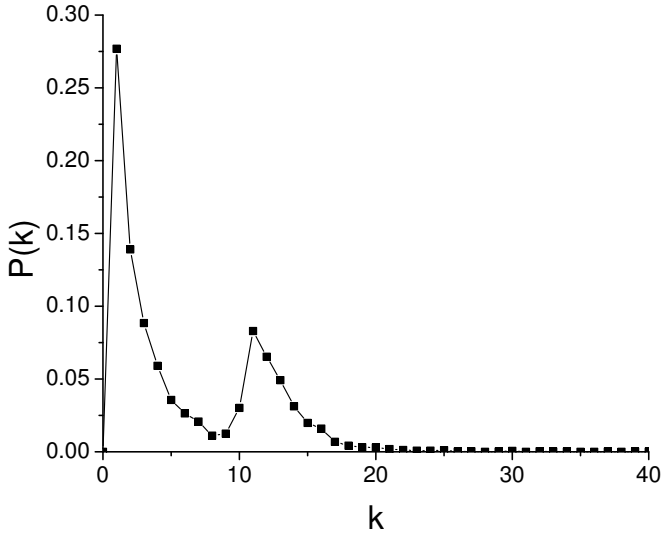

Figure E.15: Degree distribution of Gnutella08 network [230] on linear scale.

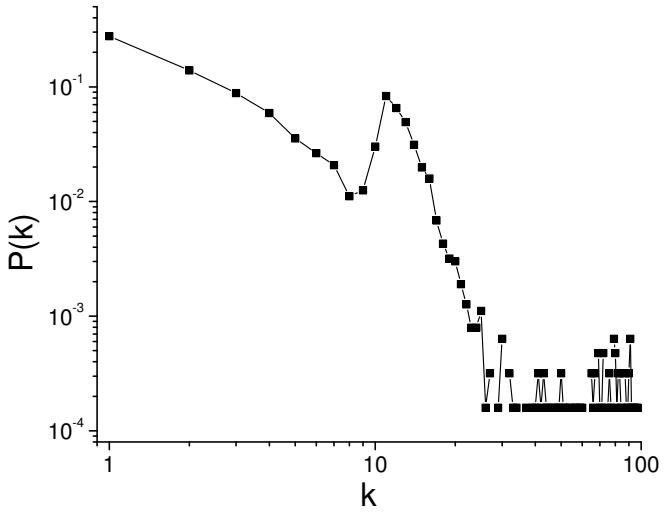

Figure E.16: Degree distribution of Gnutella08 network [230] on double-logarithmic scale.

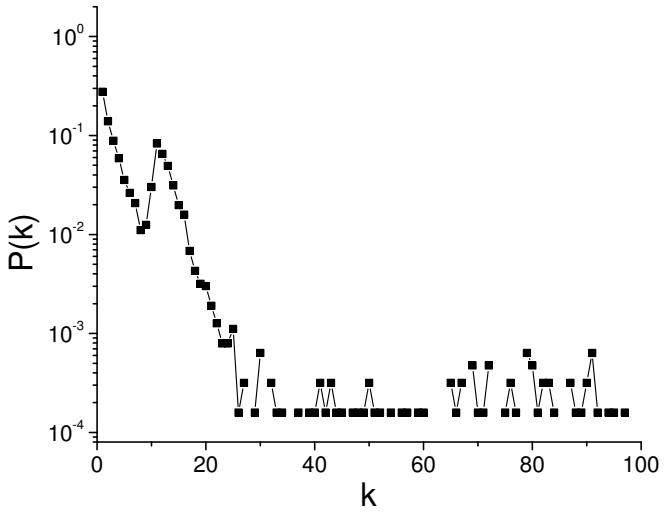

Figure E.17: Degree distribution of Gnutella08 network [230] on log-linear scale.

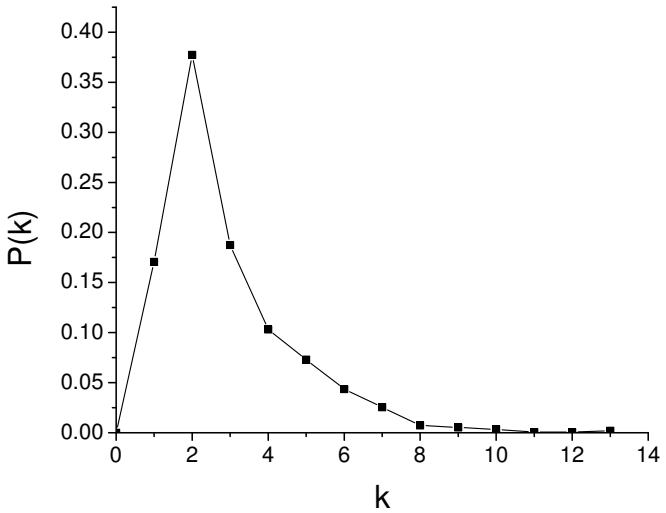

Figure E.18: Degree distribution of ENTSO-E powergrid [239, 240] on linear scale.

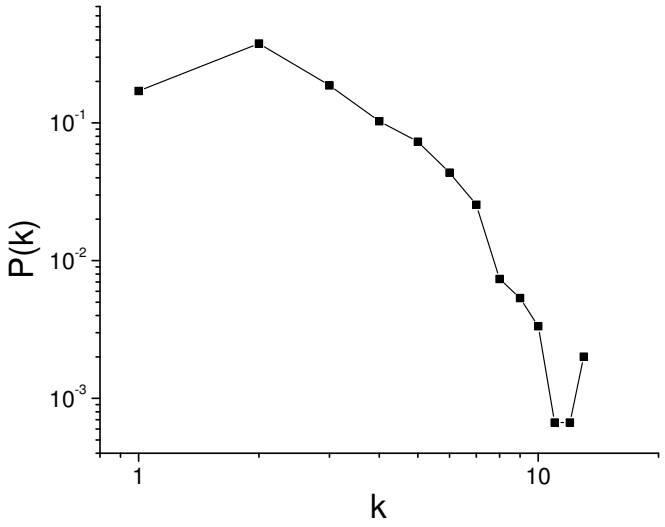

Figure E.19: Degree distribution of ENTSO-E powergrid [239, 240] on double-logarithmic scale.

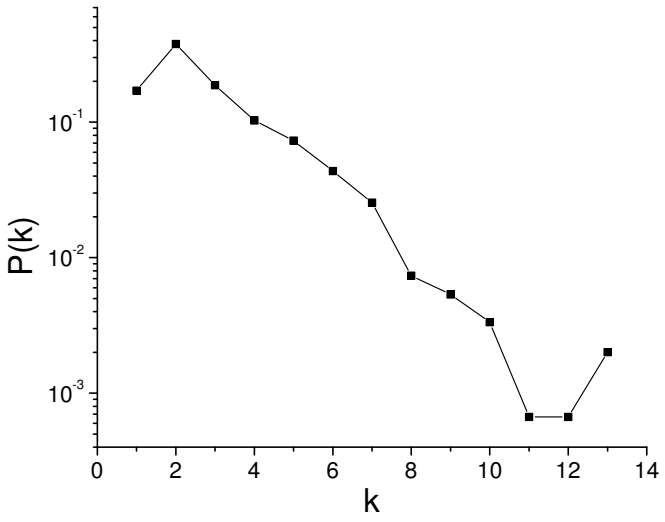

Figure E.20: Degree distribution of ENTSO-E powergrid [239, 240] on log-linear scale.

E.8 Brief Analysis of the Brain Graph Dataset

Table E.1: Sample numbers of brain graphs that we use to identify them in our figures. Graph data files are available at [241].

#	graph	#	graph
1	KKI-21_KKI2009-01_big_graph_w_inv	51	MRN114_M87114047_big_graph_w_inv
2	KKI-21_KKI2009-02_big_graph_w_inv	52	MRN114_M87114064_big_graph_w_inv
3	KKI-21_KKI2009-03_big_graph_w_inv	53	MRN114_M87115498_big_graph_w_inv
4	KKI-21_KKI2009-04_big_graph_w_inv	54	MRN114_M87115517_big_graph_w_inv
5	KKI-21_KKI2009-05_big_graph_w_inv	55	MRN114_M87117119_big_graph_w_inv
6	KKI-21_KKI2009-06_big_graph_w_inv	56	MRN114_M87117167_big_graph_w_inv
7	KKI-21_KKI2009-07_big_graph_w_inv	57	MRN114_M87120962_big_graph_w_inv
8	KKI-21_KKI2009-08_big_graph_w_inv	58	MRN114_M87121943_big_graph_w_inv
9	KKI-21_KKI2009-09_big_graph_w_inv	59	MRN114_M87121956_big_graph_w_inv
10	KKI-21_KKI2009-10_big_graph_w_inv	60	MRN114_M87122092_big_graph_w_inv
11	KKI-21_KKI2009-11_big_graph_w_inv	61	MRN114_M87123042_big_graph_w_inv
12	KKI-21_KKI2009-12_big_graph_w_inv	62	MRN114_M87123449_big_graph_w_inv
13	KKI-21_KKI2009-13_big_graph_w_inv	63	MRN114_M87123913_big_graph_w_inv
14	KKI-21_KKI2009-14_big_graph_w_inv	64	MRN114_M87124633_big_graph_w_inv
15	KKI-21_KKI2009-15_big_graph_w_inv	65	MRN114_M87124781_big_graph_w_inv
16	KKI-21_KKI2009-16_big_graph_w_inv	66	MRN114_M87124827_big_graph_w_inv
17	KKI-21_KKI2009-17_big_graph_w_inv	67	MRN114_M87125134_big_graph_w_inv
18	KKI-21_KKI2009-18_big_graph_w_inv	68	MRN114_M87128444_big_graph_w_inv
19	KKI-21_KKI2009-19_big_graph_w_inv	69	MRN114_M87129719_big_graph_w_inv
20	KKI-21_KKI2009-20_big_graph_w_inv	70	MRN114_M87129789_big_graph_w_inv
21	KKI-21_KKI2009-21_big_graph_w_inv	71	MRN114_M87131806_big_graph_w_inv
22	KKI-21_KKI2009-22_big_graph_w_inv	72	MRN114_M87134068_big_graph_w_inv
23	KKI-21_KKI2009-23_big_graph_w_inv	73	MRN114_M87135647_big_graph_w_inv
24	KKI-21_KKI2009-24_big_graph_w_inv	74	MRN114_M87136332_big_graph_w_inv
25	KKI-21_KKI2009-25_big_graph_w_inv	75	MRN114_M87136832_big_graph_w_inv
26	KKI-21_KKI2009-26_big_graph_w_inv	76	MRN114_M87139021_big_graph_w_inv
27	KKI-21_KKI2009-27_big_graph_w_inv	77	MRN114_M87139257_big_graph_w_inv
28	KKI-21_KKI2009-28_big_graph_w_inv	78	MRN114_M87141220_big_graph_w_inv
29	KKI-21_KKI2009-29_big_graph_w_inv	79	MRN114_M87141664_big_graph_w_inv
30	KKI-21_KKI2009-30_big_graph_w_inv	80	MRN114_M87141793_big_graph_w_inv
31	KKI-21_KKI2009-31_big_graph_w_inv	81	MRN114_M87141858_big_graph_w_inv
32	KKI-21_KKI2009-32_big_graph_w_inv	82	MRN114_M87141906_big_graph_w_inv
33	KKI-21_KKI2009-33_big_graph_w_inv	83	MRN114_M87141949_big_graph_w_inv
34	KKI-21_KKI2009-34_big_graph_w_inv	84	MRN114_M87142764_big_graph_w_inv
35	KKI-21_KKI2009-35_big_graph_w_inv	85	MRN114_M87143273_big_graph_w_inv
36	KKI-21_KKI2009-36_big_graph_w_inv	86	MRN114_M87144889_big_graph_w_inv
37	KKI-21_KKI2009-37_big_graph_w_inv	87	MRN114_M87144896_big_graph_w_inv
38	KKI-21_KKI2009-38_big_graph_w_inv	88	MRN114_M87145479_big_graph_w_inv
39	KKI-21_KKI2009-39_big_graph_w_inv	89	MRN114_M87145575_big_graph_w_inv
40	KKI-21_KKI2009-40_big_graph_w_inv	90	MRN114_M87146520_big_graph_w_inv
41	KKI-21_KKI2009-41_big_graph_w_inv	91	MRN114_M87146993_big_graph_w_inv
42	KKI-21_KKI2009-42_big_graph_w_inv	92	MRN114_M87147006_big_graph_w_inv
43	MRN114_M87102217_big_graph_w_inv	93	MRN114_M87148745_big_graph_w_inv
44	MRN114_M87102806_big_graph_w_inv	94	MRN114_M87149014_big_graph_w_inv
45	MRN114_M87103074_big_graph_w_inv	95	MRN114_M87149025_big_graph_w_inv
46	MRN114_M87105476_big_graph_w_inv	96	MRN114_M87150194_big_graph_w_inv
47	MRN114_M87107085_big_graph_w_inv	97	MRN114_M87150415_big_graph_w_inv
48	MRN114_M87108094_big_graph_w_inv	98	MRN114_M87150639_big_graph_w_inv
49	MRN114_M87111487_big_graph_w_inv	99	MRN114_M87151117_big_graph_w_inv
50	MRN114_M87111924_big_graph_w_inv	100	MRN114_M87151146_big_graph_w_inv

#	graph	#	graph
101	MRN114_M87151453_big_graph_w_inv	151	MRN114_M87192995_big_graph_w_inv
102	MRN114_M87152844_big_graph_w_inv	152	MRN114_M87193409_big_graph_w_inv
103	MRN114_M87153569_big_graph_w_inv	153	MRN114_M87196363_big_graph_w_inv
104	MRN114_M87154559_big_graph_w_inv	154	MRN114_M87196591_big_graph_w_inv
105	MRN114_M87155496_big_graph_w_inv	155	MRN114_M87199297_big_graph_w_inv
106	MRN114_M87155949_big_graph_w_inv	156	MRN114_M87199728_big_graph_w_inv
107	MRN114_M87156106_big_graph_w_inv	157	NKI-TRT_0021001_1_big_graph_w_inv
108	MRN114_M87157827_big_graph_w_inv	158	NKI-TRT_0021001_2_big_graph_w_inv
109	MRN114_M87158338_big_graph_w_inv	159	NKI-TRT_0021002_1_big_graph_w_inv
110	MRN114_M87158534_big_graph_w_inv	160	NKI-TRT_0021002_2_big_graph_w_inv
111	MRN114_M87159410_big_graph_w_inv	161	NKI-TRT_0021006_1_big_graph_w_inv
112	MRN114_M87159580_big_graph_w_inv	162	NKI-TRT_0021006_2_big_graph_w_inv
113	MRN114_M87160332_big_graph_w_inv	163	NKI-TRT_0021018_1_big_graph_w_inv
114	MRN114_M87160375_big_graph_w_inv	164	NKI-TRT_0021018_2_big_graph_w_inv
115	MRN114_M87161235_big_graph_w_inv	165	NKI-TRT_0021024_1_big_graph_w_inv
116	MRN114_M87161902_big_graph_w_inv	166	NKI-TRT_0021024_2_big_graph_w_inv
117	MRN114_M87162915_big_graph_w_inv	167	NKI-TRT_1427581_2_big_graph_w_inv
118	MRN114_M87164412_big_graph_w_inv	168	NKI-TRT_1793622_1_big_graph_w_inv
119	MRN114_M87164886_big_graph_w_inv	169	NKI-TRT_1793622_2_big_graph_w_inv
120	MRN114_M87165017_big_graph_w_inv	170	NKI-TRT_1961098_1_big_graph_w_inv
121	MRN114_M87165441_big_graph_w_inv	171	NKI-TRT_1961098_2_big_graph_w_inv
122	MRN114_M87166115_big_graph_w_inv	172	NKI-TRT_2475376_1_big_graph_w_inv
123	MRN114_M87168759_big_graph_w_inv	173	NKI-TRT_2475376_2_big_graph_w_inv
124	MRN114_M87174803_big_graph_w_inv	174	NKI-TRT_2799329_1_big_graph_w_inv
125	MRN114_M87176019_big_graph_w_inv	175	NKI-TRT_2799329_2_big_graph_w_inv
126	MRN114_M87176708_big_graph_w_inv	176	NKI-TRT_2842950_1_big_graph_w_inv
127	MRN114_M87178630_big_graph_w_inv	177	NKI-TRT_2842950_2_big_graph_w_inv
128	MRN114_M87179511_big_graph_w_inv	178	NKI-TRT_3201815_1_big_graph_w_inv
129	MRN114_M87179597_big_graph_w_inv	179	NKI-TRT_3201815_2_big_graph_w_inv
130	MRN114_M87179713_big_graph_w_inv	180	NKI-TRT_3313349_1_big_graph_w_inv
131	MRN114_M87181205_big_graph_w_inv	181	NKI-TRT_3313349_2_big_graph_w_inv
132	MRN114_M87181216_big_graph_w_inv	182	NKI-TRT_3315657_1_big_graph_w_inv
133	MRN114_M87182922_big_graph_w_inv	183	NKI-TRT_3315657_2_big_graph_w_inv
134	MRN114_M87183189_big_graph_w_inv	184	NKI-TRT_3795193_1_big_graph_w_inv
135	MRN114_M87183485_big_graph_w_inv	185	NKI-TRT_3795193_2_big_graph_w_inv
136	MRN114_M87184910_big_graph_w_inv	186	NKI-TRT_3808535_1_big_graph_w_inv
137	MRN114_M87185000_big_graph_w_inv	187	NKI-TRT_3808535_2_big_graph_w_inv
138	MRN114_M87186642_big_graph_w_inv	188	NKI-TRT_3893245_2_big_graph_w_inv
139	MRN114_M87187090_big_graph_w_inv	189	NKI-TRT_4176156_1_big_graph_w_inv
140	MRN114_M87187750_big_graph_w_inv	190	NKI-TRT_4176156_2_big_graph_w_inv
141	MRN114_M87187984_big_graph_w_inv	191	NKI-TRT_4288245_1_big_graph_w_inv
142	MRN114_M87188000_big_graph_w_inv	192	NKI-TRT_4288245_2_big_graph_w_inv
143	MRN114_M87188762_big_graph_w_inv	193	NKI-TRT_6471972_1_big_graph_w_inv
144	MRN114_M87190609_big_graph_w_inv	194	NKI-TRT_7055197_1_big_graph_w_inv
145	MRN114_M87190745_big_graph_w_inv	195	NKI-TRT_7055197_2_big_graph_w_inv
146	MRN114_M87191087_big_graph_w_inv	196	NKI-TRT_8574662_1_big_graph_w_inv
147	MRN114_M87191258_big_graph_w_inv	197	NKI-TRT_8735778_1_big_graph_w_inv
148	MRN114_M87192333_big_graph_w_inv	198	NKI-TRT_8735778_2_big_graph_w_inv
149	MRN114_M87192557_big_graph_w_inv	199	NKI-TRT_9630905_1_big_graph_w_inv
150	MRN114_M87192637_big_graph_w_inv	200	NKI-TRT_9630905_2_big_graph_w_inv

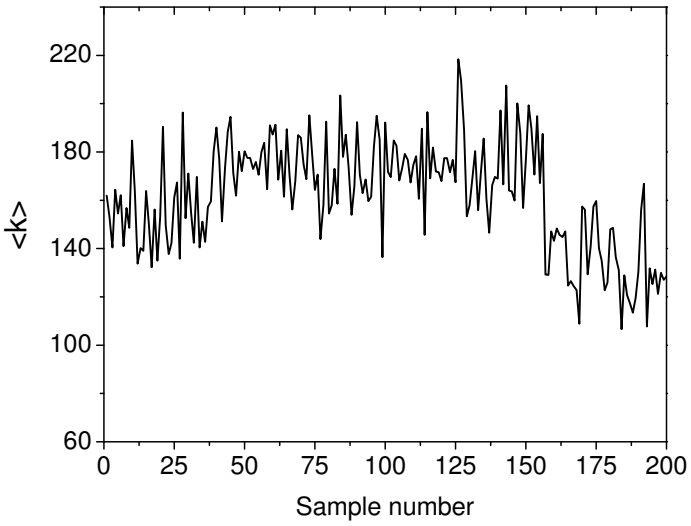

Figure E.21: Average degree in brain graphs. See Table E.1 for sample numbers.

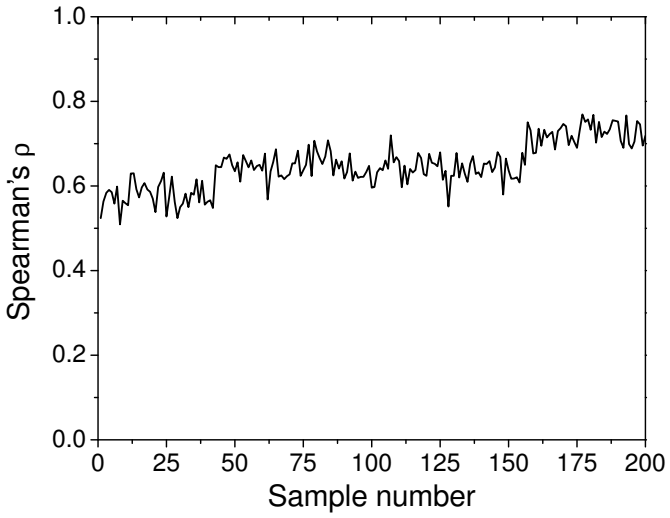

Figure E.22: Assortativity of brain graphs. See Table E.1 for sample numbers.

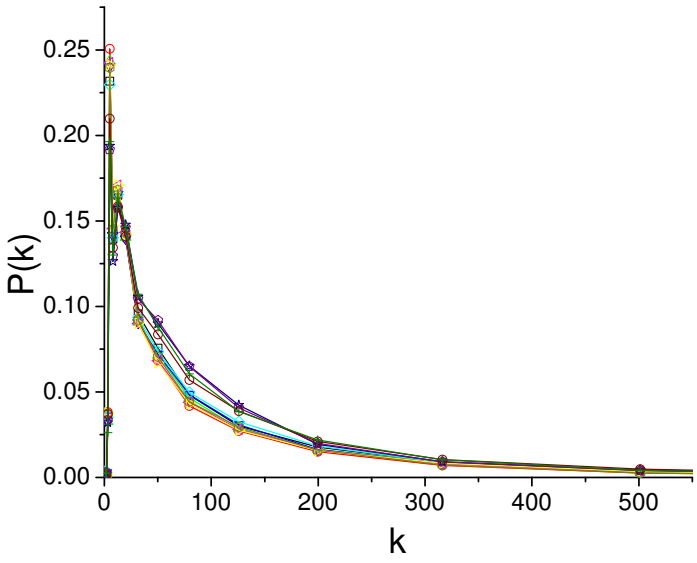

Figure E.23: Degree distributions of 12 randomly picked brain graphs on linear scale.

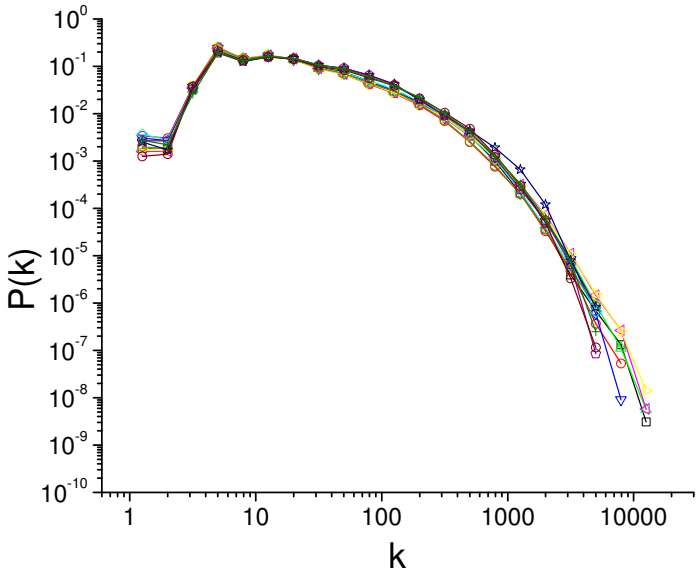

Figure E.24: Degree distributions of 12 randomly picked brain graphs on double-logarithmic scale.

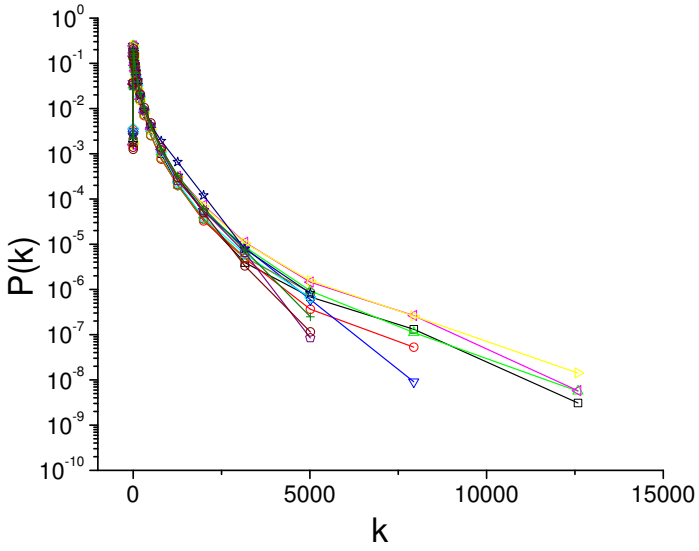

Figure E.25: Degree distributions of 12 randomly picked brain graphs on log-linear scale.

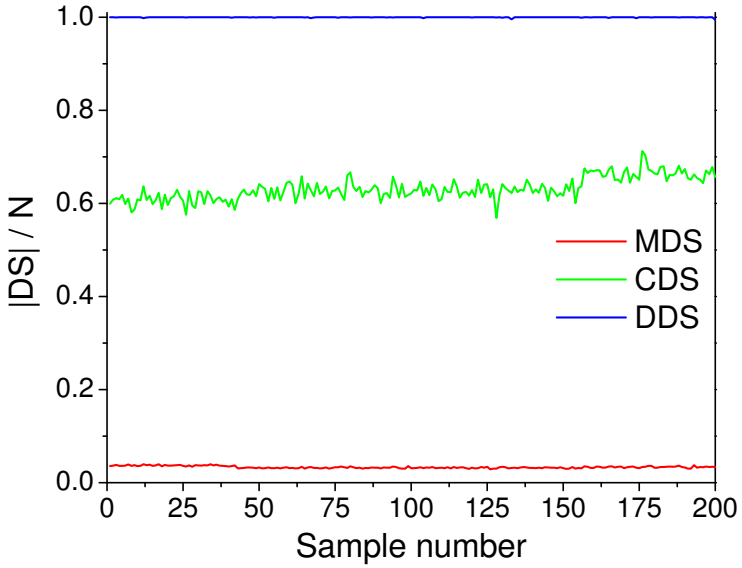

Figure E.26: Comparison of dominating set sizes in brain graphs. See Table E.1 for sample numbers.

E.9 Effects of Changing Assortativity on frDS in Real Networks

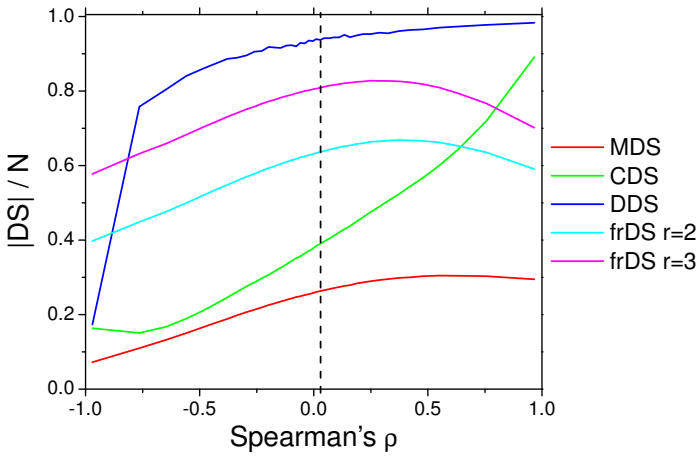

Figure E.27: Dominating set sizes vs. assortativity in Gnutella08 graph [230], achieved by random (biased) mixing of exges by double-edge swaps. The vertical dashed line indicates the assortativity of the original graph.

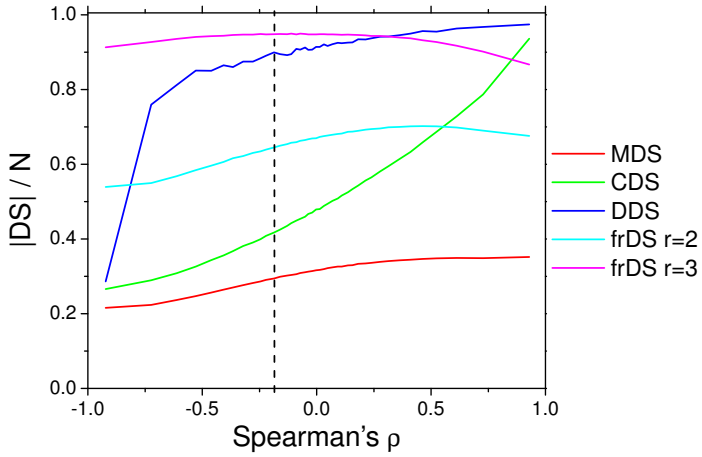

Figure E.28: Dominating set sizes vs. assortativity in powergrid graph [239, 240], achieved by random (biased) mixing of exges by double-edge swaps. The vertical dashed line indicates the assortativity of the original graph.

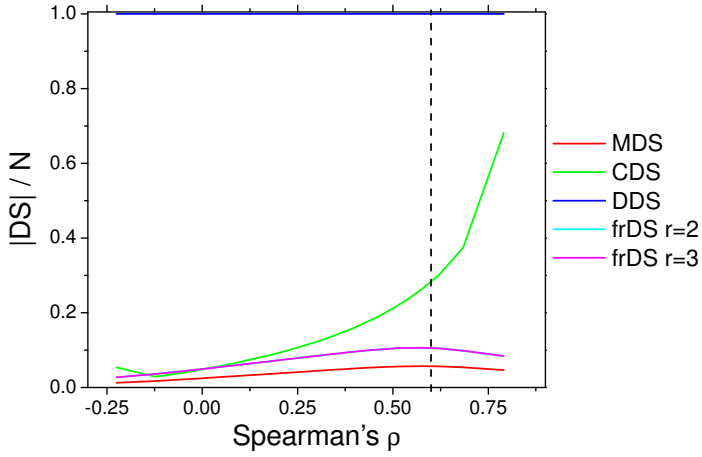

Figure E.29: Dominating set sizes vs. assortativity in brain graph KKI-21_KKI2009-19 [24, 241], achieved by random (biased) mixing of edges by double-edge swaps. The vertical dashed line indicates the assortativity of the original graph. Note that frDS curves with $r = 2$ and $r = 3$ overlap.