ABSTRACT

The cyber-physical systems (CPS) are emerging as the underpinning technology for major industries in the 21th century. This dissertation is focused on two fundamental issues in cyber-physical systems: network interdependence and information dynamics. It consists of the following two main thrusts.

The first thrust is targeted at understanding the impact of network interdependence. It is shown that a cyber-physical system built upon multiple interdependent networks are more vulnerable to attacks since node failures in one network may result in failures in the other network, causing a cascade of failures that would potentially lead to the collapse of the entire infrastructure. There is thus a need to develop a new network science for modeling and quantifying cascading failures in multiple interdependent networks, and to develop network management algorithms that improve network robustness and ensure overall network reliability against cascading failures. To enhance the system robustness, a “regular” allocation strategy is proposed that yields better resistance against cascading failures compared to all possible existing strategies. Furthermore, in view of the load redistribution feature in many physical infrastructure networks, e.g., power grids, a CPS model is developed where the threshold model and the giant connected component model are used to capture the node failures in the physical infrastructure network and the cyber network, respectively.

The second thrust is centered around the information dynamics in the CPS. One speculation is that the interconnections over multiple networks can facilitate information diffusion since information propagation in one network
can trigger further spread in the other network. With this insight, a theoretical framework is developed to analyze information epidemic across multiple interconnecting networks. It is shown that the conjoining among networks can dramatically speed up message diffusion. Along a different avenue, many cyber-physical systems rely on wireless networks which offer platforms for information exchanges. To optimize the QoS of wireless networks, there is a need to develop a high-throughput and low-complexity scheduling algorithm to control link dynamics. To that end, distributed link scheduling algorithms are explored for multi-hop MIMO networks and two CSMA algorithms under the continuous-time model and the discrete-time model are devised, respectively.
ACKNOWLEDGEMENTS

Foremost, I owe my most sincere gratitude to my advisor, Professor Junshan Zhang, for his close guidance and support throughout my PhD study. His motivation, immense knowledge, and enthusiastic attitude towards research quality have always inspired me. It is unimaginable how much time and effort he has spent in discussion and in helping my writing. I greatly appreciate his invaluable input in my research and his warm-hearted help in my personal life. I am very fortunate to have him as my advisor without whom I could not have my current achievement.

I also would like to express my sincere gratitude to Professor Douglas Cochran, with whom I had the opportunity to work on a part of my research. I am very grateful to him for all support and time he gave to me. Thanks also go to Professor Lei Ying and Professor Yanchao Zhang for serving in my doctoral committee and for their valuable suggestions on my thesis.

I also thank my collaborator Osman Yagan for his helpful input in my research work. Several technical discussions I had with him were really valuable and helped me building a strong theoretical background. Many thanks to my current and former colleagues: Dong Zheng, Qinghai Gao, Weiyen Ge, Chandra Theajaswi, Shanshan Wang, Xu Chen, Donghoon Shin, Chuan Huang, Shibo He, Yang Cao, Miao He, Lei Yang, Xiaowen Gong, Brian Proulx, David Ganger, Mojgan Hedayati, Mengyuan Zhang and Zhengyu Zhang for the pleasant and inspiring discussions. Their friendship is a valuable experience for me.

Last but not least, I want to thank my mother Jie Huang and my father Long Qian for their endless love. This thesis is dedicated to them!
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>List of Tables</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF TABLES</td>
<td>xii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xiii</td>
</tr>
</tbody>
</table>

## CHAPTER

1 INTRODUCTION ................................................. 1

1.1 Scope of the Report ......................... 1

1.1.1 Network Interdependence in Cyber-Physical Systems 2

1.1.2 Information Dynamics in Cyber-Physical Systems 3

1.2 Summary of Main Contributions ................. 6

2 OPTIMUM ALLOCATION OF INTERCONNECTING LINKS IN
CYBER-PHYSICAL SYSTEMS ................................. 10

2.1 Introduction ....................................... 10

2.1.1 Background and Related Work ............... 11

2.1.2 Summary of Main Contributions ............. 13

2.2 System Model ........................................ 15

2.3 Analysis of Cascading Failures under Regular Allocation of Inter-edges ................. 18

2.4 Optimality of Regular Allocation Strategy .......... 24

2.4.1 Analysis of Random Allocation Strategy ...... 25

2.4.2 Regular Allocation vs. Random Allocation ...... 28

2.4.3 Bi-directional Inter-Edges vs. Unidirectional Inter-Edges 31

2.5 Numerical Results: The Erdős-Rényi Networks Case .......... 34

2.5.1 Numerical Results for System 1 ............... 35

2.5.2 Numerical Results for System 2 ............... 38

2.5.3 A Comparison of System Robustness ............. 41

2.6 Conclusions ......................................... 45
CHAPTER 3  THE IMPACT OF NETWORK INTERDEPENDENCE ON RESILIENCE OF CYBER-PHYSICAL SYSTEMS: WHEN THRESHOLD MODEL MEETS GCC MODEL ................................................. 47
  3.1 Introduction ........................................................................ 47
      3.1.1 Motivation ........................................................... 47
      3.1.2 Summary of Main Results ...................................... 48
  3.2 System Setup ..................................................................... 50
      3.2.1 Model of the Cyber-Physical System ...................... 50
      3.2.2 GCC Model vs. Threshold Model ......................... 51
      3.2.3 Cascading Failures due to Network Interdependence . 54
  3.3 Analysis of Cascading Failures ......................................... 54
      3.3.1 Recursive Dynamics in Cascading Failures ............ 54
      3.3.2 Analytical Results for Cascading Failures .............. 58
  3.4 Numerical Results .......................................................... 63
      3.4.1 System 1 vs. System 2 .......................................... 63
      3.4.2 System 1 vs. System 3 .......................................... 64
  3.5 Enhancing CPS Resilience with Autonomous Nodes .......... 67
      3.5.1 Analysis of CPS Model with Autonomous Nodes ...... 67
      3.5.2 CPS Resilience with Autonomous Nodes ............... 69
  3.6 Conclusions ..................................................................... 71

CHAPTER 4  INFORMATION DIFFUSION IN OVERLAYING SOCIAL-PHYSICAL
            NETWORKS .................................................................. 73
  4.1 Introduction ..................................................................... 73
      4.1.1 Motivation and Background .................................. 73
      4.1.2 Summary of Main Contributions ............................ 74
<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.2</td>
<td>76</td>
</tr>
<tr>
<td>4.3</td>
<td>78</td>
</tr>
<tr>
<td>4.3.1</td>
<td>78</td>
</tr>
<tr>
<td>4.3.2</td>
<td>82</td>
</tr>
<tr>
<td>4.3.3</td>
<td>85</td>
</tr>
<tr>
<td>4.4</td>
<td>87</td>
</tr>
<tr>
<td>4.5</td>
<td>88</td>
</tr>
<tr>
<td>4.5.1</td>
<td>88</td>
</tr>
<tr>
<td>4.5.2</td>
<td>94</td>
</tr>
<tr>
<td>5</td>
<td>102</td>
</tr>
<tr>
<td>5.1</td>
<td>102</td>
</tr>
<tr>
<td>5.1.1</td>
<td>102</td>
</tr>
<tr>
<td>5.1.2</td>
<td>103</td>
</tr>
<tr>
<td>5.2</td>
<td>105</td>
</tr>
<tr>
<td>5.2.1</td>
<td>105</td>
</tr>
<tr>
<td>5.2.2</td>
<td>108</td>
</tr>
<tr>
<td>5.2.3</td>
<td>109</td>
</tr>
<tr>
<td>5.3</td>
<td>110</td>
</tr>
<tr>
<td>5.4</td>
<td>112</td>
</tr>
<tr>
<td>5.5</td>
<td>116</td>
</tr>
<tr>
<td>5.6</td>
<td>119</td>
</tr>
<tr>
<td>5.6.1</td>
<td>119</td>
</tr>
<tr>
<td>5 DIFFUSION OF REAL-TIME INFORMATION IN SOCIAL-PHYSICAL NETWORKS</td>
<td>102</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>5.6.2 Information Epidemics and Network Coupling</td>
<td>120</td>
</tr>
<tr>
<td>5.7 Conclusions</td>
<td>125</td>
</tr>
<tr>
<td>5.8 Appendix</td>
<td>126</td>
</tr>
<tr>
<td>5.8.1 Proof of Lemma 5.4.1</td>
<td>126</td>
</tr>
<tr>
<td>5.8.2 Proof of Lemma 5.4.2</td>
<td>128</td>
</tr>
<tr>
<td>6 CSMA-BASED DISTRIBUTED SCHEDULING IN MULTI-HOP MIMO NETWORKS</td>
<td>130</td>
</tr>
<tr>
<td>6.1 Introduction</td>
<td>130</td>
</tr>
<tr>
<td>6.2 System Setup and Related Work</td>
<td>134</td>
</tr>
<tr>
<td>6.2.1 Feasible States and Capacity Region in a MIMO Network</td>
<td>135</td>
</tr>
<tr>
<td>6.2.2 SINR Model versus Protocol Model</td>
<td>136</td>
</tr>
<tr>
<td>6.2.3 Review: CSMA Scheduling under Protocol Model</td>
<td>138</td>
</tr>
<tr>
<td>6.3 MIMO-pipe Modeling: Rates, SINR, and Interference Tolerance</td>
<td>140</td>
</tr>
<tr>
<td>6.3.1 MIMO Configurations and SINR Requirements</td>
<td>141</td>
</tr>
<tr>
<td>6.3.2 Interference Tolerance</td>
<td>142</td>
</tr>
<tr>
<td>6.4 CSMA Algorithm for MIMO-pipe Scheduling: A Continuous-time Model</td>
<td>143</td>
</tr>
<tr>
<td>6.4.1 SINR-aware Channel Probing: A Dual Band Approach</td>
<td>143</td>
</tr>
<tr>
<td>6.4.2 CSMA Algorithm for MIMO-pipe Scheduling</td>
<td>145</td>
</tr>
<tr>
<td>6.5 CSMA Algorithm for MIMO-pipe Scheduling: A Discrete-Time Model</td>
<td>151</td>
</tr>
<tr>
<td>6.5.1 CSMA Algorithm for Conservative MIMO-pipe Scheduling</td>
<td>151</td>
</tr>
<tr>
<td>6.5.2 Efficiency Ratio of Conservative MIMO-pipe Scheduling</td>
<td>158</td>
</tr>
<tr>
<td>6.6 Numerical Examples</td>
<td>161</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>Page</td>
</tr>
<tr>
<td>---------</td>
<td>------</td>
</tr>
<tr>
<td>6.6.1</td>
<td>Simulation Settings</td>
</tr>
<tr>
<td>6.6.2</td>
<td>Continuous-time Network Model</td>
</tr>
<tr>
<td>6.6.3</td>
<td>Discrete-time Network Model</td>
</tr>
<tr>
<td>6.7</td>
<td>Conclusions</td>
</tr>
<tr>
<td>6.8</td>
<td>Appendix</td>
</tr>
<tr>
<td>6.8.1</td>
<td>Proof of Lemma 6.4.4</td>
</tr>
<tr>
<td>6.8.2</td>
<td>Proof of Lemma 6.5.5</td>
</tr>
<tr>
<td>6.8.3</td>
<td>Proof of Theorem 6.5.5</td>
</tr>
<tr>
<td>6.8.4</td>
<td>Proof of Theorem 6.5.6</td>
</tr>
<tr>
<td>7</td>
<td>DYNAMICS OF COMPLEX NETWORKS: AN ENTROPIC AND MARKOVIAN VIEW</td>
</tr>
<tr>
<td>7.1</td>
<td>Introduction</td>
</tr>
<tr>
<td>7.2</td>
<td>Network Entropy Maximization: Problem formulation</td>
</tr>
<tr>
<td>7.3</td>
<td>Steady State Characteristics: NEM Degree Distribution</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Diminishing Marginal Utility and NEM Degree Distribution</td>
</tr>
<tr>
<td>7.3.2</td>
<td>From Preferential Attachment to NEM</td>
</tr>
<tr>
<td>7.4</td>
<td>Two Timescale Markov Model for Network Dynamics</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Two Timescale Markov Model: From Microscopic to Macroscopic</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Basic Markov Model at Smaller Timescale</td>
</tr>
<tr>
<td>7.4.3</td>
<td>Two Timescale Markov Model and Transient Behaviors</td>
</tr>
<tr>
<td>7.4.4</td>
<td>From Two Timescale Model to the NEM Distributions</td>
</tr>
<tr>
<td>7.5</td>
<td>Solution to the Constrained NEM Problem</td>
</tr>
<tr>
<td>7.6</td>
<td>Conclusions</td>
</tr>
<tr>
<td>CHAPTER</td>
<td>Page</td>
</tr>
<tr>
<td>-------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>8 CONCLUSIONS AND FUTURE WORK</td>
<td>193</td>
</tr>
<tr>
<td>8.1 Conclusions</td>
<td>193</td>
</tr>
<tr>
<td>8.2 Future Work</td>
<td>196</td>
</tr>
<tr>
<td>8.2.1 Network Interdependence in Cyber-Physical Systems</td>
<td>196</td>
</tr>
<tr>
<td>8.2.2 Information Dynamics in Cyber-Physical Systems</td>
<td>196</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>199</td>
</tr>
</tbody>
</table>
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Key notation in the analysis of cascading failures</td>
<td>18</td>
</tr>
<tr>
<td>5.1 The clique size distribution in four scenarios</td>
<td>120</td>
</tr>
<tr>
<td>6.1 Parameters in CSMA-based algorithm (at MIMO link $l$)</td>
<td>140</td>
</tr>
<tr>
<td>6.2 Feasible state in the example MIMO network</td>
<td>148</td>
</tr>
<tr>
<td>6.3 Tolerable set and intolerable set in the example network</td>
<td>154</td>
</tr>
<tr>
<td>7.1 Key notation in the NEM problem</td>
<td>176</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>A sketch of the regular allocation strategy of inter-edges.</td>
</tr>
<tr>
<td>2.2</td>
<td>An illustration of cascading failures in two interdependent networks.</td>
</tr>
<tr>
<td>2.3</td>
<td>$a = b = 3$, $k = 2$. $x$-axis represents $f_A$ and $y$-axis represents $f_B$.</td>
</tr>
<tr>
<td>2.4</td>
<td>The critical $p_c$ value vs. $k$ for the regular allocation strategy (System 1).</td>
</tr>
<tr>
<td>2.5</td>
<td>Experimental results for the regular allocation strategy (System 1) with $N = 5000$ nodes.</td>
</tr>
<tr>
<td>2.6</td>
<td>The critical $p_c$ value vs. $k$ for the random allocation strategy (System 2).</td>
</tr>
<tr>
<td>2.7</td>
<td>Experimental results for System 2 with $N = 5000$ nodes.</td>
</tr>
<tr>
<td>2.8</td>
<td>$p_c$ v.s. $k$ for different values of $a = b$ in System 1 and System 3.</td>
</tr>
<tr>
<td>2.9</td>
<td>$p_c$ v.s. $a = b$ for various $k$ values in System 1 and System 3.</td>
</tr>
<tr>
<td>2.10</td>
<td>$p_c$ v.s. $k$ for different values of $a = b$ in System 1 and System 2.</td>
</tr>
<tr>
<td>2.11</td>
<td>$p_c$ v.s. $a = b$ for various $k$ values in System 1, System 2 and System 3.</td>
</tr>
<tr>
<td>3.1</td>
<td>System model of the cyber-physical networks.</td>
</tr>
<tr>
<td>3.2</td>
<td>An example of cascading failures in the GCC model.</td>
</tr>
<tr>
<td>3.3</td>
<td>An example of cascading failures in the threshold model.</td>
</tr>
<tr>
<td>3.4</td>
<td>Graphical solutions to the system of equations (3.23) and (3.24), where $z_A = 1.5$, $z_B = 3$, $\theta = 0.21$ and $p = 0.999$.</td>
</tr>
<tr>
<td>3.5</td>
<td>Graphical solutions to the system of equations (3.23) and (3.24), where $z_A = 1.5$, $z_B = 3$, $\theta = 0.37$ and $p = 0.999$.</td>
</tr>
<tr>
<td>3.6</td>
<td>$p_{A\infty}$ vs. $z_A$ for System 1 and System 2.</td>
</tr>
<tr>
<td>3.7</td>
<td>$p_{A\infty}$ vs. $z_A$ for System 1 and System 3.</td>
</tr>
<tr>
<td>3.8</td>
<td>An overview of the robustness of System 1.</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>3.9 $p_{A^\infty}$ vs. $z_A$ for System 1 and System 4.</td>
<td>70</td>
</tr>
<tr>
<td>3.10 $p_{A^\infty}$ vs. $p$ for System 1 and System 4.</td>
<td>71</td>
</tr>
<tr>
<td>4.1 The minimum $\lambda_f T_f$ required for existence of a giant component (ER graphs).</td>
<td>88</td>
</tr>
<tr>
<td>4.2 The fractional size of the giant component (ER graphs).</td>
<td>89</td>
</tr>
<tr>
<td>4.3 The minimum $T_f$ required for the existence of a giant component (power-law distributions).</td>
<td>90</td>
</tr>
<tr>
<td>4.4 The fractional size of the giant component (power-law distributions).</td>
<td>91</td>
</tr>
<tr>
<td>5.1 An illustration of the coupled network models.</td>
<td>104</td>
</tr>
<tr>
<td>5.2 Structure of coupled-network model I</td>
<td>106</td>
</tr>
<tr>
<td>5.3 Equivalent graph $\mathbb{E}$.</td>
<td>111</td>
</tr>
<tr>
<td>5.4 The influenced nodes fraction in $\mathcal{N}_W$ versus $T_f$.</td>
<td>121</td>
</tr>
<tr>
<td>5.5 The influenced nodes fraction in $\mathcal{N}<em>W \cup \mathcal{N}</em>{F_2}$ versus $\alpha$.</td>
<td>123</td>
</tr>
<tr>
<td>5.6 The number of the influenced nodes in $\mathcal{N}<em>W \cup \mathcal{N}</em>{F_2}$ versus the size of social network with $T_w = 0.1$.</td>
<td>124</td>
</tr>
<tr>
<td>5.7 The number of the influenced nodes in $\mathcal{N}<em>W \cup \mathcal{N}</em>{F_2}$ versus the size of social network with $T_w = 0.3$.</td>
<td>124</td>
</tr>
<tr>
<td>5.8 The number of the influenced nodes in $\mathcal{N}<em>W \cup \mathcal{N}</em>{F_2}$ versus the size of social network with $T_w = 0.8$.</td>
<td>125</td>
</tr>
<tr>
<td>6.1 Rate-reliability tradeoff for a MIMO link with $4 \times 4$ antennas.</td>
<td>143</td>
</tr>
<tr>
<td>6.2 An example network with two $4 \times 4$ MIMO links.</td>
<td>147</td>
</tr>
<tr>
<td>6.3 State transition graph for the continuous-time Markov chain associated with the network in Fig. 6.2.</td>
<td>148</td>
</tr>
<tr>
<td>6.4 An example network with 5 links.</td>
<td>153</td>
</tr>
<tr>
<td>6.5 The conflict graph for example network in Fig. 6.4.</td>
<td>153</td>
</tr>
<tr>
<td>Figure</td>
<td>Page</td>
</tr>
<tr>
<td>--------</td>
<td>------</td>
</tr>
<tr>
<td>6.6 Continuous-time model: queueing length behavior at each MIMO link with $\rho = 0.98$.</td>
<td>163</td>
</tr>
<tr>
<td>6.7 Continuous-time model: total queue lengths of 6 MIMO links with different $\rho$ values.</td>
<td>164</td>
</tr>
<tr>
<td>6.8 Discrete-time model: total queue lengths of 6 MIMO links with different $\rho$ values.</td>
<td>165</td>
</tr>
<tr>
<td>6.9 Comparisons of total queue lengths for continuous-time model and discrete-time model.</td>
<td>165</td>
</tr>
<tr>
<td>7.1 Network evolution process under two timescale Markov model.</td>
<td>180</td>
</tr>
<tr>
<td>7.2 A birth-death Markov chain model for degree connectivity of a typical node.</td>
<td>182</td>
</tr>
<tr>
<td>7.3 The average degree obeys densification for the case of a power-law degree distribution.</td>
<td>187</td>
</tr>
<tr>
<td>7.4 Numerical simulations for a network model with a power-law degree distribution, with $n = 10000$.</td>
<td>188</td>
</tr>
<tr>
<td>7.5 Numerical simulations for a network model with a exponential degree distribution, with $n = 10000$</td>
<td>189</td>
</tr>
<tr>
<td>7.6 The average degree stays constant for the case with a exponential degree distribution.</td>
<td>189</td>
</tr>
<tr>
<td>8.1 Coupled networks with hierarchical structure.</td>
<td>197</td>
</tr>
</tbody>
</table>
Chapter 1
INTRODUCTION

1.1 Scope of the Report

Today’s worldwide network infrastructure consists a web of interacting cyber-networks (e.g., the Internet) and physical systems (e.g., power grids, water supply systems, and transportation systems). There is a consensus that integrated Cyber-Physical Systems (CPS) will emerge as the underpinning technology for major industries in the 21st century [1]. A cyber-physical system is based on a tight combination and coordination between physical elements and the cyber network, where the computing, communication and control technologies are inherently integrated. The CPS makes use of the cutting edge communication and computation technologies to enhance the control and the management of physical infrastructure networks. It is envisioned that cyber-physical systems will transform how we interact with the physical world around us, very much in the same way the Internet has transformed how human beings interact with one another. The design and optimization of cyber-physical system intimately relates to a wide range of research fields, including Internet, wireless networks, smart grids, social networks and other complex systems.

In this dissertation, we focus on two fundamental issues in cyber-physical systems: the impact of network interdependence and network information dynamics. Typically, we do not limit our study in cyber-physical systems. Our focus spans related scenarios in wireless networks, smart grids, social networks and complex networks.
1.1.1 Network Interdependence in Cyber-Physical Systems

One primary goal of this work is to understand the impact of network interdependence on the robustness of the overall cyber-physical system. One salient feature of the CPS is that it makes use of the cutting edge communication and computation technologies to enhance the control and the management of the physical infrastructure network (e.g., power grids, water supply systems, and transportation systems). Typically, many cyber-physical systems are built upon the interdependence between the cyber network and the physical network [1], whose operation requires the coordinations and interactions between the physical elements and the cyber network. One archetypal example is the smart grid, where the operation of the power grid relies on the real-time control from the communication network, which also requires power supply from the power grid.

The interdependence structure makes the cyber-physical system more intelligent and yet complex. However, it is also because of the interdependence that the cyber-physical system could be more vulnerable to random failures, natural hazards and malicious attacks [2, 3]. Simply put, in the event of external attacks, the failures in one network can cause the failures in the other network and vice versa. As a result, even a small fraction of failed nodes could trigger a recursive process of cascading failures between the cyber network and the physical network and finally lead to catastrophic damages on the whole system. Despite recent studies of cascading failures in complex networks, the dynamics of such failures and the impact across multiple networks are not well understood. There is thus a need to develop a new network science for modeling and quantifying cascading failures, and to develop network manage-
ment algorithms that improve network robustness and ensure overall network reliability against cascading failures. Specifically, it is of critical importance to obtain a fundamental understanding of the following two issues:

- How to design a model that accurately captures some key properties of practical cyber-physical systems and characterizes the recursive process of the cascading failures therein?
- How to optimize the design of interdependent systems for better robustness against cascading failures?

Our aim is to develop systematic frameworks for solving these issues.

1.1.2 Information Dynamics in Cyber-Physical Systems

Another key objective of this study is to understand the behaviors of information dynamics in cyber-physical systems. Under this theme, our work is composed of the following three main thrusts.

The first thrust is targeted at exploring the information diffusion process across multiple coupled networks. Apart from cyber physical systems, network coupling can also be observed between different types of social networks. Traditionally, people are tied together in a physical information network through old-fashioned communication media, such as face-to-face interactions. On the other hand, recent advances of Internet and mobile communication technologies have enabled people to be connected more closely through online social networks. Indeed, people can now interact through e-mail or online chatting, or communicate through web sites such as Facebook, Twitter, YouTube, etc. Clearly, the physical information network and online social networks are not completely separate since people may participate in two or more of these
networks at the same time. It is in this sense these networks are “coupled” together. For instance, a person can forward a message to her online friends via Facebook and Twitter upon receiving it from someone via face-to-face communication. One speculation is that the interconnections over multiple networks can facilitate information diffusion since information propagation in one network can trigger further spread in the other network, and may result in a possible cascade of information. Despite the important impact of network coupling on information diffusion, there been little study on information diffusion across such networks. With this insight, We aim to develop a theoretical framework of analyzing information epidemic across multiple coupled networks.

The second thrust is targeted at developing a high-throughput and low-complexity distributed scheduling algorithm in wireless networks. Nowadays, many cyber-physical systems rely on wireless networks which provide platforms of information exchange. Clearly, the QoS offered by wireless network largely impacts the performance of cyber-physical system.

One main factor that degrades the QoS of wireless communication is co-channel interference, i.e., the shared nature of wireless medium may result in transmission failure due to interference from other transmissions in the vicinity of the receiver. To resolve co-channel interference, scheduling is required to determine at each time instance which links are active for transmissions. The scheduling algorithms control the link activation dynamics and determine the whole network throughput. Different scheduling algorithms may have different ways of realizing non-conflicting link transmissions, with different complexities.

Recently, low-complexity scheduling schemes based on carrier sense multiple access (CSMA) have been proposed (see [4–8] and the references
therein). In these CSMA algorithms, nodes first sense the channel activity, and only when the channel is sensed to be idle can the nodes continue with data transmissions. When the channel is detected busy, the nodes need to backoff for a random amount of time before reattempting the transmission. Due to its simplicity, CSMA and its variants have been widely opted in practical MAC protocols (e.g., IEEE 802.11). In this study, a primary goal is to develop a high-throughput and low-complexity scheduling scheme by utilizing the idea of CSMA scheduling.

The third thrust is dedicated to understanding the dynamics behaviors in large-scale cyber networks. Existing studies have been shown that many cyber networks (e.g., Internet) can evolve into “steady states” where the connectivity degree follows some “typical” distribution, such as the power-law degree distribution in many large social networks (see, e.g., [9]), the exponential distribution in email networks [10,11], and the Weibull distribution in some IP graphs [12,13]. Along a different avenue, recent works [14,15] have explored the trend of network evolution based on the observations over time, and their findings show that some growing cyber networks can exhibit surprising transient phenomena such as “network densification” and “shrinking diameter.”

Most of the existing works focus on mechanisms that can yield some given properties, and there has been little work on developing unifying models that naturally “catch” both steady state and dynamic characteristics. It remains intriguing why the observed properties are so prevailing in the real world and under what conditions they would happen. With this motivation, we aim to develop a mathematically rigorous model towards understanding the following two fundamental issues:
• Why would cyber networks always evolve towards these degree distributions, such as power-law, exponential, and Weibull? Further, under what condition would the network evolve to a specific degree distribution?

• Is there an unifying model that naturally approximates network evolution and captures both steady state and transient characteristics?

1.2 Summary of Main Contributions

In the following, we present a brief summary of main contributions in this dissertation.

In Chapter 2, we study cascading failures in a cyber-physical system (CPS) where a cyber network overlays a physical network and the operation of one network depends heavily on the functioning of the other network. Specifically, we optimize the design of cyber-physical systems for better system robustness against cascading failures. The robustness of interdependent systems hinges heavily on the allocation of the (interconnecting) links that connect nodes in one network to nodes in the other network. By using the Giant Connected Component (GCC) model to capture the cascading failures in each single network, we characterize the optimum inter-link allocation strategy against random attacks in the case where the topology of each individual network is unknown. In particular, we analyze the “regular” allocation strategy that allots exactly the same number of bi-directional inter-network links to all nodes in the system. We show, both analytically and experimentally, that this strategy yields better performance (from a network resilience perspective) compared to all possible strategies, including strategies using random allocation, unidirectional inter-links, etc.
In Chapter 3, we study cascading failures in a cyber-physical system (CPS) from a different perspective. It is known that the threshold model is more applicable to capture node failures in the physical infrastructure network (e.g., power grids), and that the GCC model is more applicable to functioning nodes in the cyber network (e.g., the Internet). Assuming this for the cyber-physical systems under consideration, we quantify the impact of network interdependence on the robustness of the overall CPS by characterizing the functioning node fraction after the cascading failures. We show that this CPS model can naturally capture some key features of practical cyber-physical systems which are not accounted for in the existing studies using GCC models for both physical and cyber networks. Further, our results reveal that the dense connectivity in the physical network would likely make the cyber-physical system more vulnerable to cascading failures. This is somewhat surprising and indeed is in a stark contrast to some previous observations on the resilience in a single physical network. Finally, we develop a strategy to improve the system robustness by enabling a fraction of nodes to be autonomous in the sense that the nodes can support by themselves.

In Chapter 4, we study the diffusion of information in an overlaying social-physical network. Specifically, we consider the following set-up: There is a physical information network where information spreads amongst people through conventional communication media (e.g., face-to-face communication, phone calls), and conjoint to this physical network, there are online social networks which offer alternative platforms for information diffusion, such as Facebook, Twitter, FriendFeed, YouTube, etc. Capitalizing on the theory of inhomogeneous random graphs, we quantify the size and the critical threshold of information cascades in this conjoint social-physical network by assuming
that information diffuses according to the SIR epidemic model. One interesting finding is that even if there is no percolation in the individual networks, percolation (i.e., information epidemics) can take place in the conjoint social-physical network. We also show, both analytically and experimentally, that the fraction of individuals who receive the information (started from an arbitrary node) is significantly larger in the conjoint social-physical network case, as compared to the case where the networks are disjoint. These findings reveal that conjoining the physical network with online social networks can have dramatic impact on the speed and the scale of information diffusion.

In Chapter 5, we further characterize the diffusion of real-time information in a social network with clique structure. Capitalizing on the theory of inhomogeneous random networks, we show that the social network has a critical threshold above which information epidemics are very likely to happen. We also theoretically quantify the fractional size of individuals that finally receive the message.

In Chapter 6, we explore CSMA-based scheduling algorithms for MIMO networks in the follow three steps. We first take a bottom-up approach to develop the MIMO-pipe model, which consists of multiple stream configurations, each with a feasible rate and the corresponding SINR requirement. Using this model, the tradeoff between diversity and multiplexing of MIMO communications can be captured by the selection of MIMO configurations.

We next consider the CSMA algorithms for MIMO-pipe scheduling in a continuous-time network. Assuming that there is no collision of control signals, we show that the network dynamics can be captured by a continuous-time Markov chain. Further, we characterize the optimal backoff parameters of different stream configurations, for throughput-optimal scheduling.
Finally, we focus on the CSMA algorithms for MIMO-pipe scheduling in a discrete-time network, where control signals may “collide.” To tackle the collisions and the link coupling problem under the SINR model, we devise a distributed scheduling algorithm using a “conservative” strategy. We show that this conservative distributed scheduling can achieve an efficiency ratio bounded below.

In Chapter 7, we study the steady state and transient behaviors of complex networks from an entropic and Markovian view. Specifically, we first take a network entropy maximization (NEM) view to examine network steady state characteristics, in terms of degree distributions, and explore the underlying rationale connecting network entropy and widely observed phenomena, such as power law degree distributions, exponential degree distributions and Weibull degree distributions. Next, to capture the microscopic behaviors of network dynamics, we develop a two timescale Markov model where link generation and deletion takes place on a smaller timescale and new node arrivals (i.e., the network size grows) occur on a larger timescale. This two timescale model provides a natural platform to study both microscopic and macroscopic behaviors of network dynamics. Indeed, the corresponding graph dynamics offers a general framework towards understanding many transient network characteristics, such as network densification, which remain not well understood.

We summarize in Chapter 8 our proposed work along with a discussion on future research directions.
Chapter 2

OPTIMUM ALLOCATION OF INTERCONNECTING LINKS IN CYBER-PHYSICAL SYSTEMS

2.1 Introduction

Today’s worldwide network infrastructure consists a web of interacting cyber-networks (e.g., the Internet) and physical systems (e.g., the power grid). There is a consensus that integrated cyber-physical systems will emerge as the underpinning technology for major industries in the 21st century [1]. The smart grid is one archetypal example of such systems where the power grid network and the communication network that controls it are coupled together and depend on each other; i.e., they are interdependent. While interdependency allows building systems that are larger, smarter and more complex, it has been observed [3] that interdependent systems tend to be more fragile against failures, natural hazards and attacks. For example, in the event of an attack to an interdependent system, the failures in one of the networks can cause failures of the dependent nodes in the other network and vice versa. This process may continue in a recursive manner and hence lead to a cascade of failures causing a catastrophic impact on the overall cyber-physical system. In fact, the cascading effect of even a partial Internet blackout could disrupt major national infrastructure networks involving Internet services, power grids and financial markets [2]. Real-world examples include the 2003 blackout in the northeastern United States and southeastern Canada [3] and the electrical blackout that affected much of Italy on 28 September 2003 [2].
2.1.1 Background and Related Work

Despite recent studies of cascading failures in complex networks, the dynamics of such failures and the impact across multiple networks are not well understood. There is thus a need to develop a new network science for modeling and quantifying cascading failures, and to develop network management algorithms that improve network robustness and ensure overall network reliability against cascading failures. Most existing studies on failures in complex networks consider single networks only. A notable exception is the very recent work of Buldyrev et al. [2] in which a “one-to-one correspondence” model for studying the ramifications of interdependence between two networks is set forth. This model considers two networks of the same size, say network $A$ and network $B$, where each node in network $A$ depends on one and only one node in network $B$ and vice versa. In other words, each node in network $A$ has one bi-directional inter-edge connecting it to a unique node in network $B$. Furthermore, it is assumed that a node in either network can function only if it has support from the other network; i.e., it is connected (via an inter-edge) to at least one functioning node from the other network.

The robustness of the one-to-one correspondence model was studied in [2] using a similar approach to that of the works considering single networks [16, 17]. Specifically, it is assumed that a random attack is launched upon network $A$, causing the failure of a fraction $1 - p$ of the nodes; this was modeled by a random removal of a fraction $1 - p$ of the nodes from network $A$. Due to the interdependency, these initial failures lead to node failures from network $B$, which in turn may cause further failures from network $A$ thereby triggering an avalanche of cascading failures. To evaluate the robustness of the model,
the size of the functioning parts of both networks are computed at each stage of the cascading failure until a *steady-state* is reached; i.e., until the cascade of failure ends. One of the important findings of [2] was to show the existence of a critical threshold on $p$, denoted by $p_c$, above which a considerable fraction of nodes in both networks remain functional at the steady-state; on the other hand, if $p < p_c$, both networks go into a complete fragmentation and the entire system collapses. Also, it is observed in [2] that interdependent network systems have a much larger $p_c$ compared to that of the individual constituent networks; this is compatible with the observation that interdependent networks are more vulnerable to failures and attacks.

The original work of Buldyrev et al. [2] has received much attention and spurred the study of interdependent networks in many different directions; e.g., see [18–23]. One major vein of work, including [18,21,23], aims to extend the findings of [2] to more realistic scenarios than the one-to-one correspondence model. More specifically, in [18] the authors consider a one-to-one correspondence model with the difference that mutually dependent nodes are now assumed to have the same number of neighbors in their own networks; i.e., their intra-degrees are assumed to be the same. In [21] the authors consider the case where only a fraction of the nodes in network $A$ depend on the nodes in network $B$, and vice versa. In other words, some nodes in one network are assumed to be *autonomous*, meaning that they do not depend on nodes of the other network to function properly. Nevertheless, in [21] it was still assumed that a node can have at most one supporting node from the other network. More recently, Shao et al. [23] pointed out the fact that, in a realistic scenario, a node in network $A$ may depend on more than one node in network $B$, and vice versa. In this case, a node will function as long as
at least one of its supporting nodes is still functional. To address this case, Shao et al. [23] proposed a model where the inter-edges are unidirectional and each node supports (and is supported by) a random number of nodes from the other network. In a different line of work, Schneider et al. [22] adopted a design point of view and explored ways to improve the robustness of the one-to-one correspondence model by letting some nodes be autonomous. More precisely, they assume that the topologies of networks $A$ and $B$ are known and propose a method, based on degree and centrality, for choosing the autonomous nodes properly in order to maximize the system robustness.

2.1.2 Summary of Main Contributions

In this chapter, we stand in the intersection of the two aforementioned lines of work. First, we consider a model where inter-edges are allocated regularly in the sense that all nodes have exactly the same number of bi-directional inter-edges. This ensures a uniform support-dependency relationship where each node supports (and is supported by) the same number of nodes from the other network. We analyze this new model in terms of its robustness against random attacks via characterizing the steady state size of the functioning parts of each network as well as the critical fraction $p_c$. In this regard, our work generalizes the studies on the one-to-one correspondence model and the model studied by Shao et al. [23]. From a design perspective, we show analytically that the proposed method of regular inter-edge allocation improves the robustness of the system over the random allocation strategy studied in [23]. Indeed, for a given expected value of inter-degree (the number of nodes supported plus the number of nodes depended upon) per node, we show: i) it is better (in terms of robustness) to use bi-directional inter-links than unidirectional links, since this ensures that for each node the amount of support being received
and the amount of support provided are equal, and ii) it is better (in terms of robustness) to deterministically allot each node exactly the same number of bi-directional inter-edges rather than allotting each node a random number of inter-edges.

These results imply that if the topologies of network $A$ and network $B$ are unknown, then the optimum inter-link allocation strategy is to allot exactly the same number of bi-directional inter-edges to all nodes. Even if the statistical information regarding the networks is available; e.g., say it is known that network $A$ is an Erdős-Rényi [24] network and network $B$ is a scale-free network [25], regular inter-edge allocation is still the best strategy in the absence of the detailed topological information; i.e., in the case where it is not possible to estimate the nodes that are likely to be more important in preserving the connectivity of the networks, say nodes with high betweenness [26]. Intuitively, this makes sense because without knowing which nodes play a key role in preserving the connectivity of the networks, it is best to treat all nodes identically and give them equal priority in inter-edge allocation.

The theoretical results are also supported by extensive computer simulations. Numerical results are given for the case where both networks are Erdős-Rényi (ER) and the optimality of the regular allocation strategy is verified. To get a more concrete sense, assume that $A$ and $B$ are ER networks with $N$ nodes and average degree 4. When inter-edges are allocated regularly so that each node has exactly 2 bi-directional inter-edges, the critical threshold $p_c$ is equal to 0.43. However, for the same networks $A$ and $B$, if the number of inter-edges follows a Poisson distribution with mean 2, the critical $p_c$ turns out to be equal to 0.82. This is a significant difference in terms of robustness, since in the former case the system is resilient to the random failure of up to
57% of the nodes while in the latter case, the system is resilient to the random failure of up to only 18% of the nodes.

2.2 System Model

We consider a cyber-physical system consisting of two interacting networks, say network $A$ and network $B$. For simplicity, both networks are assumed to have $N$ nodes and the vertex sets in their respective graphical representations are denoted by $\{v_1, \ldots, v_N\}$ and $\{v'_1, \ldots, v'_N\}$. We refer to the edges connecting nodes within the same network as *intra-edges* and those connecting nodes from two different networks as *inter-edges*. Simply put, we assume that a node can function *only if* it is connected (via an inter-edge) to at least one functioning node in the other network [2]; and we will elaborate further on this. Clearly, the interdependency between two networks is intimately related to the inter-edges connecting them. In this study, inter-edges are assumed to be bi-directional so that it is convenient to use an $N \times N$ interdependency matrix $C$ to represent the bi-directional inter-edges between networks $A$ and $B$. Specifically, for each $n, m = 1, \ldots, N$, let

$$
(C)_{nm} = \begin{cases} 
1 & \text{if } v_n \text{ and } v'_m \text{ depend on each other} \\
0 & \text{otherwise}
\end{cases} 
$$

(2.1)

We also assume that inter-edges are allocated *regularly* so that each node has exactly $k$ inter-edges, where $k$ is an integer satisfying $k \leq N$. Without loss of generality, this strategy can be implemented in the following manner: For each $n = 1, 2, \ldots, N$, let the interdependency matrix be given by

$$
(C)_{nm} = \begin{cases} 
1 & \text{if } m = n, n \oplus 1, \ldots, n \oplus (k-1) \\
0 & \text{otherwise},
\end{cases} 
$$

(2.2)
Figure 2.1: A sketch of the regular allocation strategy of inter-edges: each node in $A$ is connected to exactly $k$ nodes in $B$, and vice versa.

Figure 2.2: An illustration of cascading failures in two interdependent networks. Network $A$ with vertices $\{v_1, v_2, \ldots, v_6\}$ and network $B$ with vertices $\{v'_1, v'_2, \ldots, v'_6\}$ are interdependent where each node has exactly 2 bi-directional inter-edges. Initially, a random attack causes the failure of nodes $v_1$ and $v_2$. In stage 1, $v_1$ and $v_2$ are removed from the system along with all the links (inter and intra) that are incident upon them. As a result, node $v_3$ becomes disconnected from the functioning giant component of network $A$, and thus fails. These failures then cause the nodes $v'_2$ and $v'_3$ to fail as they lose all their supports; i.e., all the inter-edges that are incident upon them are removed. In stage 2, we see the effect of removing $v'_2$ and $v'_3$ from network $B$: nodes $v'_1$ and $v'_6$ fail as they become disconnected from the functioning giant component. The failure of nodes $v'_1$ and $v'_6$ then leads to the failure of node $v_6$ in stage 3 since $v_6$ was being supported solely by $v'_1$ and $v'_6$. By the removal of the node $v_6$, the failures stop and the system reaches a steady-state.
where we define \( n \oplus l = n + l - 1[n + l > N] \cdot N \); see also Figure 2.1.

We are interested in evaluating the network robustness in the case of random node failures (or equivalently random attacks). Specifically, in the dynamics of cascading failures, we assume that a node is functioning at Stage \( i \) if the following conditions are satisfied [2,23]: i) The node has at least one inter-edge with a node that was functioning at Stage \( i - 1 \); ii) The node belongs to the giant (i.e., the largest) component of the sub-network formed by the nodes (of its own network) that satisfy condition i). Therefore, this model is called the Giant Connected Component (GCC) model. For both networks, a giant component consisting of functioning nodes will be referred to as a functioning giant component.

We assume that the cascade of failures is triggered by the failure of a fraction \( 1 - p \) of the nodes in network \( A \). We further assume that these \((1-p)N\) nodes are chosen (say by the attacker) uniformly at random amongst all nodes in network \( A \). By the definitions given above, it can be seen that after the initial attack, only nodes in the functioning giant component of \( A \) can operate properly. As a result of that, in the next stage, some of the nodes in network \( B \) may end up losing all of their inter-connections and turn dysfunctional. In that case, the nodes that can function properly in network \( B \) will only be those in the functioning giant component of \( B \). But, this fragmentation of network \( B \) may now trigger further failures in network \( A \) due to nodes that lose all their \( B \)-connections. Continuing in this manner, the cascade of failures propagates alternately between \( A \) and \( B \), eventually (i.e., in steady state) leading to either: 1) residual functioning giant components in both networks, or 2) complete failure of the entire system. For an illustrative example, see Figure 2.2 where the cascading failures is demonstrated for a pair
Table 2.1: Key notation in the analysis of cascading failures

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A_i, B_i$</td>
<td>collection of functioning nodes in $A$ (or $B$) at Stage $i$</td>
</tr>
<tr>
<td>$\bar{A}_i, \bar{B}_i$</td>
<td>collection of nodes in $A$ (or $B$) with support from the other network at Stage $i$</td>
</tr>
<tr>
<td>$</td>
<td>\cdot</td>
</tr>
<tr>
<td>$p_{A_i}, p_{B_i}$</td>
<td>functioning node fraction at Stage $i$, $</td>
</tr>
<tr>
<td>$p'<em>{A_i}, p'</em>{B_i}$</td>
<td>&quot;equivalent&quot; remaining node fraction in network $A$ (or $B$) at Stage $i$</td>
</tr>
</tbody>
</table>

of interdependent networks with $N = 6$ nodes and $k = 2$, $p = 2/3$.

2.3 Analysis of Cascading Failures under Regular Allocation of Inter-edges

In this section, we analyze the dynamics of cascading failures in two interacting networks. A principal objective of this study is to quantify the effectiveness of the regular allocation strategy on network robustness, by means of: i) characterizing the size of the remaining giant components in networks $A$ and $B$ after the cascade has reached a steady state, and ii) finding the corresponding critical threshold $p_c$. To that end, we will use the technique of generating functions [27,28] to analyze the sizes of functioning giant components in the two networks at each stage. For convenience, the key notation used in the calculations is summarized in Table 2.1.

**Stage 1: random failure of nodes in network $A$.** Following the failures of a fraction $1 - p$ of randomly selected nodes in network $A$, the remaining network $\bar{A}_1$ has size $pN$; since we eventually let $N$ grow large, $pN$ can be approximated as an integer. As in [2,23,27,28], we use the technique of generating functions to quantify the fraction of the functioning giant component $A_1 \subset \bar{A}_1$. Specifically, let the function $P_A(p)$ determine the fraction of the giant component in a random subgraph that occupies a fraction $p$ of the
nodes in network A (the exact calculation of $P_A(p)$ will be elaborated later).

It follows that the functioning giant component has size

$$|A_1| = pP_A(p)N = p_{A1}N. \quad (2.3)$$

As shall become apparent soon, at the end of each stage it is necessary to determine not only the size of the functioning giant component, but also the specific inter-edge distribution over the functioning nodes; i.e., the numbers of functioning nodes having particular numbers of inter-edges. Indeed, this is what makes the analysis of the regular allocation model more complicated than the models considered in [2, 21, 23]. Here, at the end of Stage 1, each node in $A_1$ has still $k$ inter-edges from network $B$ since network $B$ has not changed yet.

**Stage 2: impact of random $A$-node failures on network $B$.** As the functioning part of network $A$ fragments from $A$ to $A_1$ (in Stage 1), some of the inter-edges that were supporting $B$-nodes will be removed. Observe that the probability of removal can be approximated by $1 - |A_1|/|A| = 1 - p_{A1}$ for each inter-edge. With this perspective, a $B$-node loses $k - j$ of its inter-edges with probability \( \binom{k}{j} p_{A1}^j (1-p_{A1})^{k-j} \). Moreover, it stops functioning with probability $(1 - p_{A1})^k$ due to losing all $k$ of its inter-edges. As a result, with $\bar{B}_2$ denoting the set of nodes in $B$ that retain at least one inter-edge, we have

$$|\bar{B}_2| = (1 - (1 - p_{A1})^k) N = p'_{B2}N, \quad (2.4)$$

where $p'_{B2} = 1 - (1 - p_{A1})^k$. Also, the distribution of inter-edges over the nodes in $\bar{B}_2$ is given by

$$|\bar{B}_2|_j = \binom{k}{j} p_{A1}^j (1-p_{A1})^{k-j}N, \quad j = 1, 2 \ldots, k, \quad (2.5)$$

with $|\bar{B}_2|_j$ denoting the number of nodes in $\bar{B}_2$ that have $j$ inter-edges.
As in Stage 1, the size of the functioning giant component \( B_2 \subset \bar{B}_2 \) can be predicted by:

\[
|B_2| = p'_{B_2} P_B(p'_{B_2}) N = p_{B_2} N, \tag{2.6}
\]

where \( P_B(\cdot) \) is defined analogously to the definition of \( P_A(\cdot) \) given in Section 2.3. Obviously, each node in \( \bar{B}_2 \) can survive as a functioning node in \( B_2 \) with probability \( P_B(p'_{B_2}) \). Thus, for each \( j = 1, 2, \ldots, k \), the number of nodes in \( B_2 \) that have \( j \) inter-edges is given (in view of (2.5)) by

\[
|B_2|_j = P_B(p'_{B_2}) \binom{k}{j} p_{A_1}^j (1 - p_{A_1})^{k-j} N. \tag{2.7}
\]

**Stage 3: further A-nodes failures due to B-node failures.** Due to the fragmentation of the functional part of network \( B \) from \( \bar{B}_2 \) to \( B_2 \) (not \( B \) to \( B_2 \)), some of the nodes in \( A_1 \) may now lose all their inter-edges and stop functioning. To compute the probability of this event, first observe that each inter-edge from \( \bar{B}_2 \) to \( A_1 \) will be removed with an approximate probability of \( 1 - |\bar{B}_2|/|\bar{B}_2| = 1 - P_B(p'_{B_2}) \). Hence the probability that a node in \( A_1 \) will lose all of its inter-edges is given by \( (1 - P_B(p'_{B_2}))^k \). It also follows that the size of the network \( \bar{A}_3 \subset A_1 \) comprised of the nodes that did not lose all their inter-connections is given via

\[
|\bar{A}_3| = p_{A_1} \left( 1 - (1 - P_B(p'_{B_2}))^k \right) N. \tag{2.8}
\]

In other words, from \( A_1 \) to \( \bar{A}_3 \), a fraction \( 1 - |\bar{A}_3|/|A_1| = (1 - P_B(p'_{B_2}))^k \) of the nodes have failed. As previously, the next step is to compute the size of the functioning giant component \( A_3 \subset \bar{A}_3 \). However, this a challenging task as noted in [2]. Instead, we view the joint effect of the node failures in Stage 1 and Stage 3 as equivalent (in terms of the size of the resulting functional giant component; i.e., \( |A_3| \)) to the effect of an initial random attack which
targets an appropriate fraction of the nodes that will later be determined.

Intuitively, the node failures in $A_1$ at Stage 3 (i.e., the removal of a fraction $(1 - P_B(p'_{B2}))^k$ of nodes from $A_1$) have the same effect as taking out the same portion from $\bar{A}_1$ [2]. In other words, it is equivalent to the removal of a fraction $p(1 - P_B(p_{B2}))^k$ of the nodes from $A$. Recalling also that a fraction $1 - p$ of the nodes in network $A$ failed as a result of the initial attack at Stage 1, we find that the fragmentation of $A$ to $\bar{A}_3$ can as well be modeled (with respect to the size of $A_3$) by an initial attack targeting a fraction

$$1 - p + p(1 - P_B(p_{B2}))^k = 1 - p\left(1 - (1 - P_B(p'_{B2}))^k\right)$$

of the nodes. It is now a standard step to conclude that, with

$$p_{A3}' = p\left(1 - (1 - P_B(p'_{B2}))^k\right),$$

the size of the functioning giant component $A_3$ is given by

$$|A_3| = p_{A3}'P_A(p_{A3})N = p_{A3}N. \quad (2.9)$$

**Stage 4: further fragmentation of network $B$.** Due to the network fragmentation from $\bar{A}_3$ to $A_3$ in Stage 3, each inter-edge supporting a $B_2$-node will be disconnected with probability that equals the proportion nodes in $\bar{A}_3$ that did not survive to $A_3$; i.e., $1 - |A_3|/|\bar{A}_3| = 1 - P_A(p_{A3}')/P_A(p)$ by (2.8) and (2.9). Consequently, a node in $B_2$ with $j$ inter-edges will stop functioning with probability $(1 - P_A(p_{A3}')/P_A(p))^j$. Recalling also the inter-edge distribution
the fraction $L$ of node failures in $B_2$ is given by
\[
L = \frac{1}{N} \sum_{j=1}^{k} |B_2| \left( 1 - \frac{P_A(p'_A)}{P_A(p)} \right)^j
\]
\[
= P_B(p'_B) \sum_{j=1}^{k} \binom{k}{j} p'_{A1} (1 - p_{A1})^{k-j} \left( 1 - \frac{P_A(p'_A)}{P_A(p)} \right)^j
\]
\[
= P_B(p'_B) \left( \left( 1 - p_{A1} \frac{P_A(p'_A)}{P_A(p)} \right)^k - (1 - p_{A1})^k \right)
\]
\[
= P_B(p'_B) \left( (1 - pP_A(p'_A))^k - (1 - p_{A1})^k \right).
\]
Since $|\bar{B}_4| = |B_2| - LN$, we now obtain
\[
|\bar{B}_4| = P_B(p'_B) \left( 1 - (1 - pP_A(p'_A))^k \right) N. \tag{2.10}
\]

In order to compute the size of the functioning giant component $B_4 \subset \bar{B}_4$, we proceed as in Stage 3. Specifically, we view the joint effect of node removals in Stage 2 and Stage 4 as equivalent to that of an initial random attack which targets an appropriate fraction of the nodes. To determine this fraction, first observe that the failures in Stage 3 have triggered further node failures in $B_2$ resulting a fraction
\[
1 - |\bar{B}_4|/|B_2| = 1 - \left( 1 - (1 - pP_A(p'_A))^k \right) / p'_B \tag{2.11}
\]
of the nodes’ failure. Next, note that the effect of these failures on $|B_4|$ is equivalent to that of taking out the same fraction of nodes from $\bar{B}_2$ [2]. Moreover, it has the same effect as taking out a fraction $p'_B \left\{ 1 - \left( 1 - (1 - pP_A(p'_A))^k \right) / p'_B \right\}$ of the nodes in $B$. Now, recalling that a fraction $1 - p'_B$ of nodes in $B$ have failed in Stage 2, we conclude that the joint effect of cascading failures in Stage 2 and Stage 4 (on $|B_4|$) is identical to that of an initial random attack which targets a fraction
\[
1 - p'_B + p'_B \left( 1 - \frac{1 - (1 - pP_A(p'_A))^k}{p'_B} \right)
\]
of nodes. As previously, with \( p_{B4}' = 1 - (1 - p_P A(p_{A3}')^k \) we conclude that the size of the functioning giant component \( B_4 \) is given by \( |B_4| = p_{B4}' P_B(p_{B4}') N = p_{B4} N \).

As mentioned earlier, the main goal of this section is to characterize the size of the functional giant components in steady state. Indeed, along the lines outlined above, one can obtain the sizes of all functioning giant components \( A_1 \supset A_3 \supset \ldots \supset A_{2m+1} \) and \( B_2 \supset B_4 \supset \ldots \supset B_{2m} \) for any integer \( m \). However, it is easy to observe the pattern in the expressions obtained so far and conclude that with \( p_{A1}' = p \) the size of all giant components are given by the recursive relations:

\[
p_{Ai} = p_{Ai}' P_A(p_{Ai}'),
\]

\[
p_{Ai}' = p \left( 1 - (1 - P_B(p_{Bi}' - 1))^k \right), \quad i = 3, 5, 7 \ldots \tag{2.12}
\]

and

\[
p_{Bi} = p_{Bi}' P_B(p_{Bi}'),
\]

\[
p_{Bi}' = 1 - (1 - p_P A(p_{Ai}' - 1))^k, \quad i = 2, 4, 6, \ldots \tag{2.13}
\]

This recursive process stops at an “equilibrium point” where we have \( p_{B2m-2}' = p_{B2m}' \) and \( p_{A2m-1}' = p_{A2m+1}' \) so that neither network A nor network B fragments further. Setting \( x = p_{A2m+1}' \) and \( y = p_{B2m}' \), this yields the transcendental equations

\[
x = p \left( 1 - (1 - P_B(y))^k \right) \quad y = 1 - (1 - p_P A(x))^k. \tag{2.14}
\]

The analysis carried out up to this point is valid for all networks irrespective of their intra-structures. In principle, for specific intra-structures of networks \( A \) and \( B \) (which determine the functions \( P_A \) and \( P_B \), respectively), the system (2.14) of equations can be solved for given \( p \) and \( k \). The steady-state fraction of nodes in the giant components can then be computed by using
the relations \( \lim_{i \to \infty} p_{Ai} := P_{A \infty} = xP_A(x) \) and \( \lim_{i \to \infty} p_{Bi} := P_{B \infty} = yP_B(y) \).

Indeed, in Section 2.5, we consider a special case where both networks \( A \) and \( B \) are Erdős-Rényi (ER) graphs [24] and give solutions of the system (2.14) for several values of \( p \) and \( k \).

### 2.4 Optimality of Regular Allocation Strategy

In this section, we show analytically that the regular allocation strategy always yields stronger robustness than other possible strategies and thus it is optimal in the absence of intra-topology information. In the following, we refer to the system that uses the regular allocation strategy as System 1. Specifically, we consider two networks \( A \) and \( B \) where each node is uniformly supported by \( k \) bi-directional inter-edges. For convenience, we denote the fractions in the recursive relations (2.12)-(2.13) as \( p'_{Ai}(p; k) \) and \( p'_{Bi}(p; k) \), where \( 1 - p \) is the initially failed fraction of nodes in network \( A \). Also, we let \( P_{A \infty}(p; k) \) and \( P_{B \infty}(p; k) \) be the steady-state fractions of functional giant components of the two networks, respectively. Finally, we use \( p_{c_1}(k) \) to denote the critical threshold associated with System 1.

In what follows, we first investigate the dynamics of cascading failures in the auxiliary System 2, where bi-directional inter-edges are distributed randomly amongst nodes. The analysis is carried out under a generic inter-degree distribution so that all possible (bi-directional) inter-link allocation strategies are covered. By making use of the convexity property and Jensen’s inequality, we show that for a fixed mean inter-degree, System 2 achieves the highest robustness against random attacks when its inter-degree distribution degenerates, i.e., when all nodes have exactly the same number of inter-edges so that System 2 is equivalent to System 1. Therefore, we conclude that regular
allocation yields the strongest robustness amongst all possible (bi-directional) inter-link allocation strategies. Next, we show that systems with bi-directional inter-edges can better combat the cascading failures compared to the systems with unidirectional inter-edges [23]. Together, these results prove the optimality of the inter-link allocation strategy in System 1; i.e., regular allocation of bi-directional inter-edges.

2.4.1 Analysis of Random Allocation Strategy

We now introduce the auxiliary System 2. Consider two arbitrary networks $A$ and $B$, each with $N$ nodes, and a discrete probability distribution $F : \mathbb{N} \rightarrow [0, 1]$ such that

$$F(j) = \alpha_j, \quad j = 0, 1, \ldots, \quad (2.15)$$

with $\sum_{j=0}^{\infty} \alpha_j = 1$.

To allocate the interdependency links, we first partition each network randomly into subgraphs with sizes $\alpha_0 N, \alpha_1 N, \alpha_2 N, \ldots$. By doing so, we can obtain subgraphs $\{S_{A0}, S_{A1}, S_{A2}, \ldots\}$ and $\{S_{B0}, S_{B1}, S_{B2}, \ldots\}$, such that

$$|S_{A0}| = |S_{B0}| = \alpha_0 N, \quad j = 0, 1, \ldots.$$ 

Then, for each $j = 0, 1, \ldots$, assume that each node in the subgraphs $S_{A0}$ and $S_{B0}$ is assigned $j$ bi-directional inter-edges. This ensures that the inter-degree of each node is a random variable drawn from the distribution $F$, i.e., an arbitrary node will have $j$ inter-edges with probability $\alpha_j$, for each $j = 0, 1, \ldots$. It is worth noting that the inter-degrees of the nodes are not mutually

---

$^1$For $N$ large enough, each of these subgraph sizes can be well approximated by an integer.
independent since the total number of inter-edges is fixed at $E = \sum \alpha_j jN$ for both networks.

We have a few more words on the possible implementation of the above random allocation strategy. Observe that each bi-directional edge can be treated equivalently as two unidirectional edges. In this way, there are a total of $2E$ unidirectional inter-edges in the system, where $E$ edges are going outward from network $A$, and the other $E$ edges are going outward from network $B$. We randomly match each unidirectional edge going outward from $A$ to a unique edge going outward from $B$ and combine them into a single bi-directional edge. To this end, let the edges going outward from $A$ and $B$ be separately labeled as $e = \{e_1, \ldots, e_E\}$ and $e' = \{e'_1, \ldots, e'_E\}$, respectively. Next, use the Knuth shuffle algorithm [29] to obtain random permutations $\bar{e} = \{\bar{e}_1, \ldots, \bar{e}_E\}$ and $\bar{e}' = \{\bar{e}'_1, \ldots, \bar{e}'_E\}$ of the vectors $e$ and $e'$, respectively. Finally, for each $i = 1, \ldots, E$, match the unidirectional inter-edges $\bar{e}_j$ and $\bar{e}'_j$ to obtain $E$ bi-directional inter-edges.

We now analyze the dynamics of cascading failures in System 2 using an iterative approach similar to that in Section 2.3. For brevity, we skip most of the details and give only an outline of the arguments that lead to the sizes of functional giant components. The main difference from the analysis of Section 2.3 is that the fractions of nodes in $A$ and $B$ retaining at least one inter-edge, i.e., the fractions $\bar{A}_i$ and $\bar{B}_i$, need to be calculated differently from (2.8) and (2.10) due to the random inter-degree per node.

Owing to the fragmentation from $\bar{B}_{i-1}$ to $B_{i-1}$, each inter-edge supporting $A$ could be disconnected with probability $1 - |B_{i-1}|/|\bar{B}_{i-1}|$, triggering further failures in network $A$ at step $i$. With this insight, the aggregated effect of the failures in $B$ up to stage $i$ can be equivalently treated (with respect
to the size of $A_i$) as removing each inter-edge supporting $A$ with probability $1 - u_i$. According to Section 2.3, $u_i$ can be derived as follows:

$$u_i = \prod_{\ell=1}^{(i-1)/2} \frac{|B_{2\ell}|}{|B_{2\ell}|} = P_B(p'_{Bi-1}) \quad i = 3, 5, 7, \ldots \quad (2.16)$$

Similarly, the aggregated effect of node failures in $A$ before step $i$ can be viewed as equivalent to removing each inter-edge supporting $B$ with probability $1 - v_i$ (with respect to the size of $B_i$) such that:

$$v_i = \frac{|A_1|}{|A|} \prod_{\ell=1}^{i/2-1} \frac{|A_{2\ell+1}|}{|A_{2\ell+1}|} = pP_A(p'_{Ai-1}) \quad i = 2, 4, 6, \ldots \quad (2.17)$$

In System 2, each node is supported by $j$ inter-edges with probability $\alpha_j$. In view of this, at step $i$, a node in network $A$ would retain at least one inter-edge with probability $1 - \sum_{j=0}^{\infty} \alpha_j (1 - u_i)^j$. Recalling also that a fraction $1 - p$ of the nodes had already failed before the onset of the cascading failure, the equivalent remaining fraction of network $A$ at stage $i$ is given by:

$$p'_{Ai} = p(1 - \sum_{j=0}^{\infty} \alpha_j (1 - u_i)^j)$$

$$= p(1 - \sum_{j=0}^{\infty} \alpha_j (1 - P_B(p'_{Bi-1}))^j).$$

Similarly, the equivalent remaining fraction of network $B$ turns out to be:

$$p'_{Bi} = 1 - \sum_{j=0}^{\infty} \alpha_j (1 - v_i)^j$$

$$= 1 - \sum_{j=0}^{\infty} \alpha_j (1 - pP_A(p'_{Ai-1}))^j.$$

Hence, the fractional sizes of the giant components at each stage are given (with $p'_{A1} = p$) by

$$p_{Ai} = p'_{Ai} P_A(p'_{Ai}),$$

$$p'_{Ai} = p(1 - \sum_{j=0}^{\infty} \alpha_j (1 - P_B(p'_{Bi-1}))^j),$$

$$p_{Bi} = p(1 - \sum_{j=0}^{\infty} \alpha_j (1 - P_B(p'_{Bi-1}))^j).$$

(2.18)
for \( i = 3, 5, 7 \ldots \), and by

\[
\begin{align*}
p_{Bi} &= p'_{Bi} P_B(p'_{Bi}), \\
p'_{Bi} &= 1 - \sum_{j=0}^{\infty} \alpha_j (1 - p P_A(p'_{Ai-1}))^j,
\end{align*}
\]  

(2.19)

for \( i = 2, 4, 6, \ldots \). We next show that System 1 is always more robust than System 2 against random attacks by comparing the recursive relations (2.12)-(2.13) and (2.18)-(2.19).

2.4.2 Regular Allocation vs. Random Allocation

We now compare Systems 1 and 2 in terms of their robustness against random attacks. For convenience, we use a vector \( \alpha = (\alpha_0, \alpha_1, \ldots) \) to characterize the inter-degree distribution \( F \), where \( F(j) = \alpha_j \). Next, we denote the fractions in the recursive relations (2.18)-(2.19) as \( p_{Ai}(p; \alpha), p'_{Ai}(p; \alpha) \) and \( p_{Bi}(p; \alpha), p'_{Bi}(p; \alpha) \). Also, we let \( P_{A2}(p; \alpha) \) and \( P_{B2}(p; \alpha) \) be the steady-state fractions of the functional giant components in two networks where \( 1 - p \) is the initially failed fraction of nodes in network \( A \). In other words, we set \( \lim_{i \to \infty} p_{Ai}(p; \alpha) := P_{A2}(p; \alpha) \) and \( \lim_{i \to \infty} p_{Bi}(p; \alpha) := P_{B2}(p; \alpha) \). Finally, we define the critical threshold associated with System 2 as \( p_{c2}(\alpha) \).

Assume that network \( A \) (respectively network \( B \)) of Systems 1 and 2 have the same size \( N \) and the same intra-degree distribution such that the functions \( P_A \) (respectively \( P_B \)) are identical for both systems. The next result shows that if the two systems are matched through their mean inter-degrees; i.e., if \( k = \sum_{j=0}^{\infty} \alpha_j j \), System 1 always yields stronger robustness than System 2 against random node failures.

**Theorem 2.4.1.** Under the condition

\[ k = \sum_{j=0}^{\infty} \alpha_j j, \]  

(2.20)
it follows that
\begin{align*}
P_{A_k}(p; k) &\geq P_{A_k}(p; \alpha), \\
P_{B_k}(p; k) &\geq P_{B_k}(p; \alpha),
\end{align*}
(2.21)
Furthermore, we have that
\begin{equation}
p_{c_1}(k) \leq p_{c_2}(\alpha).
\end{equation}
(2.22)

Proof. Since \( P_A \) and \( P_B \) are monotonically increasing functions \([28]\), a sufficient condition ensuring (2.21) will hold is
\begin{align*}
p'_{A_i}(p; k) &\geq p'_{A_i}(p; \alpha), \quad i = 3, 5, 7 \ldots, \\
p'_{B_i}(p; k) &\geq p'_{B_i}(p; \alpha), \quad i = 2, 4, 6 \ldots,
\end{align*}
(2.23)
where \( p'_{A_i}(p; k) \), \( p'_{B_i}(p; k) \), and \( p'_{A_i}(p; \alpha) \), \( p'_{B_i}(p; \alpha) \) denote the fractions in the recursive relations (2.12)-(2.13), and (2.18)-(2.19), respectively. We will establish (2.23) by induction. First observe that \( p'_{A_1}(p; k) = p'_{A_1}(p; \alpha) = p \) and the inequality (2.23) is satisfied for \( i = 1 \). In view of (2.12)-(2.13) and (2.18)-(2.19), condition (2.23) for \( i = 2 \) will be satisfied if
\begin{equation}
(1 - pP_A(p'))^k \leq \sum_{j=0}^{\infty} \alpha_j (1 - pP_A(p'))^j,
\end{equation}
or equivalently
\begin{equation}
(1 - pP_A(p))^k \leq \sum_{j=0}^{\infty} \alpha_j (1 - pP_A(p))^j.
\end{equation}
(2.24)
Under (2.20), the convexity of \( (1 - pP_A(p))^x \) implies (2.24) by Jensen’s inequality. Hence, we get \( p'_{B_2}(p; k) \geq p'_{B_2}(p; \alpha) \) and the base step is completed.

Suppose that the condition (2.23) is satisfied for each \( i = 1, 2, \ldots, 2m - 1, 2m \). We need to show that (2.23) holds also for \( i = 2m + 1 \) and \( i = 2m + 2 \). For \( i = 2m + 1 \), the first inequality will be satisfied if it holds that
\[(1 - P_B(p_{2m}^{B}(p; \alpha))^k \leq \sum_{j=0}^{\infty} \alpha_j (1 - P_B(p_{2m}^{B}(p; \alpha)))^j \]

By the induction hypothesis, we have \( P_B(p_{2m}^{B}(p; k)) \geq P_B(p_{2m}^{B}(p; \alpha)) \) since \( p_{2m}^{B}(p; k) \geq p_{2m}^{B}(p; \alpha) \). As a result, the above inequality will be satisfied if we show that

\[(1 - u)^k \leq \sum_{j=0}^{\infty} \alpha_j (1 - u)^j \quad (2.25)\]

with \( u = P_B(p_{2m}^{B}(p; \alpha)) \). As before, under (2.20), (2.25) is ensured by the convexity of \((1 - u)^x\) in view of Jensen’s inequality. The condition \( p_{2m+1}^{A}(p; k) \geq p_{2m+1}^{A}(p; \alpha) \) is now established.

Now let \( i = 2m + 2 \). The desired condition \( p_{2m+2}(p; k) \geq p_{2m+2}(p; \alpha) \) will be established if we show that

\[
(1 - p p_{A}^{A}(p_{2m+1}^{A}(p; k)))^k \leq \sum_{j=0}^{\infty} \alpha_j (1 - p p_{A}^{A}(p_{2m+1}^{A}(p; \alpha)))^j,
\]

or equivalently

\[(1 - v)^k \leq \sum_{j=0}^{\infty} \alpha_j (1 - v)^j, \quad (2.26)\]

where we set \( v = p p_{A}^{A}(p_{2m+1}^{A}(p; \alpha)) \). The last step follows from the previously obtained fact that \( p_{2m+1}^{A}(p; k) \geq p_{2m+1}^{A}(p; \alpha) \). Once more, (2.26) follows by the convexity of \((1 - v)^x\) and Jensen’s inequality. This establishes the induction step and the desired conclusion (2.21) is now obtained.

We next prove the inequality \( p_{c_1}(k) \leq p_{c_2}(\alpha) \) by way of contradiction. Assume towards a contradiction that \( p_{c_2}(\alpha) < p_{c_1}(k) \) and fix \( p \) such that \( p_{c_2}(\alpha) < p < p_{c_1}(k) \). Then, let a fraction \( 1 - p \) of the nodes randomly fail in network \( A \) of both systems. Since \( p \) is less than the \( p_{c_1} \), the node failures will eventually lead to complete fragmentation of the two networks in System 1.
i.e., we get \( P_{A\infty}(p; k) = P_{B\infty}(p; k) = 0 \). On the other hand, the fact that \( p \) is larger than the critical threshold \( p_c \) ensures \( P_{A\infty}(p; \alpha) > 0 \) and \( P_{B\infty}(p; \alpha) > 0 \) by definition. This clearly contradicts (2.21) and therefore it is always the case that \( p_c(k) \leq p_c(\alpha) \) under (2.20).

We have now established that the regular allocation of bi-directional inter-edges always yields stronger robustness than any possible random allocation strategy that uses bi-directional links. In the following section, we show that using bidirectional inter-edges leads to a smaller critical threshold and better robustness than using unidirectional inter-edges.

2.4.3 Bi-directional Inter-Edges vs. Unidirectional Inter-Edges

We now compare the robustness of System 2 with that of the model considered in [23], hereafter referred to as System 3. As mentioned earlier, the model considered in [23] is based on the random allocation of unidirectional inter-edges and can be described as follows. As with System 2, consider two arbitrary networks \( A \) and \( B \), each with \( N \) nodes, and a discrete probability distribution \( F : N \rightarrow [0, 1] \) such that (2.15) holds. Assume that each node is associated with a random number of supporting nodes from the other network, and that this random number is distributed according to \( F \). In other words, for each \( j = 0, 1, \ldots \), a node has \( j \) inward inter-edges with probability \( \alpha_j \). The supporting node for each of these inward edges is selected randomly amongst all nodes of the other network ensuring that the number of outward inter-edges follows a binomial distribution for all nodes.

System 3 was studied in [23] using similar methods to those of Section 2.3 and Section 2.4.1. This time, after an initial failure of \( 1 - p \) fraction of nodes in network \( A \), the recursive relations for the fractions of giant components at
each stage turns \([23]\) out to be (with \(p'_{A_1} = p\)):

\[ p_{A_i} = p'_{A_i} P_A(p'_{A_i}), \]
\[ p'_{A_i} = p(1 - \sum_{j=0}^{\infty} \alpha_j (1 - p'_{B_{i-1}} P_B(p'_{B_{i-1}}))^j), \]  
\(2.27\)

for \(i = 3, 5, 7, \ldots\), and

\[ p_{B_i} = p'_{B_i} P_B(p'_{B_i}), \]
\[ p'_{B_i} = 1 - \sum_{j=0}^{\infty} \alpha_j (1 - p'_{A_{i-1}} P_A(p'_{A_{i-1}}))^j, \]  
\(2.28\)

for \(i = 2, 4, 6, \ldots\).

We now compare System 2 and System 3 using the recursive relations (2.18)-(2.19) and (2.27)-(2.28). In doing so, we use the same notations to define the fractions in the recursive relations (2.18)-(2.19) as in Section 2.4.2, while the fractions in (2.27)-(2.28) will be denoted by \(p^3_{A_i}(p; \alpha), p'^3_{A_i}(p; \alpha)\) and \(p^3_{B_i}(p; \alpha), p'^3_{B_i}(p; \alpha)\). We let \(P_{A^3}(p; \alpha)\) and \(P_{B^3}(p; \alpha)\) be the steady-state fractions of functional giant components in System 3 if a fraction \(1 - p\) of the nodes initially fail in network \(A\). In other words, we set \(\lim_{i \to \infty} p^3_{A_i}(p; \alpha) = P_{A^3}(p; \alpha)\) and \(\lim_{i \to \infty} p^3_{B_i}(p; \alpha) = P_{B^3}(p; \alpha)\). Finally, we let \(p_{c3}(\alpha)\) be the critical threshold of System 3.

The next result shows that System 2 is always more robust than System 3 against random node failures.

**Theorem 2.4.2.** We have that

\[ P_{A^3}(p; \alpha) \geq P_{A^3}(p; \alpha), \]
\[ P_{B^3}(p; \alpha) \geq P_{B^3}(p; \alpha), \]  
\(2.29\)

and furthermore,

\[ p_{c3}(\alpha) \leq p_{c3}(\alpha). \]  
\(2.30\)
Proof. Since \( P_A(x) \) and \( P_B(x) \) are monotonically increasing [28], a sufficient condition ensuring (2.29) is given by

\[
P'_A(p; \alpha) \geq p^3 A_1(p; \alpha), \quad i = 1, 3, 5, \ldots, \quad \text{(2.31)}
\]

\[
P'_B(p; \alpha) \geq p^3 B_i(p; \alpha), \quad i = 2, 4, 6, \ldots
\]

We establish (2.31) by induction. First, observe that for \( i = 1 \), \( p'_A(p; \alpha) = p^3 A_1(p; \alpha) = p \) and condition (2.31) is satisfied. Next, for \( i = 2 \), we see from (2.19) and (2.28) that the inequality

\[
p'_B(p; \alpha) \geq p^3 B_i(p; \alpha)
\]

will hold if

\[
\sum_{j=0}^{\infty} \alpha_j (1 - p P_A(p'_A(p; \alpha)))^j \leq \sum_{j=0}^{\infty} \alpha_j (1 - p^3 A_1(p; \alpha) P_B(p^3 A_1(p; \alpha)))^j.
\]

(2.32)

Since \( p'_A(p; \alpha) = p^3 A_1(p; \alpha) = p \), it is immediate that (2.32) is satisfied with equality and this completes the base step of the induction.

Suppose now that condition (2.31) is satisfied for all \( i = 1, 2, \ldots, 2m - 1, 2m \). We will establish (2.31) for \( i = 2m + 1 \) and \( i = 2m + 2 \) as well. Comparing (2.18) and (2.27), it is easy to check that for \( i = 2m + 1 \), (2.31) will hold if

\[
\sum_{j=0}^{\infty} \alpha_j (1 - p B_B(p'_B(p; \alpha)))^j \leq \sum_{j=0}^{\infty} \alpha_j (1 - p^3 B_2m(p; \alpha) P_B(p^3 B_2m(p; \alpha)))^j.
\]

(2.33)

By the induction hypothesis, (2.31) holds for \( i = 2m \) so that \( P_B(p^3 B_{2m}(p; \alpha)) \leq P_B(p^3 B_{2m}(p; \alpha)) \). It is now immediate that (2.33) holds since we always have \( p^3 B_{2m}(p; \alpha) \leq 1 \). This establishes (2.31) for \( i = 2m + 1 \); i.e., that

\[
p^3 A_{2m+1}(p; \alpha) \leq p^3 A_{2m+1}(p; \alpha).
\]

(2.34)
For $i = 2m + 2$, we see from (2.19) and (2.28) that condition (2.31) will be satisfied if

$$\sum_{j=0}^{\infty} \alpha_j (1 - pP_A(p_{A2m+1}(p; \alpha)))^j$$

(2.35)

$$\leq \sum_{j=0}^{\infty} \alpha_j (1 - p_{A2m+1}^3(p; \alpha)P_A(p_{A2m+1}^3(p; \alpha)))^j.$$

In view of (2.34) and the fact that $p_{A2m+1}^3(p; \alpha) \leq p$, we immediately obtain (2.35) and the induction step is now completed. This establishes the condition (2.31) for all $i = 1, 2, \ldots$ and we get (2.29).

The fact that (2.29) implies (2.30) can be shown by contradiction, as in the proof of Theorem 2.4.1. \qed

Summarizing, it can be seen from Theorem 2.4.2 that using bi-directional inter-edges (System 2) always yields stronger system robustness compared to using unidirectional inter-edges (System 3). This being valid under an arbitrary distribution $\alpha$ of inter-edges, we conclude that the regular allocation of bi-directional inter-edges leads to the strongest robustness (amongst all possible strategies) against random attacks as we recall Theorem 2.4.1.

### 2.5 Numerical Results: The Erdős-Rényi Networks Case

To get a more concrete sense of the above analysis results, we next look at some special cases of network models. In particular, we assume both networks are Erdős-Rényi networks [24], with mean intra-degrees $a$ and $b$, respectively. For this case, the functions $P_A(x)$ and $P_B(y)$ that determine the size of the giant components can be obtained as follows [28]:

$$P_A(x) = 1 - f_A \quad \text{and} \quad P_B(y) = 1 - f_B,$$

(2.36)
where \( f_A, f_B \) are the unique solutions of

\[
\begin{align*}
  f_A &= \exp\{ax(f_A - 1)\} \quad \text{and} \quad f_B = \exp\{by(f_B - 1)\}. 
\end{align*}
\]

(2.37)

In what follows, we derive numerical results for the steady-state giant component sizes as well as critical \( p_c \) values. Specifically, we first study System 1 by exploiting the recursive relations (2.12)-(2.13) using (2.36) and (2.37). Similarly, we derive the numerical results for System 2 by using the recursive relations (2.18)-(2.19). For both cases, we use extensive simulations to verify the validity of the numerical results obtained theoretically.

### 2.5.1 Numerical Results for System 1

Reporting (2.36) into (2.14), we get

\[
\begin{align*}
  x &= p(1 - f_B^k) \quad y = 1 - (1 - p(1 - f_A))^k. 
\end{align*}
\]

(2.38)

It follows that the giant component fractions at the steady-state are given by

\[
\begin{align*}
  P_{A\infty} &= p(1 - f_B^k)(1 - f_A), \\
  P_{B\infty} &= \left(1 - (1 - p(1 - f_A))^k\right)(1 - f_B). 
\end{align*}
\]

(2.39)

Next, substituting (2.38) into (2.37) we obtain

\[
\begin{align*}
  f_A &= \exp\{ap(1 - f_B^k)(f_A - 1)\}, \\
  f_B &= \exp\{b \left(1 - (1 - p(1 - f_A))^k\right)(f_B - 1)\}. 
\end{align*}
\]

(2.40)

We note that the system of equations (2.40) always has a trivial solution \( f_A = f_B = 1 \), in which case the functional giant component has zero fraction for both networks. More interesting cases arise for large values of \( p \) when there exists non-trivial solutions to (2.40). In particular, we focus on determining the critical threshold \( p_c \), i.e., the minimum \( p \) that yields a non-trivial solution of the system. To dig into this further, we see by elementary algebra that, (2.40) is equivalent to
\[
\begin{align*}
  f_B &= \sqrt{1 - \frac{\log f_A}{(f_A - 1)ap}} & \text{if } 0 \leq f_A < 1; \quad \forall f_B \text{ if } f_A = 1 \\
  f_A &= 1 - \frac{1 - \sqrt{1 - \frac{\log f_B}{(f_B - 1)p}}}{p} & \text{if } 0 \leq f_B < 1; \quad \forall f_A \text{ if } f_B = 1.
\end{align*}
\]

(2.41)

In general, it is not possible to derive an explicit expression for \( p_c \). Instead, we can solve (2.41) graphically for a given set of parameters \( a, b, k, p \) and infer the critical threshold \( p_c \) from the plots. For instance, Figure 2.3 shows the possible solutions of the system for several different \( p \) values when \( a = b = 3, k = 2 \). In Figures 2.3 (a-c), we have \( p < p_c \) and there is only the trivial solution \( f_A = f_B = 1 \) so that both networks go into a complete fragmentation at the steady state. In Figure 2.3(d), we have \( p = p_c \) and there exist one non-trivial solution, since the two curves intersect tangentially at one point. In Figures 2.3(e-f), we have \( p > p_c \) and there exist two non-trivial intersection points corresponding to two sets of giant component sizes. In these cases, the solution corresponding to the cascading failures should be the point that yields the larger giant component size. In other words, the solution corresponds to the intersection point that is closer to the starting point of the iterative process (see (2.39)).

In the manner outline above, we can find the critical threshold \( p_c \) for any fixed values of the parameters \( a, b \) and \( k \). As illustrated in Figure 2.3, we can further add the tangential condition

\[
\frac{df_A}{df_B} \times \frac{df_B}{df_A} = 1
\]

(2.42)

to the equations (2.41) since the critical \( p_c \) value corresponds to the tangent point of the two curves given by (2.41). For any given parameters, the system of equations through which the critical values \( f_{A_c}, f_{B_c} \) and \( p_c \) can be computed
is given as below:

\[
\begin{align*}
    f_B &= \sqrt{1 - \frac{\log f_A}{(f_A - 1)ap}} \text{ if } 0 \leq f_A < 1; \\
    f_A &= 1 - \frac{1 - \sqrt{1 - \frac{\log f_B}{(f_B - 1)b}}}{p} \text{ if } 0 \leq f_B < 1; \\
    \left|\frac{df_A}{df_B}\right|_{\text{Eq.(2.44)}} \times \left|\frac{df_B}{df_A}\right|_{\text{Eq.(2.43)}} &= 1.
\end{align*}
\]

The analysis results are now corroborated by simulations. In Figure 2.4, we show the variation of \(p_c\) with respect to \(k\) for different values of \(a = b\), where the critical \(p_c\) values are obtained by solving the system (2.45) graphically. To verify these findings, we pick a few sets of values \(a, b\) and \(k\) from the curves in Figure 2.4 and run simulations with \(N = 5000\) nodes to estimate the probability \(p_{inf}\) of the existence of a functional giant component in steady state. As expected [2], in all curves we see a sharp increase in \(p_{inf}\) as \(p\) approaches to a critical threshold \(p_c\). It is clear that the estimated \(p_c\) values from the sharp transitions in Figure 2.5 are in good agreement with the analysis results given in Figure 2.4.
Figure 2.4: The critical $p_c$ value vs. $k$ for the regular allocation strategy (System 1). The plots are obtained by solving the system (2.45) graphically for various $a, b$ values. It can be seen that as $k$ increases the robustness of the system increases and the critical fraction $p_c$ approaches that of a single network; i.e., $\frac{1}{a}$ [24].

2.5.2 Numerical Results for System 2

As in System 1, the recursive process (2.18)-(2.19) of System 2 stops at an “equilibrium point” where we have $p'_{B2m-2} = p'_{B2m} = x$ and $p'_{A2m-1} = p'_{A2m+1} = y$. This yields the transcendental equations

$$x = p(1 - \sum_{j=0}^{\infty} \alpha_j (1 - P_B(y))^j),$$
$$y = 1 - \sum_{j=0}^{\infty} \alpha_j (1 - pP_A(x))^j. \quad (2.46)$$

The steady-state fraction of nodes in the giant components can be computed by using the relations $\lim_{i \to \infty} p_{Ai} := P_{A_{\infty}} = xP_A(x)$ and $\lim_{i \to \infty} p_{Bi} := P_{B_{\infty}} = yP_B(y)$.

In particular, we assume that the inter-degree distribution $F$ at each
Figure 2.5: Experimental results for the regular allocation strategy (System 1) with $N = 5000$ nodes. A $1 - p$ fraction of the nodes are randomly removed (from network $A$) and the corresponding empirical probability $p_{\text{inf}}$ for the existence of a functional giant component at the steady state is plotted. As expected, in all cases there is a sharp increase when $p$ approaches to a critical threshold $p_c$; for $(a = b = 3, k = 3)$, $(a = b = 3, k = 5)$ and $(a = b = 6, k = 3)$, the critical $p_c$ values are roughly equal to 0.47, 0.41 and 0.23, respectively. Clearly, these $p_c$ values are in perfect agreement with the corresponding ones of Figure 2.4 which are obtained analytically.

Node is a Poisson distribution with mean $k$, and hence

$$\alpha_j = e^{-k} \frac{k^j}{j!}, \quad j = 0, 1, 2, \ldots, \infty. \quad (2.47)$$

Substituting (2.36) and (2.47) into (2.46), we get

$$x = p \left( 1 - \sum_{j=0}^{\infty} \frac{k^j}{j!} e^{-k f_B^j} \right) = p \left( 1 - e^{-k(1-f_B)} \right), \quad (2.48)$$

and

$$y = 1 - \sum_{j=0}^{\infty} \frac{k^j}{j!} e^{-k (1 - p(1 - f_A))^j} = 1 - e^{-kp(1-f_A)}. \quad (2.49)$$
Next, putting (2.48) and (2.49) into (2.37), we find

\[
\begin{align*}
    f_A &= 1 + \frac{1}{pk} \ln \left( 1 + \frac{\ln f_B}{b(1-f_B)} \right), \text{ if } 0 \leq f_B < 1; \\
    f_B &= 1 + \frac{1}{k} \ln \left( 1 + \frac{\ln f_A}{ap(1-f_A)} \right), \text{ if } 0 \leq f_A < 1; \\
    \forall f_A & \text{ if } f_B = 1; \quad \forall f_B & \text{ if } f_A = 1.
\end{align*}
\]

(2.50)

As in the case for System 1, the critical threshold \( p_c \) for System 2 corresponds to the tangential point of the curves given by (2.50), and can be obtained by solving (2.50) graphically.

We now check the validity of the analytical results via simulations. In Figure 2.6, we show the variation of analytically solved \( p_c \) values with respect to average inter-degree \( k \) for different values of \( a = b \). To verify these results, we pick a few sets of values \( a, b \) and \( k \) from the curves in Figure 2.6 and
Figure 2.7: Experimental results for System 2 with $N = 5000$ nodes. A $1 - p$ fraction of the nodes are randomly removed (from network $A$) and the corresponding empirical probability $p_{inf}$ for the existence of a functional giant component at the steady state is plotted. As expected, in all cases there is a sharp increase when $p$ approaches to a critical threshold $p_c$; for $(a = b = 4, k = 2)$, $(a = b = 4, k = 3)$ and $(a = b = k = 4)$, the critical $p_c$ values are roughly equal to 0.480, 0.380 and 0.335, respectively. Clearly, these $p_c$ values are in perfect agreement with the corresponding ones of Figure 2.6 which are obtained analytically.

run simulations with $N = 5000$ nodes to estimate the probability $p_{inf}$ of the existence of a functional giant component in steady state. As expected [2], in all curves we see a sharp increase in $p_{inf}$ as $p$ approaches to a critical threshold $p_c$. It is also clear from Figure 2.7 that for all parameter sets such sharp transition occurs when $p$ is close to the corresponding $p_c$ value given in Figure 2.6.

2.5.3 A Comparison of System Robustness

In Section 2.4.2 and 2.4.3, we have analytically proved that the regular allocation of bi-directional inter-edges leads to the strongest robustness against random attacks. To get a more concrete sense, we now numerically compare the system robustness of these strategies in term of their critical thresholds
Figure 2.8: $p_c$ v.s. $k$ for different values of $a = b$ in System 1 and System 3.

Figure 2.9: $p_c$ v.s. $a = b$ for various $k$ values in System 1 and System 3. In all cases, we see that the regular allocation of bi-directional inter-edges yields a smaller $p_c$ than the Poisson distribution of unidirectional inter-edges with the same mean value $k$.

$p_c$. Specifically, we consider coupled Erdős-Rényi networks with mean intra-degrees $a$ and $b$. For the sake of fair comparison, we assume that the mean inter-degree is set to $k$ for all systems; and in both Systems 2 and 3, the inter-degree distribution $F$ at each node is assumed to be Poisson. The critical
Figure 2.10: $p_c$ v.s. $k$ for different values of $a = b$ in System 1 and System 2. It is clear that System 1 yields a lower $p_c$ (and thus a higher robustness) than System 2 in all cases.

Figure 2.11: $p_c$ v.s. $a = b$ for various $k$ values in System 1, System 2 and System 3. In all cases System 1 yields the lowest $p_c$ (i.e., highest robustness), while System 3 has the highest $p_c$ (i.e., lowest robustness) and System 2 stands in between.

threshold value $p_c$ corresponding to all three strategies are compared under a variety of conditions. For Systems 1 and 2, we use the numerical results derived in Section 2.5.1 and Section 2.5.2, respectively, while for System 3 we
use the numerical results provided in [23].

First, we compare System 1 with System 3 to see the difference between the proposed regular inter-edge allocation strategy with the strategy in [23]. Figure 2.8 depicts $p_c$ as a function of mean inter-degree $k$ for various values of $a = b$, whereas Figure 2.9 depicts the variation of $p_c$ with respect to $a = b$ for different $k$ values. In all cases, it is seen that the regular allocation of bi-directional inter-edges yields a much smaller $p_c$ (and thus, a more robust system) than the random allocation of unidirectional inter-edges. For instance, for $a = b = k = 4$, System 3 [23, Figure 2] gives $p_c = 0.43$, whereas as seen via Figure 2.8, System 1 yields a critical threshold at 0.317. This is a significant difference since it means that System 3 can have a functioning giant component despite a random failure of at most 57% of the nodes, whereas System 1, which is proposed in this study, is resistant to a random failure of up to 68% of the nodes. Indeed, in some cases, our strategy can outperform that in [23] even with half (mean) inter-degree per node. For instance, when $a = b = 4$, our strategy yields $p_c = 0.414$ with only $k = 2$ as compared to $p_c = 0.43$ of the System 3 with $k = 4$.

We also compare System 1 with System 2 in order to see the improvement in allocating bi-directional edges regularly rather than randomly. Figure 2.10 depicts $p_c$ as a function of mean inter-degree $k$ for various values of $a = b$. It is seen that in all cases System 1 yields a lower $p_c$ (and thus a more robust system) than System 2. For example, when $a = b = 3$ and $k = 2$, we get $p_c = 0.56$ for System 1, while for System 2, we find that $p_c = 0.68$. The difference is significant in that it corresponds to a resiliency against a random failure of up to 44% of the nodes in System 1 as compared to that of 32% in System 2.
Finally, in order to better illustrate the optimality of System 1 in term of system robustness, we depict in Figure 2.11 the variation of \( p_c \) with respect to \( a = b \) for different \( k \) values in all three systems. It is clear that the proposed regular allocation strategy in System 1 always yields the lowest \( p_c \) and thus provides the best resiliency against random attacks. We also see that System 2 always outperforms System 3 showing the superiority (in terms of robustness) of using bi-directional inter-edges rather than unidirectional ones.

We believe that the drastic improvement in the robustness against random attacks seen in System 1 can be attributed to the following reasons: First, in the lack of the intra-topology information, it is difficult to tell which nodes play more important roles in preserving the connectivity of the networks. Thus, in order to combat against random attacks, it is reasonable to treat all nodes the same and give them equal priority in inter-edge allocation. Secondly, in Systems 2 and 3, there may exist a non-negligible fraction of nodes with no inter-edge support from the other network; and those nodes are automatically non-functional even if they are not attacked. But, the regular allocation scheme promises a guaranteed support in terms of inter-edges, for all nodes in both networks. Finally, using bi-directional inter-edges ensures that the amount of support provided is equal to the amount of support being received for each node. Thus, bi-directional inter-edges increases the regularity of the support-dependency relationship over unidirectional inter-edges, and this may help improve the system robustness.

2.6 Conclusions

We studied the robustness of a cyber-physical system in which a cyber-network overlays a physical-network. To improve network robustness against random
node failures, we developed and studied a regular allocation strategy that allots a fixed number of inter-network edges to each node. Our findings reveal that the proposed regular allocation strategy yields the optimal robustness amongst all strategies when no information regarding the intra-topologies of each individual network is available. We expect that in the presence of such information, the topology of the networks can be exploited to improve further the robustness of cyber-physical systems against cascading failures.
3.1 Introduction

3.1.1 Motivation

In the previous chapter, we studied the cascading failures across multiple interdependent networks. For each individual network, the Giant Connected Component (GCC) model [27, 28] is used to characterize the failure propagation process. Specifically, only the nodes in the giant connected component are assumed to be operational, whereas the other isolated components would not work. However, such assumptions may not be applicable to many real-world cyber-physical systems, since the GCC model is incapable to accurately characterize the node failures in many physical infrastructure networks. Specifically, for physical systems such as power grids, the propagation of cascading failures is mainly due to the load redistribution. For example, the outage of one power station could lead to the redistribution of the load from the failed power station to its nearby neighbors. Further failures could happen only if the carried load exceeds the maximum capacity due to the load redistribution. This key feature, however, cannot be captured by the GCC model. Furthermore, the GCC model assumes that the whole network stops working once it is fragmented into many isolated components. In many physical networks, however, the isolated components can still function independently. For example, with the advent of the micro-grid technology, the power grid can be separated into multiple isolated components [30, 31]. Another example is the defensive
islanding technique [32, 33] that can block the propagation of node failures by intentionally disintegrating the power grid into multiple isolated “islands,” many of which can continue to work normally.

In more recent studies on complex networks, one plausible model for the physical network is the threshold model [34–36]. In contrast to the GCC model, the threshold model is developed based on the load redistribution feature, and each node has a threshold that represents its load capacity. A node would fail only if its failed neighbor fraction exceeds the threshold, indicating that the redistributed load from the failed neighbors exceeds its load capacity. As elaborated in Section 3.2, the threshold model offers a more accurate characterization of the cascading failures in a physical network, particularly for the power grid. To the best of our knowledge, the threshold model has not been applied to study the interdependent networks.

### 3.1.2 Summary of Main Results

In this chapter, we study the dynamics of cascading failures in the CPS, with focus on quantifying the impact of network interdependence on network resilience. To this end, we consider a CPS model with two interacting networks that stand for a cyber network and a physical infrastructure network, respectively. For the cyber network, the GCC model is utilized to characterize the node failures, since the operation of the cyber network (e.g., the Internet) relies on the existence of the giant component [25]. For the physical network, the threshold model is used to capture the unique physical feature as aforementioned. Based on the above models, we quantify the impact of network interdependence on the robustness of the overall CPS by characterizing the functioning node fraction after the cascading failures. We show that the CPS model proposed above can naturally capture some key features of real-world
cyber-physical systems. As in the practical cases, e.g., the 2003 blackout in Italy [37], our model shows that even a very small initial failures could trigger a cascade of failures that finally damage the whole system. It is also observed that the sparse connectivity in the physical network can slow down the propagation of the cascading failures, which agrees with the design principle of the islanding techniques [32,33]. Note that these features are not captured in the existing studies [2].

We also compare the robustness of the proposed CPS model with that of a single physical network. Our findings reveal that the dense connectivity in the physical network would likely make the cyber-physical system more vulnerable to cascading failures. This is somewhat surprising and indeed is in a stark contrast to some previous observations on the resilience of a single physical network. As elaborated in Section 3.4.2, this discrepancy can be attributed to the joint “effect” of the dense connectivity in the physical network and the network interdependence in the CPS.

Finally, we develop a strategy to improve the robustness for our CPS model by enabling a fraction of node to be autonomous, i.e., some nodes can support by themselves (with no support from the other network). We show that the functioning node fraction could exhibit a more “smooth” transition in contrast to the “sharp” transition for the case without autonomous nodes, which indicates that the autonomous nodes can significantly enhance the robustness of cyber-physical systems under heavy attacks.

To the best of our knowledge, this study is the first work that characterizes the robustness of cyber-physical systems with the threshold model for the physical network and the GCC model for the cyber network. We believe that our findings along this line can shed light on the impact of network inter-
dependence on CPS resilience and future development of protection strategies that would improve the robustness of cyber-physical systems.

3.2 System Setup

3.2.1 Model of the Cyber-Physical System

We consider a CPS model with two interconnecting networks, say network $A$ and network $B$, each with $N$ nodes, with a topology structure illustrated in Fig. 3.1. In particular, network $A$ stands for a physical network and network $B$ stands for a cyber network. As in \cite{2,18,38–41}, the nodes in each individual network are connected by intra-edges. We assume that in network $A$, one node can randomly connect to $k^A$ other nodes in $A$, where $k^A$ is a random variable following the distribution \{\(p^A_k, k = 0, 1, \ldots\)\}. Similarly, we define $k^B$ and \{\(p^B_k, k = 0, 1, \ldots\)\} for network $B$.

In this study, we assume that each node in $A$ depends on only one node in $B$ and vice versa. We model the network interdependence by inter-edges, each connecting a node from network $A$ and a node from network $B$. The
connected two nodes provide mutual supports to each other. Therefore, the failure of a node in $A$ would incur the further failure of the “dependent” node in $B$ and vice versa. The connections of inter-edges can be represented by using an $N \times N$ interdependency matrix $C$. Specifically, for $n, m = 1, \ldots, N$, let

$$
(C)_{nm} = \begin{cases} 
1 & \text{if } n = m \\
0 & \text{otherwise}
\end{cases}
$$

(3.1)

where $C_{nm} = 1$ means that node $n$ in $A$ and node $m$ in $B$ are connected by an inter-edge; otherwise they are independent.

### 3.2.2 GCC Model vs. Threshold Model

Given the topology structure, the next key question is how to accurately characterize the spreading of cascading failures therein. To this end, we focus on two candidate models, the GCC model and the threshold model.

The GCC model [2] assumes that only the nodes in the giant connected component that occupies a substantial fraction of the network are functioning, whereas the other isolated components would stop working, since they lose the connections to the majority of the network. In other words, the collection of the nodes in the giant component is also that of the functioning nodes. For example, in Fig. 3.2, once node $c$ is attacked, its connections $c-f$, $c-e$, $c-a$, $c-d$ and $c-b$ will be taken out. It follows that the network would separate into three isolated components, i.e., $(a,b), (c), (e,f,g,d,h)$ and only the component $(e,f,g,d,h)$ can still operate.

Under the GCC model, the existence of the giant component hinges heavily on the network connectivity. According to the random graph theory [28], only when the network has a dense connectivity above a critical threshold [28], will a giant connected component exist. Otherwise, the net-
work will fragment into isolated components. In this case, no node can still function under the GCC model. Note that such property is compatible with the communication network, which requires the existence of the giant component to maintain the global communication. However, as we mentioned before, it is not necessarily the case in many physical networks where isolated components can still work, e.g., power grids.

Along a different avenue, the threshold model is developed based on the load redistribution feature [34–36]. Generally speaking, once a node is attacked, its load would be redistributed to nearby nodes, which may cause further failures if their aggregated loads exceed capacity. Therefore, a node is more likely to be toppled down if a large fraction of its neighbors have failed. To capture this feature, the threshold model assumes that each node has a threshold \( \theta \in (0, 1] \) corresponding to its load capacity. In the event of attacks, an arbitrary node \( i \) would fail if its failed neighbor fraction exceeds \( \theta \); in other words, the number of failed neighbors is larger than \( d_i \theta \), where \( d_i \) is the degree of node \( i \). As illustrated in Fig. 3.3, the failure of node \( c \) would cause the further failures of nodes \( a, b, d \), while nodes \( f, e, g \) remain functioning.
Under the threshold model, the network operation does not rely on the existence of the giant component. In fact, in a sparsely connected network, an isolated component can still operate independently. As aforementioned, this feature is more relevant to some practical physical networks such as power grids. Worth noting is that the threshold $\theta$ indicates the strength of network resilience against cascading failures. Besides, the resilience of each individual node also depends on its degree [36]. For instance, if $\theta$ is 0.45 and node $i$ has only two neighbors, then the failure of either neighbor can topple down node $i$. In case that node $i$ has 5 neighbors, it can still operate when two neighbors have failed.

In our study for cyber-physical systems, we capitalize on the features of both models. Particularly, we use the threshold model and the GCC model to characterize the cascading failures in physical network $A$ and cyber network $B$, respectively. For convenience, we refer to our proposed CPS model as System 1, the interdependent networks based on two GCC models [2] as System 2, and the single physical network under the threshold model [36] as System 3.
3.2.3 Cascading Failures due to Network Interdependence

The cascading failures are triggered by the failure of a fraction $1 - p$ of randomly selected nodes in network $A$. It is possible that even if the fraction $1 - p$ is close to 0 such that a few nodes fail in a very large network, the subsequent cascading failures would propagate recursively between networks $A$ and $B$, and finally lead to catastrophic damages on the whole system. Specifically, at Stage 1, after the initial failures, further failures will happen in the remaining $pN$ nodes in network $A$. As a result, at Stage 2, some nodes in network $B$ may end up losing their supports from network $A$ and then become dysfunctional. Within the remaining part of network $B$, only the nodes in the giant component can still operate. Consequently, at Stage 3, the fragmentation of network $B$ would cause further failure propagation in network $A$. Continuing in this recursive manner, the cascading failures propagate alternatively between two networks until no further failures would happen.

3.3 Analysis of Cascading Failures

3.3.1 Recursive Dynamics in Cascading Failures

In this section, we analyze the dynamics of cascading failures in System 1. The main focus is on quantifying the functioning node fractions in both networks at each stage of the cascading failures. For ease of exposition, we first define the following two functions:

- Under the threshold model, suppose that the cascading failures start with a fraction $1 - p$ of node failures in a network with $N$ nodes. Let the function $G_A(p)$ denote the functioning node fraction after the cascading failures; accordingly, the number of remaining functioning nodes
is $G_A(p)N$.

- Under the GCC model, suppose that a fraction $1 - p$ of the nodes have been taken down in a network with $N$ nodes. Let the function $G_B(p)$ denote the functioning giant component fraction out of the remaining $pN$ nodes; accordingly, the giant component contains $pG_B(p)N$ nodes.

The detailed calculation of $G_A(p)$ and $G_B(p)$ will be elaborated in Section 3.3.2.

Our first main result is given in the following proposition. For convenience, we use the same notation summarized in Table 2.1.

**Proposition 3.1.** At stage $i$, the functioning node fractions in networks $A$ and $B$, denoted by $p_{A_i}$ and $p_{B_i}$, are given as follows. For $i = 1$, $p_{A_1} = G_A(p'_{A_1})$, where $p'_{A_1} = p$. For $i = 2$, $p_{B_2} = p'_{B_2}G_B(p'_{B_2})$, where $p'_{B_2} = G_A(p)$. For $i \geq 3$, the fractions can be found by using the following recursive equations:

- For $i = 3, 5, 7, \cdots$,
  \[
  p_{A_i} = G_A(p'_{A_i}), \quad p_{B_i} = p_{B_{i-1}},
  \]
  \[
  p'_{A_i} = pG_B(p'_{B_{i-1}}).
  \]

- For $i = 4, 6, 8, \cdots$,
  \[
  p_{A_i} = p_{A_{i-1}}, \quad p_{B_i} = p'_{B_i}G_B(p'_{B_i}),
  \]
  \[
  p'_{B_i} = \frac{G_A(p'_{A_{i-1}})}{G_B(p'_{B_{i-2}})}.
  \]

In what follows, we present the proof outline of Proposition 3.1. We first analyze of the initial stages of the cascading failures.

**Stage 1: initial node failures in network A.** The cascading failures start with an initial attack at a fraction $1 - p$ of randomly selected nodes in
network $A$. We denote the set of nodes still functioning after the initial attack as $\bar{A}_1$ whose size can be given by $|\bar{A}_1| = pN$.  

Under the threshold model, the initial attack would trigger further failures in $\bar{A}_1$. It follows that the number of remaining functioning nodes in network $A$ can be given by

$$|A_1| = G_A(p)N. \quad (3.4)$$

**Stage 2: node failures in $B$ due to the failures in Stage 1.** The node failures in network $A$ would cause further failures of the dependent nodes in network $B$. Particularly, a fraction $1 - G_A(p)$ of nodes in $B$ would turn dysfunctional due to the one-to-one support between two networks. Let $\bar{B}_2$ denote the collection of nodes in $B$ still receiving support from network $A$, the size of which can be given by

$$|\bar{B}_2| = G_A(p)N = p'_{B_2}N, \quad (3.5)$$

where $p'_{B_2} = G_A(p)$. Note that any node in $\bar{B}_2$ can work if and only if it is also in the giant component $B_2$, whose size can be calculated by

$$|B_2| = p_{B_2}N = p'_{B_2}G_B(p'_{B_2})N. \quad (3.6)$$

**Stage 3: further node failures in $A$ due to the failures in Stage 2.** The fragmentation from $B_2$ to $B_2$ will cause further failures in $A_1$, since the failed nodes in $\bar{B}_2 \setminus B_2$ no longer provide support to the nodes in $A_1$. \footnote{We assume that $N$ is very large and $pN$ can be approximated as an integer.} As a result, a fraction $1 - |B_2|/|\bar{B}_2| = 1 - G_B(p'_{B_2})$ of the nodes in $A_1$ will become dysfunctional. Let $\bar{A}_3 \subset A_1$ denote the collection of nodes in $A_1$ that still receive support from network $B$, whose size can be given by

$$|\bar{A}_3| = p_{A_1}G_B(p'_{B_2})N.$$
As in Stage 1, the failures of the nodes in $A_1 \setminus \bar{A}_3$ would continue to cause further failures in $\bar{A}_3$ at Stage 3, resulting in a smaller collection of functioning nodes $A_3 \subset \bar{A}_3$. Worth noting is that we cannot directly utilize the function $G_A(p)$ to quantify the size of $A_3$ since $G_A(p)$ is only applicable to the original network $A$ (as we shown in Section 3.3.2, the exact expression of $G_A(p)$ depends on the topology structure of network $A$). With this insight, we treat the failure of the nodes in $A_1 \setminus \bar{A}_3$ as an initial attack on $A$ that yields the equivalent effect on the size of $A_3$. Specifically, the failure of the nodes in $A_1 \setminus \bar{A}_3$ (the failure of a fraction $1 - G_B(p'_{B_2})$ of the nodes in $A_1$) has the same effect as attacking the same fraction of nodes in $\bar{A}_1$ [38]. In other words, it is equivalent to the failure of a fraction $p(1 - G_B(p_{B_2}' \bar{A}_3))$ of the nodes in $A$. Recall that a fraction $1 - p$ of nodes in network $A$ have failed already in the initial attack at Stage 1. Together, we conclude that the joint effect of node failures in Stages 1 and 3 is equivalent to attacking at a fraction

$$1 - p + p(1 - G_B(p_{B_2}')) = 1 - pG_B(p_{B_2}')$$

of the nodes in network $A$. With this insight, we define $p'_{A_3}$ as an “equivalent” remaining node fraction in network $A$, where

$$p'_{A_3} = pG_B(p'_{B_2}).$$  \hspace{1cm} (3.7)

It follows that the number of functioning nodes in $A_3$ can be given by

$$|A_3| = p_{A_3}N = G_A(p_{A_3}')N.$$  \hspace{1cm} (3.8)

**Stage 4: further failures in network B due to the failures in**

**Stage 3.** As previously discussed, the failure of the nodes in $\bar{A}_3 \setminus A_3$ can cause subsequent fragmentation in $B_2$. Along the same line as in Stage 2, a fraction

$$1 - |A_3|/|\bar{A}_3| = 1 - G_A(p_{A_3}')/(G_B(p_{B_2}')G_A(p))$$

of nodes in $B_2$ will lose the
supports from network $A$ and thus become dysfunctional. Let $\bar{B}_4$ denote the set of nodes in $B_2$ that still receive supports from $A$. It follows that

$$|\bar{B}_4| = |B_2| \frac{|A_3|}{|A_3|} = \frac{p'_B B_2 G_B(p'_B B_2) G_A(p'_A A_3) N}{G_B(p'_B B_2) p'_B B_2} \quad (3.9)$$

$$= G_A(p'_A A_3) N. \quad (3.10)$$

As in Stage 3, the fragmentation from $B_2$ to $\bar{B}_4$ has the same effect as taking out a fraction $p'_B B_2 \left(1 - \frac{|A_3|}{|A_3|}\right)$ of nodes in $B$ with respect to the size of the giant component inside $\bar{B}_4$. Recall that a fraction $1 - p'_B B_2$ of nodes in $B$ have failed already at Stage 2. Therefore, the joint effect of the node failures in Stages 2 and 4 is equivalent to taking out a fraction $1 - p'_B B_4$ of nodes in $B$, where

$$p'_B B_4 = 1 - \left(1 - p'_B B_2 + p'_B B_2 \left(1 - \frac{|A_3|}{|A_3|}\right)\right)$$

$$= \frac{G_A(p'_A A_3)}{G_B(p'_B B_2)}. \quad (3.11)$$

Accordingly, the giant component fraction can be given by

$$p_{B_4} = p'_B B_4 G_B(p'_B B_4). \quad (3.12)$$

Along the line outlined above, we observe that the cascading failures take place alternatively between networks $A$ and $B$ at each stage. As in [38], based on the preliminary results obtained so far, we conclude that the functioning node fractions in two networks can be quantified by the recursive relations (3.2) and (3.3). Note that starting with the base steps (3.4)-(3.12), we can rigorously show (3.2) and (3.3) by using induction method.

### 3.3.2 Analytical Results for Cascading Failures

To obtain the functioning node fraction after the cascading failures, we can directly explore the “fixed point” of the recursive process in Proposition 3.1,
which corresponds to the final stage of the cascading failures [2,38]. Specifically, we have the following results.

**Proposition 3.2.** After the cascading failures, the functioning node fractions in networks A and B, denoted by \( p_{A\infty} \) and \( p_{B\infty} \), are given by

\[
p_{A\infty} = p_{B\infty} = G_A(x) = yG_B(y),
\]

where \( x \) and \( y \) are the solutions to the following system of equations:

\[
x = pG_B(y), \quad (3.14)
\]

\[
y = \frac{G_A(x)}{G_B(y)}, \quad (3.15)
\]

**Proof.** The cascading failures would finally stop when the system is either totally destroyed or still with a fraction of nodes remaining functioning, due to the finite network size. Mathematically speaking, the recursion of (3.2) and (3.3) will finally stop at a fixed point where no further node failures will happen, i.e., \( p'_{A_i} = p'_{A_{i-2}} \) and \( p'_{B_i} = p'_{B_{i-2}} \). Clearly, due to the one-to-one support between two networks, after the cascading failures, the functioning node fractions in two networks should be equal, i.e., \( p_{A\infty} = p_{B\infty} \). As in [2,38], at the fixed point, we can set \( x = p'_{A_i} = p'_{A_{i-1}} = p'_{A_{i-2}} \) and \( y = p'_{B_i} = p'_{B_{i-1}} = p'_{B_{i-2}} \) in (3.2) and (3.3) which directly gives rise to (3.13)-(3.15).

We next consider the specific expressions of \( G_A(x) \) and \( G_B(y) \). Under the threshold model [36], \( G_A(x) \) can be determined by

\[
G_A(x) = x - x \sum_{k=1}^{\infty} p_k^A \sum_{m=0}^{k} \binom{k}{m} f_A^m (1 - f_A)^{k-m} F(m,k), \quad (3.16)
\]
where $f_A$ is the solution to

$$f_A = 1 - x + x \sum_{k=1}^{\infty} \frac{k p_A^k}{z_A} \left( \sum_{m=0}^{k-1} \binom{k-1}{m} f_A^m (1 - f_A)^{k-m-1} F(m, k) \right).$$  

(3.17)

In (3.16) and (3.17), \( \{p_A^k, k = 0, 1, \ldots\} \) is the degree distribution of network $A$, whose mean degree is denoted by $z_A$. $F(m, k)$ denotes the failure probability of a $k$-degree node if $m$ out of its $k$ neighbors have failed already. In this study, we assume that the threshold $\theta$ is the same across all the nodes and it follows that

$$F(m, k) = \begin{cases} 1 & \text{if } m \geq \theta k \\ 0 & \text{if } m < \theta k. \end{cases}$$

Note that the above results are applicable to any topology structure of networks $A$ and $B$. In what follows, we consider a special case that both networks $A$ and $B$ are Erdős-Rényi networks [24], with average intra-degrees $z_A$ and $z_B$, respectively. Therefore, the distribution $\{p_A^k, k = 0, 1, \ldots\}$ follows Poisson distribution with mean $z_A$ [24] and $G_B(y)$ can be obtained from [28]

$$G_B(y) = 1 - f_B,$$  

(3.18)

where $f_B$ is the solution to

$$f_B = \exp\{z_B y (f_B - 1)\}.$$  

(3.19)

In this case, the functioning node fraction after the cascading failures can be characterized by the following corollary. For ease of exposition, we define $M_1$ and $M_2$ as the functions of $f_A$:

$$M_1(f_A) = \sum_{k=1}^{\infty} p_A^k \left( \sum_{m=0}^{k} \binom{k}{m} f_A^m (1 - f_A)^{k-1} F(m, k) \right)$$  

(3.20)
\[ M_2(f_A) = \sum_{k=1}^{\infty} \left( \sum_{m=0}^{k-1} \binom{k-1}{m} f_A^m (1 - f_A)^{k-m-1} F(m, k) \right) \cdot \] (3.21)

**Corollary 3.1.** Suppose that both networks A and B are Erdős-Rényi networks, with average intra-degrees \( z_A \) and \( z_B \), respectively. The remaining functioning node fraction after the cascading failures is given by

\[ p_{A\infty} = p_{B\infty} = \frac{(1 - f_A) (1 - M_1(f_A))}{1 - M_2(f_A)} , \] (3.22)

where \( f_A \) can be obtained by solving the following system of equations:

\[
    f_B = 1 - \frac{1 - f_A}{p(1 - M_2(f_A))}, \\
    f_B = \exp \left( z_B \frac{(1 - M_1(f_A)) (f_A - 1)}{(1 - M_2(f_A))} \right).
\] (3.23) (3.24)

**Proof.** In view of (3.20) and (3.21), we can rewrite (3.16) and (3.17) as follows:

\[ G_A(x) = x - xM_1(f_A), \] (3.25)
\[ f_A = 1 - x + xM_2(f_A). \] (3.26)

By reporting (3.26) and (3.14) into (3.18), we can obtain (3.23). Furthermore, in view of (3.15) (3.18) and (3.19), we have

\[ f_B = \exp \left( -z_B G_A(x) \right) . \] (3.27)

Reporting (3.25) and (3.26) into (3.27), we can get (3.24). Finally, we can show (3.22) by substituting (3.25) and (3.26) into (3.13).

According to Corollary 3.1, one key step in calculating the functioning node fraction is to figure out the value of \( f_A \). However, the closed forms of \( f_A \) are difficult to derive if not possible. Instead, \( f_A \) can be obtained by graphically solving the system of equations (3.23) and (3.24). Figs. 3.4 and 61
3.5 illustrate the graphical solutions for the cases with $\theta = 0.21$ and $\theta = 0.37$, where $z_A = 1.5$, $z_B = 3$ and $p = 0.999$. Two curves correspond to the values of $f_B$ versus $f_A$ based on (3.23) and (3.24). The intersection of two curves corresponds to the solution to (3.23) and (3.24) and indicates the fixed point where the recursive process (3.2) and (3.3) will stop. Obviously, the above system of equations always yield a trivial solution $f_A = f_B = 1$, which indicates $p_{A\infty} = p_{B\infty} = 0$, i.e., the whole system is destroyed. For $\theta = 0.37$, there exists another nontrivial solution $f_A = 0.25$ and $f_B = 0.13$, indicating that a certain fraction of network is still operational. The nontrivial solution corresponds to the fixed point where the process of cascading failures would actually stop. With the obtained $f_A$, $p_{A\infty}$ can be computed through (3.22).
3.4 Numerical Results

In Section 3.3, we have analyzed the dynamics of cascading failures in System 1. To get a more concrete sense of the findings, we first compare, via numerical examples, the robustness of System 1 and System 2. Further, we compare the robustness between System 1 and a single physical network (System 3), aiming at understanding the impact of the network interdependence in cyber-physical systems. For System 1, we use the analytical results obtained in Section 3.3.2, while for System 2 and System 3, we use the results provided in [2] and [36], respectively. In particular, we assume that both networks $A$ and $B$ are Erdős-Rényi (ER) graphs [28] with mean intra-degree $z_A$ and $z_B$, respectively. Note that many cascading failures in practical cyber-physical systems are triggered by a few malfunctioned nodes [42]. With this insight, we here focus on the case that the cascading failures start with a very small initial failures.

3.4.1 System 1 vs. System 2

We first compare the robustness of System 1 and System 2 in terms of $p_{A_{\infty}}$, i.e., the functioning node fraction after the cascading failures. In Fig. 3.6, we depict $p_{A_{\infty}}$ versus $z_A$ for both System 1 and System 2, whose behaviors are specified in the following.

**System 1**: In Fig. 3.6, the solid curves illustrate how $p_{A_{\infty}}$ evolves with $z_A$ under different $\theta$. For the case with median and high thresholds ($\theta = 0.51$ and $\theta = 0.71$), $p_{A_{\infty}}$ remains above 0.9, which means that the system is robust such that less than 10% of nodes are failed after the cascading failures. In contrast, for $\theta = 0.18$, the corresponding $p_{A_{\infty}}$ quickly deceases to zero as $z_A$ increases, indicating that the system is vulnerable to even a very small initial failures. One related real-world example is the blackout in Italy in 2003.
that an outage at a few power stations finally damaged a large part of the power grid. The curve corresponding to $\theta = 0.18$ further indicates that the cascading failures can only cause limited damage when the physical network is very sparsely connected with low $z_A$. In other words, a sparse network connectivity essentially makes the cascading failures difficult to propagate.

**System 2:** The dashed curve illustrates that $p_{A^\infty}$ monotonically increases with $z_A$ in System 2. We observe that $p_{A^\infty}$ remains at 0 when $z_A$ is below 1.8, which is due to the assumption under the GCC model that all nodes are dysfunctional when the sparsely connected network only consists of isolated components. As we noted in Section 3.2, this assumption is incompatible with practical physical systems. When $z_A$ exceeds 1.8, $p_{A^\infty}$ quickly approaches to 1, where System 2 exhibits a similar behavior as System 1 with median and high thresholds ($\theta = 0.51$ and $\theta = 0.71$). However, System 2 falls short of capturing the case that the cascading failures totally damage a cyber-physical system with weak node resilience (e.g., System 1 with $\theta = 0.18$). In a nutshell, compared to System 1, System 2 gives an incomplete characterization of the robustness of cyber-physical systems.

### 3.4.2 System 1 vs. System 3

We compare the robustness between our CPS model (System 1) and a single physical network (System 3). Particularly, we focus on the case with low threshold $\theta$ in the physical network.$^3$

Fig. 3.7 depicts $p_{A^\infty}$ versus $z_A$ for both systems, where the values of $p_{A^\infty}$ for both systems quickly decrease to 0 as the average degree $z_A$ increases. Nevertheless, two systems exhibit significant discrepancy, when $z_A$

---

$^3$With high $\theta$, it is expected that a single network (System 3) would exhibit strong resilience to the cascading failures, since it is the case in a more vulnerable System 1 in Fig. 3.6. Therefore, Systems 1 and 3 would yield similar behaviors with high $\theta$. 

64
Figure 3.6: $p_{A\infty}$ vs. $z_A$ for System 1 (solid curves) and System 2 (dashed curve), where $z_B = 4$ and $p = 0.999$.

Figure 3.7: $p_{A\infty}$ vs. $z_A$ for System 1 (solid curves) and System 3 (dashed curves), where $z_B = 3$ and $p = 0.999$. 
is sufficiently high. Specifically, System 3 exhibits a sharp increase of $p_{A\infty}$ at $z_{A} = 4.2$ (for the case $\theta = 0.25$) and $z_{A} = 6.3$ (for the case $\theta = 0.18$), indicating that System 3 exhibits strong resilience to cascading failures with high $z_{A}$. Worth noting is that such sharp transition disappears in System 1 and $p_{A\infty}$ still remains at 0, indicating the high vulnerability of our CPS model in high $z_{A}$ region. Based on the results in Figs. 3.6 and 3.7, we summarize the robustness of System 1 in Fig. 3.8.

We believe that this discrepancy can be attributed to the joint effect of the dense connectivity in the physical network and the interdependence in the CPS. Generally speaking, when $\theta$ is low, the dense network connectivity yields a double-edged effect on network robustness, which not only enhances the resilience of each individual node as we mentioned in Section 3.2, but also accelerates the failure propagation when a node fails. We observe that the dense connectivity can gradually become harmful to network resilience as the failed node fraction increases. Typically, even if the initial failures in Fig. 3.7 could not further propagate in a single physical network with high connectivity, they still can be recursively “amplified” between the two networks of
the CPS due to the interdependence therein, which could finally trigger an avalanche of catastrophic damages by taking advantage of the dense connectivity. Therefore, in contrast to a single physical network, the high $z_A$ would make our CPS model more vulnerable to cascading failures.

3.5 Enhancing CPS Resilience with Autonomous Nodes

Due to the network interdependence, the CPS tends to be more vulnerable to node failures than each individual network. One strategy to enhance the system robustness is to reduce the network interdependence by enabling a fraction of nodes to be autonomous in the sense that the nodes can support by themselves. Capitalizing on previous studies on autonomous nodes [43], we investigate this strategy in our CPS model. Specifically, we define the CPS with autonomous nodes as System 4, with $N_A$ and $N_B$ nodes in networks $A$ and $B$, respectively. In System 4, only a fraction $r_A$ of nodes in $A$ and a fraction $r_B$ of nodes in $B$ depend on each other, while the other nodes are autonomous. Note that System 4 degenerates to System 1, when $N_A = N_B = N$ and $r_A = r_B = 1$.

3.5.1 Analysis of CPS Model with Autonomous Nodes

As in Section 3.3, we study the robustness of System 4 by characterizing the functioning node fraction after the cascading failures. Here, we use the same notation as in Section 3.3. Worth noting is that the propagation of the cascading failures hinges heavily on the dependent nodes in both networks. Therefore, at each stage $i$, we need to quantify the “densities” of dependent nodes in networks $A$ and $B$, denoted by $r_{Ai}$ and $r_{Bi}$. The dynamics of the cascading failures in System 4 follow the similar recursive process as in System 1. At Stage 1, the number of functioning nodes in network $A$ can be obtained by
Since only the dependent nodes in network $B$ are impacted by the node failures at Stage 1, we rewrite (3.5) as $|\bar{B}_2| = p'_{B_2}N = (1 - r_B(1 - G_A(p)))N$. Further, at Stage 2, the number of functioning nodes $|B_2|$ can be calculated by (3.6).

As in System 1, we treat the failures in $A_1$ at Stage 3 as an “equivalent” initial failures on network $A$. Recall that in System 1, such failed fraction is $p(1 - G_B(p'))$. It follows that the corresponding fractional size in System 4 is reduced to $r_A p(1 - G_B(p'))$ since only the dependent nodes are impacted. Combining with a fraction $1 - p$ of the failed nodes in Stage 1, we have

$$p'_{A_3} = 1 - (1 - p + r_A p(1 - G_B(p'))))$$

$$= p(1 - r_A(1 - G_B(p'))).$$

We can further obtain the functioning node fraction $p_{A_4}$ by (3.8).

Since a fraction $1 - p'_{B_2}$ of dependent nodes in network $B$ have failed at Stage 2, the dependent node fraction in network $B$ at Stage 4 can be given by

$$r_{B_4} = \frac{r_B - (1 - p'_{B_2})}{p'_{B_2}}$$

$$= \frac{r_B G_A(p)}{1 - r_B(1 - G_A(p))}.$$  

As in System 1, the fragmentation from $\bar{A}_3$ to $A_3$ can result in an equivalent effect as taking out a fraction $r_{B_4} p'_{B_2} (1 - \frac{A_3}{|A_3|})$ of nodes in $B$. Combining with a fraction $1 - p'_{B_2}$ of the failed nodes in Stage 2, we have that

$$p'_{B_4} = 1 - r_B \left(1 - \frac{pG_A(p'_{A_3})}{p'_{A_3}}\right)$$

and the functioning node fraction at Stage 4 can be given by (3.12). As in Section 3.3.1, we have the following results.

68
Proposition 5.1. At stage \(i, i \geq 3\), the functioning node fractions in System 4 can be found by using the following recursive equations:

- For \(i = 3, 5, 7, ...\),
  \[
  p_A_i = G_A(p'_A_i), \quad p_B_i = p_{B_{i-1}},
  \]
  \[
  p'_A_i = p \left(1 - r_A(1 - G_B(p'_{B_{i-1}}))\right), \quad (3.32)
  \]

- For \(i = 4, 6, 8, ...\),
  \[
  p_A_i = p_{A_{i-1}}, \quad p_B_i = p'_{B_i}G_B(p'_{B_i}),
  \]
  \[
  p'_B_i = 1 - r_B + r_Bp\frac{G_A(p'_{A_{i-1}})}{p'_{A_{i-1}}}, \quad (3.33)
  \]

After the cascading failures, the functioning node fractions in networks \(A\) and \(B\) can be given by \(p_A^\infty = G_A(x)\) and \(p_B^\infty = yG_B(y)\), respectively, where \(x\) and \(y\) can be obtained by solving the following system of equations:

\[
\begin{align*}
  x &= p \left(1 - r_A(1 - G_B(y))\right), \quad (3.34) \\
  y &= 1 - r_B + r_Bp\frac{G_A(x)}{x}. \quad (3.35)
\end{align*}
\]

The above system of equations can be solved graphically as in Section 3.3.2. For brevity, we skip the details here.

3.5.2 CPS Resilience with Autonomous Nodes

To get a more concrete sense of analytical results, we compare the robustness of System 1 and System 4. We first consider the case that the cascading failures start with a very small initial failures. In Fig. 3.9, we depict \(p_A^\infty\) versus \(z_A\) for both systems. We observe that \(p_A^\infty\) in System 4 decreases more slowly than that of System 1 as \(z_A\) increases. Similar to System 3, System 4 exhibits a sharp increase of \(p_A^\infty\) at \(z_A = 7.5\), indicating its strong resilience to cascading failures with high \(z_A\). This is due to the low network interdependence in
Figure 3.9: $p_{A\infty}$ vs. $z_A$ for System 1 (solid curves) and System 4 (dashed curves) where $\theta = 0.25$ and $p = 0.999$. For system 4, we set $r_A = 0.4$ and $r_B = 0.4$.

System 4 which makes the cascading failures difficult to propagate between the two networks of CPS.

We next consider a heavy initial attack that destroys a substantial fraction of the physical network. Fig. 3.10 depicts $p_{A\infty}$ versus $p$ for both Systems 1 and 4. It is observed that $p_{A\infty}$ in System 1 suddenly decreases to zero at $p = 0.85$ (for the case $\theta = 0.51$) and $p = 0.73$ (for the case $\theta = 0.71$), indicating that even the physical network has strong node resilience (high $\theta$), System 1 is still highly vulnerable to a heavy attack. On the contrary, $p_{A\infty}$ in System 4 descends smoothly as $p$ decreases, which means that a substantial part of the network is still functioning even if a large fraction of nodes have failed initially under a heavy attack. This “smooth” transition implies that autonomous nodes can significantly enhance the robustness of cyber-physical systems, especially under heavy attacks.
3.6 Conclusions

We studied cascading failures in a cyber-physical system where a cyber network overlays a physical network and the operation of one network depends heavily on the functioning of the other network. Specifically, we used the threshold model to capture the node failures in the physical infrastructure network and the GCC model for that in the cyber network. We showed that the developed CPS model can naturally capture some key features of practical cyber-physical systems which are not captured in the existing studies. Further, our results revealed that the dense connectivity in the physical network would likely make the cyber-physical system more vulnerable to cascading failures. Finally, we developed a strategy to improve the system robustness by enabling a fraction of nodes to be autonomous in the sense that the nodes can support by themselves. We showed that the autonomous nodes can significantly enhance the robustness of cyber-physical systems under heavy attacks. We believe that our findings presented here shed light on the impact of network interdepen-
dence on CPS resilience, which in turn helps future development of protection strategies that would improve the robustness of cyber-physical systems.
Chapter 4

INFORMATION DIFFUSION IN OVERLAYING SOCIAL-PHYSICAL NETWORKS

4.1 Introduction

4.1.1 Motivation and Background

In the previous chapters, we have studied the network interdependence in cyber-physical systems. In fact, such coupling structure can also be observed between different types of social networks. Traditionally, people are tied together in a physical information network through old-fashioned communication media, such as face-to-face interactions. On the other hand, recent advances of Internet and mobile communication technologies have enabled people to be connected more closely through online social networks. Indeed, people can now interact through e-mail or online chatting, or communicate through web sites such as Facebook, Twitter, YouTube, etc. Clearly, the physical information network and online social networks are not completely separate since people may participate in two or more of these networks at the same time. For instance, a person can forward a message to her online friends via Facebook and Twitter upon receiving it from someone via face-to-face communication. As a result, the information spread in one network may trigger the propagation in another network, and may result in a possible cascade of information.

One conjecture is that due to this coupling between the physical and online social networks, today’s breaking news (and information in general) can spread at an unprecedented speed throughout the population, and this is the main subject of the current study. Information cascades over coupled networks can deeply influence the patterns of social behavior. In fact, people have be-
come increasingly aware of the fundamental role of the coupled social-physical network as a medium for the spread of not only information, but also ideas and influence. Twitter has emerged as an ultra-fast source of news and Facebook has attracted major businesses and politicians for advertising products or candidates. Several music groups or singers have gained international fame by uploading videos to YouTube. In almost all cases, a new video uploaded to YouTube, a rumor started in Facebook or Twitter, or a political movement advertised through online social networks, either dies out quickly or reaches a significant proportion of the population. In order to fully understand the extent to which these events happen, it is of great interest to consider the joint behavior of the physical information network and online social networks.

4.1.2 Summary of Main Contributions

We aim to develop a new theoretic framework towards understanding the characteristics of information diffusion across multiple networks. We model the physical network and the online social network as random graphs with different topology properties. We assume that each individual in the population is a member of the physical network, and becomes a member of the social network independently with a certain probability.

The problem under consideration is intricate since the relevant random graph model corresponds to a union of coupled random graphs, and the techniques employed in [27, 28] for single networks fall short of characterizing its phase transition properties. Capitalizing on the recent progress in inhomogeneous random graphs [44, 45], we show that the overlaying social-physical network exhibits a “critical point” above which information epidemics are possible; i.e., a single node can spread an item of information (a rumor, an advertisement, a video, etc.) to a positive fraction of individuals in the asymptotic
limit. Below the critical point, only small information outbreaks can occur and the fraction of influenced individuals always tends to zero.

Specifically, we consider two different models for the individual networks. First, we assume that both the physical information network and the online social network are Erdős-Rényi (ER) graphs [24], and then we consider the case where both networks are generated according to the configuration model [24,28] with specified degree distributions. In each case, we quantify the aforementioned critical point by computing the phase transition threshold of the conjoined random graph model, and show that it depends on both the degree distributions of the networks and the number of individuals that are members of the online social network. Further, we compute the probability that information originating from an arbitrary individual will yield an epidemic along with the resulting fraction of individuals that are influenced; this is done for both cases by computing the giant component size of the corresponding models.

The results show that the conjoint social-physical network can spread an item of information to a significantly larger fraction of the population as compared to the case where the two networks are disjoint. For instance, consider a physical information network $W$ and an online social network $F$ that are ER graphs with respective mean degrees $\lambda_w$ and $\lambda_f$, and assume that each node in $W$ is a member of $F$ independently with probability $\alpha$. If $\lambda_w = 0.6$ and $\alpha = 0.2$, we show that information epidemics are possible in the overlaying social-physical network $H = W \cup F$ whenever $\lambda_f \geq 0.77$. In stark contrast, this happens only if $\lambda_w > 1$ or $\lambda_f > 1$ when the two networks are disjoint. Furthermore, in a single ER network $W$ with $\lambda_w = 1.5$, an information item originating from an arbitrary individual gives rise to an epidemic with prob-
ability 0.58 (i.e., can reach at most 58% of the individuals). However, if the
same network $\mathcal{W}$ is conjoined with an ER network $\mathcal{F}$ with $\alpha = 0.5$ and $\lambda_f = 1.5$,
the probability of an epidemic becomes 0.82 (indicating that up to 82% of the
population can be influenced).

4.2 System Model

We consider the following model for an overlaying social-physical network. Let
$\mathcal{W}$ stand for the physical information network of human beings on the node
set $\mathcal{N} = \{1, \ldots, n\}$. We assume that the graph $\mathcal{W}$ characterizes the possible
spread of information amongst people through old-fashioned communication
media; e.g., face-to-face communication, phone calls, etc. Next, let $\mathcal{F}$ stand
for the network that characterizes the information spread through an online
social networking web site, e.g., Facebook. We assume that each node in $\mathcal{N}$ is
a member of this auxiliary network with probability $\alpha \in (0, 1]$ independently
from any other node. In other words, we have

$$
P[i \in \mathcal{N}_F] = \alpha, \quad i = 1, \ldots, n, \tag{4.1}$$

with $\mathcal{N}_F$ denoting the set of human beings that are members of Facebook.
With this assumption, it is clear that the vertex set $\mathcal{N}_F$ of $\mathcal{F}$ satisfies $\frac{|\mathcal{N}_F|}{n} \xrightarrow{a.s.} \alpha$
by the law of large numbers.

In order to study information diffusion amongst human beings, a key
step is to characterize an overlaying graph $\mathcal{H}$ that is constructed by taking the
union of $\mathcal{W}$ and $\mathcal{F}$. In other words, for any distinct pair of nodes $i, j$, we say
that $i$ and $j$ are adjacent in the network $\mathcal{H}$, denoted $i \sim \mathcal{H} j$, as long as at least
one of the conditions $\{i \sim \mathcal{W} j\}$ or $\{i \sim \mathcal{F} j\}$ holds. This is intuitive since a
node $i$ can forward information to another node $j$ either by using old-fashioned
communication channels (i.e., links in $W$) or by using Facebook (i.e., links in $F$). Of course, for the latter to be possible, both $i$ and $j$ should be Facebook users.

The information is assumed to spread among the population according to the SIR model. In this context, an individual is either susceptible meaning that he/she has not yet received a particular item of information, or infectious meaning that he/she is aware of the information and is capable of spreading it to his/her contacts, or recovered meaning that he/she is no longer spreading the information. As in [27], we assume that an infectious individual $i$ transmits the information to a susceptible contact $j$ with probability $T_{ij} = 1 - e^{-r_{ij} \tau_i}$. Here, $r_{ij}$ denotes the average rate of being in contact over the link from $i$ to $j$, and $\tau_i$ is the time $i$ keeps spreading the information; i.e., the time it takes for $i$ to become recovered.

It is expected that the information propagates over the physical and social networks at different speeds, which manifests from different probabilities $T_{ij}$ across links in this case. Specifically, let $T_{ij}^w$ stand for the probability of information transmission over a link (between and $i$ and $j$) in $W$ and let $T_{ij}^f$ denote the probability of information transmission over a link in $F$. For simplicity, we assume that $T_{ij}^w$ and $T_{ij}^f$ are independent for all distinct pairs $i, j = 1, \ldots, n$. Furthermore, we assume that the random variables $r_{ij}^w$ and $\tau_i^w$ are independent and identically distributed (i.i.d.) with probability densities $P_w(r)$ and $P_w(\tau)$, respectively. In that case, it was shown in [27] that the information propagates over $W$ as if all transmission probabilities were equal to $T_w$, where $T_w$ is the mean value of $T_{ij}^w$; i.e.,

$$T_w = \langle T_{ij}^w \rangle = 1 - \int_0^\infty \int_0^\infty e^{-r \tau} P_w(r)P_w(\tau) dr d\tau.$$
We refer to $T_w$ as the transmissibility of the information over the physical network $\mathbb{W}$ and note that $0 \leq T_w \leq 1$. In the same manner, we assume that $r_{ij}^f$ and $\tau_i^f$ are i.i.d. with respective densities $P_f(r)$ and $P_f(\tau)$ leading to a transmissibility $T_f$ of information over the online social network $\mathbb{F}$ (in most practical scenarios we expect to see that $T_f > T_w$).

Under these assumptions, information diffusion becomes equivalent to (bond) percolation in the conjoint network $\mathbb{H} = \mathbb{W} \cup \mathbb{F}$. More specifically, we say (as in [27]) that each edge in $\mathbb{W}$ is occupied, meaning that it can be used in spreading the information, with probability $T_w$ independently from all other edges. Similarly, each edge in $\mathbb{F}$ is deemed occupied (independently) with probability $T_f$. Then, the size of the information epidemic in $\mathbb{H}$ is equal to the number of individuals that can be reached from an arbitrary node by using only the occupied links of $\mathbb{H}$. Hence, the threshold and the size of the information epidemic can be achieved by studying the phase transition properties in $\mathbb{H}(T_w, T_f)$ where $\mathbb{H}(T_w, T_f)$ is the random graph containing only the occupied edges of $\mathbb{H}$.

4.3 Conjoining Speeds up Information Diffusion in Overlaying Networks

4.3.1 Information Diffusion in Two Coupled ER Graphs

We first consider a basic scenario where both the physical information network $\mathbb{W}$ and the online social network $\mathbb{F}$ are Erdős-Rényi graphs [24]. More specifically, let $\mathbb{W} = \mathbb{W}(n; \lambda_w/n)$ be an ER network on the vertices $\{1, \ldots, n\}$ such that there exists an edge between any pair of distinct nodes $i, j = 1, \ldots, n$ with probability $\lambda_w/n$; this ensures that mean degree of each node is asymptotically equal to $\lambda_w$. Next, obtain a set of vertices $\mathcal{N}_F$ by picking each node $1, \ldots, n$ independently with probability $\alpha \in (0, 1]$. Now, let $\mathbb{F} = \mathbb{F}(n; \alpha, \lambda_f/(\alpha n))$ be
an ER graph on the vertex set \( \mathcal{N}_F \) with edge probability given by \( \frac{\lambda_f}{\alpha n} \). The mean degree of a node in \( \mathcal{F} \) is given (asymptotically) by \( \lambda_f \). Assume further that each edge in \( \mathcal{W} \) (resp. in \( \mathcal{F} \)) is occupied with probability \( T_w \) (resp. with probability \( T_f \)), independently from all other edges. Under these conditions, the online social network \( \mathcal{F} \) is still an ER graph, but with average degree \( T_f \lambda_f \), whereas the physical network \( \mathcal{W} \) is an ER graph with average degree \( T_w \lambda_w \).

The overall system model \( \mathcal{H} \) can now be obtained by conjoining the physical information network \( \mathcal{W} \) and the online social network \( \mathcal{F} \). In other words, \( \mathcal{H} \) is constructed on the vertices \( 1, \ldots, n \) by conjoining the occupied edges of \( \mathcal{W} \) and \( \mathcal{F} \), i.e., we have \( \mathcal{H}(n; \alpha, T_w \lambda_w, T_f \lambda_f) = \mathcal{W}(n; T_w \lambda_w/n) \cup \mathcal{F}(n; \alpha, T_f \lambda_f/(\alpha n)) \).

We now present the first main result that characterizes the critical threshold and the size of the information epidemic in \( \mathcal{H} \).

Let \( \lambda_f^\ast \) be defined by
\[
\lambda_f^\ast = \frac{1}{2} (T_f \lambda_f + T_w \lambda_w) + \frac{1}{2} \sqrt{(T_f \lambda_f + T_w \lambda_w)^2 - 4(1 - \alpha)T_f \lambda_f T_w \lambda_w}. \tag{4.2}
\]

Also, let \( \rho_1 \) be the largest solution of the equation
\[
(1 - \alpha)T_w \lambda_w ((1 - \rho_1)e^{\rho_1 T_f \lambda_f} - 1) - \log(1 - \rho_1) = \rho_1 (T_f \lambda_f + \alpha T_w \lambda_w) \tag{4.3}
\]
with \( \rho_1 \) in \([0, 1]\), and let \( \rho_2 \) be given by
\[
\rho_2 = \frac{-\log(1 - \rho_1) - \rho_1 (\alpha T_w \lambda_w + T_f \lambda_f)}{(1 - \alpha)T_w \lambda_w}. \tag{4.4}
\]

**Theorem 4.3.3.** With the above assumptions, we have that

(i) If \( \lambda_f^\ast \leq 1 \), then whp the size of the largest component satisfies
\[
C_1(\mathcal{H}(n; \alpha, T_w \lambda_w, T_f \lambda_f)) = O(\log n);
\]
in contrast, if \( \lambda_f^\ast > 1 \) we have
\[
C_1(\mathcal{H}(n; \alpha, T_w \lambda_w, T_f \lambda_f)) = \Theta(n)
\]
whp, while the size of the second largest component satisfies

\[ C_2(\mathbb{H}(n; \alpha, T_w \lambda_w, T_f \lambda_f)) = O(\log n). \]

(ii) Moreover, \( \frac{1}{n} C_1(\mathbb{H}(n; \alpha, T_w \lambda_w, T_f \lambda_f)) \xrightarrow{p} \alpha \rho_1 + (1 - \alpha) \rho_2. \)

A proof of Theorem 4.3.3 is given in Section 4.5.1. Theorem 4.3.3 quantifies the number of individuals in the overlaying social-physical network that are likely to receive an item of information which starts spreading from a single individual. Specifically, the “critical point” of the information epidemic is marked by \( \lambda_{fw}^* = 1, \) with the critical threshold \( \lambda_{fw}^* \) given by (4.2). We conclude from Theorem 4.3.3 that for any parameter set that yields \( \lambda_{fw}^* \leq 1 \) (the subcritical regime), the largest number of individuals who receive the information is \( O(\log n) \), meaning that only small (non-epidemic) information outbreaks can take place. On the other hand, if \( \lambda_{fw}^* > 1 \) (the supercritical regime), the information has a positive probability of reaching a linear fraction of the individuals; i.e, information epidemics can occur. In that case, an information item originating from an arbitrary individual gives rise to an information epidemic with probability \( \alpha \rho_1 + (1 - \alpha) \rho_2 \) and reaches a fraction \( \alpha \rho_1 + (1 - \alpha) \rho_2 \) of individuals; here \( \rho_1 \) is given by the largest solution of (4.3) and \( \rho_2 \) is given by (4.4).

We observe that the threshold function \( \lambda_{fw}^* \) is symmetric in \( T_f \lambda_f \) and \( T_w \lambda_w \), meaning that both networks have identical roles in carrying the conjoint network to the supercritical regime where information can reach a linear fraction of the nodes. To get a more concrete sense, we depict in Figure 4.1 the minimum \( \lambda_f T_f \) required to have a giant component in \( \mathbb{H}(n; \alpha, T_w \lambda_w, T_f \lambda_f) \) versus \( \lambda_w T_w \) for various \( \alpha \) values. Each curve in the figure corresponds to a phase
transition boundary above which information epidemics are possible; below the boundary all outbreaks are limited to \( O(\log n) \) nodes. This clearly shows how two networks that are in the subcritical regime can yield an information epidemic when they are conjoined. For instance, we see that for \( \alpha = 0.1 \) it suffices to have \( \lambda_f T_f = \lambda_w T_w = 0.76 \) for the existence of an information epidemic. Yet, if the two networks are disjoint, this would be require [24] either \( \lambda_f T_f > 1 \) or \( \lambda_w T_w > 1 \). See Figure 4.2 for additional simulation results for the giant component sizes.

We elaborate further on Theorem 4.3.3. First, we note from the classical results [24] that ER graphs have a giant component whenever average node degree exceeds one. This is compatible with part (i) of Theorem 4.3.3, since the condition for giant component existence reduces to \( T_f \lambda_f > 1 \) if \( T_w \lambda_w = 0 \) and \( T_w \lambda_w > 1 \) when \( T_f \lambda_f = 0 \). Also, it is easy to see that the random graph \( \mathbb{H} \) will have a giant component whenever \( F \) or \( W \) has a giant component. This fact is also captured by the threshold function (4.2) as we see that \( \lambda_{fw}^\star > 1 \) whenever \( T_f \lambda_f > 1 \) or \( T_w \lambda_w > 1 \). Finally, in the case where \( \alpha = 1 \) (i.e., when everyone in the population is a member of Facebook), the graph \( \mathbb{H} \) reduces to an ER graph with edge probability \( \frac{T_f \lambda_f + T_w \lambda_w - T_f T_w \lambda_f \lambda_w}{n} \) leading to a mean node degree of \( T_f \lambda_f + T_w \lambda_w \) in the asymptotic regime. As expected, for the case \( \alpha = 1 \), Theorem 4.3.3 reduces to classical results for ER graphs as we see that \( \lambda_{fw}^\star = T_f \lambda_f + T_w \lambda_w \) and \( \frac{1}{n} \mathbb{C}_1(\mathbb{H}) \xrightarrow{d} \rho_1 \) where \( \rho_1 \) is the largest solution of \( \rho_1 = 1 - e^{-\rho_1(T_f \lambda_f + T_w \lambda_w)} \).

It is of interest to state whether or not Theorem 4.3.3 can be deduced from the phase transition results for random graphs with arbitrary degree distributions (e.g., see [27, 28, 46]). It is well known [46] that for these graphs
the critical point of the phase transition is given by

$$\frac{E[d_i(d_i - 1)]}{E[d_i]} = 1$$

where \(d_i\) is the degree of an arbitrary node. We next show that this condition is not equivalent (and, indeed is not even a good approximation) to \(\lambda_{fw}^* = 1\). First, note that the degree of an arbitrary node \(i\) in \(H\) follows a Poisson distribution with mean \(T_w \lambda_w\) if \(i \notin N_F\) (which happens with probability \(1 - \alpha\)), and it follows a Poisson distribution with mean \(T_f \lambda_f + T_w \lambda_w - \frac{T_f \lambda_f T_w \lambda_w}{n}\) if \(i \in N_F\) (which happens with probability \(\alpha\)). When \(n\) becomes large this leads to

$$\frac{E[d_i(d_i - 1)]}{E[d_i]} = \frac{\alpha (T_f \lambda_f + T_w \lambda_w)^2 + (1 - \alpha) (T_w \lambda_w)^2}{\alpha T_f \lambda_f + T_w \lambda_w}.$$  \hspace{1cm} (4.5)

It can be seen that the above expression is not equal to \(\lambda_{fw}^*\) given by (4.2). For instance, with \(\alpha = 0.2, T_w \lambda_w = 0.6\) and \(T_f \lambda_f = 0.8\), we have \(\lambda_{fw}^* = 1.03\) while (4.5) yields 0.89 signaling a significant difference between the exact threshold \(\lambda_{fw}^*\) and the mean field approximation given by (4.5). We conclude that the results established above go beyond the classical results for random graphs with arbitrary degree distributions.

### 4.3.2 Information Diffusion in Two Coupled Graphs with Arbitrary Degree Distributions

We now expand the previous result to a more general and in fact more practically relevant class of graphs usually known as random graphs with arbitrary degree distribution [27,28]. In particular, we specify a degree distribution that gives the properly normalized probabilities \(\{p^w_k, k = 0, 1, \ldots\}\) that an arbitrary node in \(\mathbb{W}\) has degree \(k\). Namely, we let each node \(i = 1, \ldots, n\) in \(\mathbb{W} = \mathbb{W}(n; \{p^w_k\})\) have a random degree drawn from the distribution \(\{p^w_k\}\) independently from any other node. Similarly, we assume that the degrees of all
nodes in $\mathcal{F}$ are drawn independently from the distribution $\{p_f^k, k = 0, 1, \ldots \}$; see [27, 28, 45] for details about the construction of random graphs with given degree distributions. Finally, the vertex set of $\mathcal{F} = \mathcal{F}(n; \alpha, \{p_f^k\})$ is obtained in the usual manner by picking each node 1, $\ldots$, $n$ independently with probability $\alpha$. In what follows, we shall assume that the degree distributions are well-behaved in the sense that all moments of arbitrary order are finite.

As in the previous section, let $T_w$ be the information transmissibility (i.e., the mean probability of information transfer between any two nodes) in the physical network $\mathcal{W}$, and let $T_f$ be the information transmissibility in the social network $\mathcal{F}$. Namely, each edge in $\mathcal{W}$ is deemed occupied, meaning that it can be used in spreading the information, independently with probability $T_w$. Similarly, we let each edge in $\mathcal{F}$ be occupied with probability $T_f$ independently from all the other edges. The overall system model can now be obtained by taking a union of the occupied edges of $\mathcal{W}$ and $\mathcal{F}$. That is, we let $\mathbb{H}(n; \alpha, \{p_w^k\}, T_w, \{p_f^k\}, T_f) = \mathcal{W}(n; \{p_w^k\}, T_w) \cup \mathcal{F}(n; \alpha, \{p_f^k\}, T_f)$ be the corresponding social-physical network over which the information diffuses.

We now present the second main result. For notational convenience, let $k_f$ and $k_w$ be random variables independently drawn from the distributions $\{p_f^k\}$ and $\{p_w^k\}$, respectively, and let $< k_f > = \lambda_f$ and $< k_w > = \lambda_w$. Further, assume that $\beta_f$ and $\beta_w$ are given by

$$\beta_f = \frac{< k_f^2 > - \lambda_f}{\lambda_f} \quad \text{and} \quad \beta_w = \frac{< k_w^2 > - \lambda_w}{\lambda_w}, \quad (4.6)$$

and define the threshold function $\sigma_{fw}^*$ by

$$\sigma_{fw}^* = T_f \beta_f + T_w \beta_w + \sqrt{(T_f \beta_f - T_w \beta_w)^2 + 4\alpha T_f T_w \lambda_f \lambda_w}. \quad (4.7)$$

Finally, let $h_1, h_2$ in $(0, 1]$ be given by the pointwise smallest solution of the
recursive equations

\[
\begin{align*}
    h_1 &= \frac{1}{\lambda_f} E[k_f (1 + T_f(h_1 - 1))^{k_f-1}] E[(1 + T_w(h_2 - 1))^{k_w}] \\
    h_2 &= \frac{1}{\lambda_w} E[\alpha (1 + T_f(h_1 - 1))^{k_f} + 1 - \alpha] E[k_w (1 + T_w(h_2 - 1))^{k_w-1}]
\end{align*}
\] (4.8) (4.9)

**Theorem 4.3.4.** Under the assumptions just stated, we have that

(i) If \( \sigma_{fw}^* \leq 1 \), then with high probability we have

\[
C_1 \left( \mathcal{H}(n; \alpha, \{p_k^w\}, T_w, \{p_k^f\}, T_f) \right) = o(n).
\]

On the other hand, if \( \sigma_{fw}^* > 1 \), then

\[
C_1 \left( \mathcal{H}(n; \alpha, \{p_k^w\}, T_w, \{p_k^f\}, T_f) \right) = \Theta(n) \text{ whp.}
\]

(ii) Also,

\[
\frac{1}{n} C_1 \left( \mathcal{H}(n; \alpha, \{p_k^w\}, T_w, \{p_k^f\}, T_f) \right) \xrightarrow{p} 1 - E[\alpha (1 + T_f(h_1 - 1))^{k_f} + 1 - \alpha] \\
\times E[(1 + T_w(h_2 - 1))^{k_w}].
\] (4.10)

A proof of Theorem 4.3.4 is given in Section 4.5.2. Theorem 4.3.4 can be viewed as a counterpart of Theorem 4.3.3. It quantifies the number of individuals in the overlaying social-physical network likely to receive a particular information item when the physical network \( \mathcal{W} \) and the social network \( \mathcal{F} \) have arbitrary degree distributions \( \{p_k^w\} \) and \( \{p_k^f\} \), respectively. Specifically, for \( \{p_k^w\} \) and \( \{p_k^f\} \) with finite moments, Theorem 4.3.4 shows that the critical point of the information epidemic is marked by \( \sigma_{fw}^* = 1 \), with the critical threshold \( \sigma_{fw}^* \) given by (4.7). In other words, for any parameter set that yields \( \sigma_{fw}^* > 1 \) (supercritical regime), an item of information has a positive probability of giving rise to an epidemic; i.e., reaching a linear fraction of
the individuals. In that case, the asymptotic fraction of the individuals who receive the information can be found by first solving the recursive equations (4.8)-(4.9) for the smallest \( h_1, h_2 \) in \((0, 1]\) and then computing the expression given in (4.10). On the other hand, whenever it holds that \( \sigma_{fw}^* \leq 1 \) (subcritical regime), we conclude from Theorem 4.3.4 that the largest number of individuals who receive the information will be \( o(n) \) with high probability, meaning that all outbreaks are non-epidemic.

We have some further remarks on the applicability of Theorem 4.3.4 and Theorem 4.3.3. Consider the case where both \( \mathbb{W} \) and \( \mathbb{F} \) are ER graphs; i.e., let \( p_k^w = e^{-\lambda_w \frac{k^w}{\lambda_w}} \) and \( p_k^f = e^{-\lambda_f \frac{k^f}{\lambda_f}} \). We have that \( \beta_f = \lambda_f, \beta_w = \lambda_w \), and it is easy to check that \( \sigma_{fw}^* = \lambda_{fw}^* \) so that part (i) of Theorem 4.3.4 is compatible with part (i) of Theorem 4.3.3. Also, we find (numerically) that the second parts of Theorems 4.3.3 and 4.3.4 yield the same asymptotic giant component size. Nevertheless, it is worth noting that, although ER graphs constitute a special case of random graphs with arbitrary degree distributions, Theorem 4.3.3 is not a corollary of Theorem 4.3.4. This is because, through a different technique used in the proofs, Theorem 4.3.3 provides sharper bounds \( C_1(H(n; \alpha, T_w \lambda_w, T_f \lambda_f)) = O(\log n) \) (subcritical case) and \( C_2(H(n; \alpha, T_w \lambda_w, T_f \lambda_f)) = O(\log n) \) (supercritical case) that go beyond Theorem 4.3.4.

### 4.3.3 Numerical Results

In order to gain more insight on the consequences of Theorem 4.3.4, we now consider a specific example of information diffusion over the physical information network \( \mathbb{W} \) and the online social network \( \mathbb{F} \) with degree distributions \( \{p^w_k\} \) and \( \{p^f_k\} \), respectively. Here, we use power-law distributions with exponential
cutoff. Specifically, we let $p_0^w = p_0^f = 0$, and

$$p_k^w = \left( \text{Li}_{\gamma_w}(e^{-1/\Gamma_w}) \right)^{-1} k^{-\gamma_w} e^{-k/\Gamma_w}, \quad k = 1, 2, \ldots \quad (4.11)$$

$$p_k^f = \left( \text{Li}_{\gamma_f}(e^{-1/\Gamma_f}) \right)^{-1} k^{-\gamma_f} e^{-k/\Gamma_f}, \quad k = 1, 2, \ldots \quad (4.12)$$

where $\gamma_w, \gamma_f, \Gamma_w$ and $\Gamma_f$ are positive constants and the normalizing constant $\text{Li}_m(z)$ is the $m$th polylogarithm of $z$; i.e., $\text{Li}_m(z) = \sum_{k=1}^{\infty} \frac{z^k}{k^m}$. Power law distributions with exponential cutoff are chosen here because they are applied to a variety of real-world networks (e.g., see [47] for a detailed empirical study on the degree distributions of real-world networks). Moreover, the distributions (4.11)-(4.12) ensure that all moments of arbitrary order are finite as required by Theorem 4.3.4.

In view of Theorem 4.3.4 (and the epidemic threshold given by (4.7)) we depict in Figure 4.3 the minimum $T_f$ value required to have a giant component in $\mathbb{H}(n; \alpha, \{p_k^w\}, T_w, \{p_k^f\}, T_f)$ versus $T_w$, for various $\alpha$ values. In other words, each curve in the figure corresponds to a phase transition boundary above which information epidemics are possible, in the sense that an information has a positive probability of reaching out to a linear fraction of individuals in the network.

Figure 4.3 illustrates (in the arbitrary distribution case) how conjoining two networks can speed up the information diffusion. It can be seen that even for small $\alpha$ values, two networks, albeit having no giant component individually, can yield an information epidemic when they are conjoined. As an example, we see that for $\alpha = 0.1$, it suffices to have that $\beta_f T_f = \beta_w T_w = 0.774$ for the existence of an information epidemic in the conjoint network $\mathbb{H}$, whereas if the networks $\mathbb{W}$ and $\mathbb{F}$ are disjoint, an information epidemic can occur only if $\beta_w T_w > 1$ or $\beta_f T_f > 1$ [27].

86
We also compute the corresponding giant component sizes via Theorem 4.3.4. Figure 4.4 depicts the fractional size of the giant component in $\mathbb{H}(n; \alpha, \{p_k^w\}, T_w, \{p_k^f\}, T_f)$ versus $T_f \beta_f = T_w \beta_w$, for various $\alpha$ values. In other words, the plots stand for the largest fraction of individuals in the social-physical network who can receive an information item that has started spreading from a single individual. In Figure 4.4, the curves were obtained analytically via Theorem 4.3.4 whereas the marked points stand for the experimental results obtained with $n = 20,000$ nodes by averaging over 200 experiment results for each parameter set. We see that there is good agreement between theory and experiment even for such a small number of vertices; the discrepancy near the phase transition is clearly due to the finite size effect. Moreover, according to the experiments, the critical threshold for the existence of a giant component (i.e., an information epidemic) appears at $T_f \beta_f = T_w \beta_w = 0.78$ when $\alpha = 0.1$, $T_f \beta_f = T_w \beta_w = 0.61$ when $\alpha = 0.5$, and $T_f \beta_f = T_w \beta_w = 0.53$ when $\alpha = 0.9$. These values are in perfect agreement with the theoretically obtained critical threshold $\sigma_{f,w}^*$ given in (4.7).

4.4 Conclusions

We characterized the critical threshold and the asymptotic size of information epidemics in an overlaying social-physical network. To capture the spread of information, we considered a physical information network that characterizes the face-to-face interactions of human beings, and some overlaying online social networks (e.g., Facebook, Twitter, etc.) that are defined on a subset of the population. Assuming that information is transmitted between individuals according to the SIR model, we have shown that the critical point and the size of information epidemics on this overlaying social-physical network can be
Figure 4.1: The minimum $\lambda_f T_f$ required for existence of a giant component in $\mathbb{H}(n; \alpha, T_w \lambda_w, T_f \lambda_f)$ versus $\lambda_w T_w$ for various $\alpha$ values. In other words, each curve corresponds to the boundary of the phase transition for the corresponding $\alpha$ value. Above the boundary there exists a giant component, but below it all components have $O(\log n)$ nodes.

precisely determined by employing the approaches on inhomogeneous random graphs. We believe that our findings here shed light on the further studies on information (and influence) propagation across social-physical networks.

4.5 Appendix

4.5.1 Proof of Theorem 4.3.3

Inhomogeneous Random Graphs: We first summarize the technical tools that will be used in the proof of Theorem 4.3.3. Recently, Bollobas, Janson and Riordan [44] have developed a new theory of inhomogeneous random graphs that would allow studying a very broad class of complex networks rigorously. Here, we summarize their tools with focus on the results used in this study.

At the outset, assume that a graph is defined on vertices $\{1, \ldots, n\}$, where each vertex $i$ is assigned randomly or deterministically a point $x_i$ in a metric
Figure 4.2: The fractional size of the giant component in $\mathbb{H}(n; \alpha, T_w \lambda_w, T_f \lambda_f)$ versus $T_f \lambda_f = T_w \lambda_w$. The curves correspond to analytical results obtained from Theorem 4.3.3, whereas marked points stand for the experimental results obtained with $n = 20,000$ by averaging over 200 experiment results for each parameter set. There is good agreement between the theoretical and experimental results; the small discrepancy in the subcritical regime is clearly due to the finite size effect.

Consider random graphs $G^V(n, \kappa)$ for which the kernel $\kappa$ is bounded and continuous a.e. on $S \times S$. In fact, in this study it suffices to consider only
Figure 4.3: The minimum $T_f$ required for the existence of a giant component in $H(n; \alpha, \{p^w_k\}, T_w, \{p^f_k\}, T_f)$ versus $T_w$. The distributions $\{p^w_k\}$ and $\{p^f_k\}$ are given by (4.11) and (4.12), with $\gamma_f = \gamma_w = 2.5$ and $\Gamma_f = \Gamma_w = 10$. The $T_f$ and $T_w$ values are multiplied by the corresponding $\beta_f$ and $\beta_w$ values to provide a fair comparison with the disjoint network case where it is required [27] to have $\beta_w T_w > 1$ (or $\beta_f T_f > 1$) for the existence of an epidemic; under the current setting we have $\beta_f = \beta_w = 1.545$.

the cases where the metric space $S$ consists of finitely many points, i.e., $S = \{1, \ldots, r\}$; this special case is equivalent to the model studied by Söderberg [48]. Under these assumptions, the kernel $\kappa$ reduces to an $r \times r$ matrix, and $G^V(n, \kappa)$ becomes a random graph with vertices of $r$ different types; e.g., vertices with/without Facebook membership, etc. Two nodes (in $G^V(n, \kappa)$) of type $i$ and $j$ are joined by an edge with probability $n^{-1}\kappa(i,j)$ and the condition (4.13) reduces to

$$\frac{n_i}{n} \xrightarrow{p} \mu_i, \quad i = 1, \ldots, r,$$  \hspace{1cm} (4.14)

where $n_i$ stands for the number of nodes of type $i$ and $\mu_i$ is equal to $\mu(\{i\})$.

As usual, the phase transition properties of $G^V(n, \kappa)$ can be studied by exploiting a related branching process. The survival probability $\rho(\kappa)$ of the
Figure 4.4: The fractional size of the giant component in $\mathbb{H}(n; \alpha, \{p^w_k\}, T_w, \{p^f_k\}, T_f)$ versus $T_f \beta_f = T_w \beta_w$. The distributions $\{p^w_k\}$ and $\{p^f_k\}$ are given by (4.11) and (4.12), with $\gamma_f = \gamma_w = 2.5$ and $\Gamma_f = \Gamma_w = 10$. The $T_f$ and $T_w$ values are multiplied by the corresponding $\beta_f$ and $\beta_w$ values for fair comparison with the disjoint network case; under the current setting we have $\beta_f = \beta_w = 1.545$. The curves were obtained analytically via Theorem 4.3.4, whereas the marked points stand for the experimental results obtained with $n = 20,000$ by averaging over 200 experiment results for each parameter set. We see that there is good agreement between theory and experiment even for such a small number of vertices; the small discrepancy in the subcritical regime is attributed to the finite size effect.

branching process is then given by

$$\rho(\kappa) = \sum_{i=1}^{r} \rho(\kappa; i) \mu_i. \quad (4.15)$$

In analogy with the classical results for ER graphs [24], it can be shown [44,48] that $\rho(\kappa; i), i = 1, \ldots, r$ satisfy the recursive equations

$$\rho(\kappa; i) = 1 - \exp \left\{ - \sum_{j=1}^{r} \kappa(i, j) \mu_j \cdot \rho(\kappa; j) \right\}, \quad i = 1, \ldots, r. \quad (4.16)$$

The value of $\rho(\kappa)$ can be computed via (4.15) by characterizing the stable fixed point of (4.16) reached from the starting point $\rho(\kappa; 1) = \cdots = \rho(\kappa; r) = 0$. It is a simple matter to check that, with $M$ denoting an $r \times r$ matrix given by
\( M(i, j) = \kappa(i, j) \cdot \mu_j \), the iterated map (4.16) has a non-trivial solution (i.e., a solution other than \( \rho(\kappa; 1) = \cdots = \rho(\kappa; r) = 0 \)) if and only if

\[
\sigma(M) = \max\{|\lambda_i| : \lambda_i \text{ is an eigenvalue of } M\} > 1. \tag{4.17}
\]

For a square matrix \( M \), its largest eigenvalue in absolute value, \( \sigma(M) \), defines its \textit{spectral radius}. Thus, we see that if the spectral radius of \( M \) is less than or equal to one, the branching process is subcritical with \( \rho(\kappa) = 0 \) and the graph \( G^V(n, \kappa) \) has no giant component; i.e., we have that \( C_1(G^V(n, \kappa)) = o(n) \) whp. On the other hand, if \( \sigma(M) > 1 \), then the branching process is supercritical and there is a non-trivial solution \( \rho(\kappa; i) > 0, i = 1, \ldots, r \) that corresponds to a stable fixed point of (4.16). In that case, \( \rho(\kappa) > 0 \) corresponds to the probability that an arbitrary node belongs to the giant component, which asymptotically contains a fraction \( \rho(\kappa) \) of the vertices. In other words, if \( \sigma(M) > 1 \), we have that \( C_1(G^V(n, \kappa)) = \Omega(n) \) whp, and \( \frac{1}{n}C_1(G^V(n, \kappa)) \xrightarrow{p} \rho(\kappa) \).

Bollobás et al. [44, Theorem 3.12] have shown that the bound \( C_1(G^V(n, \kappa)) = o(n) \) in the subcritical case can be improved under some additional conditions: They established that whenever \( \sup_{i,j} \kappa(i, j) < \infty \) and \( \sigma(M) \leq 1 \), then \( C_1(G^V(n, \kappa)) = O(\log n) \) whp as in the case of ER graphs. They have also shown that if either \( \sup_{i,j} \kappa(i, j) < \infty \) or \( \inf_{i,j} \kappa(i, j) > 0 \), then in the supercritical regime (i.e., when \( \sigma(M) > 1 \)) the second largest component satisfies \( C_2(G^V(n, \kappa)) = O(\log n) \) whp.

A Proof of Theorem 4.3.3: First assume that \( T_w = T_f = 1 \). We will study the phase transition in \( \mathcal{H} = \mathcal{H}(n; \alpha, \lambda_w, \lambda_f) \) by using the techniques summarized in the previous section. Let \( S = \{1, 2\} \) stand for the space of vertex types, where vertices with Facebook membership are referred to as
type 1 while vertices without Facebook membership are said to be of type 2. Assume that the metric space $S$ is equipped with a probability measure $\mu$ that satisfies the condition (4.14); i.e., $\mu(\{1\}) = \mu_1 = \alpha$ and $\mu(\{2\}) = \mu_2 = 1 - \alpha$. We can compute the appropriate kernel $\kappa$ such that, for each $i,j = \{1,2\}$, $\kappa(i,j)/n$ gives the probability that two vertices of type $i$ and $j$ are connected. Clearly, we have $\kappa(1,1) = n \left( 1 - \left( 1 - \frac{\lambda_w}{n} \right) \left( 1 - \frac{\lambda_f}{\alpha n} \right) \right) = \lambda_w + \frac{\lambda_f}{\alpha} - \frac{\lambda_w \lambda_f}{\alpha n}$, whereas $\kappa(1,2) = \kappa(2,1) = \kappa(2,2) = \lambda_w$.

We are now in a position to derive the critical point of the phase transition in $\mathbb{H}(n; \alpha, \lambda_w, \lambda_f)$. First, we compute the matrix $M(i,j) = \kappa(i,j)\mu_j$ and get

$$M = \begin{bmatrix} \alpha \lambda_w + \lambda_f - \frac{\lambda_w \lambda_f}{n} & (1 - \alpha) \lambda_w \\ \alpha \lambda_w & (1 - \alpha) \lambda_w \end{bmatrix}$$

It is a simple matter to check that the spectral radius of $M$ is given by

$$\sigma(M) = \frac{\lambda_f + \lambda_w - \frac{\lambda_f \lambda_w}{n}}{2} + \frac{1}{2} \sqrt{ \left( \lambda_f + \lambda_w - \frac{\lambda_f \lambda_w}{n} \right)^2 - 4(1 - \alpha) \lambda_f \lambda_w } \quad (4.18)$$

In fact, it is clear that the term $\frac{\lambda_w \lambda_f}{n}$ has no effect on the results as we eventually let $n$ go to infinity. This leads to the conclusion that the random graph $\mathbb{H}(n; \alpha, \lambda_w, \lambda_f)$ has a giant component if and only if

$$\frac{\lambda_f + \lambda_w + \sqrt{(\lambda_f + \lambda_w)^2 - 4(1 - \alpha) \lambda_f \lambda_w}}{2} > 1 \quad (4.19)$$

as we recall (4.17). If condition (4.19) is not satisfied, then we have $C_1(\mathbb{H}(n; \alpha, \lambda_w, \lambda_f)) = O(\log n)$ as we note that $\sup_{i,j} \kappa(i,j) < \infty$. From [44, Theorem 3.12], we also get that $C_2(\mathbb{H}(n; \alpha, \lambda_w, \lambda_f)) = O(\log n)$ whenever (4.19) is satisfied.

Next, we compute the size of the giant component whenever it exists. Let $\rho(\kappa; 1) = \rho_1$ and $\rho(\kappa; 2) = \rho_2$. In view of (4.15) and the arguments presented in the previous section, the asymptotic fraction of nodes in the
giant component is given by

$$\rho(\kappa) = \alpha \rho_1 + (1 - \alpha) \rho_2,$$  \hspace{1cm} (4.20)

where $\rho_1$ and $\rho_2$ constitute a stable simultaneous solution of the transcendental equations

$$\rho_1 = 1 - \exp\{-\rho_1(\alpha \lambda_w + \lambda_f) - \rho_2(1 - \alpha)\lambda_w\}$$

$$\rho_2 = 1 - \exp\{-\rho_1 \alpha \lambda_w - \rho_2(1 - \alpha)\lambda_w\}$$  \hspace{1cm} (4.21)

By easy algebra, we see that $\rho_1$ is given by the largest solution of the equation

$$(1 - \alpha)\lambda_w ((1 - \rho_1)e^{\rho_1\lambda_f} - 1) - \log(1 - \rho_1) = \rho_1(\lambda_f + \alpha \lambda_w)$$  \hspace{1cm} (4.22)

on the interval $0 \leq \rho_1 \leq 1$, while $\rho_2$ can be computed via

$$\rho_2 = \frac{-\log(1 - \rho_1) - \rho_1(\alpha \lambda_w + \lambda_f)}{(1 - \alpha)\lambda_w}.$$  \hspace{1cm} (4.23)

So far, we have established the epidemic threshold and the size of the information epidemic when $T_w = T_f = 1$. In the more general case where there is no constraint on the transmissibilities, we see that the online social network $\mathbb{F}$ becomes an ER graph with average degree $T_f \lambda_f$, whereas the physical network $\mathbb{W}$ becomes an ER graph with average degree $T_w \lambda_w$. Therefore, the critical threshold and the size of the information epidemic can be found by substituting $T_f \lambda_f$ for $\lambda_f$ and $T_w \lambda_w$ for $\lambda_w$ in the relations (4.19), (4.21), (4.22), and (4.23). This establishes Theorem 4.3.3.

4.5.2 Proof of Theorem 4.3.4

The approach outlined in Section 4.5.1 allows one to distinguish between different types of vertices in the random graph $\mathbb{H}$ (e.g., vertices with and without Facebook memberships), and assign probabilities to edges according to the types of vertices that they are connecting together. This leads to a mixed
Poisson distribution for the degree of a randomly chosen vertex in \( H \). However, in the case where the underlying random graphs \( \mathbb{W} \) and \( \mathbb{F} \) have arbitrary degree distributions, it is more useful to differentiate between different types of edges, i.e., edges in Facebook and edges in the physical network. In fact, Söderberg has followed this approach and studied the phase transition in the so-called “colored degree-driven random graphs” [45]. Here, we give a simplified version of their model and summarize the main results.

Let \( \{1, \ldots, r\} \) be the possible types of edges in the graph. The colored degree of a node \( i \) is then represented by an integer vector \( d_i = [d_i^1 \cdots d_i^r] \), where \( d_i^j \) stands for the number of edges of type \( j \) that are incident on node \( i \). It is assumed that the colored degrees (i.e., \( d_1, \ldots, d_n \)) are drawn independently from a colored degree distribution \( \{p_m\} \) such that for any \( i \)

\[
P[d_i^j] = m_j = p_m
\]  

(4.24)

whenever \( m = (m_1, \ldots, m_r) \). Given that the colored degrees are picked such that \( \sum_{i=1}^n d_i^j \) is even for each \( j = 1, \ldots, r \), we construct the graph as in [27]. To do this, each node \( i = 1, \ldots, n \) is first given the appropriate number \( d_i^j \) of stubs of type \( j \) for each \( j = 1, \ldots, r \). Then, pairs of these stubs that are of the same type are picked randomly and connected together to form complete edges; in this study we assume that two stubs can be connected together only if they are of the same type. Pairing of stubs continues until none are left. It is also assumed that the colored degree distribution \( \{p_m\} \) has finite moments of arbitrary order.

This random stub-pairing mechanism will be encoded through an \( r \times r \) color preference matrix \( C \) where \( C(i, j)/n \) is equal to the asymptotic probability that two arbitrary stubs of types \( i \) and \( j \) are connected. Here, this amounts
to setting $C(i,j) = \frac{1}{<d_j>}$ if $i = j$ and $C(i,j) = 0$ if $i \neq j$, where $<d_j>$ is the expected number of edges of type $j$ for an arbitrary node. Let $G(n,p_m,C)$ define the random graph constructed in the manner outlined above. In order to study the phase transition in $G(n,p_m,C)$, we use generating function of the colored degree distribution \{p_m\}. This distribution can be represented by a multivariate generating function, $H(x) = \sum_m p_m x^m$, where $x^m = \prod_{i=1}^r x_i^{m_i}$. The multivariate combinatorial moments can be derived by repeated differentiation at $x = 1$, e.g. $\partial_i H(x = 1) = <d_i>$, $\partial_i \partial_j H(x = 1) = <d_i d_j - d_i \delta_{ij}>$, etc.

Now, let $P_k$ ($0 \leq k < \infty$) denote the size distribution of the largest connected component that can be reached from a randomly chosen initial vertex in $G(n,p_m,C)$. Also, let $g(z)$ denote the generating function of $P_k$; i.e., $g(z) = \sum_k P_k z^k$. We can express $g(z)$ in terms of $h(z) = [h_1(z) \cdots h_r(z)]$, where $h_i(z)$ stands for the generating function of the size distribution of the component reached by following a stub of type $i$. In fact, it was shown [45] that

$$g(z) = z \sum_m p_m \prod_{i=1}^r h_i(z)^{m_i} = z H(h(z)),$$

while $h(z)$ satisfy the self-consistency equations

$$h_i(z) = z \sum_{j=1}^r C(i,j) \partial_j H(h(z)) = \frac{z}{<d_i>} \partial_i H(h(z)), \quad i = 1, \ldots, r,$$

where $\partial_i$ denotes the partial derivative with respect to $h_i$, for each $i = 1, \ldots, r$.

We are interested in the solution of the recursive relations (4.26) for the case $z = 1$. This case exhibits a trivial fixed point $h(1) = 1$ which yields $g(1) = 1$ meaning that the underlying branching process is in the subcritical regime and that all components have finite size as understood from the conservation of probability. However, the fixed point $h(1) = 1$ corresponds to the physical
solution only if it is an *attractor*; i.e., a stable solution to the recursion (4.26).

The stability of this fixed point can be checked via linearization of (4.26) around \( h(1) = 1 \), which (upon taking the derivative in (4.26)) yields the Jacobian \( J \) given by

\[
J(i, j) = \frac{1}{\langle d_i \rangle} \partial_i \partial_j H(h)|_{h=1} := \frac{\langle d_i d_j - d_i \delta_{ij} \rangle}{\langle d_i \rangle}, \quad i, j = 1, \ldots, r. \tag{4.27}
\]

If all the eigenvalues of \( J \) are less than one in absolute value, then the solution \( h(1) = 1 \) is an attractor and \( g(1) = 1 \) becomes the physical solution, meaning that \( G(n, p_m, C) \) does not possess a giant component whp. On the other hand, if the spectral radius of \( J \), \( \sigma(J) \) is larger than one, then the fixed point \( h(1) = 1 \) is unstable pointing out that the asymptotic branching process is supercritical, with a positive probability of producing infinite trees. In that case, a nontrivial fixed point exists and becomes the attractor of the recursions (4.26), yielding a solution with \( h_i(1) < 1, i = 1, \ldots, r \). In view of (4.25) this implies \( g(1) < 1 \) and the corresponding probability deficit \( 1 - g(1) \) is attributed to the existence of a giant component. In fact, the quantity \( 1 - g(1) \) is equal to the probability that a randomly chosen vertex belongs to the giant component, which contains asymptotically a fraction \( 1 - g(1) \) of the vertices.

Consider random graphs \( \mathbb{W}(n, \{p^w_k\}) \) and \( \mathbb{F}(n; \alpha, \{p^f_k\}) \) as in Section 4.3. In order to study the phase transition in \( \mathbb{H} = \mathbb{W} \cup \mathbb{F} \), we use the tools just outlined. As before, we start with the case where \( \mathcal{T}_f = \mathcal{T}_w = 1 \). To adopt the notation, let \( \{1, 2\} \) be the set of edge types in \( \mathbb{H} \) where type 1 stands for the edges of \( \mathbb{F} \) and type 2 refers to the edges in \( \mathbb{W} \). The colored degree of node \( i \) is given by \( d_i = [d^f_i \ d^w_i] \), where \( d^f_i \) and \( d^w_i \) are the number of edges incident upon \( i \) in the graphs \( \mathbb{F} \) and \( \mathbb{W} \), respectively. Of course, if \( i \) is not a member of the vertex set \( \mathcal{N}_F \) of \( \mathbb{F} \) (which happens with probability \( 1 - \alpha \)), \( d^f_i \) becomes
automatically zero. When considering the colored degrees of an arbitrary node in $H$, we write $d_f$ (for the number of edges in $F$) and $d_w$ (for the number of edges in $W$) for convenience. Throughout, we shall distinguish between the random variables $k_f$ and $d_f$, where $k_f$ is a random variable drawn from the distribution $\{p_f^k\}$, while $d_f$ is statistically equivalent to $k_f$ with probability $\alpha$, and equal to zero otherwise. For notational convenience, we also define $k_w$ in an analogous manner, but note that $k_w \equiv d_w$ since each node in $H$ is automatically a member of the vertex set of $W$.

We now determine the appropriate colored degree distribution $\{p_m\}$. First, observe that $p(i,j) = P[d_f = i, d_w = j] = P[d_f = i]P[d_w = j]$ due to independence of $W$ and $F$. This yields that

$$p(i,j) = \left(\alpha p_f^i + (1 - \alpha)p_w^j\right)$$

as we recall (4.1) and the fact that a node can have $i > 0$ edges in $F$ only if it belongs to $N_F$.

Let $<d_f>$ be the mean number of the online social network (i.e., type 1) edges for an arbitrary node; i.e., $<d_f> = \sum_{i=0}^{\infty} iP[d_f = i]$. Similarly, let $<d_w>$ be the mean number of physical connections (i.e., type 2 edges) for an arbitrary node. In view of (4.28), we have $<d_f> = \alpha \lambda_f$ and $<d_w> = \lambda_w$, where $\lambda_f = <k_f> = \sum_{k=0}^{\infty} kp_f^k$ and $\lambda_w = <k_w> = \sum_{k=0}^{\infty} kp_w^k$ are mean node degrees in $F$ and $W$ respectively. Next, let $<d_f^2>$ and $<d_w^2>$ denote the second moments of the number of Facebook and physical connections of an arbitrary node, respectively; i.e, let $<d_f^2> = \sum_{i=0}^{\infty} i^2P[d_f = i]$ and $<d_w^2> = \sum_{i=0}^{\infty} i^2P[d_w = i]$. In view of these, the Jacobian $J$ defined in (4.27)
is given by

\[
J = \begin{bmatrix}
\frac{<d_f^2> - \alpha \lambda f}{\alpha \lambda f} & \frac{\lambda w}{\lambda w} \\
\frac{<d_w^2> - \lambda w}{\alpha \lambda f} & \frac{<d_f^2> - \lambda w}{\lambda w}
\end{bmatrix} = \begin{bmatrix}
\frac{<k_f^2> - \lambda f}{\lambda f} & \frac{\lambda w}{\lambda w} \\
\frac{<k_w^2> - \lambda w}{\lambda w} & \frac{<k_f^2> - \lambda w}{\lambda w}
\end{bmatrix}
\]
we have
\[ \frac{1}{n} C \left( \mathbb{H}(n; \alpha, \{p^w_k\}, \{p^f_k\}) \right) \xrightarrow{p} 1 - E[\alpha h_1^{k_f} + 1 - \alpha]E[h_2^{k_w}] \] (4.31)

The above results reveal the epidemic threshold and the size of the information epidemic when \( T_f = T_w = 1 \). We now consider the more general case where there is no such constraint, i.e., the case where \( T_f, T_w < 1 \). Let \( \{\tilde{p}^w_k\} \) and \( \{\tilde{p}^f_k\} \) be the occupied degree distributions obtained from the original distributions \( \{p^w_k\} \) and \( \{p^f_k\} \) by deleting each edge with probability \( 1 - T_w \) and \( 1 - T_f \), respectively. For each \( k = 0, 1, \ldots, \) we can compute \( \tilde{p}^w_k \) and \( \tilde{p}^f_k \) from the generating functions of the distributions \( \{p^w_k\} \) and \( \{p^f_k\} \). First let \( G_w(x) \) and \( G_f(x) \) be the respective generating functions of \( \{p^w_k\} \) and \( \{p^f_k\} \); i.e., \( G_w(x) = \sum_{k=0}^{\infty} p^w_k x^k \) and \( G_f(x) = \sum_{k=0}^{\infty} p^f_k x^k \). Similarly, let \( \tilde{G}_w(x) \) and \( \tilde{G}_f(x) \) be defined by \( \tilde{G}_w(x) = \sum_{k=0}^{\infty} \tilde{p}^w_k x^k \) and \( \tilde{G}_f(x) = \sum_{k=0}^{\infty} \tilde{p}^f_k x^k \). It is a simple matter to check that [27], we have \( \tilde{G}_w(x) = G_w(1 - (x - 1)T_w) \) and \( \tilde{G}_f(x) = G_f(1 + (x - 1)T_f) \).

In view of these, we now derive the critical point of the phase transition for arbitrary \( 0 \leq T_f, T_w \leq 1 \). Let \( \tilde{k}_f, \tilde{k}_w \) be random variables drawn from the distributions \( \{\tilde{p}^w_k\} \) and \( \{\tilde{p}^f_k\} \), respectively. Furthermore, we let \( \tilde{\lambda}_w \) and \( \tilde{\lambda}_f \) define the corresponding mean values, while as before we use \( \tilde{\beta}_f = \frac{<k^2_f> - \tilde{\lambda}_f}{\tilde{\lambda}_f} \) and \( \tilde{\beta}_w = \frac{<k^2_w> - \tilde{\lambda}_w}{\tilde{\lambda}_w} \) \( \text{.} \) We find \( \tilde{\lambda}_f = \tilde{G}''_f(1) = T_f G''_f(1) = T_f \lambda_f \) and \( < \tilde{k}_f^2 > - \tilde{\lambda}_f = \tilde{G}''_f(1) = T_f^2 G''_f(1) = T_f^2 (< k^2_f > - \lambda_f) \) so that \( \tilde{\beta}_f = T_f \beta_f \). Similarly, we have \( \tilde{\lambda}_w = T_w \lambda_w \) and \( \tilde{\beta}_w = T_w \beta_w \). We then conclude that the spectral radius of the corresponding matrix \( \tilde{J} \) is given by (4.29) with \( \beta_f \) replaced by \( T_f \beta_f \), \( \beta_w \) replaced by \( T_w \beta_w \), \( \lambda_f \) replaced by \( T_f \lambda_f \), and \( \lambda_w \) replaced by \( T_w \lambda_w \). As a result, the critical point of the phase transition in \( \mathbb{H}(n; \alpha, \{p^w_k\}, T_w, \{p^f_k\}, T_f) \) is given by \( \sigma_{fw}^* = 1 \) with \( \sigma_{fw}^* \) defined as in (4.7).
Next, we compute the size of the giant component. It is clear that the relations (4.30)-(4.31) are valid with $k_f, k_w, \lambda_f$ and $\lambda_w$ are replaced by $\tilde{k}_f, \tilde{k}_w, T_f \lambda_f$, and $T_w \lambda_w$, respectively. Furthermore, we see that $E[h_1^{\tilde{k}_f}] = \tilde{G}_f(h_1) = G_f(1+(h_1-1)T_f) = E[(1 + T_f(h_1 - 1))^{\tilde{k}_f}]$ and similarly $E[k_f h_1^{\tilde{k}_f-1}] = \tilde{G}_f'(h_1) = T_f G_f'(1 + (h_1 - 1)T_f) = T_f E[k_f(1 + T_f(h_1 - 1))^{\tilde{k}_f-1}]$. In the same manner, we can compute $E[h_1^{\tilde{k}_w}]$ and $E[\tilde{k}_w h_1^{\tilde{k}_w-1}]$ in terms of $k_w$ and $T_w$. Reporting these expressions into (4.30)-(4.31), we see that, with $h_1, h_2$ in $(0, 1]$ corresponding to the pointwise smallest solution of the recursive equations (4.8)-(4.9), the desired condition (4.10) holds. This establishes part $(ii)$ of Theorem 4.3.4 and the proof is now complete.
5.1 Introduction

5.1.1 Motivation and Background

In the previous chapter, we have shown that the coupling between the social-physical network could greatly facilitate the diffusion of information. In this chapter, we turn our attention to the case of real-time information. Typically, the real-time information is valuable only for a limited time duration [49] and hence needs to be delivered before its deadline. For example, once a time-limited coupon is released from Groupon or Dealsea.com, people can share this news either by talking to friends or posting it on Facebook. However, the interest on this deal would die down after it expires.

Clearly, due to the timeliness requirement, the influence of real-time information depends on its propagation speed. The faster the message passes from one to another, the more people can learn this news before it expires, indicating that its diffusion behavior hinges heavily on how fast the message can spread along different social connections.

In this study, we assume that information could spread amongst people through both face-to-face contacts and online communications. Observe that the efficiency of face-to-face communications depends on the physical distance between individuals, but in an online social network, message spreading depends mainly on online connections (not on physical distance). Recent works [50,51] have explored the structure of physical information network by
tracking in-person interactions over the population, and their findings indicate that such interactions would give rise to a social graph consisting of a large number of small cliques, which are somewhat loosely connected to each other. Each clique therein stands for a group of people who are close to each other. The message can spread quickly within a clique via frequent face-to-face interactions, but takes longer time to spread across cliques separated by longer distances. Clearly, constrained by its limited propagation time, the real-time information is less likely to propagate across cliques via face-to-face contacts. Needless to say, in order to characterize the diffusion behavior of real-time information, we need to consider the impact of the clique structure, which is missing in other related works [27, 28, 40].

5.1.2 Summary of Main Contributions

We study the diffusion of real-time information in an overlaying social-physical network. In this study, we consider a physical information network where the message could spread amongst people through face-to-face contacts. Furthermore, the information could also propagate via an online social network conjoint to this physical information network. For convenience, we refer to the physical information network simply as the physical network and refer to the online social network simply as the social network. Hence, the overall system is termed as the coupled (or overlaying) social-physical network.

Specifically, we investigate the information diffusion under two scenarios, namely, coupled-network model I and coupled-network model II, as illustrated in Fig. 5.1. In model I, we assume that all nodes in the social network are also in the physical network, i.e., the collection of online users is a proper subset of the individuals in the physical network. In model II, we consider a more general case where the social network also has online users
Figure 5.1: An illustration of two models. The blank ellipse and the dotted ellipse stand for the physical network and the social network, respectively. In coupled-network model II, the overlapping fraction between two networks represents the collection of online users who are also in the physical network, while the other fraction in the social network represents the collection of online users who are outside the physical network.

who do not belong to the physical network. As illustrated in Fig. 5.1, the social and physical networks are “partially overlapping” and the overlapping fraction represents the collection of online users who are also in the physical network. As we will elaborate in Sections 5.2 and 5.5, the assumptions of these two models correspond to different practical scenarios.

In both models, we characterize the information diffusion process by studying the phase transition behaviors of the underlying random graph models (see Section 5.2.3 for details). Specifically, we show that the system model has a critical threshold above which information epidemics can take place, i.e., the information can reach a non-trivial fraction of individuals. We also quantify the number of individuals that finally receive the message by computing the size of the giant component in the induced random graph model. One interesting finding is that a larger size social network may not always yield
a larger size of information epidemic in this coupled social-physical network. Specifically, we show that given the fixed degree distribution, the growing size of the social network could essentially reduce the coupling strength between two networks. Under certain conditions, due to the reduction in the network coupling, the size of information epidemic could decrease with the growing size of the social network while fixing its degree distribution. This is in stark contrast to the information diffusion behavior in a single network.

In related work, it is assumed [27,28] that the message propagates at the same speed along different social relationships. Clearly, this assumption is not appropriate for the diffusion of real-time information, where propagation speeds play a key role. Very recent work [40] considered online connections and face-to-face connections for general information diffusion, but did not study the impact of the clique structure on information diffusion. To the best of our knowledge, our work is the first attempt on the diffusion of real-time information while considering the clique structure in social networks. We believe that our work will offer initial steps towards understanding the diffusion behaviors of real-time information in a coupled social-physical network.

5.2 Coupled-Network Model I

5.2.1 Illustration of Model Structure

Fig. 5.2 illustrates the structure of model I. We consider an overlaying social-physical network $\mathcal{H}$ that consists of a physical network $\mathcal{W}$ and a social network $\mathcal{F}$. The collection of the nodes $\mathcal{N}_W = \{1, 2, ..., N\}$ in the physical network $\mathcal{W}$ stand for the human beings in the real world. Meanwhile, each node in $\mathcal{W}$ is also a member of the social network $\mathcal{F}$ with probability $\alpha$, and the collection
of nodes in $\mathcal{F}$, denoted by $\mathcal{N}_F$, stand for their online memberships\(^1\). We also refer to the nodes in $\mathcal{W}$ and $\mathcal{F}$ as “individuals” and “online users,” respectively. Clearly, Model I describes the practical scenario where the collection of online users is a subset of the individuals in the physical network. For example, we consider a very large population living in a city (physical network) where some of the individuals can use Internet to reach out to each other (social network). We use model I to capture the diffusion of information among the residents of this city.

**Cliques in the physical information network.** Based on empirical studies in [50,51], we assume that the physical network has $N$ nodes which are gathered into many cliques with different sizes. Each clique represents a group of people with frequent face-to-face interactions, e.g., family in a house or colleagues in an office. It is assumed that the clique size follows the distribution $\{\mu^w_n, n = 1, 2, \ldots, D\}$, where $D$ is the largest possible size. Therefore, an arbitrary clique could contain $n$ nodes with probability $\mu^w_n$. We

\(^1\)Throughout, we use “nodes in $\mathcal{W}$” and “nodes in $\mathcal{N}_W$” interchangeably. So it is the same with the social network $\mathcal{F}$ and $\mathcal{N}_F$. 

Figure 5.2: Structure of coupled-network model I
generate these cliques as follows: at step $t = 1$, we randomly choose $n$ nodes from the collection $\mathcal{N}_W$ and create a clique with the selected $n$ nodes, where $n$ is a random number following the distribution $\{\mu_n^w, n = 1, 2, \ldots, D\}$. We also denote the collection of the remaining nodes in $\mathcal{N}_W$ by $\mathcal{N}_1$. At each step $t$, we repeat the above procedure to create a new clique from the collection $\mathcal{N}_{t-1}^2$, and assume that we can finally generate $N_c$ cliques in $\mathcal{W}^3$. It follows that $N = N_c \sum_n n \mu_n^w$. Generally speaking, the existence of large size cliques indicates that many individuals are close to each other.

As we elaborate in the following, the links connecting the nodes in $\mathcal{W}$ stand for traditional face-to-face connections, while the links in $\mathcal{F}$ represent online connections.

**Type-0 (intra-clique) links in $\mathcal{W}$**. Since the nodes within the same cliques could interact to each other frequently, we assume these nodes are fully connected by type-0 links. Note that in this study, the concept of clique is different from the well-studied “community” in social networks [52], in the sense that the nodes in a clique are fully connected to each other.

**Type-1 (inter-clique) links in $\mathcal{W}$**. We assume that a face-to-face interaction is still possible to happen between cliques, e.g., a person may talk to a remote friend by walking across a long distance. Suppose each node can randomly connect to $k^w$ nodes from other cliques through type-1 links where $k^w$ is a random variable drawn independently from the distribution $\{p^w_k, k = 0, 1, \ldots\}$.

\[ \text{Note that the last generated clique may not follow the expected size distribution, since there would be only too few nodes left to choose. However, such impact on clique size distribution would be negligible if the number of cliques is large enough.} \]

\[ \text{Throughout, we use “clique in $\mathcal{W}$” and “clique in $\mathcal{H}$” interchangeably, in the sense that the network $\mathcal{W}$ is also a part of system model $\mathcal{H}$.} \]
Online users and type-2 (online) links. The nodes in the social network \(\mathbb{F}\) represent the online users. As in [40], we assume each online user randomly connects to \(k_f\) online neighbors in \(\mathbb{F}\), where \(k_f\) is a random variable whose distribution is drawn independently from \(\{p_k, k = 0, 1, \ldots\}\). We denote such online connection as type-2 link. Furthermore, we draw a virtual type-3 link from an online user in \(\mathbb{F}\) to the actual person it corresponds to in the physical information network \(\mathbb{W}\); this indicates that the two nodes actually correspond to the same person.

Online users associated with a clique. To avoid confusions, we say “size-\(n\) clique with \(m\) online members” when referring to the case that a clique contains \(n\) individuals and only \(m\) of them participate in the social network \(\mathbb{F}\). Specifically, for the collection of size-\(n\) cliques with \(m\) online members, \(m \leq n \leq D\), we assume that their fractional size in the whole collection of cliques is \(\mu_{nm}\). It is easy to see that

\[
\mu_{nm} = \mu_n^w \binom{n}{m} \alpha^m (1 - \alpha)^{n-m} \quad \text{and} \quad \mu_n^w = \sum_{m=1}^{n} \mu_{nm}. \quad (5.1)
\]

5.2.2 Information Transmissibility

The message can propagate at different speeds along different types of social connections in \(\mathbb{H}\). Due to timeliness requirement, the real-time information is easier to pass over a link with fast propagation speed. With this insight, we assign each link with a transmissibility as in [27,40], i.e., the probability that the message can successfully pass through.

For ease of exposition, we set the transmissibility along type-0 link as \(T_c = 1\) since the message spreads quickly within a clique. We also define the transmissibilities along type-1 and type-2 links as \(T_w\) and \(T_f\), respectively. Throughout, we say a link is occupied if the message can successfully pass.
through that link. Hence, in $\mathbb{H}$ each type-1 link is occupied independently with probability $T_w$, whereas each type-2 link is occupied independently with probability $T_f$.

5.2.3 Information Cascade

We give a brief description of the information diffusion process in the following. All individuals in network $\mathcal{W}$, i.e., the collection of nodes in $\mathcal{N}_W$, are potential information recipients. Suppose that the message starts to spread from an arbitrary node $i$ in a clique of $\mathcal{W}$. Then, the other nodes in this clique will quickly receive that message through type-0 links. The message can also propagate to nodes in other cliques through occupied type-1 and type-2 links. This process could continue iteratively in this manner and may eventually lead to an information epidemic; i.e., a non-zero fraction of individuals may receive the information in the limit $N \to \infty$ [40].

Clearly, an arbitrary individual can spread the information to nodes that are reachable from itself via the occupied edges of $\mathbb{H}$. Hence, the size of an information outbreak (i.e., the number of individuals that are informed) is closely related to the size of the largest connected components of $\mathbb{H}$, which contains only the occupied type-1 and type-2 links [27, 28, 40] of $\mathbb{H}$. Thus, the information diffusion process considered here is equivalent to a heterogeneous bond-percolation process over $\mathbb{H}$; the corresponding bond percolation is heterogeneous since the occupation probabilities are different for type-1 and type-2 links. In this study, we will exploit this relation and find the condition and the size of information epidemics by studying the phase transition behaviors of $\mathbb{H}$. A key observation is that the system $\mathbb{H}$ exhibits a phase transition behavior at a critical threshold [40]. Specifically, a giant connected component $G_H$ that covers a non-trivial fraction of $\mathbb{H}$ is likely to appear above the critical
threshold, meaning that information epidemics are possible. Below that critical threshold, all the connected components in $\mathbb{H}$ are small indicating that the influenced individuals fraction tends to zero in the large network size limit.

It is easy to see that the influenced individuals and cliques correspond to the nodes and cliques in $\mathbb{W}$ that are contained inside $G_H$. Hence, we introduce two parameters to evaluate the size of information epidemic:

- $S_c$: The fractional size of the influenced cliques in $\mathbb{W}$. Specifically, $S_c$ is the ratio of the number of the cliques contained in $G_H$ to the total number of cliques in $\mathbb{W}$.

- $S_n$: The fractional size of the influenced individuals in $\mathbb{W}$. Specifically, $S_n$ is the ratio of the number of the individuals contained in $G_H$ to the total number of nodes in $\mathcal{N}_W$.

With this insight, we can explore the information diffusion process by characterizing the phase transition behavior of the giant component $G_H$.

5.3 Equivalent Graph: a Clique Level Approach

In this study, we are particularly interested in the following two questions:

- What is the critical threshold of $\mathbb{H}$? In other words, under what condition, the information reaches a non-trival fraction of the network rather than dying out quickly?

- What is the expected size of an information epidemic? In other words, to what nodes fraction and cliques fraction does the information reach? Or, equivalently, what are the sizes $S_c$ and $S_n$?
These two questions can be answered by quantifying the phase transition behaviors of $H$. Due to the clique structure in our system model, the techniques employed in existing works [27, 28, 40] cannot be directly applied here. To tackle this challenge, we develop an equivalent random graph $E$ that exhibits the same phase transition behavior as $H$. Then, we characterize the phase transition behaviors in the graph $E$ by capitalizing on the recent results in inhomogeneous random graph [48].

We first construct an equivalent graph $E$ based on the structure of $H$. Since the nodes within the same clique can immediately share the message, we treat each clique including affiliated online users as a single virtual node in $E$. Furthermore, we assign type-1 and type-2 links between two virtual nodes according to the original connections in $H$. To get a more concrete sense, we depict the equivalent graph in Fig. 5.3 that corresponds to the original model I in Fig. 5.2. It is easy to see that the (type-1 and type-2) link degree of a virtual node equals the total number of (type-1 and type-2) links that are incident on the nodes within the corresponding clique. The equivalent graph $E$ is expected to exhibit the same phase transition behavior as the original model.
since both graphs have the similar connectivity structure. In particular, the fractional size of the giant component $G_E$ in the equivalent graph $\mathbb{E}$ (the ratio of the number of nodes in $G_E$ to the number of nodes in $\mathbb{E}$) is equal to the aforementioned fraction $S_c$. Thus, with a slight abuse of notation, we use $S_c$ to denote the fractional size of $G_E$.

The degree of an arbitrary node in $\mathbb{E}$ can be represented by a two-dimensional vector $d = [d^w \ d^f]$ where $d^w$ and $d^f$ correspond to the numbers of type-1 and type-2 links incident on that node, respectively. For a node in $\mathbb{E}$ that corresponds to a size-$n$ clique in $\mathbb{W}$, we use $K^w_n$ to denote its type-1 link degree, where $K^w_n$ is a random variable following the distribution $\{P^w_{nk}, k = 0, 1, 2, \ldots; n = 1, 2, \ldots, D\}$. Similarly, for a node in $\mathbb{E}$ that corresponds to a clique with $m$ online users, we use $K^f_m$ to denote its type-2 link degree where $K^f_m$ follows the distribution $\{P^f_{mk}, k = 0, 1, 2, \ldots; m = 0, 1, \ldots, D\}$. It is clear to see that an arbitrary node in $\mathbb{E}$ has link degree $[i, j]$ with probability

$$p(i, j) = \sum_{n=1}^{D} \sum_{m=0}^{n} \mu_{nm} P^w_{ni} P^f_{mj} \quad i, j \in N. \quad (5.2)$$

Let $E[d^w]$ and $E[d^f]$ be the mean numbers of type-1 and type-2 links for a node in $\mathbb{E}$, i.e., $E[d^w] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p(i, j) i$ and $E[d^f] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p(i, j) j$. We also define $E[d^w d^f] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p(i, j) ij$. Furthermore, let $E[(d^w)^2]$ and $E[(d^f)^2]$ denote the second moments of the number of type-1 and type-2 links for a node in $\mathbb{E}$, respectively; i.e., $E[(d^w)^2] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p(i, j) i^2$ and $E[(d^f)^2] = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p(i, j) j^2$.

5.4 Analytical Solutions

In this section, we analyze information diffusion process by characterizing the phase transition behaviors in the equivalent random graph $\mathbb{E}$. We present our
analytical results in the following two steps. We first quantify the conditions for the emergence of a giant component as well as the fractional sizes $S_c$ and $S_n$ for the special case $T_w = 1$ and $T_f = 1$. We next show that these results can be easily extended to a more general case with $0 \leq T_w \leq 1$ and $0 \leq T_f \leq 1$.

In what follows, we characterize the phase transition behavior of the giant component in $E$ by capitalizing on the theory of inhomogeneous random graphs [45,48]. Specifically, we define $a_{11} = E[(d_w)^2]/E[d_w] - 1$, $a_{12} = E[d_w d_f]/E[d_w]$, $a_{21} = E[d_w d_f]/E[d_f]$ and $a_{22} = E[(d_f)^2]/E[d_f] - 1$. Along the same line in [40,45,48], we have the following result.

**Lemma 5.4.1.** Let

$$\sigma = \frac{1}{2} \left( a_{11} + a_{22} + \sqrt{(a_{11} - a_{22})^2 + 4a_{12}a_{21}} \right)$$

(5.3)

if $\sigma > 1$, with high probability (whp) there exists a giant component in $E$, i.e., a non-trival fraction of nodes in $E$ are connected; otherwise, a giant component does no exist in $E$ whp.

The proof of Lemma 5.4.1 is relegated to Section 5.8.1. As we discussed in Section 5.2.3, the existence of a giant component in $E$ indicates that the information can reach a non-trival fraction of cliques in $H$ rather than dying out quickly.

Next, let $h_1$ and $h_2$ in $(0, 1]$ be given by the smallest solution to the following recursive equations:

$$h_1 = \frac{1}{E[d_w]} \sum_{n=1}^{D} \sum_{m=0}^{n} \mu_{nm} E[K_{n}^{w} h_1^{K_{n}^{w} - 1}] E[h_2^{K_{n}^{f}}],$$

(5.4)

$$h_2 = \frac{1}{E[d_f]} \sum_{n=1}^{D} \sum_{m=0}^{n} \mu_{nm} E[h_1^{K_{n}^{w}}] E[K_{n}^{f} h_2^{K_{n}^{f} - 1}].$$

(5.5)
We have the following results on the size and probability of an information epidemic.

**Lemma 5.4.2.** The fractional size of the giant component in $E$ (equivalently, the fractional size of influenced cliques in $W$) is given by

$$S_c = \sum_{n=1}^{D} \sum_{m=0}^{n} \mu_{nm} \left( 1 - E[h_1^{K_w} | E[h_2^{K_f}]] \right). \quad (5.6)$$

The fractional size of influenced individuals in $W$ (equivalently, the influenced nodes fraction in $N_W$) is given by

$$S_n = \frac{1}{C} \sum_{n=1}^{D} \sum_{m=0}^{n} n\mu_{nm} \left( 1 - E[h_1^{K_w} | E[h_2^{K_f}]] \right), \quad (5.7)$$

with the normalization term $C = \sum_{n=1}^{D} n\mu_n$.

The proof of Lemma 5.4.2 is relegated to Section 5.8.2. For any given set of parameters, Lemma 5.4.2 reveals the individuals fraction and cliques fraction that are likely to receive an information that is started from an arbitrary individual.

We next generalize Lemma 5.4.1 and Lemma 5.4.2 to the case $0 \leq T_w \leq 1$ and $0 \leq T_f \leq 1$. We first break down the first/second moments of $d_w$ and $d_f$ from the condition (5.3) in Lemma 5.4.1 into the linear combinations of the first/second moments of $k^w$ and $k^f$ as follows:

$$E[d_w] = \sum_{n=1}^{D} \mu_n n E[k^w] \quad E[d_f] = \sum_{m=1}^{D} \mu_m m E[k^f], \quad (5.8)$$

$$E[d_w d_f] = \sum_{n=1}^{D} \sum_{m=0}^{n} \mu_{nm} nm E[k^w] E[k^f], \quad (5.9)$$

$$E[(d_w)^2] = \sum_{n=1}^{D} \mu_n \left( n E[(k^w)^2] + (n^2 - n)(E[k^w])^2 \right), \quad (5.10)$$

114
\[ \mathbb{E}[(d_f)^2] = \sum_{m=1}^{D} m \mathbb{E}[(k_f)^2] + (m^2 - m) \mathbb{E}[k_f]^2. \] (5.11)

Similarly, \( \mathbb{E}[h_1^{K_n^w}] \), \( \mathbb{E}[K_n^w h_1^{K_n^w-1}] \), \( \mathbb{E}[h_2^{K_n^f}] \) and \( \mathbb{E}[K_m^f h_2^{K_m^f-1}] \) in (5.4)-(5.7) can boil down to the integrals with respect to the distributions of \( k_w \) and \( k_f \) by utilizing the following transformations:

\[ \mathbb{E}[h_1^{K_n^w}] = \mathbb{E}[h_1^{k_w}] \quad \mathbb{E}[h_2^{K_n^f}] = \mathbb{E}[h_2^{k_f}], \] (5.12)

\[ \mathbb{E}[K_n^w h_1^{K_n^w-1}] = n \mathbb{E}[h_1^{k_w}]^{n-1} \mathbb{E}[k_w h_1^{k_w-1}], \] (5.13)

\[ \mathbb{E}[K_m^f h_2^{K_m^f-1}] = m \mathbb{E}[h_2^{k_f}]^{m-1} \mathbb{E}[k_f h_2^{k_f-1}]. \] (5.14)

In this way, the calculations in (5.3)-(5.7) can be simplified by utilizing (5.8)-(5.14).

As aforementioned in Section 5.2.2, for the case with \( 0 \leq T_w \leq 1 \) and \( 0 \leq T_f \leq 1 \), the original degree distributions \( k_w \) and \( k_f \) should be replaced by the degree distributions of occupied links. Specifically, we maintain the occupied links in the equivalent graph \( \mathbb{E} \) by deleting each type-1 and type-2 link with probability \( 1 - T_w \) and \( 1 - T_f \), respectively. Let \( \tilde{k}_w \) and \( \tilde{k}_f \) be the occupied link degrees (instead of \( k_w \) and \( k_f \)) with the distributions \( \{ \tilde{p}_k^w, k = 0, 1, \ldots \} \) and \( \{ \tilde{p}_k^f, k = 0, 1, \ldots \} \). According to [27], the generating functions corresponding to \( \tilde{k}_w \) and \( \tilde{k}_f \) can be given by

\[ \tilde{g}(x) = g \left( 1 + T_w (x - 1) \right) \quad \tilde{q}(x) = q \left( 1 + T_f (x - 1) \right). \] (5.15)

From (5.8)-(5.14), we observe that the critical threshold and the giant component size are determined by the distributions of \( k_w \) and \( k_f \). Therefore, Lemma 5.4.1 and Lemma 5.4.2 still hold if we replace the terms associated with \( k_w \) and \( k_f \) in (5.8)-(5.14) by those associated with \( \tilde{k}_w \) and \( \tilde{k}_f \), respectively.
this end, by using the generating functions (5.15), we find

\[ E[\tilde{k}^w] = T_w E[k^w], \]

\[ E[(\tilde{k}^w)^2] = T_w^2 (E[(k^w)^2] - E[k^w]) + T_w E[k^w]. \]

In the same manner, we can compute \( E[\tilde{k}^f] \) and \( E[(\tilde{k}^f)^2] \). The critical threshold (in the general case) can now be computed by replacing \( E[k^w], E[k^f], E[(k^w)^2], E[(k^f)^2] \) with \( E[\tilde{k}^w], E[\tilde{k}^f], E[(\tilde{k}^w)^2], E[(\tilde{k}^f)^2] \), respectively, in (5.8)-(5.11).

In order to compute the giant component size, we only need to replace the corresponding terms in (5.12)-(5.14) with \( E[h_1 \tilde{k}^w], E[h_2 \tilde{k}^f], E[\tilde{k}^w h_1^{\tilde{k}^w - 1}] \) and \( E[\tilde{k}^f h_2^{\tilde{k}^f - 1}] \). By using (5.15), we have

\[ E[h_1 \tilde{k}^w] = \tilde{g}(h_1) = E[(1 + T_w(h_1 - 1))^{k^w}], \]

\[ E[\tilde{k}^w h_1^{\tilde{k}^w - 1}] = [\tilde{g}(h_1)']_1 = T_w E[k^w (1 + T_w(h_1 - 1))^{k^w - 1}]. \]

Similar relations can be obtained for \( E[h_1 \tilde{k}^f] \) and \( E[h_1 \tilde{k}^f] \). The size of the giant component (in the general case) can now be computed by reporting the updated (5.12)-(5.14) into (5.4)-(5.7).

5.5 Coupled-Network Model II

In practical scenarios, the social networks, e.g., Facebook and Twitter, enable the message to reach remote online users from other cities or countries. With this insight, we consider a more realistic model II, where the social network also has online users outside the physical network. Specifically, we assume that the physical network has the same clique structure as in model I. A fraction \( \alpha \) of the total \( N \) nodes in the physical network are also online users in social network \( \mathcal{F} \). For convenience, we denote the collection of these online users from the physical network as \( \mathcal{N}_{F_1} \) and hence \( |\mathcal{N}_{F_1}| = \alpha N \). In contrast to model I,
we assume that the nodes in $\mathcal{N}_F$ only occupy a fraction $\beta$ of the total online users, $\beta \in (0, 1]$, since the social network also has online users who do not belong to physical network $\mathcal{W}$. Therefore, the size of social network $\mathcal{F}$ turns out to be $|\mathcal{N}_F| = \alpha N / \beta$. Moreover, we use $\mathcal{N}_{F_2}$ to denote the collection of the online users outside the physical network and hence $|\mathcal{N}_{F_2}| = \alpha N (1 - \beta) / \beta$. Clearly, when $\beta = 1$, model II reduces to model I as a special case.

By doing so, the collection of the potential information recipients extends from $\mathcal{N}_W$ to $\mathcal{N}_W \cup \mathcal{N}_{F_2}$. We are particularly interested to see the size of information epidemic among the overall population, equivalently, the number of the influenced nodes in $\mathcal{N}_W \cup \mathcal{N}_{F_2}$. We tackle this problem by transforming the coupled networks into an equivalent graph defined in Section 5.3. Specifically, we assume that each node in $\mathcal{N}_{F_2}$ has a virtual counterpart node in $\mathcal{W}$, which has no type-1 links. Each pair of these two nodes can be treated as a single virtual node in the equivalent graph $\mathcal{E}$ and this node only has type-2 links with degree distribution $\{p^f_k, k = 0, 1, \ldots\}$. By definition of the equivalent graph in Section 5.3, the fractional size of such virtual nodes can be given by $\nu_w^0 = \frac{|\mathcal{N}_{F_2}|}{N_c + |\mathcal{N}_{F_2}|}$. Furthermore, we use $\nu_n^w$ to denote the fractional size of the virtual nodes in $\mathcal{E}$ which correspond to size-$n$ cliques and it follows that $\nu_n^w = \frac{N_c}{N_c + |\mathcal{N}_{F_2}|} \mu_n^w$. We also use $\nu_{nm}^w$ to denote the fractional size of the virtual nodes which correspond to the size-$n$ cliques with $m$ online users, following the similar definitions as (5.1).

The equivalent graph $\mathcal{E}$ is an inhomogeneous random graph with two types of links. As in Section 5.3, an arbitrary node in $\mathcal{E}$ has link degree $[i, j]$ with probability

$$p(i, j) = \sum_{n=1}^{D} \sum_{m=0}^{n} \nu_{nm}^w p^w_{m} p^f_{nj} + \nu_0^w p^w_{0} p^f_{1j}, \quad i, j \in N,$$  \hspace{1cm} (5.16)
where $P_{ni}$ and $P_{mj}$ follow the same definitions in (5.2) and

\[
P^w_{0i} = \begin{cases} 
1 & \text{if } i = 0, \\
0 & \text{if } i > 0. 
\end{cases}
\]

Based on the equivalent graph, the phase transition behaviors in model II can be characterized in the same way as in model I, only with a different degree distribution in $E$. Therefore, we can still use Lemma 5.4.1 to characterize the existence condition of the giant component in $E$, which indicates the outbreak of the information epidemic.

Next, let $h_1$ and $h_2$ in $(0, 1]$ be given by the smallest solution to the following recursive equations:

\[
h_1 = \frac{1}{E[d_w]} \sum_{n=1}^{D} \sum_{m=0}^{n} \nu_{nm} E[K^w_n h_1^{K^w_n - 1}] E[h_2^{K^f_m}], \tag{5.17}
\]

\[
h_2 = \frac{1}{E[d_f]} \left( \sum_{n=1}^{D} \sum_{m=0}^{n} \nu_{nm} E[h_1^{K^w_n}] E[K^f_n h_2^{K^f_m - 1}] + \nu^w_0 E[K^f_m h_2^{K^f_m - 1}] \right). \tag{5.18}
\]

We have the following results on the size of an information epidemic.

**Lemma 5.5.3.** The fractional size of the giant component in $E$ is given by

\[
S_c = \sum_{n=1}^{D} \sum_{m=0}^{n} \nu_{nm} \left( 1 - E[h_1^{K^w_n}] E[h_2^{K^f_m}] \right) + \nu^w_0 \left( 1 - E[h_2^{K^f_m}] \right). \tag{5.19}
\]

The fractional size of influenced nodes in $N_W \cup N_{F_2}$ is given by

\[
S_n = \frac{1}{C} \left( \sum_{n=1}^{D} \sum_{m=0}^{n} n \nu_{nm} \left( 1 - E[h_1^{K^w_n}] E[h_2^{K^f_m}] \right) + \nu^w_0 \left( 1 - E[h_2^{K^f_m}] \right) \right), \tag{5.20}
\]

with the normalization term $C = \sum_{n=1}^{D} n \nu_{n} + \nu_0$.

Lemma 5.5.3 can be proved in the same way as in Lemma 5.4.2. Note that in model II the collection of information recipients becomes $N_W \cup N_{F_2}$. Therefore, the total number of the influenced nodes turns out to be $(N + |N_{F_2}|) S_n$. 

118
Furthermore, Lemma 5.5.3 can be generalized to the case with $0 \leq T_w \leq 1$ and $0 \leq T_f \leq 1$ along the same line as in model I.

5.6 Numerical Results

In Section 5.4 and 5.5 we analyzed the critical threshold and the size of information epidemic in both model I and model II. To get a more concrete sense of the analytical results, we study the diffusion behavior of real-time information via numerical examples. Particularly, we focus on two main features in our models, i.e., the clique structure and the network coupling, and their impacts on information diffusion.

5.6.1 Information Epidemics and Clique Structure

We first investigate how the clique structure could impact information diffusion. Particularly, we compare four scenarios in model I, each with different clique size distribution as illustrated in Table 5.1.

For the sake of fair comparison, the total number of nodes in the physical network is fixed at 12000 in each scenario. From scenario 1 to scenario 4, we can see that the average clique size increases from 1 to 2, indicating that individuals are getting closer to each other. We assume that the type-1 link degree for each node in $\mathcal{W}$ follows a poisson distribution, i.e., $p^w_k = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$, $k = 0, 1, 2, \ldots$, where $\lambda$ is the average type-1 link degree. Meanwhile, the type-2 link degree for each online user in the social network follows a power-law distribution with exponential cutoff, i.e., $p^f_0 = 0$ and

$$p^f_k = \frac{1}{C} k^{-\gamma} e^{-\frac{k}{\bar{\tau}}}, \quad k = 1, 2, \ldots, \tag{5.21}$$

with the normalization factor $C = \sum_{k=1}^{\infty} k^{-\gamma} e^{-\frac{k}{\bar{\tau}}}$.
We compare the sizes of information epidemic in terms of the influenced nodes fraction in $\mathbb{W}$. For each scenario, we plot the fractional size of the nodes that will receive the information versus $T_f$ in Figure 5.4. The curves stand for analytical results obtained by (5.7), while the marked points stand for the simulation results obtained by averaging 200 experiments for each set of parameter. We set $T_w = 0.3$, $\lambda = 2$, $\alpha = 0.3$, $\gamma = 3$ and $\Gamma = 10$. It is easy to check that the analytical results are in good agreement with the simulations. Obviously, the information is much easier to propagate when larger size cliques exist. For instance, when $T_f = 1$, as the average clique size increases from 1 (scenario 1) to 2 (scenario 4), the fractional size of individuals that receive the message grows sharply from 14% to 80%. The above results agree with a natural conjecture that the messages are more influential (i.e., more likely to reach a large portion of the population) when people are close to each other.

### 5.6.2 Information Epidemics and Network Coupling

We next investigate how the coupling between the social and physical networks could facilitate the information diffusion. As illustrated in Fig. 5.1, we say the social and physical networks are coupled in the sense that a fraction of nodes are in both networks. Generally speaking, the coupling strength between two networks depends on the fractional size of the overlapping part in Fig. 5.1 (determined by $\alpha$ and $\beta$) and the number of links therein [21]. Clearly, the strong network coupling enables the information propagation in one network

<table>
<thead>
<tr>
<th>scenario</th>
<th>size-1</th>
<th>size-2</th>
<th>size-3</th>
<th>average clique size</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>100%</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>66.7%</td>
<td>33.3%</td>
<td>0</td>
<td>1.333</td>
</tr>
<tr>
<td>3</td>
<td>33.3%</td>
<td>66.7%</td>
<td>0</td>
<td>1.666</td>
</tr>
<tr>
<td>4</td>
<td>33.3%</td>
<td>33.3%</td>
<td>33.3%</td>
<td>2</td>
</tr>
</tbody>
</table>

Table 5.1: The clique size distribution in four scenarios
Figure 5.4: The influenced nodes fraction in $N_W$ versus $T_f$. The curves stand for analytical results obtained by (5.7) and the marked points stand for the simulation results. The analytical results are in good agreement with the simulations. For comparison, we also plot the fractional size of influenced cliques in scenario 1 where the each clique has only one node.

more likely to trigger further propagation in the other network and hence facilitates the diffusion process. To get a more concrete sense, we study the following two cases in model II.

In the first case, we change the network coupling strength by choosing different $\alpha$, i.e., the fractional size of the nodes in the physical network which are also online users, while fixing the total number of individuals. Specifically, the size of the physical network is fixed at $N = 12000$. We select the clique size distribution in scenario 2 in Table 5.1 with $\lambda = 2$. It is also assumed that the social network has a fixed power-law degree distribution (5.21) with $\gamma = 3$ and $\Gamma = 10$. In Fig. 5.5, we plot the influenced nodes fraction versus $\alpha$. Meanwhile, we let $\beta = \frac{\alpha}{\alpha + 1}$ so that the number of online users outside the physical network and the total number of individuals are fixed at 12000 and 24000, respectively. The curves in Fig. 5.5 show that as the coupling strength between two networks increases with the growing $\alpha$, the influenced nodes fraction could increase monotonically.
In the second case, we change both the network coupling strength and the total number of individuals by increasing the size of the social network while fixing its degree distribution. Specifically, we fix the size of the physical network at \( N = 12000 \) and let \( \alpha = 0.2 \). It follows that the number of online users who are also in the physical network is fixed at \( |\mathcal{N}_F| = \alpha N \). We select the clique size distribution in scenario 2 in Table 5.1 with \( \lambda = 2 \). It is also assumed that the social network has a fixed power-law degree distribution. We increase the size of the social network, i.e., \( |\mathcal{N}_F| = \alpha N / \beta \) by decreasing \( \beta \) from 1 to 0. At the same time, the number of online users outside the physical network and the total number of individuals \( |\mathcal{N}_W \cup \mathcal{N}_F| \) would increase as well. In what follows, we evaluate the size of the information epidemic in terms of the number of the influenced nodes in \( \mathcal{N}_W \cup \mathcal{N}_F \).

Note that in a single social network, the influenced nodes fraction depends on the degree distribution [27, 28]. Therefore the number of the influenced nodes would increase monotonically with the growing network size while fixing its degree distribution. On the contrary, it is more intricate in a coupled social-physical network. One key observation is that the growing size of the social network with the fixed degree distribution could yield two opposite effects on information diffusion. Clearly, the information could spread to more recipients as the network size grows. On the other hand, since the degree distribution is fixed and online users randomly connect to each other [27], there would be fewer links in the overlapping part in Fig. 5.1 (online users who are also in the physical network) as the non-overlapping fraction in the social network increases (online users outside the physical network). This essentially amounts to reducing the coupling strength between the two networks. Simply put, the growing size of the social network increases the number of total
individuals on one hand, but on the other hand reduces the network coupling strength which makes the information more difficult to propagate between two networks. Clearly, the size of information epidemic may either increase or decrease depending on which effect (increase in total number of individuals or reduction in the network coupling strength) is dominant. In what follows, we study the overall impact of these two conflicting effects on a case-by-case basis.

As illustrated in Figs. 5.6-5.8, as the size of the social network grows, the number of influenced nodes exhibits different behaviors under different values of transmissibility $T_w$. For the case with low and high $T_w$ ($T_w = 0.1$ and $T_w = 0.8$), the number of the influenced nodes increases with the growing size of the social network. In contrast, for the case with median $T_w = 0.3$, the number of the influenced nodes decreases as the social network size grows.

We believe that such diverse behaviors can be attributed to the following reasons. For low transmissibility ($T_w = 0.1$ in Fig. 5.6), the information is difficult to spread through type-1 link, indicating that the propagation in the...
Figure 5.6: The number of the influenced nodes in $N_W \cup N_{F2}$ versus the size of social network $|N_F|$ with $T_w = 0.1$, $\gamma = 2.6$ and $\Gamma = 10$. The size of social network increases from $2.4 \times 10^3$ to $2.4 \times 10^4$ as $\beta$ decreases from 1 to 0.1.

Figure 5.7: The number of the influenced nodes in $N_W \cup N_{F2}$ versus the size of social network $|N_F|$ with $T_w = 0.3$, $\gamma = 3$ and $\Gamma = 10$. The size of social network increases from $2.4 \times 10^3$ to $2.4 \times 10^4$ as $\beta$ decreases from 1 to 0.1.

Social network is less likely to trigger further propagation in the physical network. While for high transmissibility ($T_w = 0.8$ in Fig. 5.8), most of the nodes in the physical network could already receive the information through type-1 link and hence the online contacts are not necessary. Therefore, in both cases, the network coupling does not contribute much in facilitating the information diffusion. In other words, the impact of reducing the coupling strength would
Figure 5.8: The number of the influenced nodes in $\mathcal{N}_W \cup \mathcal{N}_F^2$ versus the size of social network $|\mathcal{N}_F|$ with $T_w = 0.8$, $\gamma = 3$ and $\Gamma = 10$. The size of social network increases from $2.4 \times 10^3$ to $2.4 \times 10^4$ as $\beta$ decreases from 1 to 0.1.

be trivial. As the size of the social network grows, the increase in the total number of individuals becomes the dominant impact which makes the total number of the influenced nodes keep growing up. On the contrary, for the median transmissibility $T_w = 0.3$, only a limited fraction of nodes can receive the information purely through type-1 link (in contrast to the case with $T_w = 0.8$), indicating that in this case the network coupling would have a great potential to enhance the information diffusion by triggering the propagation between two networks. In other words, the reduction in the coupling strength could result in a substantial negative impact on information diffusion and makes the total number of the influenced nodes decrease as illustrated in Fig. 5.7.

5.7 Conclusions

In this study, we explored the diffusion of real-time information in a coupled social-physical networks. We developed a model that consists of an online social network and a physical information network with clique structure. One interesting finding is that a larger size online social network, with the same
degree distribution, may not necessarily yield a larger size of information epidemic. In fact, under certain conditions, the size of information epidemic could even decrease with the growing size of the online social network. This is in stark contrast to that in a single network. We believe that our studies will offer initial steps towards understanding the diffusion behaviors of real-time information.

5.8 Appendix

5.8.1 Proof of Lemma 5.4.1

In [45] Söderberg studied the phase transition behaviors of inhomogeneous random graphs where nodes are connected by different types of edges. Such graphs are also called “colored degree-driven random graphs” in the sense that different types of edges correspond to different colors. In a graph with \( r \)-types of edges, the edge degree of an arbitrary node can be represented by a \( r \)-dimension vector \( \mathbf{d} = [d^1 \cdots d^r] \), where \( d^j \) stands for the number of type-\( j \) edges incident on that node. In our study, the equivalent graph \( \Xi \) has two types of edges and the degree distribution of an arbitrary node is denoted by \( p(i,j) = P[d_w = i, d_f = j] \). Also, the generating function of degree distribution \( \{p(i,j)\} \) can be defined by \( H(x_1, x_2) = \sum_i \sum_j p(i,j)x_1^ix_2^j \). Clearly, the multivariable combinatorial moments can be achieved by partial differentiation at \( x_1 = 1 \) and \( x_2 = 1 \), i.e.,

\[
\begin{align*}
\mathbb{E}[d_w] &= \partial_1 H(x_1, x_2) |_{x_1=x_2=1}, \\
\mathbb{E}[d_f] &= \partial_2 H(x_1, x_2) |_{x_1=x_2=1}, \\
\mathbb{E}[d_wd_f] &= \partial_1 \partial_2 H(x_1, x_2) |_{x_1=x_2=1}, \\
\mathbb{E}[(d_w)^2] &= \partial_1^2 H(x_1, x_2) |_{x_1=x_2=1}, \\
\mathbb{E}[(d_f)^2] &= \partial_2^2 H(x_1, x_2) |_{x_1=x_2=1}.
\end{align*}
\]
Let \( \{a_k\} \) denote size distribution of the largest connected component that can be reached from an arbitrary node in \( \mathbb{E} \), whose generating function is defined by \( g(z) = \sum_k a_k z^k \). Furthermore, we define a two-dimension vector \( h(z) = [h_1(z), h_2(z)] \), where \( h_i(z) \) stands for the generating function of size distribution of the component connected by type-\( i \) edges. According to the existing results in [40, 45], we have that

\[
g(z) = z \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p(i, j) h_1(z)^i h_2(z)^j = z H(h(z)), \tag{5.22}
\]

where \( h(z) \) satisfies the following recursive equations

\[
h_1(z) = \frac{z}{E[d_w]} \partial_1 H(h(z)), \tag{5.23}
\]

\[
h_2(z) = \frac{z}{E[d_f]} \partial_2 H(h(z)). \tag{5.24}
\]

The emergence of giant component in \( \mathbb{E} \) can be checked by examining the stability of the recursive equations (5.23)-(5.24) at the point \( h_1 = h_1(1) = 1 \) and \( h_2 = h_2(1) = 1 \). Along the same line as in [45], we define a 2 \( \times \) 2 Jacobian matrix \( J \), i.e.,

\[
J = \begin{bmatrix}
a_{11} & a_{12} \\
a_{21} & a_{22}
\end{bmatrix},
\]

where

\[
a_{11} = \frac{1}{E[d_w]} \partial_1^2 H(h(z))|_{h_1=h_2=1} = E[(d_w)^2 - d_w]/E[d_w],
\]

\[
a_{12} = \frac{1}{E[d_w]} \partial_1 \partial_2 H(h(z))|_{h_1=h_2=1} = E[d_w d_f]/E[d_w],
\]

\[
a_{21} = \frac{1}{E[d_f]} \partial_1 \partial_2 H(h(z))|_{h_1=h_2=1} = E[d_w d_f]/E[d_f],
\]

\[
a_{22} = \frac{1}{E[d_f]} \partial_2^2 H(h(z))|_{h_1=h_2=1} = E[(d_f)^2 - d_f]/E[d_f].
\]

The spectral radius of \( J \) is given by

\[
\sigma = \frac{1}{2} \left( a_{11} + a_{22} + \sqrt{(a_{11} - a_{22})^2 + 4a_{12}a_{21}} \right).
\]
By [40, 45, 48], if $\sigma > 1$, with high probability there exist a giant component in the graph $E$; otherwise, a giant component is very less likely to exist in $E$. Therefore, the condition (5.3) in Lemma 5.4.1 is achieved.

5.8.2 Proof of Lemma 5.4.2

Recall that the fraction size $S_c$ equals $1 - g(1)$ [40]. By (5.22), we have that

$$
S_c = 1 - g(1) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p(i, j) \left(1 - h_1^i h_2^j\right)
$$

$$
= \sum_{n=1}^{D} \sum_{m=0}^{n} \mu_{nm} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P_{ni}^w P_{mj}^f \left(1 - h_1^i h_2^j\right) \quad (5.25)
$$

$$
= \sum_{n=1}^{D} \sum_{m=0}^{n} \mu_{nm} \left(1 - \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P_{ni}^w P_{mj}^f h_1^i h_2^j\right)
$$

$$
= \sum_{n=1}^{D} \sum_{m=0}^{n} \mu_{nm} \left(1 - E[h_1^{K_{w}^{n}}]E[h_2^{K_{f}^{m}}]\right).
$$

In view of (5.23) and (5.24), we have that

$$
h_1 = \frac{1}{E[d_w]} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p(i, j) i h_1^{i-1} h_2^j
$$

$$
= \frac{1}{E[d_w]} \sum_{n=1}^{D} \sum_{m=0}^{n} \mu_{nm} E[K_{w}^{n} h_1^{K_{w}^{n}-1}] E[h_2^{K_{f}^{m}}],
$$

$$
h_2 = \frac{1}{E[d_f]} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} p(i, j) j h_1^{i} h_2^{j-1}
$$

$$
= \frac{1}{E[d_f]} \sum_{n=1}^{D} \sum_{m=0}^{n} \mu_{nm} E[K_{w}^{n} h_2^{K_{f}^{m}-1}].
$$

Furthermore, (5.25) can be rewritten in the following form:

$$
S_c = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{n=1}^{D} \sum_{m=0}^{n} \mu_{nm} P_{ni}^w P_{mj}^f \left(1 - h_1^i h_2^j\right).
$$

Clearly, the term in parentheses gives the probability that a node with colored degree $[d_w = i, d_f = j]$ belongs to the giant component. In other words,
the term in parentheses is the expected number of cliques added to the giant cluster by a degree \([d_w = i, d_f = j]\) clique. Hence, summing over all such \(i, j\)'s we get an expression for the expected size of the giant cluster (in terms of number of cliques).

In order to compute the expected giant component size in terms of the number of nodes, namely to compute \(S_n\), we can modify the above expression such that the term \(n(1 - h_i^1 h_j^2)\) gives the expected number of nodes to be included in the giant cluster by a degree \([d_w = i, d_f = j]\) clique. In other words, with probability \((1 - h_i^1 h_j^2)\) the clique under consideration will belong to the giant component \(G_H\) and will bring \(n\) nodes to the actual giant size \(S_n\).

This yields

\[
\tilde{S}_n = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \sum_{n=1}^{D} \sum_{m=0}^{n} n \mu_{nm} P_{n}^{w} P_{m}^{f} n (1 - h_i^1 h_j^2) \\
= \sum_{n=1}^{D} \sum_{m=0}^{n} n \mu_{nm} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P_{n}^{w} P_{m}^{f} (1 - h_i^1 h_j^2) \\
= \sum_{n=1}^{D} \sum_{m=0}^{n} n \mu_{nm} \left(1 - \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P_{n}^{w} P_{m}^{f} h_i^1 h_j^2\right) \\
= \sum_{n=1}^{D} \sum_{m=0}^{n} n \mu_{nm} \left(1 - E[h_1^K w] E[h_2^K f]\right).
\]

We next have

\[
S_n = \frac{1}{C} \tilde{S}_n, \quad C = \sum_{n=1}^{D} n \mu_n,
\]

where the normalized term \(C\) makes \(S_n = 1\) at \(h_1 = h_2 = 0\). Therefore, the conclusions (5.6) and (5.7) in Lemma 5.4.2 have been obtained.
Chapter 6

CSMA-BASED DISTRIBUTED SCHEDULING IN MULTI-HOP MIMO NETWORKS

6.1 Introduction

We study distributed scheduling in multi-hop networks with MIMO links, where each node is equipped with an antenna array. There has been a tremendous body of work on the multiple-input multiple-output (MIMO) technology from a PHY-layer communication perspective. For single-user wireless channels, it has been shown that using the MIMO technique can lead to dramatic improvements on capacity and link reliability [53, 54]. Recent studies have explored the fundamental tradeoffs and relations between the different gains in single-user MIMO systems [55]. In contrast to the extensive studies on the single-user settings, however, there has been little work on exploring multi-hop MIMO networks. Obtaining a rigorous understanding of the tradeoffs between the possible MIMO gains therein has remained a largely open problem.

Leveraging MIMO gains in a multi-hop network is intimately related to link scheduling, because the intrinsic rate-reliability tradeoff hinges heavily on the SINR values of the coupled MIMO links due to mutual interference (see, e.g., [56, 57]). In this study, we will take two steps to explore the scheduling in multi-hop MIMO networks:

- Step 1: Develop a link abstraction that can capture the rate-reliability tradeoff in MIMO communications;
- Step 2: Pursue a deep understanding of throughput-optimal scheduling
under the SINR model\(^1\), and use this as a basis for studying distributed MIMO link scheduling.

More specifically, to facilitate the development of low-complexity scheduling, we propose an appropriate “MIMO-pipe” model that provides an abstraction of the rate-reliability tradeoff in MIMO communications. Clearly, choosing the highest rate for a given MIMO link may not be optimal for the network, since it may prevent other links from being simultaneously active and degrade the overall network throughput. Instead, we model a MIMO-link using a set of achievable “configurations,” under which a link can transmit multiple data streams at the same time; and different configurations have different SINR requirements for reliable communication. Each MIMO link can select one among a set of configurations according to its SINR requirement. Observe that the MIMO communications expands the space of possible network states, and if not designed intelligently it would further complicate scheduling schemes that are already very complex [58].

Recently, low-complexity scheduling schemes based on carrier sense multiple access (CSMA) have been proposed (see [4–8] and the references therein). In these CSMA algorithms, nodes first sense the channel activity, and only when the channel is sensed to be idle can the nodes continue with data transmissions. When the channel is detected busy, the nodes need to backoff for a random amount of time before reattempting the transmission. Due to its simplicity, CSMA and its variants have been widely opted in practical MAC protocols (e.g., IEEE 802.11). It has been shown in [4,7] that under an idealized CSMA model, where the backoff time is continuous and collisions

\(^1\)A scheduling algorithm is said to be throughput-optimal if it can achieve every point in the capacity region [58].
neither happen, the network state dynamics can be captured by a continuous-time Markov Chain (CTMC)\(^2\). The throughput-optimal scheduling algorithm is developed based on the Markov chain modeling of the CSMA network. However, in practical scenarios, collisions could not be avoided completely. Recent work [5] has proposed a discrete-time CSMA scheduling algorithm where the evolution of network states follows a discrete-time Markov Chain (DTMC).

A common theme in these works is to capture the network dynamics by a time-reversible Markov chain, and to drive, via adaptive scheduling, the corresponding stationary distribution to achieve the throughput-optimality. Note that all the algorithms noted above have been developed under protocol-based interference models where two links cannot transmit simultaneously if one link is within a certain range (or hops) of the other link.

In this chapter, we study CSMA-based scheduling in a multi-hop MIMO network, under the SINR interference model. Different from protocol-based models, the rate-reliability tradeoff of a MIMO link hinges heavily on its SINR value. More specifically, under the SINR model, a link transmission is said to be successful if its SINR value is greater than a pre-determined threshold for a given rate. A critical observation is that a successful link transmission under the SINR model depends on its aggregated interference level, and not on the activity of a particular link. As we will elaborate in Section 6.2, the SINR model induces intrinsic global coupling, making it challenging to develop distributed scheduling schemes. In general, it has been largely open on how to design distributed scheduling algorithms under the SINR model (even for the SISO case), and a primary goal of this study is to take some steps in this direction.

\(^2\)Strictly speaking, the algorithms in [4,5] are CSMA/CA. We use the term CSMA to refer to a class of algorithms based on the CSMA mechanism.
We will explore the CSMA algorithms for MIMO-pipe scheduling, for both continuous-time and discrete-time networks. We summarize below the main contributions in this study.

1. We take a bottom-up approach to develop the MIMO-pipe model, which consists of multiple stream configurations, each with a feasible rate and the corresponding SINR requirement. Using this model, the tradeoff between diversity and multiplexing of MIMO communications can be captured by the selection of MIMO configurations. In a nutshell, we treat each configuration as a virtual link with a fixed rate and the corresponding SINR requirement, and each MIMO link is mapped to multiple virtual links with different rates and SINR requirements.

2. We consider the CSMA algorithms for MIMO-pipe scheduling in a continuous-time network. To tackle the intrinsic challenge in the “aggregate interference effect” under the SINR model, we propose to separate the control channel for signal exchanges from that for data transmissions. Assuming that there is no collision of control signals, we show that the network dynamics can be captured by a continuous-time Markov chain. Further, we characterize the optimal backoff parameters of different stream configurations, for throughput-optimal scheduling.

3. We then focus on the CSMA algorithms for MIMO-pipe scheduling in a discrete-time network, where control signals may “collide.” To tackle the collisions and the link coupling problem under the SINR model, we devise a distributed scheduling algorithm using a “conservative” strategy. Specifically, we impose a more stringent SINR constraint to ensure that the transitions of the network states only happen in the feasible state.
region, at the cost of reduced network throughput. We then systematically quantify the performance gap between the optimal scheduling and the conservative scheduling approach. We show that this conservative distributed scheduling can achieve an efficiency ratio bounded below.

6.2 System Setup and Related Work

Consider a multi-hop MIMO network consisting of $K$ links, where each link employs $N_t$ transmit antennas and $N_r$ receive antennas. The received signal at the $i$-th receiver can be given by

$$ y_i = \sqrt{\frac{P}{N_t d_{ii}}}{\bf H}_{ii}{\bf s}_i + \sum_{j \neq i} \sqrt{\frac{P}{N_t d_{ji}}}{\bf H}_{ji}{\bf s}_j + {\bf n}_i, \quad (6.1) $$

where $P$ is the total transmission power at each transmitter; $s_i$ is the $N_t \times 1$ transmitted signal from the $i$-th transmitter, with normalized power at each antenna array to be 1, in each symbol period; $\alpha$ is the path loss exponent; $d_{ji}$ is the distance from the $j$-th transmitter to the $i$-th receiver. We consider a frequency flat fading MIMO channel $^3$ such that $H_{ji}$ is the $N_r \times N_t$ channel matrix between the $j$-th transmitter to the $i$-th receiver, where the entries of each matrix are i.i.d. complex circular symmetric Gaussian with unit variance. Furthermore, the entries of $H_{ji}$ are independent from those of $H_{ji'}$ if $i \neq i'$; $n_i$ is the additive White Gaussian noise with $\sigma^2 = E[||n_i^2||]/N_r$.

The first term in (6.1) is the desired data signal for link $i$, while the last two terms are co-channel interference and noise, respectively. As is standard, we assume that the channel matrix $H_{ii}$ is known at the receiver but unknown at the transmitter of link $i$ (CSI at the receiver) [59]. Moreover, in practical systems, it is difficult, if not impossible, to obtain the MIMO channel matrices.

$^3$As in [59], shadow fading is not considered in this channel model.
\{H_{ji}, j \neq i\} from the interferers, simply because the signals are not intended for the desired link and it is infeasible to estimate and track these complex matrices. Based on the above signal model, it is clear that unlike single-user MIMO systems, multi-hop networks are interference-limited, and MIMO communications are intimately tied to the SINR values that are coupled across the links.

As in [59], let \(I_i\) denote average power level of interference-plus-noise at the receiver of link \(i\), i.e.,

\[ I_i = \sum_{j \neq i} \frac{P}{N_t d_{ji}^\alpha} \frac{E[\text{Tr}(H_{ji}H_{ji}^H)]}{N_r} + \sigma^2, \]  

(6.2)

and let \(\text{SINR}_i\) denote the SINR at the receiver of link \(i\), i.e.,

\[ \text{SINR}_i = \frac{P_{d_i}^{-\alpha}}{\sum_{j \neq i} P_{d_{ji}}^{-\alpha} \frac{E[\text{Tr}(H_{ji}H_{ji}^H)]}{N_t N_r}} + \sigma^2. \]  

(6.3)

Since the entries of \(H_{ji}\) are identically distributed with unit power, we have \(E[\text{Tr}(H_{ji}H_{ji}^H)] = N_t N_r\). Then, the SINR value at \(i\)-th link receiver can be given by

\[ \text{SINR}_i = \frac{P_{d_i}^{-\alpha}}{\sum_{j \neq i} P_{d_{ji}}^{-\alpha} + \sigma^2}. \]  

(6.4)

The SINR value plays a critical role in link scheduling.

### 6.2.1 Feasible States and Capacity Region in a MIMO Network

Throughout, we say that two active links can coexist if they can make successful transmissions at the same time. An interference model specifies the link coexistence constraint. We say that the network is in a feasible state if the set of active links satisfy the coexistence constraint of the interference model. In a network with \(K\) links, we use a binary vector \(x_i = \{0, 1\}^K\) to describe a feasible state. We define that \(x_i^l = 1\), if link \(l\) is active in state \(i\); \(x_i^l = 0\) otherwise.
With some abuse of notation, we also treat $x^i$ as the set of active links in state $i$, i.e., $l \in x^i$ if $x^i_l = 1$. In SISO networks, it suffices to use a binary vector $x$ to represent the data rate of each link, if each link transmits at unit rate $[4,5]$. In contrast, each MIMO link has multiple stream configurations with different transmission rates. Hence, to describe a feasible state in a MIMO network, we also need to specify the configuration and the corresponding transmission rate of each active link. Without loss of generality, we consider a MIMO network with $K$ links, where each link has $J$ configurations. We use $z^i = (z^i_1, z^i_2, ..., z^i_K)$ to denote the configuration of each link at feasible state $i$, where $z^i_l \in [1...J]$ indicates the configuration of link $l$. We also use $c^i = (c^i_1, c^i_2, ..., c^i_K)$ to denote the data rates, where $c^i_l$ is the data rate at link $l$ at state $i$. Furthermore, we define $\Theta(\cdot)$ as the mapping from the configuration index to the corresponding normalized transmission rate, i.e., $c^i_l = \Theta(z^i_l)$. Finally, we set $c^i_l = 0$ and $z^i_l = 0$ if link $l$ is not active at state $i$.

Let $S$ be the set of rate vectors corresponding to the feasible states of a MIMO network. By definition $[58]$, the capacity region $\Lambda$ is the convex hull of the vectors in $S$. Assume that the traffic load at link $l$ is represented by the normalized arrival rate $\lambda_l \geq 0$. The scheduling algorithm is said to be throughput-optimal if it can keep the network stable at any arrival rate vector $\lambda = (\lambda_1, \lambda_2, ..., \lambda_K)$ within the capacity region $\Lambda$ $[58]$.

### 6.2.2 SINR Model versus Protocol Model

Clearly, different interference models yield different link coexistence constraints and hence different sets of feasible states. Roughly speaking, existing interference models can be classified into two categories: the protocol model and the SINR model $[60]$. Under the protocol model, the transmission of link $l$ is deemed successful if no other links within a certain transmission range are
active. Therefore, the coexistence relationship between two links is mainly determined by the geometry, and hence is “static” and “binary.” Due to its simplicity, the protocol model has been widely used.

In contrast, under the SINR model, the coexistence relationship is neither static nor binary, and the success of a transmission depends on its own channel condition and the level of the aggregated interference. Specifically, a transmission of a link is said to be successful if its SINR value (6.4) is greater than a pre-determined threshold for a given rate. The SINR model, built upon recent advances in PHY-layer communication theory, opens a new avenue for more efficient resource allocation in wireless networks.

As noted before, one significant challenge under the SINR model is that multiple links can transmit successfully through a common channel, even if they observe some interference signal from each other, which is drastically different from that under the protocol model. Furthermore, link relationship is a function of distance to the neighboring links and their status that may change over time. Therefore, the link coexistence relationship under the SINR model is “multi-lateral” and “dynamic.” As a result, link scheduling under the SINR model is much more complicated.

In principle, every link in the network can contribute interference to an active receiver under the SINR model. However, when the links are sparsely located and the interference power level decreases over distance due to the free space path loss as in [61] and [62], it is reasonable to assume the aggregated interference from the transmitters beyond certain distance can be upper bounded by a threshold [63]. Specifically, we define a “close-in” radius for each link \( l \) such that the aggregated interference power to \( l \) from the transmitters beyond the close-in range is no more than a given parameter \( \sigma_{\text{int}}^2 \). Denote
\( N(l) \) as the set of links whose transmitters are in the close-in range of link \( l \),
called interfering links of link \( l \) and \( N(l)^c \) as the set of links whose transmitters
are outside the close-in range of link \( l \). It follows that \( \sum_{k \in N(l)^c} Pd_{kl}^{-\alpha} < \sigma_{int}^2 \).
Based on \( \sigma_{int} \), the close-in range of each link can be obtained in an initialization
stage before link scheduling, where each link informs its incurring interference
power level to neighbors by broadcasting a dummy packet sequentially. Next,
each link \( l \) ranks its neighboring links in an ascending order based on their
interference. A neighboring link \( k \) (staring from the link incurring the lowest
interference to the highest) is deemed to be outside the close-in range of \( l \) as
long as the aggregated interference from the links beyond the close-in radius and
link \( k \) is lower than \( \sigma_{int}^2 \).

For ease of exposition, we approximately treat the aggregated interference from active links in \( N(l)^c \) as white noise with power \( \sigma_{int}^2 \). By doing so, we
define the following “nominal” SINR constraint, where link \( l \) can successfully
transmit if the following condition holds:

\[
\text{SINR}_l = \frac{Pd_{ll}^{-\alpha}}{I_{l}^m + \sigma^2 + \sigma_{int}^2} \geq \beta_l \tag{6.5}
\]

where \( I_{l}^m \) is the aggregated interference from the active links in \( N(l) \); \( \sigma^2 \) is the
power of Gaussian noise; \( \beta_l \) is the threshold of successful transmission. In the
following study, unless otherwise specified, the SINR model is defined based
on the nominal SINR constraint in (6.5) \(^4\).

6.2.3 Review: CSMA Scheduling under Protocol Model

We provide below a brief review of [4,5], which are perhaps the most related
works to our study here.

\(^4\)In Section 6.5, we also defined a conservative SINR constraint that is more stringent
than the nominal SINR constraint.
Under the protocol model, an “idealized” CSMA scheduling algorithm is proposed in [4] for a continuous-time network. It is assumed that random backoff time and data transmission time follow continuous distributions. It also takes the assumption that the range of carrier-sensing is large enough and signal propagation delay is zero, which remove potential hidden terminal problem (see [64] for further discussions on hidden terminal problems). Therefore, the probability for two conflicting links to start transmission at the same time is 0 and the collisions can be ignored. Under these assumptions, the state transitions of the CSMA network can be modeled as a continuous-time Markov chain, where transitions only occur between the feasible states that differ from each other by only one link status. It follows that the stationary distribution of feasible states $x^i$ can be characterized by

$$p(x^i) = \frac{1}{C} \prod_{l \in x^i} R_l,$$

where $R_l$ is defined as backoff rate and $C$ is the normalization term satisfying $\sum_i p(x^i) = 1$. In [5], the idea has been extended to a time-slotted system, where simultaneous transmissions in a time slot may collide. It has been shown that the network states can be modeled as a discrete-time Markov chain, and the corresponding stationary distribution can also be written in a product-form:

$$p(x^i) = \frac{1}{C} \prod_{l \in x^i} \bar{p}_l,$$

where $p_l$ is defined as link activation probability in [5] and $\bar{p}_l = 1 - p_l$. Furthermore, it has been shown that adaptive CSMA scheduling algorithms that adjust link parameter based on local queue information can achieve throughput-optimality. We extend the results to more general MIMO scenarios. To this end, we define similar parameters for each MIMO configuration $v$ of link $l$ as shown in Table 6.1.
Table 6.1: Parameters in CSMA-based algorithm (at MIMO link $l$)

<table>
<thead>
<tr>
<th>Continuous time case</th>
<th>Discrete time case</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_{lv}$</td>
<td>$p_{lv}$</td>
</tr>
<tr>
<td>backoff rate of configuration $v$ at link $l$</td>
<td>link activation probability of configuration $v$ at link $l$</td>
</tr>
<tr>
<td>$r_{lv}$</td>
<td>$\bar{p}_{lv}$</td>
</tr>
<tr>
<td>$r_{lv} = \log(R_{lv})$</td>
<td>$\bar{p}<em>{lv} = 1 - p</em>{lv}$</td>
</tr>
</tbody>
</table>

6.3 MIMO-pipe Modeling: Rates, SINR, and Interference Tolerance Levels

A first key step in our study on MIMO scheduling is to develop a PHY-based tractable model that captures the rate-reliability tradeoff for a single MIMO link, which we call the “MIMO-pipe” model.

In MIMO networks, every MIMO link can offer stream multiplexing by opening up multiple spatial data streams in the same frequency channel, and achieve spatial multiplexing gain. The number of data streams depends on the stream configuration of the link. Given the number of antennas and the total transmission power at each node,\footnote{In this study, the transmission power is assumed to be fixed. Dynamic power control is beyond the scope of this study.} we assume that the transmission power is equally split among the transmit antennas. Clearly, the greater the number of data streams there are at each MIMO link, the lower the reliability and the interference tolerance capability per stream. Accordingly, the required average SINR per receive antenna [59], called SINR requirement, is more stringent. In the following, we will elaborate the tradeoff between stream multiplexing gain and interference tolerance capability (determined by the corresponding SINR requirements).
6.3.1 MIMO Configurations and SINR Requirements

Without loss of generality, suppose that each link has $J$ configurations, and for configuration $v$, $v \in [1...J]$, there are $\Theta(v)$ date streams. For simplicity, we set the transmission rate of each stream to be the same, denoted as $R_s$, and hence the link rate is $R_s \Theta(v)$ at configuration $v$. Without loss of generality, we assume the stream rate $R_s$ is fixed at 1 in this study. The SINR requirement of stream $r$ at configuration $v$, can be in general given as

$$\beta_{vr} = f(v, r, H, P_e), \quad (6.8)$$

which depends on the channel matrix $H$ and the average BER requirement $P_e$ for reliable communication. The function $f$ depends on the physical-layer techniques, such as coding and modulation.

To guarantee the decodability of all data streams, the SINR requirement of configuration $v$ should be set as $\beta_v = \max\{\beta_{v1}, \beta_{v2}, ..., \beta_{v\Theta(v)}\}$, i.e., the highest SINR requirement corresponding to the bottleneck stream. Such bottleneck stream usually has the least number of transmit antennas. Therefore, it is reasonable to consider a subset of configurations in which transmit antennas are equally divided for each stream. Clearly, the collection of configurations for a MIMO link with $N_t$ transmit antennas corresponds to an integer set $\{n_v \mid n_v \text{ is a divisor of } N_t, \ v = 1, 2, 3..., J\}$ and the number of configurations equals the number of divisors of $N_t$. Specifically, the configuration $v$ has $n_v$ data streams and each stream has $\frac{N_t}{n_v}$ transmit antennas. For example, for the $4 \times 4$ MIMO link, we consider three configurations: 1-transmit antenna per stream, 2-transmit antennas per stream, and 4-transmit antennas per stream, with data rates $4R_s, 2R_s, R_s$, and SINR requirements $\beta_1 > \beta_2 > \beta_3$, respectively.

141
6.3.2 Interference Tolerance

Under the SINR model, the successful transmission depends on the current SINR value at the MIMO receiver. By definition of the nominal SINR constraint in (6.5), we assume that the MIMO link \( l \) can successfully transmit with \( v \)-th configuration at time \( t \) if the following condition holds:

\[
\text{SINR}_l(t) = \frac{P_d^{-\alpha}}{I_l^{in}(t) + \sigma^2 + \sigma_{int}^2} \geq \beta_{lv},
\]

where \( I_l^{in}(t) \) is the aggregated interference from the active links in \( N(l) \); \( \beta_{lv} \) is the SINR requirement of \( v \)-th configuration at link \( l \); other items follow the same definitions as in (6.5). Given a link activation setting, we define the interference tolerance level as the interference power that the receiver can further tolerate without violating the SINR requirement. By (6.9), for the \( v \)-th configuration of link \( l \), its interference tolerance at time \( t \) can be given by:

\[
T_{lv}(t) = \frac{P_d^{-\alpha}}{\beta_{lv}} - I_l^{in}(t) - \sigma^2 - \sigma_{int}^2.
\]

Clearly, the interference tolerance can be calculated by the receiver based on the interference power level \( I_l^{in}(t) \) that the receiver currently experiences. Note that the interference tolerance level depends on the aggregated interference from the neighbors, and will change dynamically over time according to the on/off status of nearby links.

Fig. 6.1 illustrates the relationship between interference tolerance (reliability) and rate of a single \( 4 \times 4 \) MIMO link. We emphasize that the stream configurations here correspond to a few points on the rate-reliability tradeoff curve, and that the rates are set to multiplications of the basic rate \( R_s \) to reflect the multiplexing gain. In general, one can find multiple pairs of (rate, interference tolerance level) of a MIMO link.
Scheduling problem under the MIMO-pipe model is to decide which link to transmit and which configuration to use in data transmission. Clearly, the configuration with more data streams (higher multiplexing degree) can achieve a higher data rate, but in the meanwhile, fewer transmit antennas are assigned to each stream which results in a lower interference tolerance level. Once a link chooses a higher rate configuration, it would not be able to co-exist with many nearby links. Hence, there exists an intrinsic tradeoff between the throughput for a single link and overall network.

6.4 CSMA Algorithm for MIMO-pipe Scheduling: A Continuous-time Model

In this section, we study the CSMA algorithm for a continuous-time network, under the SINR model. For ease of exposition, we first focus on the distributed scheduling for SISO case and further generalize our study to the MIMO-pipe model.

6.4.1 SINR-aware Channel Probing: A Dual Band Approach

We aim to develop the scheduling algorithm under the SINR model by utilizing the Markov chain structure of a CSMA network, where the network states
evolve as a continuous-time Markov chain and each state in the Markov chain corresponds to a feasible link activation. According to [4], a CSMA network can be described by a continuous-time Markov chain when it satisfies the following requirements:

**(R1)** Network state transitions only occur between the feasible states that differ from each other by only one link status.

**(R2)** For each link, the backoff time and the data transmission time are both exponentially distributed.

To meet the first requirement, a key challenge is to ensure that the CSMA network always stays in a feasible state under the SINR model. In other words, the scheduling algorithm can guarantee the coexistence of active links under the SINR model. Specifically, when a link is activated, it should tolerate the aggregated interference from other active links, and meanwhile, its incurring interference would not violate the SINR requirements of other on-going transmissions.

To tackle this issue, we propose the following “SINR-aware” channel probing approach. This mechanism enables each link to assess its coexistence relationship with other active links under the SINR model by utilizing carrier-sensing and control messages exchange. The key idea is that each receiver keeps sensing the channel and broadcasts its interference tolerance level to the neighbors. With that information, when an inactive link, say $k$, is about to be active, the transmitter of link $k$ can decide whether its potential transmission will violate the SINR requirements of any ongoing transmission. Simply put, for each active link $l$, the receiver calculates its interference tolerance $T_l(t)$ according to (6.10). Then, it broadcasts $T_l(t)$ in the control message to its nearby links, i.e., to any link $k$ with $k \in N(l)$. Based on the interference
the transmitter of link \( k \) can estimate how much interference it would incur to other receivers. By doing so, link \( k \) can judge its coexistence feasibility with the existing active links and avoid possible violations to the nominal SINR requirements.

To ensure that the data transmission would not collide with the control signal, we consider a dual-band approach where we separate the frequency band into data channel and control channel for each signal. By doing so, a receiver can broadcast control message and receive data packets at the same time. From the idealized CSMA assumption as in [4], the transmissions of control signal can be completed instantaneously (i.e., zero propagation delay) and do not collide in the control channel. The details of the channel probing mechanism are summarized in Algorithm 1. Note that the channel probing is a sub-step of CSMA-based scheduling that will be explained in Algorithm 2.

Note that the continuous backoff time ensures that no more than one link decides to transmit at the same instance. Therefore, only one link can change its state during each transition. By using the proposed SINR-aware channel probing approach, the state transitions of the CSMA network only take place among the feasible states under the SINR model. Furthermore, both the backoff time and data transmission time can be designed to follow exponential distributions, which will be shown in the following section. Building on these, the CSMA network can satisfy the requirements \( R1 \) and \( R2 \), and its dynamics can be captured by a continuous-time Markov chain.

### 6.4.2 CSMA Algorithm for MIMO-pipe Scheduling

We next devise the CSMA scheduling algorithm for MIMO links. Recall that under the MIMO-pipe model, each link has multiple stream configurations,
Algorithm 1 SINR-aware channel probing (at link $l$)

At the receiver

- Idle period
  - The receiver keeps sensing the data channel and updating its current $T_l(t)$ by (6.10).

- Data transmission period
  - When link $l$ starts transmission, its receiver broadcasts $T_l(t)$ through the control channel.
  - When receiver senses “new” interference during data receiving, $T_l(t)$ will be updated and broadcasted again through the control channel.
  - When link $l$ finishes transmission, its receiver broadcasts $T_l(t) = \infty$.

At the transmitter

- Keeps overhearing the control messages from the control channel.
- Once receiving a control message from the receiver of link $k$, the transmitter can estimate its possible interference incurring to $k$ based on the interference information acquired at initialization stage.

Check the link coexistence requirements

At time $t$, link $l$ can coexist with nearby active links without violations to the SINR requirements (assuming other existing active links can also coexist) under the following two necessary conditions:

1. $T_l(t) > 0$.
2. For any active link $k \in N(l)$, the interference from link $l$ to $k$ is no great than $T_k(t)$.
and can choose a feasible configuration as long as it satisfies the SINR requirement. Therefore, the MIMO network will have a much larger set of feasible states compared to the SISO case. We develop CSMA scheduling for MIMO-pipe links such that the network state transitions still can be captured by a continuous-time Markov chain, using our SINR-aware channel probing.

We model each MIMO configuration as a “virtual link,” with separate mean backoff time and interference tolerance. Specifically, letting \( l_v \) denote a virtual link with configuration \( v \) at link \( l \), the backoff time of \( l_v \) is exponentially distributed with mean \( 1/R_{l_v} \), where \( R_{l_v} \) is called “backoff rate.” With some abuse of notation, we treat \( z^i \) as the set of active virtual links at state \( i \). At state \( i \), if link \( l \) transmits at stream configuration \( v \), then \( l_v \in z^i \) and \( z^i_l = v \).

Along the same line as in conventional CSMA, each virtual link contends for transmission using the backoff timer. However, the timer freezes when the virtual link cannot make transmission because it would violate any existing transmission of nearby links. This feasibility test can be done with the information obtained from the SINR-aware channel probing. When the virtual link starts data transmission, it should broadcast its interference tolerance level though the control channel. The details of the CSMA algorithm for MIMO link scheduling are summarized in Algorithm 2.
Figure 6.3: State transition graph for the continuous-time Markov chain associated with the network in Fig. 6.2.

Table 6.2: Feasible state

<table>
<thead>
<tr>
<th>feasible state</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
</tr>
</thead>
<tbody>
<tr>
<td>( z )</td>
<td>0,0</td>
<td>0.1</td>
<td>1.0</td>
<td>0.2</td>
<td>2.0</td>
<td>3.0</td>
<td>0.3</td>
<td>1.1</td>
<td>1.2</td>
</tr>
<tr>
<td>( c )</td>
<td>0,0</td>
<td>0.1</td>
<td>1.0</td>
<td>0.2</td>
<td>2.0</td>
<td>4.0</td>
<td>0.4</td>
<td>1.1</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Note: The link configurations and link rates for each feasible state are represented by \( z = (z_1, z_2) \) and \( c = (c_1, c_2) \) as defined in Section 6.2.1. The feasible states in the table are given for illustration purpose only.

With the help of the SINR-aware channel probing, the MIMO network remains in feasible states and can be modeled as a Markov chain as in the SISO case. To get a more concrete sense, we consider an example network with two \( 4 \times 4 \) MIMO links in Fig. 6.2. The feasible states in Table 6.2 are given for illustration purpose only. The network states transition can be captured by a continuous-time Markov chain whose state transition graph is depicted in Fig. 6.3, where each cycle corresponds to a feasible state \( (z_1, z_2) \) and \( z_1 \) and \( z_2 \) represent the configuration of link 1 and link 2, respectively. In the state transition graph in Fig. 6.3, we denote the transition between two states by a directional line with the transition rate. For any two connecting states, the left state transits to the right state with a rate of \( R_{l_v} \), and the right state
transits to the left state with a rate of 1. The stationary distribution of the feasible state $z^i$ can be obtained as

$$p(z^i) = \frac{1}{C} \prod_{l_v \in z^i} R_{l_v}, \quad (6.11)$$

where $C$ is the normalization term. For each link $l$, let $R_i^l$ denote the backoff rate of the active virtual link at state $i$, i.e.,

$$R_i^l = \begin{cases} 
R_{l_v} & \text{if } z_i^l = v \\
1 & \text{if } z_i^l = 0 \text{ (i.e., link } l \text{ is inactive).} 
\end{cases} \quad (6.12)$$

Then, we can rewrite (6.11) as:

$$p(z^i) = \exp \left( \sum_{l=1}^{K} r_i^l \right) \sum_j \exp \left( \sum_{l=1}^{K} r_j^l \right), \quad (6.13)$$

where $r_i^l = \log(R_i^l)$ for each virtual link. The normalized throughput of link $l$ is given by

$$\theta_l = \sum_{i} \Theta(z_i^l) \cdot p(z^i). \quad (6.14)$$

The next key step is to optimize the backoff time of each virtual link, so that the corresponding adaptive CSMA algorithm can converge to the throughput-optimal one. A central problem is how to use local information to adapt the backoff time so as to meet the throughput requirement of each link, i.e., $\theta_l \geq \lambda_l$. Along the lines in [4], we have the following result.

**Lemma 6.4.4.** Under the time-scale-separation assumption [4]\(^6\), the CSMA algorithm for MIMO scheduling can achieve any throughput $\lambda$ in the capacity region, by adjusting the backoff rate of each virtual link as follows:

For link $l$,

$$y_l(t + 1) = [y_l(t) + \xi(\lambda_l - \theta_l(t))]^+, \quad (6.15)$$

\(^6\) As shown in [65], it is possible to achieve the throughput-optimality under certain conditions without the time-scale-separation assumption.
Algorithm 2 Continuous-time CSMA scheduling under the MIMO-pipe model (at link $l$)

Transmission initiation

- For each virtual link $l_v$, $v \in [1...J]$, the transmitter checks its coexistence with the active nearby links using Algorithm 1.

- When a virtual link $l_v$ satisfies the link coexistence constraint, it waits for a period of time (backoff) that is exponentially distributed with mean $1/R_{l_v}$.

Random backoff

When a nearby link begins transmission, $l_v$ updates its interference tolerance level and checks the link coexistence constraint using Algorithm 1. If $l_v$ can no longer coexist with the current active links, $l_v$ would suspend its backoff and resume it after the coexistence constraint is satisfied, i.e., after some nearby active link finishes its transmission.

Data transmission

- Once the back-off time of virtual link $l_v$ expires, link $l$ would launch the data transmission at the stream configuration $v$. The transmission time is exponentially distributed with mean 1.

- Other virtual links of link $l$ suspend the backoff and would resume it until link $l$ finishes data transmission.

where $y_l$ is shown to be proportional to the queue length at link $l$ [4], and $\xi > 0$ denotes a small constant (step size). Each virtual link adapts its backoff time according to

$$R_{l_v} = \exp(y_l \Theta(v)), \quad v \in [1...J],$$

where $\Theta(v)$ is the data rate of configuration $v$.

The proof of Lemma 6.4.4 is relegated to Section 6.8.1.

In the idealized CSMA network, it is assumed that control messages have zero propagation delay, and would never collide. The proposed channel probing approach is based on such “collision-free” assumption.
6.5 CSMA Algorithm for MIMO-pipe Scheduling: A Discrete-Time Model

In the following, we extend our distributed MIMO-pipe scheduling approach to a synchronized time-slotted network.

6.5.1 CSMA Algorithm for Conservative MIMO-pipe Scheduling

We study the CSMA algorithms for link scheduling under the SINR model in a discrete-time network, where the time is slotted. At each time slot \( t \), the scheduling algorithm decides a transmission schedule \( z(t) \).

In [5], the authors develop a CSMA scheduling scheme for the protocol model, which operates as follows: let \( z(t-1) \) denote the transmission schedule in time slot \( t-1 \). At the beginning of time slot \( t \), a feasible schedule denoted by *decision schedule* \( \mathcal{M}(t) \) is calculated. A subset of links in \( \mathcal{M}(t) \) is discarded if they interfere with any link in \( z(t-1) \). Each link in the remaining \( \mathcal{M}(t) \) independently determines whether it will be active in time slot \( t \) or not using its own link information, and all the other links remain in the same state as in time slot \( t-1 \). Finally, links in \( z(t) \) transmit data packets in time slot \( t \). It is required all the links in \( \mathcal{M}(t) \oplus z(t-1) \) can coexist satisfying the underlying interference constraints. Such requirement is not difficult to be satisfied under the protocol model, due to the static link coexistence relationship [5]. However, under the SINR model, the coexistence relationship between two links becomes dynamic and depends on the states of the neighboring links within their close-in radius. Therefore, a key challenge here is to ensure the coexistence of the links in \( \mathcal{M}(t) \oplus z(t-1) \) under the SINR model.

To tackle the above challenge, we impose a more stringent requirement for link coexistence beyond the previously discussed “nominal” SINR con-
straint so that the link coexistence relationship becomes static again. Under this “conservative” SINR constraint, we further develop the “conservative” CSMA link scheduling algorithm. For ease of exposition, we first consider a SISO network. Specifically, for each link \( l \), we rank its interfering links \( N(l) \) (the links within its close-in radius), in an ascending order based on the interference they incur to link \( l \). We partition the interfering links in \( N(l) \) into two disjoint sets \( N_a(l) \) and \( N_b(l) \), i.e., \( N_b(l) = N(l) \setminus N_a(l) \). Let \( N_a(l) \) contain all the neighboring links (starting from the link incurring the lowest interference to the highest) such that their potential aggregated interference to link \( l \) is no greater than \( T^o_l \), where \( T^o_l \) is defined as the initial interference tolerance level when no other neighboring links of \( l \) are active, i.e., \( T^o_l = P_{d_{ll}}^{-\alpha} / \beta_l - \sigma^2 - \sigma_{\text{int}}^2 \) and \( \sum_{k \in N_a(l)} P_{d_{kl}}^{-\alpha} < T^o_l \). For convenience, we call \( N_a(l) \) the “tolerable set” and \( N_b(l) \) the “intolerable set.” The partition of these two sets depends on the estimation of interference power levels, which requires the information of channel gains between link \( l \) and the neighboring links. As in the continuous-time case, such information can be acquired in the initialization stage. Clearly, for each link \( l \), the sets \( N_a(l) \) and \( N_b(l) \) are independent with the states of nearby links. Given a fixed network topology, the \( N_a(l) \) and \( N_b(l) \) will not change over time.

Using the above definitions, we impose the following more stringent coexistence constraint:

**Conservative coexistence constraint for SISO links:** \( \forall k \in N(l) \) and \( \forall l \in N(k) \), links \( l \) and \( k \) can coexist if and only if \( k \in N_a(l) \) and \( l \in N_a(k) \).

Thanks to this new coexistence condition, the link coexistence relationship between two links becomes static again, so that the complexity of scheduling can be greatly reduced. In the meanwhile, the conservative model
still takes into account the “aggregate interference effect,” and provides a more realistic characterization of co-channel interference compared to the protocol model. As elaborated in Section 6.5.2, despite the throughput loss due to the conservative coexistence constraint, the conservative scheduling can at least achieve a guaranteed fraction of the optimal throughput region.

Due to the static coexistence relationship, we can now depict a conflict graph $G$ for the network, where each vertex corresponds to a link, and there is an edge between two vertexes if they conflict with each other. For convenience, we say that link $l$ and link $k$ are “severely conflicting” if they cannot satisfy the conservative coexistence constraint. Since only the links in $N_a(l)$ are allowed to transmit simultaneously with $l$, the aggregated interference from $N_a(l)$ is guaranteed to be lower than $T_l^o$, so that the nominal SINR requirement is certainly satisfied.
Table 6.3: Tolerable set and intolerable set

<table>
<thead>
<tr>
<th>link</th>
<th>tolerable sets</th>
<th>intolerable sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3, 4</td>
<td>2, 5</td>
</tr>
<tr>
<td>2</td>
<td>4, 5</td>
<td>1, 3</td>
</tr>
<tr>
<td>3</td>
<td>1, 5</td>
<td>2, 4</td>
</tr>
<tr>
<td>4</td>
<td>1, 2</td>
<td>3, 5</td>
</tr>
<tr>
<td>5</td>
<td>2, 3</td>
<td>1, 4</td>
</tr>
</tbody>
</table>

Fig 6.4 depicts an example network with 5 links under the conservative coexistence constraint. We assume that the tolerable sets and the intolerable sets of each link are predetermined as shown in Table 6.3. According to the conservative coexistence constraint, only the following link pairs can coexist: (1, 3), (1, 4), (2, 4), (2, 5), (3, 5). The corresponding conflict graph of this network is shown in Fig. 6.5.

Next, we generalize the above constraint to the MIMO-pipe case by using the concept of “virtual link” introduced in the previous section. Let \( V(l) \) be the set of virtual links corresponding to link \( l \), and \( l_v \in V(l) \) be the virtual link corresponding to the \( v \)-th configuration of link \( l \). As before, we use \( z(t) \) to denote the active virtual links at time slot \( t \), where \( l_v \in z(t) \) and \( z_l(t) = v \), if link \( l \) chooses configuration \( v \) in the slot \( t \).

For virtual link \( l_v \in V(l) \), it has a unique SINR requirement, and thus has a unique initial interference tolerance level \( T_{\text{Io}} \). We also define its tolerable set of virtual links as \( \hat{N}_a(l_v) \) and intolerable set of virtual links as \( \hat{N}_b(l_v) \) in the similar way. We impose the conservative SINR constraint under the MIMO-pipe model as follows:

**Conservative coexistence constraint for the MIMO-pipe model:**

- At each slot, only one virtual link in \( V(l) \) can transmit data.
• For two links $l$ and $k$, their virtual links $l_v$ and $k_j$ can coexist if and only if $l_v \in \hat{N}_a(k_j)$ and $k_j \in \hat{N}_a(l_v)$.

We next devise CSMA algorithm for MIMO link scheduling by using the above conservative coexistence constraint. We combine channels for control message and data transmission, by dividing a time slot into a control slot and a data slot, each with multiple mini-slots as in [5]. During the control slot, each link contends to be included in the decision schedule $\mathcal{M}$ by broadcasting a control message. To ensure that the links in $\mathcal{M}$ can conform the conservative constraints, each virtual link includes the information of its intolerable set in the control message. Once a virtual link $l_v$ sends the control message and successfully joins $\mathcal{M}$, the interfering links in $N(l)$ can check its coexistence relationship with $l_v$ based on the information of $\hat{N}_b(l_v)$, and will give up contending if the coexistence constraint fails to hold. Staring from an empty set, and adding links to $\mathcal{M}$ one-by-one, we can obtain the decision schedule $\mathcal{M}$ such that all the links included in $\mathcal{M}$ can conform the conservative coexistence constraint.

A complication may occur when there is a “collision” during the control slot, i.e., more than one link sends control packet to contend for channel at the same mini-slot, and they conflict under the conservative constraint. For example, suppose link $l_v$ and link $k_j$ that conflict under the conservative constraints contend for channel at the same mini-slot. It is possible that each link can decode its own control packet but fails to decode the packet from the other link. As a result, both links would include themselves in the decision schedule $\mathcal{M}$ independently even they conflict under the conservative constraints. To avoid this situation, we assume that once there is a collision in the control channel (the receiver can detect the collision from the SINR level), each link
will give up joining decision schedule and no one can be included in $\mathcal{M}$ in that slot. Once we obtain a decision schedule $\mathcal{M}$, we remove some links in $\mathcal{M}$ that conflict any link in $\mathbf{z}(t-1)$ and change the status of the rest links in $\mathcal{M}$ with certain probability. The proposed scheduling algorithm is summarized in Algorithm 3.

Algorithm 3 Discrete-time CSMA scheduling under the MIMO-pipe model (at link $l$)

**Initialization:** Find $\hat{\mathcal{N}}_a(l_v)$ and $\hat{\mathcal{N}}_b(l_v)$ for every virtual link $l_v$.

**Selection of decision schedule $\mathcal{M}$**

1. Virtual link $l_v$ selects a random backoff time uniformly in $[1, W_l]$ mini-slots, and begins backoff.

2. Virtual link $l_v$ stops the backoff timer and will not be included in the decision schedule, if one of the following two conditions is valid: (1) $l_v$ hears an INTENT message from virtual link $k_j$, and link $l_v$ and $k_j$ are severely conflicting links, or (2) other virtual links in $\mathcal{V}(l)$ send INTENT messages.

3. After the backoff timer expires, virtual link $l_v$ sends INTENT message to announce its intention to be included in the decision schedule.

4. After $l_v$ sends INTENT message, it keeps sensing the channel. If its INTENT message collides with other control messages, $l_v$ will not be included in $\mathcal{M}(t)$ in this control slot. Otherwise, $l_v$ will join in the decision schedule.

**Setup of the transmission state**

- If virtual link $l_v$ satisfies both the following conditions: 1) $l_v \in \mathcal{M}$; 2) $l_v \notin \hat{\mathcal{N}}_b(k_j)$ and $k_j \notin \hat{\mathcal{N}}_b(l_v)$ for all $k_j \in \mathbf{z}(t-1)$, it will change its state: active ($z_l(t) = v$) with activation probability $p_{l_v}$, and inactive ($z_l(t) = 0$) with probability $\bar{p}_{l_v} = 1 - p_{l_v}$. Otherwise, $l_v$ remains in the same state as in previous time slot, i.e., $z_l(t) = z_l(t-1)$.

**Data transmission**

- If $z_l(t) = v$, $l$ will transmit using configuration $v$ in the data slot.
- If $z_l(t) = 0$, $l$ will not transmit in the data slot.

---

<sup>7</sup>INTENT message has the similar definitions as in [5]. The index of links in $\hat{\mathcal{N}}_b(l_v)$
Observe that in Algorithm 3, each virtual link can make decisions on its transmission state independently. It is clear that the network state $z(t)$ can be modeled as a discrete-time Markov chain, since the state transition probability depends on the selection probability of decision schedule $\mathcal{M}$ and the activation probability of each virtual link. As in [5], the transition probability from $z$ to $z'$ is given as:

$$p(z, z') = \sum_{\mathcal{M} \in A(z, z')} \epsilon(\mathcal{M}) \prod_{i \in a} \tilde{p}_i \cdot \prod_{k \in b} p_k \cdot \prod_{i \in c} p_i \cdot \prod_{j \in d} \tilde{p}_j,$$

(6.15)

where $A(z, z')$ denotes the set of possible decision schedules $\mathcal{M}$ that include all links differ in $z$ and $z'$. Furthermore, $\epsilon(\mathcal{M}) > 0$ is the probability that the decision schedule $\mathcal{M}$ will be chosen in the control slot. For all virtual links included in $\mathcal{M}$ with no severely conflicting links active in the previous slot, they can be classified into four sets: set $a$ denotes the virtual links active in $z$ and inactive in $z'$; set $b$ denotes the virtual links inactive in $z$ and active in $z'$; set $c$ denotes the virtual links which keep active in two states; and set $d$ denotes the virtual links which keep inactive in two states. Also, $p$ and $\tilde{p}$ are the corresponding activation probabilities specified in Algorithm 2. It can be verified that the stationary distribution of feasible state $z^i$ is given by:

$$p(z^i) = \frac{1}{C} \prod_{i \in z^i} \frac{p_i}{\tilde{p}_i},$$

(6.16)

where $C$ is the normalization term satisfying $\sum_i p(z^i) = 1$.

As in the continuous-time case, each $p_l$ can be adapted using local queue information.

**Lemma 6.5.5.** Under the time-scale-separation assumption [4], the CSMA algorithm for MIMO scheduling can achieve any network throughput $\lambda$ in the is included in the INTENT message, so any link $k_j$ receiving this INTENT message can examine if $l_v$ and $k_j$ can coexist.
capacity region corresponding to the conservative coexistence constraint, by adjusting the activation probability of virtual links as follows:

For link $l$,

$$y_{l}(t + 1) = \left[y_{l}(t) + \xi(\lambda_{l} - \theta_{l}(t))\right]^{+},$$

where $y_{l}$ is shown to be proportional to the queue length at link $l$ [4], and $\xi > 0$ is a small constant (step size). Each virtual link can update its activation probability according to

$$p_{lv} = \frac{e^{y_{l}\Theta(v)}}{1 + e^{y_{l}\Theta(v)}}$$

where $\Theta(v)$ is the data rate of configuration $v$.

We provide the proof of Lemma 6.5.5 in Section 6.8.2.

Note that each link may not fully utilize its initial interference tolerance due to the conservative coexistence constraint. Since the feasible states under the conservative SINR constraint will be a subset of those under the nominal SINR constraint, it is clear that the capacity region corresponding to the conservative coexistence constraint is only a fraction of that under the nominal SINR constraint. Hence, the “conservative scheduling” achieves a suboptimal performance. In the following, we will show that the conservative scheduling at least achieves a guaranteed fraction of the optimal throughput region.

6.5.2 Efficiency Ratio of Conservative MIMO-pipe Scheduling

In this section, we characterize the throughput performance achieved by the conservative SINR-based scheduling. Specifically, we provide a lower bound of $\gamma \in [0, 1]$ such that for any traffic arrival rate $\lambda$ in the capacity region under the nominal SINR constraint, $\gamma \lambda$ is supported by the conservative scheduling. The fraction $\gamma$ is called as the efficiency ratio.
Recall that the throughput region of our suboptimal scheduling algorithm is the convex hull of the set of feasible states under the conservative SINR constraint. To compare the throughput region of CSMA algorithm under different interference constraints, it suffices to compare the convex hulls formed by their feasible states.

For convenience, let $S$ and $C$ be the sets of the rate vectors obtained from the feasible states under the nominal SINR model and the conservative SINR model, respectively. For a MIMO-pipe model with $K$ links, we use a $K$-dimension vector to denote the feasible rates, where each element is the link transmission rate at the corresponding state. For each feasible rate $s \in S$, there exists a subset $C \subset C$ such that the set of the active virtual links in $s$, can be “covered” by the union of the sets of the active virtual links for the feasible rate in $C$, i.e.,

$$\{l \in 1, 2, \cdots, K : s_l = r \text{ in State } s\} \subset \bigcup_{c \in C} \{l \in 1, 2, \cdots, K : c_l = r \text{ in State } c\}. \quad \text{(6.17)}$$

Note that there may exist multiple different subsets $C \subset C$ that “cover” the set of the active links of $s$. Nevertheless, we will show that only the subsets with the least cardinality are closely related to the efficiency ratio.

Let $V_k^* \subset C$ be the minimal covering set for state $s_k$ in the sense that 1) $V_k^*$ satisfies (6.17), and 2) for any other subset $V \subset C$ that satisfies (6.17), we have that the cardinality of $V_k^*$ is no larger than that of $V$, i.e., $|V_k^*| \leq |V|$.

Define the effective interference number as the maximum of the cardinalities among the minimal covering set for all the feasible rates in $S$, i.e.,

$$N(S, C) \triangleq \max_{\{k: s_k \in S\}} |V_k^*|.$$
Under the conservative SINR model, any $s_k$ in $S$ can be decomposed into no more than $N(S, C)$ states in $C$, where $N(S, C)$ depends on the coexistence relationship of links.

**Theorem 6.5.5.** The conservative MIMO-pipe scheduling results in an efficiency ratio $\gamma \geq 1/N(S, C)$.

The proof is given in Section 6.8.3.

The above result reveals that the efficiency ratio is bounded from below by the reciprocal of the effective interference number. Note that determining the effective interference number requires globe information of all the feasible states in general. In the following, we develop a local search algorithm to find an upper bound on the effective interference number.

Observe that for any virtual link $l_v$, there may exist a set of virtual links $L = \{l_v\} \cup \{N|N \subset \hat{N}(l_v)\}$, such that the virtual links in $L$ can coexist under the nominal SINR constraint, where $\hat{N}(l_v) = N_a(l_v) \cup N_b(l_v)$. We call $L$ a “local feasible state,” and clearly virtual link $l_v$ can have multiple local feasible states. We use $L(l_v, j)$ to denote the $j$-th local feasible state of $l_v$, and $n_v(l_v, j)$ to denote the number of links in $L(l_v, j)$ severely conflicting with $l_v$ under the conservative SINR constraint, i.e., $n_v(l_v, j) = |L(l_v, j) \cap N_b(l_v)|$. We further define

$$n_e \triangleq \max_{l_v} \max_{L(l_v, j)} n_v(l_v, j).$$

It follows that for any virtual link, $n_v(l_v, j)$ would be no greater than $n_e$. Detailed algorithm to find $n_e$ is provided in Algorithm 4. We next have the following result.

**Theorem 6.5.6.** The effective interference number is upper bounded by $n_e + 1$, i.e., $N(S, C) \leq n_e + 1$. 

160
The proof is given in Section 6.8.4.

Algorithm 4 Local search algorithm

\begin{verbatim}
let \( n_e = 0 \);
for \( l = 1 \) to \( K \) do
  For link \( l \), let \( n_l = 0 \)
  for \( v = 1 \) to \( J \) do
    For virtual link \( l_v \), let \( n_v(l_v) = 0 \)
    repeat
      For local feasible state \( L(l_v, j) \)
      if \( n_v(l_v, j) \geq n_v(l_v) \), then
        \( n_v(l_v) = n_v(l_v, j) \)
      end if
    until all local feasible states of \( l_v \) has been enumerated
    if \( n_v(l_v) \geq n_l \), then
      \( n_l = n_v(l_v) \)
    end if
  end for
  if \( n_l \geq n_e \), then
    \( n_e = n_l \)
  end if
end for
\end{verbatim}

Combining Theorems 6.5.5 and 6.5.6, we conclude that

\[
\gamma \geq \frac{1}{N(S, \mathcal{C})} \geq \frac{1}{n_e + 1}.
\]

(6.18)

6.6 Numerical Examples

In this section, we illustrate, via numerical examples, the performance of the proposed CSMA algorithms in a multi-hop MIMO-pipe network. We explore the cases for both continuous-time model and discrete-time model.

6.6.1 Simulation Settings

Specifically, we study a network with six 4 × 4 MIMO links. Assume that each link has three possible configurations, with data rate 1 (data unit/ms),
2 (data units/ms) and 4 (data units/ms), respectively. We construct the network topology as follows. Consider an area of $20 \times 20$ square unit, we randomly deploy six transmitter-receiver pairs, such that each receiver is within distance 3 from the corresponding transmitter. According to (6.1), the signal power from the transmitter attenuates as it propagates through space. In the simulations, the path loss exponent $\alpha$ is fixed at 2 and the transmission power $P$ is set to 1 unit. For the white noise, we set $SNR_{dB} = 10 \log \frac{P}{\sigma^2} = 20$dB. We also choose $\sigma^2_{int} = 0$, and hence the close-in range of each link includes other 5 links. The SINR requirements corresponding to three configurations are 8dB, 16dB and 24dB, respectively.

We illustrate the queue length behaviors of MIMO-pipe scheduling under different traffic loads. To illustrate the throughput optimality, we first find an arrival rate vector at the boundary of capacity region, denoted as $\lambda$. Then, we consider a “load factor” $\rho$, $\rho > 0$, and set the traffic load at $\lambda = \rho \lambda$ as in [66]. Clearly, the traffic load is in the capacity region if $\rho < 1$ and outside the capacity region if $\rho > 1$. We build up $\lambda$ by using a set of feasible states under the nominal SINR constraints. Specifically, for feasible state $i$, let $c^i$ denote the rate vector of active links, and let $s^i$ denote the summation of the active link rates, i.e., $s^i = \|c^i\|_1$. Among all the feasible rate vectors, let $M$ be the set of vectors with maximal value of $s^i$, i.e., $M = \{c^i : s^i = \max_j s^j\}$. Clearly, a convex combination of a set of rate vectors in $M$ corresponds to a point on the boundary of the capacity region. In the simulations, we simply choose $\lambda = \frac{1}{|M|} \sum c^i$, $c^i \in M$.

6.6.2 Continuous-time Network Model

To illustrate the throughput-optimality, we compare the queue behaviors of continuous-time CSMA algorithm under different traffic loads. Specifically,
the queue length usually keeps increasing if the network throughput cannot meet the traffic demands. Note that a scheduling algorithm is said to be throughput-optimal if it can yield stable queue length behaviors at any traffic load in the capacity region, corresponding to $\rho < 1$ [67]. We first consider $\rho = 0.98$ such that the traffic arrival rate vector $\lambda = \rho \bar{\lambda}$ is in the interior of capacity region. As shown in Fig. 6.6, the scheduling algorithm yields stable queue length behavior at each link, indicating it can achieve network throughput $\lambda$. Fig. 6.7 exemplifies the throughput-optimality by comparing the total queue length under various $\rho$. As expected, the total queue length tends to be stable under traffic load in the capacity region ($\rho < 1$). However, while $\rho > 1$, the queue length grows rapidly, and the system will become unstable, which means the scheduling algorithm fails to support the traffic loads beyond the capacity region.

![Figure 6.6: Continuous-time model: queueing length behavior at each MIMO link with $\rho = 0.98$.](image)

6.6.3 Discrete-time Network Model

We evaluate the conservative CSMA scheduling scheme under the discrete-time model. Due to its throughput sub-optimality, the conservative scheduling
scheme can only achieve a fraction of the capacity region and cannot support all the traffic loads with $\rho < 1$. We illustrate its throughput performance by comparing the total queue lengths under various $\rho$ in Fig. 6.8. We observe that when $\rho \geq 0.6$ the queue length keeps increasing, indicating that the scheme can no longer support the traffic loads with $\rho \geq 0.6$. We also compare the queue behaviors for the continuous-time case and the discrete-time case in Fig. 6.9. In this figure, we depict the total queue lengths averaged over the period from 1600ms to 2000ms. We observe that the queue length corresponding to the discrete-time case grows rapidly at a smaller $\rho$ than that of the continuous-time case, indicating its inferior performance to the continuous-time scheduling scheme.

For this scenario, we find that the effective interference number $N(S, C)$ is no more than 2 by using Algorithm 4 and hence the efficiency ratio $\gamma$ is no less than 0.5 by Theorem 5.1. It follows that the conservative scheduling can at least achieve a $\frac{1}{2}$ fraction of the capacity region, which is confirmed by Fig. 6. Indeed, the network remains stable under traffic load with $\rho = 0.55$. 

Figure 6.7: Continuous-time model: total queue lengths of 6 MIMO links with different $\rho$ values.
Figure 6.8: Discrete-time model: total queue lengths of 6 MIMO links with different $\rho$ values.

Figure 6.9: Comparisons of total queue lengths for continuous-time model and discrete-time model

6.7 Conclusions

We investigated CSMA algorithms in multi-hop MIMO networks under the SINR interference model. To this end, we first developed a MIMO-pipe model that provides the upper layers a set of rates and SINR requirements, which capture the rate-reliability tradeoffs in MIMO communications. We then focused on developing distributed scheduling for MIMO-pipe networks under
the SINR model. Specifically, we explored the CSMA algorithms for MIMO-pipe scheduling in both a continuous-time system and a discrete-time system. Particularly, in the idealized continuous-time CSMA network, we proposed a dual-band approach to facilitate the message passing on interference tolerance levels, and showed that the CSMA scheduling algorithm can achieve throughput optimality under the SINR model. For the more difficult discrete-time case, we developed a “conservative” scheduling algorithm in which a more stringent SINR constraint is imposed. We showed that an efficiency ratio bounded below can be achieved by our distributed scheduling algorithm.

We believe that the studies here on SINR-based distributed scheduling scratch only the tip of the iceberg. Clearly, there are still many open issues in the MIMO network scheduling. We are currently investigating these issues along this avenue.

6.8 Appendix

6.8.1 Proof of Lemma 6.4.4

Following the same lines as in [4], we study the backoff time adaption algorithm based on the following entropy maximization problem:

\[
\text{max} \quad - \sum_i u_i \log u_i \\
\text{s.t.} \quad \sum_i u_i \cdot c_i^l \geq \lambda_l, \\
\quad u_i \geq 0, \quad \sum_i u_i = 1.
\] (6.19)

Assume that each \( i \) relates to a feasible state in the MIMO network. In contrast to the binary data rate in the SISO link case [4], the MIMO link rate \( c_i^l \) can take multiple values depending on the link configuration. If this problem is feasible, the optimal point \( \mathbf{u}^* \) would satisfy the constraint \( \sum_i u_i^* \cdot c_i^l \geq \lambda_l \). That is to say, as long as the optimal value \( u_i^* \) equals the stationary distribution of feasible
states (6.13), then each MIMO link will meet the throughput requirement $\theta_l \geq \lambda_l$ according to (6.14). With this insight, a key challenge is to find a sufficient condition for the equivalence of these two distributions, i.e., $p(z^i) = u_i^*$. The Lagrangian of (6.19) can be written as

$$L_1 = -\sum_i u_i \log u_i + \sum_l y_l (\sum_i u_i \cdot c^i_l - \lambda_l) + \mu (\sum_i u_i - 1) + \sum_i w_i u_i,$$

(6.20)

where $y$, $\mu$ and $w$ are dual variables. Based on the KKT condition, we obtain that

$$u_i^* = \exp(\sum^K_{l=1} y_l c^i_l) \sum_j \exp(\sum^K_{l=1} y_l c^j_l).$$

(6.21)

With (6.13), it can ensure $p(z^i) = u_i^*$ if the following condition holds:

$$\exp\left(\sum^K_{l=1} y_l c^i_l\right) = \exp\left(\sum^K_{l=1} r^i_l\right), \forall i.$$

(6.22)

From $c^i_l = \Theta(z^i_l)$ and $r^i_l = r_{lv}$ when $l_v$ is the active link for state $i$, a sufficient condition for (6.22) is

$$r_{lv} = y_l \Theta(v), \forall v \in [1...J].$$

This condition can also be rewritten as:

$$R_{lv} = \exp(y_l \Theta(v)), \forall v \in [1...J].$$

(6.23)

As in [4], the optimal dual variable $y^*_l$ is essentially proportional to queue length at link $l$, and can be achieved by using the following gradient method:

$$y_l(t + 1) = [y_l(t) + \xi (\lambda_l - \theta_l(t))]^+. $$

Meanwhile, each virtual link can adjust its backoff time according to (6.23). Note that the above adaptive algorithm depends on accurate estimation of link throughput $\theta_l(t)$. As in [4], we take the same time-scale-separation assumption,
i.e., the variable $y_l$ changes slowly enough so that the CSMA Markov chain can converge to its stationary distribution within each duration $t$ and $t + 1$. By doing so, we can always obtain a good estimation of the link throughput.

6.8.2 Proof of Lemma 6.5.5

Based on the Markov chain modeling, the activation probability of each virtual link can be obtained by the same gradient method as in Section 6.4.2. The only additional requirement is that the stationary distribution of the feasible states in the discrete-time network (6.16) equals the distribution (6.21). A sufficient condition for this requirement turns out to be:

$$
\frac{p_{lv}}{p_{lv}} = \exp(y_l\Theta(v)), \quad \forall l, v,
$$

and equivalently

$$
p_{lv} = \frac{e^{y_l\Theta(v)}}{1 + e^{y_l\Theta(v)}},
$$

(6.24)

where $\Theta(v)$ is the data rate of configuration $v$. Clearly, $y_l$ can be achieved along the same line as in the continuous-time network, and each virtual link can update its activation probability according to (6.24). It follows that the adaptive algorithm also requires the time-scale-separation assumption in [4].

6.8.3 Proof of Theorem 6.5.5

For any feasible traffic arrival rate $\lambda = \{\lambda_1, \lambda_2, \cdots, \lambda_K\}^T$ under the SINR model, there exists a state probability vector $P = \{P_1, P_2, \cdots, P_{|S|}\}^T$ such that $\sum_{i=1}^{\vert S \vert} P_i = 1$, and

$$
P^T A^S \succeq \lambda,
$$

(6.25)

where $A^S$ is a $|S| \times K$ matrix, with

$$
A^S_{k,l} \triangleq (\text{Transmission rate of link } l \text{ in state } s_k), \forall k, l.
$$

(6.26)
To show that $\gamma \geq \frac{1}{N(S,C)}$, it suffices to show that there exists a state probability vector $Q = \{Q_1, Q_2, \cdots, Q_{|C|}\}^T$ such that
\[
Q^T A^C \geq \gamma \lambda,
\] (6.27)
where $A^C$ is defined in the same way as $A^S$ in (6.26). We use induction on $|S|$ to show that (6.27) is valid for some $Q$, for any given $P$ satisfying (6.25).

It is easy to verify when $|S| = 1$. Assume that the conclusion holds when $|S| = n$. Now we consider the case $|S| = n + 1$, pick the state $s_k$ in $S$ such that $|V^*_k| = N(S,C)$. Without the loss of generality, suppose $k = n + 1$.

It follows from (6.25) that for $l = 1, 2, \cdots, K$,
\[
\sum_{i=1}^{n} P_i A^S_{i,l} + P_{n+1} A^S_{n+1,l} \geq \lambda_l,
\] (6.28)
which indicates that for $l = 1, 2, \cdots, K$,
\[
\sum_{i=1}^{n} P'_i A^S_{i,l} \geq \lambda'_l,
\] (6.29)
where
\[
P'_i \triangleq \frac{P_i}{1 - P_{n+1}}, \quad \lambda'_l \triangleq \frac{\lambda_l - P_{n+1} A^S_{n+1,l}}{1 - P_{n+1}}.
\] (6.30)

By induction, based on (6.29), there exists $Q'$ such that $\sum_{j=1}^{|C|} Q'_j = 1$, and for $l = 1, 2, \cdots, K$,
\[
\sum_{j=1}^{|C|} Q'_j A^C_{j,l} \geq \gamma' \lambda'_l,
\] (6.31)
where $\gamma' \triangleq \frac{1}{N(S',C)}$ and $S' = s_k, k = 1, 2, \cdots, n$. It is clear that, $\gamma' \geq \gamma$, and it follows that
\[
\sum_{j=1}^{|C|} Q'_j A^C_{j,l} \geq \gamma \lambda'_l, \forall l = 1, 2, \cdots, K.
\] (6.32)

Similar, we can find $Q''$ such that $\sum_{j=1}^{|C|} Q''_j = 1$, and
\[
\sum_{j=1}^{|C|} Q''_j A^C_{j,l} \geq \gamma A^S_{n+1,l}, \forall l = 1, 2, \cdots, K.
\] (6.33)
Define
\[ Q_j \triangleq Q'_j(1 - P_{n+1}) + Q''_j P_{n+1}, \forall j = 1, 2, \cdots, |C|. \tag{6.34} \]

Observe that \( Q = \{Q_1, Q_2, \cdots, Q_{|C|}\}^T \) defined above is a state probability vector, i.e.,
\[ \sum_j Q_j = (\sum_j Q'_j)(1 - P_{n+1}) + (\sum_j Q''_j) P_{n+1} = 1. \tag{6.35} \]

Furthermore, multiplying (6.32) with \((1 - P_{n+1})\) on both sides yields that
\[ \sum_{j=1}^{|C|} Q'_j(1 - P_{n+1}) A_{j,l}^C \geq \gamma (\lambda_l - P_{n+1} A_{n+1,l}^S), \forall l = 1, 2, \cdots, K. \tag{6.36} \]

Further, multiplying (6.33) with \(P_{n+1}\) yields that
\[ \sum_{j=1}^{|C|} Q''_j P_{n+1} A_{j,l}^C \geq \gamma P_{n+1} A_{n+1,l}^S, \forall l = 1, 2, \cdots, K. \tag{6.37} \]

Adding the above two equations together, we see that \( Q \) defined in (6.34) satisfies (6.27), and the proof is concluded.

\section*{6.8.4 Proof of Theorem 6.5.6}

Under the conservative SINR model, we can build a conflict graph \( G \) associated with the MIMO-pipe network, where each vertex corresponds to a virtual link. The feasible state, under the SINR model, \( s_k \in S \) corresponds to a subgraph of \( G \), and the feasible state under the conservative SINR model corresponds to an independent set of \( G \). Let \( G(s_k) \) be the subgraph of \( G \), which only contains the vertexes corresponding to the active virtual links in \( s_k \) and their associated edges.

The value \( |V_k^*| \) relating to state \( s_k \) can be interpreted as the minimum number of independent sets to construct the subgraph \( G(s_k) \). The problem of finding these independent sets boils down to a graph coloring problem. According to graph theory, we can decompose any subgraph \( G(s_k) \) into no
more than $\Delta(G(s_k)) + 1$ independent sets, where $\Delta(G(s_k))$ is the maximum degree of $G(s_k)$.

Next, we establish the relationship between $\Delta(G(s_k))$ and $n_e$ from local search algorithm in Section 6.5. In the conflict graph, let $v(l_v)$ denote the vertex corresponding to virtual link $l_v$. Define $\deg(l_v, G(s_k))$ as the degree of vertex $v(l_v)$ in subgraph $G(s_k)$. Then we have the following result:

$$\max_{v(l_v) \in G(s_k)} \deg(l_v, G(s_k)) = \Delta(G(s_k)).$$  \hspace{1cm} (6.38)

Recall that $n_e$ is the maximum number of links severely conflicting with $l_v$ in any local feasible state under the conservative SINR constraint, where there is no interference from links other than $l_v \cup N(l_v)$. If any link other than $l_v \cup N(l_v)$ is active, some links in $L(l_v, j)$ may no longer satisfy the nominal SINR constraint. Hence, the number of conflicting links which can be active simultaneously with any virtual link $l_v$, under the nominal SINR constraint, must be no greater than $n_e$. Therefore, we conclude that

$$n_e \geq \deg(l_v, G(s_k)), \quad \forall \ l_v \in s_k, \forall \ s_k \in S.$$  \hspace{1cm} (6.39)

It follows that

$$n_e \geq \max_{l_v \in s_k} \deg(l, G(s_k)), \forall \ s_k \in S,$$

$$\quad = \Delta(G(s_k)), \forall \ s_k \in S.$$  \hspace{1cm} (6.40)

In conclusion, $n_e + 1$ is an upper bound for $|V^*_k|$ for $\forall s_k \in S$, and hence an upper bound for $N(S, C)$ as well.
7.1 Introduction

There has recently been considerable interest in complex networks across many domains, such as physics, computer science, biology, economics. In the existing studies, a popular approach is to explore the empirical observations from real-world networks and then to identify some common characteristics of complex networks. It has been shown that many complex systems can evolve into “steady states” where the connectivity degree follows some “typical” distribution, such as the power-law degree distribution in many large networks (see, e.g., [9]), the exponential distribution in email networks [10,11], and the Weibull distribution in some IP graphs [12,13]. Along a different avenue, recent works [14,15] have explored the trend of network evolution based on the observations over time, and their findings show that some growing networks can exhibit surprising transient phenomena such as “network densification” and “shrinking diameter”.

Notably, much effort has been devoted to developing graph models to generate these phenomena. For instance, the steady state characteristics, such as the power-law degree distribution and its variants, could be “regenerated” by the family of models based on the preferential attachment mechanism [68,69]. For the dynamic characteristics, very recent work [15] proposed a kronecker multiplication model, which leads to both network densification and shrinking diameter.
Most of the existing works focus on mechanisms that can yield some given properties, and there has been little work on developing unifying models that naturally “catch” both steady state and dynamic characteristics. It remains intriguing why the observed properties are so prevailing in the real world and under what conditions they would happen. With this motivation, this study aims to develop a mathematically rigorous model towards understanding the following two fundamental issues:

- Why would complex networks always evolve towards these degree distributions, such as power-law, exponential, and Weibull? Further, under what condition would the network evolve to a specific degree distribution?

- Is there an unifying model that naturally approximates network evolution and captures both steady state and transient characteristics?

Specifically, we study the steady state and transient network behaviors from an entropic and Markovian view. Analogous to the Second Law of Thermodynamics, the complex network is viewed as a dynamic system evolving towards a state with maximum network entropy under both external and internal constraints. In particular, we take a network entropy maximization (NEM) view to examine the macroscopic behaviors at the steady state, such as degree distributions, and explore the rationale of their emergence. Further, to characterize the microscopic behaviors of network dynamics, we devise a two timescale Markov model, which could capture both steady state and transient network behaviors. Particularly, it yields a general characterization of the average degree, and hence provides a natural explanation of the surprising network densification phenomenon. We also quantify the conditions under
which the Markov model leads to the NEM degree distributions at the steady state. We elaborate further on this in what follows.

**Network entropy maximization.** In this study, we call the steady state degree distribution the NEM degree distribution, in the sense that it corresponds to maximum network entropy under certain constraints. Specifically, the constraints are quantified by *network satisfaction* and *network cost*. We assume that the new connection brings in some satisfaction to the network while incurring some cost. Intuitively speaking, there are minimum requirements on network satisfaction, subject to given constraints on the corresponding cost. Accordingly, the NEM degree distribution is the solution to a constrained NEM problem, and it takes different forms under different constraints. We show that the NEM degree distribution could be power-law or exponential or Weibull, corresponding to different diminishing network satisfaction effects (cf. [70]). We note that although entropic interpretation has been used to explain some physical and chemical events, the approach of treating the degree distribution as the solution to a constrained NEM problem is new.

**Two timescale Markov model for network dynamics.** To approximate the microscopic behaviors of network dynamics, we next develop a two timescale Markov model where link generation and deletion takes place on a smaller timescale and new node arrivals occur on a larger timescale. As shown in Section 7.4, this two timescale model offers a general framework to study transient behaviors of network dynamics, and particularly, it provides a natural explanation of the surprising network densification phenomenon. Further, it also leads to the NEM degree distributions at the steady state.
In this section, we first focus on the formulation of the NEM problem, based on which we quantify the general form of the NEM degree distribution.

**Steady state objective: network entropy maximization.** In our study, the network entropy is defined as the Shannon entropy of degree distribution, as the measurement of network robustness [71]. Roughly speaking, the network robustness here points to “network survivability” over a wide range of random conditions. That is, without any a priori information of the condition, the network would work best when it “accommodates” the uncertainties of all possible conditions. As expected, it is the maximum entropy degree distribution that exhibits high robustness to environment change and hence high chance to survive. It is in this sense that we argue that the degree distribution corresponding to maximum network entropy is more popular in practical systems.

**External and internal constraints: network satisfaction and cost.** In this study, each node’s satisfaction (or reward, interchangeably) is tied to its connectivity degree; and furthermore, there is a cost associated with building a new connection. For instance, when a node builds its $j^{th}$ connection, it will receive additional satisfaction $x_j$ at the cost $y_j$. Accordingly, for a node with degree $d$, the total satisfaction it can obtain is $f_d = \sum_{j=1}^{d} x_j$, and meanwhile, the associated cost would be $g_d = \sum_{j=1}^{d} y_j$. More generally, there are $K$ types of satisfactions and $L$ types of costs. For each type of satisfaction, there is a minimum requirement on its average value as an external constraint; and furthermore, the network is subject to internal constraints on each type of cost.
Table 7.1: Key notation in the NEM problem

<table>
<thead>
<tr>
<th>Notation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>$x_{jk}$</td>
<td>the $k^{th}$-type satisfaction from the $j^{th}$ connection</td>
</tr>
<tr>
<td>$y_{jl}$</td>
<td>the $l^{th}$-type cost associated with the $j^{th}$ connection</td>
</tr>
<tr>
<td>$f_{dk}$</td>
<td>$k^{th}$-type satisfaction from $d$ connections, $f_{dk} = \sum_{j=1}^{d} x_{jk}$</td>
</tr>
<tr>
<td>$g_{dl}$</td>
<td>$l^{th}$-type cost associated with $d$ connections, $g_{dl} = \sum_{j=1}^{d} y_{jl}$</td>
</tr>
<tr>
<td>$\theta_k$</td>
<td>requirement on the $k^{th}$-type satisfaction</td>
</tr>
<tr>
<td>$\rho_l$</td>
<td>constraint on the $l^{th}$-type cost</td>
</tr>
</tbody>
</table>

Note that the terms “connection” and “degree” refer to inward connections and in-degrees throughout this study (further studies are needed to extend this to out-degree distributions). The key notation for the NEM problem can be found in Table 7.1.

With these definitions, the NEM degree distribution is the solution to the following entropy maximization problem:

$$\text{max} \quad -\sum_d p_d \log p_d$$

s.t. $\sum_d f_{ik} p_d \geq \theta_k, \quad k = 1..K,$

$$\sum_d g_{dl} p_d \leq \rho_l, \quad l = 1..L.$$ (7.1)

This problem can be solved by standard convex programming as shown in the Appendix. The NEM degree distribution turns out to be

$$p_d^* = \frac{\exp(\sum_k \lambda_k^* f_{dk} - \sum_l \nu_l^* g_{dl})}{\sum_{j=1}^{d_{\text{max}}} \exp(\sum_k \lambda_k^* f_{jk} - \sum_l \nu_l^* g_{jl})},$$ (7.2)

where $\lambda^*$ and $\nu^*$ denote the optimal dual variables. Recall that $f_{dk} = \sum_{j=1}^{d} x_{jk}$, $k = 1..K$, and $g_{dl} = \sum_{j=1}^{d} y_{jl}$, $l = 1..L$, the NEM distribution (7.2) can be
rewritten as follows:

\[ p_d^* = \frac{\exp\left(\sum_k \lambda_k^* \sum_j x_{jk} - \sum_l \nu_l^* \sum_j y_{jk}\right)}{\sum_{m=1}^{d_{\text{max}}} \exp\left(\sum_k \lambda_k^* \sum_j x_{jk} - \sum_l \nu_l^* \sum_j y_{jk}\right)}. \]

By defining \( r_d = \exp\sum_k \lambda_k x_{dk} \) and \( c_d = \exp\sum_l \nu_l y_{dl} \), the NEM degree distribution can be rewritten in the following product form:

\[ p_d^* = \frac{1}{C} \prod_{j=1}^{d} \frac{r_j}{c_j}, \quad 1 \leq d \leq d_{\text{max}}, \tag{7.3} \]

where the normalization term \( C \) is \( C = \frac{d_{\text{max}}}{\sum_{d=1}^{d_{\text{max}}} \prod_{j=1}^{d} \frac{r_j}{c_j}} \).

Under different conditions, this NEM distribution could take different forms, including power-law, exponential and Weibull, which will be elaborated further in the following.

7.3 Steady State Characteristics: NEM Degree Distribution

7.3.1 Diminishing Marginal Utility and NEM Degree Distribution

A related work [70] studied a special case of the NEM problem with a single constraint \( \sum_i f_d p_d = \theta \), and it has shown that the NEM distribution could become power-law or exponential under different assumptions on \( f_d \). Specifically, if \( f_d \) follows the “diminishing marginal utility” effect, the satisfaction a node can receive from a new connection would be decreasing with the its degree, following \( x_j = \frac{1}{j} \), from the \( j^{th} \) connection. In this case, \( f_d \) is proportional to \( \log(d) \), and the NEM degree distribution turns out to be power-law.

On the other hand, if each connection brings in the same satisfaction, then \( f_d \) would be proportional to the number of connection \( d \), and the NEM degree distribution becomes exponential.
A natural question to ask is what happens in the continuum between these two extreme cases. To this end, we model the network with the satisfaction diminishing at rate $\alpha$, $0 \leq \alpha \leq 1$, i.e., we define the marginal satisfaction from the $j^{th}$ connection as $x_j = \frac{1}{j^\alpha}$. Clearly, $\alpha = 1$ corresponds to the previous case with $f_d \sim \log(d)$ and $\alpha = 0$ points to the case with $f_d \sim d$. It follows that

$$f_d = \sum_{m=1}^{d} \frac{1}{m^\alpha} \approx \frac{d^{1-\alpha}}{1-\alpha},$$

and the NEM problem in this case boils down to

$$\max \quad -\sum_d p_d \log p_d$$

$$\text{s.t.} \quad \frac{1}{1-\alpha} \sum_d d^{1-\alpha} \cdot p_d = \theta. \quad (7.4)$$

To solve this problem, we define $y = \frac{1}{1-\alpha} d^{1-\alpha}$, and obtain its distribution $P_r(y = \frac{1}{1-\alpha} d_0^{1-\alpha}) = P_r(d = d_0)$. It can be shown that the distribution of $y$ under the maximum entropy principle is an exponential distribution. Accordingly, we have that

$$F_d(d_0) = P_r(d \leq d_0)$$

$$= P_r(y \leq \frac{1}{1-\alpha} d_0^{1-\alpha})$$

$$= 1 - e^{-\left(\frac{d_0^{1-\alpha}}{(1-\alpha)d_0}\right)}. \quad (7.5)$$

We conclude that the PDF of the NEM degree distribution is

$$p_d = \frac{\partial F_d}{\partial d} = \frac{1}{\theta} d^{-\alpha} e^{-\left(\frac{d^{1-\alpha}}{(1-\alpha)d_0}\right)},$$

i.e., it is a Weibull distribution with shape parameter $1 - \alpha$. That is to say, the resulted Weibull distribution exhibits a heavier and heavier tail behavior, as $\alpha$ ranges from 0 to 1.
Summarizing, when the diminishing rate of the satisfaction increases from 0 to 1, the NEM degree distribution “transists” from exponential to Weibull to power-law.

7.3.2 From Preferential Attachment to NEM

It has been shown that the above NEM degree distributions could be generated by existing graph models based on the “generalized” preferential attachment [72]. Simply put, in this mechanism, a number of new links are generated at each time slot. The attachment probability \( q(i) \), i.e., the probability the new link would connect to node \( i \), follows \( q(i) \sim d_i^\beta \). When \( \beta = 1 \), this mechanism becomes the “original” preferential attachment and would generate a graph with a power-law degree distribution [73]. When \( \beta = 0 \), it will lead to an exponential degree distribution [68]. When \( 0 < \beta < 1 \), the generated degree distribution can be approximated by a Weibull distribution \( p_d \sim d^{-\beta}e^{-c_d^{1-\beta}} \) with shape parameter \( 1 - \beta \) [69,72]. In a nutshell, the parameter \( \beta \) plays a critical role in determining the form of the generated degree distribution.

Now, suppose that we use this mechanism to construct a graph under the condition where the network satisfaction diminishes at rate \( \alpha \). Specifically, for an arbitrary node \( i \), the new satisfaction it can receive depends on both the attachment probability \( q(i) \) and the reward from a new connection, i.e., \( x_{d_i} = 1/d_i^\alpha \). Therefore, for node \( i \), the average marginal satisfaction is \( x_{d_i} \cdot q(i) = d_i^{\beta-\alpha} \), which depends on its degree \( d_i \). Intuitively speaking, every node receives satisfaction in a “proportional fairness” manner. Recall that the diminishing rate \( \alpha \) also determines the form of the NEM degree distribution. We conclude that the graph constructed by the generalized preferential attachment with \( \beta = \alpha \) would exhibit the NEM degree distribution corresponding to the condition with diminishing rate \( \alpha \).
7.4 Two Timescale Markov Model for Network Dynamics

In this section, we develop a two timescale Markov model to approximate the network evolution. We show that this model provides a general framework towards understanding both microscopic and macroscopic behaviors of network dynamics.

7.4.1 Two Timescale Markov Model: From Microscopic to Macroscopic

To capture the microscopic behaviors of network dynamics, we propose a two timescale Markov model where link generation and deletion takes place on a smaller timescale and new node arrivals occur on a larger timescale. The network evolution process corresponds to this model is illustrated in Fig. 7.1.

It can be seen that the evolution of network states follows a discrete time Markov chain. The Markov chain structure makes it tractable to analyze the corresponding macroscopic behaviors. For ease of exposition, we first develop a basic Markov model to study the behaviors for a given size network. Based on this, we further explore the transient behaviors for a growing size network corresponding to the two timescale Markov model. We will show that this two timescale Markov model offers new understandings of some transit
network behaviors, such as “network densification.” Further, it also leads to the NEM degree distributions at the steady state.

7.4.2 Basic Markov Model at Smaller Timescale

The basic Markov model is used to approximate the microscopic behaviors for a fixed size network, where the network is modeled as a directed graph with a given maximum in-degree. At each time slot, the graph randomly generates new links as well as deletes existing links. A node with in-degree \( d \) can be connected by a random node with probability \( q_{d,d+1} \) or deletes one of its existing inward links with probability \( q_{d,d-1} \).

Here, we are interested in the evolution of degree connectivity. For ease of exposition, we first look into the degree behavior of an arbitrary node \( k \). Denote \( X_k(t) \) as the in-degree of node \( k \) at time \( t \). Accordingly, the sequence \( \{X_k(t), t = 1, 2, \ldots\} \) is a birth-death process with the state space \( \Omega = \{1, 2, \ldots, d_{\text{max}}\} \), as illustrated in Fig. 7.2. It is clear that this birth-death process is positive recurrent, and hence it has a unique state stationary distribution \( \pi \).

Note that the global-balance equations are given by

\[
\pi_1 \cdot q_{1,2} = \pi_2 \cdot q_{2,1},
\]

and for \( 2 \leq d \leq d_{\text{max}} - 1 \),

\[
\pi_d \cdot (q_{d,d+1} + q_{d,d-1}) = \pi_{d-1} \cdot q_{d-1,d} + \pi_{d+1} \cdot q_{d+1,d}.
\]

Then, the stationary distribution can be shown to be

\[
\pi_1 = (1 + \sum_{k=2}^{d_{\text{max}}} \prod_{j=2}^{k} q_{j-1,j})^{-1},
\]

181
and for $2 \leq d \leq d_{\text{max}},$

$$\pi_d = \frac{\prod_{j=2}^{d} \frac{q_{j-1,j}}{q_{j,j-1}}}{1 + \sum_{k=2}^{d_{\text{max}}} \prod_{j=2}^{k} \frac{q_{j-1,j}}{q_{j,j-1}}}.$$  \hspace{1cm} (7.9)

### 7.4.3 Two Timescale Markov Model and Transient Behaviors

We next study the two timescale model to account for node arrivals. In this model, the graph randomly generates new links as well as deletes old links on a smaller timescale, as captured in the basic Markov model. Furthermore, on a larger timescale, new node arrivals can occur and the network size would increase accordingly. Without loss of generality, we assume that the maximum degree is increasing with the network size $n$, following $d_{\text{max}}(n) = f(n)$. The details for this model are summarized in Algorithm 5. Simply put, due to node arrivals and hence the increasing maximum degree, the corresponding Markov chain in Fig. 7.2 would be time nonhomogeneous, which means that the degree distribution varies over time. However, since the new node arrives on a larger timescale, the degree distribution would vary slowly between the arrival intervals, and can be approximated by a distribution changing with the network size $n$, denoted as $p_d(n)$.

We next study the transient properties corresponding to this two timescale model, and particularly focus on the average degree $\bar{e}(n)$ with the growing net-
Algorithm 5 Two timescale Markov model for a growing network

**Initialization**

The graph starts with \( n = n_0 \) nodes. Denote the set of nodes as \( V \). Each node generates \( m_0 \) links and each link randomly connects to other nodes.

**Network dynamics with node arrivals**

1. **Link generation and deletion on a smaller timescale**
   
   a) Set \( S_0 = \emptyset \) and \( S_1 = \emptyset \).
   
   b) Uniformly select \( m \) nodes in \( V \), and put them into \( S_0 \). Uniformly select \( m \) nodes in \( V \setminus S_0 \), and put them into \( S_1 \).
   
   c) **link generation** For each node \( l \in S_0 \), randomly select a node \( k \) in \( V \), and create a link from \( k \) to \( l \) with probability \( q_{d_l, d_l+1} \).
   
   d) **link deletion** For any node \( i \in S_1 \), randomly select one of its inward link and delete it with probability \( q_{d_i, d_i-1} \).

2. **Node arrivals on a larger timescale**

   Increase the network size by \( n = n + 1 \) and update the maximum degree following \( d_{\text{max}}(n) = \hat{f}(n) \).

work size. Clearly, \( \bar{e}(n) \) depends on the degree distribution \( p_d(n) \). It has been shown in [74, 75] that when the network size grows larger and larger, the degree distribution is stretched over a wider support range but the shape remains approximately the same. In what follows, we assume that this is the case, and explore the corresponding dynamic behaviors of the average degree with the increasing network size.

**Case 1: Exponential degree distribution.** Suppose \( p_d(n) \sim e^{-\lambda d} \), \( \lambda > 0 \), \( 1 \leq d \leq d_{\text{max}}(n) \). It follows that the average degree is given by

\[
\bar{e}(n) = \frac{1}{C} \sum_{d=1}^{d_{\text{max}}(n)} d e^{-d\lambda} \\
= \frac{1}{C} \cdot \left\{ \frac{2e^{-2\lambda} - 3e^{-3\lambda}}{(1 - e^{-\lambda})^2} - \Theta(d_{\text{max}}(n)e^{-d_{\text{max}}(n)\lambda}) \right\},
\]

183
where the normalized term \( C \) is
\[
C = \sum_{d=1}^{d_{\max}(n)} e^{-d\lambda} = 1 - \Theta(e^{-\lambda(d_{\max}(n)+1)}).
\]
In practical scenarios, \( d_{\max}(n) \) is typically large (see, e.g., [11]), indicating that the term \( \Theta(d_{\max}(n)e^{-\lambda d_{\max}(n)}) \) is very small and can be safely neglected. It follows that the average degree \( \bar{e}(n) \) stays more or less constant with the increasing network size.

**Case 2: Weibull degree distribution.** When \( p_d(n) \) follows a Weibull degree distribution, i.e., \( p_d(n) \sim d^{\gamma-1}e^{-d^\gamma}, \gamma > 0, 1 \leq d \leq d_{\max}(n) \), it is difficult to quantify the relationship between \( d_{\max}(n) \) and \( \bar{e}(n) \) in a closed-form. Nevertheless, we can still estimate the general trend of \( \bar{e}(n) \) when \( d_{\max}(n) \) grows large. That is, when \( d_{\max}(n) \to \infty \), the average degree \( \bar{e}(n) \) could be well approximated by the following integral:
\[
\lim_{d_{\max}(n) \to \infty} \bar{e}(n) \approx \gamma \int_{x=0}^{\infty} x^{\gamma}e^{-x^\gamma}dx = \Gamma(1 + \frac{1}{\gamma}). \tag{7.10}
\]
It can be seen from (7.10) that the characteristic of average degree depends on the tail behaviors of the Weibull distribution. For a heavy-tailed Weibull distribution with \( 0 < \gamma < 1 \), \( \Gamma(1 + \frac{1}{\gamma}) \) can be very large, e.g., \( \Gamma\left(1 + \frac{1}{\gamma}\right) \approx 3 \cdot 10^6 \) when \( \gamma = 0.1 \). That is to say, the average degree would increase dramatically as \( d_{\max}(n) \) increases. Roughly speaking, when the network size increases, the corresponding increment in the average degree is significant. In contrast, when the Weibull distribution is light-tailed with \( \gamma > 1 \), based on the properties of Gamma function, \( \Gamma(1 + \frac{1}{\gamma}) \) cannot be greater than 2, indicating that the average degree is nearly constant with increasing \( d_{\max}(n) \) in the presence of network growth.

**Case 3: Power-law degree distribution.** For the case of power-law degree distributions, we have the following results.
Proposition 4.1. For a network with a power-law degree distribution \( p_d(n) = \frac{1}{C} \cdot d^{-\gamma}, 1 \leq d \leq d_{\text{max}}(n), \) and \( \gamma > 0, \) we have that

\[
\bar{e}(n) \sim \begin{cases} 
  d_{\text{max}}(n) & \text{if } 0 < \gamma \leq 1, \\
  d_{\text{max}}(n)^{2-\gamma} & \text{if } 1 < \gamma < 2, \\
  \text{constant} & \text{if } \gamma > 2.
\end{cases}
\]

(7.11)

Proof. In the case \( 0 < \gamma \leq 1, \) the normalization term is \( C = \sum_{d=1}^{d_{\text{max}}(n)} d^{-\gamma} = \Theta(d_{\text{max}}(n)^{1-\gamma}). \) The average degree \( \bar{e}(n) \) can be derived as

\[
\bar{e}(n) = \sum_{d=1}^{d_{\text{max}}(n)} d \cdot p_d = \frac{1}{C} \sum_{d=1}^{d_{\text{max}}(n)} d^{-\gamma+1} = \Theta(d_{\text{max}}(n)).
\]

If \( \gamma > 1, \) the series \( \sum_{d=1}^{\infty} d^{-\gamma} \) converges, and hence \( C \) can be treated as a constant. Furthermore, in case \( 1 < \gamma < 2, \) we obtain

\[
\bar{e}(n) = \sum_{d=1}^{d_{\text{max}}(n)} d \cdot p_d = \frac{1}{C} \sum_{d=1}^{d_{\text{max}}(n)} d^{-\gamma+1} = \Theta(d_{\text{max}}(n)^{2-\gamma}).
\]

In the case \( \gamma > 2, \) the series \( \sum_{d=1}^{\infty} d^{-\gamma+1} \) converges, which indicates that the average degree \( \bar{e}(n) \) converges to a constant value.

\[\square\]

Simply put, the average degree increases with the network growth if the degree distribution takes a heavy-tailed Weibull distribution or a power-law distribution with \( \gamma < 2; \) and it would stay constant over time when the degree distribution follows exponential or light-tailed Weibull or power law with \( \gamma > 2. \) In a nutshell, the transient behavior of the average degree hinges heavily on the tail behavior of the degree distributions. Roughly speaking, the average degree would increase more likely if the degree distribution has a more heavy

\[1\] We here assume the minimum degree to be 1. This study can be further generalized to the cases of other values of the minimum degree.
tail behavior. This general characterization of the average degree offers natural explanations of some network phenomena, such as network densification, for which we turn our attention to next.

Towards understanding network densification. As shown above, when \( p_d(n) \) follows a power-law distribution with \( p_d(n) \sim d^{-\gamma} \), \( 0 < \gamma < 2 \), and \( d_{\text{max}}(n) = \varepsilon n^{\sigma} \), \( 0 < \sigma < 1 \), the average degree is given as

\[
\bar{e}(n) \sim \begin{cases} 
  n^\sigma & \text{if } 0 < \gamma \leq 1, \\
  n^{(2-\gamma)\sigma} & \text{if } 1 < \gamma < 2.
\end{cases} 
\] (7.12)

Clearly, the average degree \( \bar{e}(n) \) increases as a function of the network size \( n \), proportional to \( n^{\delta} \), \( 0 < \delta < 1 \). We note that this characteristic corroborates the recently observed “network densification” phenomenon, i.e., the total number of edges in the network grows superlinearly with the network size [14, 76]. Therefore, Equation (7.12) and the related two timescale model provide a natural explanation for the emergence of “network densification” phenomenon.

In the simulations, we consider a network growing from \( 10^3 \) nodes to \( 10^4 \) nodes. The network takes a power-law degree distribution with \( \gamma = 1.2 \) and the maximum degree follows \( d_{\text{max}}(n) = \varepsilon n^{0.6} \). According to (7.12), the average degree \( \bar{e}(n) \) should increase proportional to \( n^{0.48} \). As illustrated in the Log-log plot of Fig. 7.3, the average degree from the simulation exhibits the network densification and its slope corroborates the analytical result well.
Figure 7.3: The average degree obeys densification for the case of a power-law degree distribution.

7.4.4 From Two Timescale Model to the NEM Distributions

Based on the basic Markov model, the degree distribution corresponding to the two timescale Markov model can be shown as

\[
p_1(n) = (1 + \sum_{k=2}^{d_{\text{max}}(n)} \prod_{j=2}^{k} \frac{q_{j-1,j}}{q_{j,j-1}})^{-1}, \quad (7.13)
\]

and for \(2 \leq d \leq d_{\text{max}}(n)\),

\[
p_d(n) = \frac{\prod_{j=2}^{d} q_{j-1,j}}{1 + \sum_{k=2}^{d_{\text{max}}(n)} \prod_{j=2}^{k} \frac{q_{j-1,j}}{q_{j,j-1}}}. \quad (7.14)
\]

Recall that the NEM degree distribution \(p^*\) also takes the similar product form in (7.3). Specifically, when the transition probabilities satisfy the following conditions

\[
\frac{q_{d-1,d}}{q_{d,d-1}} = \frac{r_d}{c_d}, \quad 2 \leq d \leq d_{\text{max}}(n), \quad (7.15)
\]

this Markov model can finally lead to a graph with a given NEM degree distribution \(p^*\) at the steady state.
Figure 7.4: Numerical simulations for a network model with a power-law degree distribution, with $n = 10000$.

We next show the emergence of NEM degree distributions, such as power-law and exponential via the following simulations. We consider a network growing from 1000 nodes to 10000 nodes, and the maximum degree follows $d_{\text{max}}(n) = \epsilon n$, $\epsilon = 0.005$.

**Special case with power-law degree distribution.** The NEM degree distribution is $p_d \sim d^{-\gamma}$, which can be rewritten in the following product form:

$$p_d \sim \prod_{j=2}^{d} \left( \frac{j-1}{j} \right)^\gamma, \quad d \geq 2.$$  

Accordingly, as long as the condition $\frac{q_d-1}{q_d,d-1} = \left( \frac{d-1}{d} \right)^\gamma$ always holds, the model finally leads to the power-law degree distribution. Suppose $\gamma = 3$ in the simulation. As depicted in Fig. 7.4, when the network size reaches 10000, the degree distribution from the simulation is close to the expected $p_d \sim d^{-3}$.

**Special case with exponential degree distribution.** As depicted in Fig. 7.5, the model also leads to a exponential degree distribution, following $p_d \sim \exp(-\lambda d)$, where we set $\lambda = 0.3$ in this simulation. Further, Fig. 7.6 indicates that the average degree stays constant with the growing network size. It demonstrates our analytical result on the transient behavior of the average
Figure 7.5: Numerical simulations for a network model with a exponential degree distribution, with $n = 10000$

Figure 7.6: The average degree stays constant for the case with a exponential degree distribution.

degree with a exponential degree distribution.

7.5 Solution to the Constrainted NEM Problem

The problem (7.1) can be solved by the dual decomposition. Denote the dual variable for the constraint on satisfaction as $\lambda_k$ and the dual variable for the
constraint on cost as \( \nu_l \). It follows that the corresponding Lagrange is:

\[
L(\mathbf{p}; \lambda_k, \nu_l, z) = -\sum_d p_d \log p_d + \sum_k \lambda_k (\sum_d f_{dk} p_d - \theta_k) \\
- \sum_l \nu_l (\sum_d g_{dp} - \rho_l) + z (\sum_d p_d - 1).
\]  \hspace{1cm} (7.16)

For given dual variables, the variable \( \mathbf{p} \) that maximizes the Lagrange can be obtained as follows: setting \( \frac{\partial L}{\partial p_d} = 0 \), we get

\[
\log p_d = \sum_k \lambda_k f_{dk} - \sum_l \nu_l g_{dl} + z - 1.
\]

Since \( \sum_d p_d = 1 \), it follows that

\[
p_d = \frac{\exp(\sum_k \lambda_k f_{dk} - \sum_l \nu_l g_{dl})}{C}, \hspace{1cm} (7.17)
\]

where \( C = \sum_{d=1}^{D} \exp(\sum_k \lambda_k f_{dk} - \sum_l \nu_l g_{dl}) \) is the normalization item. The next step is to decide the dual variable \( \lambda \) and \( \nu \), plugging (7.17) into Lagrange (7.16), and we can further obtain the Lagrange dual function:

\[
g(\lambda, \nu) = \max_{\mathbf{p}} L(\mathbf{p}; \lambda_k, \nu_l, z) \\
= \log C - \sum_k \lambda_k \theta_k + \sum_l \nu_l \rho_l \hspace{1cm} (7.18)
\]

\[
= \log \sum_d \exp(\sum_k \lambda_k f_{dk} - \sum_l \nu_l g_{dl}) - \sum_k \lambda_k \theta_k + \sum_l \nu_l \rho_l.
\]

Accordingly, the dual problem becomes

\[
\min_{\lambda, \nu} g(\lambda, \nu) \hspace{1cm} (7.19)
\]

s.t. \( \lambda \succeq 0, \nu \succeq 0 \).

This dual problem can be solve by using the gradient algorithm as follows:

\[
\lambda_k(t + 1) = \lambda_k(t) - a(t) \cdot \frac{\partial g(\lambda, \nu)}{\partial \lambda_k} \\
= \lambda_k(t) - a(t) \cdot \left( \frac{1}{C} \sum_d f_{dk} \exp(\sum_k \lambda_k f_{dk} - \sum_l \nu_l g_{dl} - \theta_k) \right) \\
= \lambda_k(t) - a(t) \cdot \left( \sum_d f_{dk} p_d - \theta_k \right),
\]
\[ \nu_l(t + 1) = \nu_l(t) - a(t) \cdot \frac{\partial g(\lambda, \nu)}{\partial \nu_l} \]
\[ = \nu_l(t) - a(t) \cdot \left( \rho_l - \frac{1}{C} \sum_d g_{dl} \exp\left( \sum_k \lambda_k f_{dk} - \sum_l \nu_l g_{dl} \right) \right) \]
\[ = \nu_l(t) - a(t) \cdot \left( \rho_l - \sum_d g_{dl} p_d \right), \]

where \( a(t) \) is a small positive step size, and \( p_d \) is calculated from (7.17) based on the current \( \lambda \) and \( \nu \). By this algorithm, we can finally achieve the optimal dual variable \( \lambda^* \) and \( \nu^* \), and hence the optimal degree distribution of problem (7.1) is
\[ p_d^* = \frac{\exp(\sum_k \lambda^*_k f_{dk} - \sum_l \nu^*_l g_{dl})}{\sum_{j=1}^{d_{\text{max}}} \exp(\sum_k \lambda^*_k f_{jk} - \sum_l \nu^*_l g_{jl})}. \]  

7.6 Conclusions

We studied the dynamics of complex networks, aiming to seek answers to the following two questions: 1) Why would complex networks always evolve towards these degree distributions, such as power-law, exponential, and Weibull; and further, under what condition would the network evolve to a specific distribution? 2) Is there an unifying model that naturally approximates network evolution and captures both steady state and transient characteristics? To that end, we studied the network dynamics from an entropic and Markovian perspective. We first took a network entropy maximization (NEM) view to examine network steady state characteristics, in terms of degree distributions, and explored the underlying rationale connecting network entropy and widely observed phenomena, such as power law degree distributions, exponential degree distributions and Weibull degree distributions. Next, to capture the microscopic behaviors of network dynamics, we developed a two timescale
Markov model. This two timescale model offers a general framework to study transient behaviors of network dynamics, and particularly it provides a natural explanation of the surprising network densification phenomenon. Further, it also leads to the NEM degree distributions at the steady state.

We believe that the multi-scale modeling, built on a cohesive synergy of Markovian and entropic views, has great potential to yield a fundamental understanding of complex network dynamics, and the studies here on network dynamics scratch only the tip of the iceberg. There are still many open questions on complex networks, and we are currently investigating related issues along this avenue.
Chapter 8

CONCLUSIONS AND FUTURE WORK

8.1 Conclusions

We have studied two fundamental issues in cyber-physical systems: network interdependence and information dynamics. The first thrust is targeted at understating the impact of network interdependence. The second thrust is focused on the information dynamics in complex networks.

Chapter 2 and Chapter 3 studied the robustness of a cyber-physical system in which a cyber-network overlays a physical-network. In Chapter 2, to improve network robustness against random node failures, we developed and studied a regular allocation strategy that allots a fixed number of inter-network edges to each node. Our findings revealed that the proposed regular allocation strategy yields the optimal robustness amongst all strategies when no information regarding the intra-topologies of each individual network is available. We expect that in the presence of such information, the topology of the networks can be exploited to improve further the robustness of cyber-physical systems against cascading failures.

In Chapter 3, we used the threshold model to capture the node failures in the physical infrastructure network and the GCC model for that in the cyber network. We showed that the developed CPS model can naturally capture some key features of practical cyber-physical systems which are not captured in the existing studies. Further, our results revealed that the dense connectivity in the physical network would likely make the cyber-physical system more vulnerable to cascading failures. Finally, we developed a strategy to improve
the system robustness by enabling a fraction of nodes to be autonomous in the sense that the nodes can support by themselves. We showed that the autonomous nodes can significantly enhance the robustness of cyber-physical systems under heavy attacks.

In Chapter 4, we characterized the critical threshold and the asymptotic size of information epidemics in an overlaying social-physical network. To capture the spread of information, we considered a physical information network that characterizes the face-to-face interactions of human beings, and some overlaying online social networks (e.g., Facebook, Twitter, etc.) that are defined on a subset of the population. Assuming that information is transmitted between individuals according to the SIR model, we showed that the critical point and the size of information epidemics on this overlaying social-physical network can be precisely determined by employing the approaches on inhomogeneous random graphs. We believe that our findings here shed light on the further studies on information (and influence) propagation across social-physical networks.

In Chapter 5, we explored the diffusion of real-time information in social networks. We developed an overlaying social-physical network that consists of an online social network and a physical information network with clique structure. We theoretically quantified the condition and the size of information epidemics. To the best of our knowledge, our study is the first work on the diffusion of real-time information with consideration on the clique structure in social networks. We believe that our findings will offer initial steps towards understanding the diffusion behaviors of real-time information.

In Chapter 6, we investigated CSMA algorithms in multi-hop MIMO networks under the SINR interference model. To this end, we first developed
a MIMO-pipe model that provides the upper layers a set of rates and SINR requirements, which capture the rate-reliability tradeoffs in MIMO communications. We then focused on developing distributed scheduling for MIMO-pipe networks under the SINR model. Specifically, we explored the CSMA algorithms for MIMO-pipe scheduling in both a continuous-time system and a discrete-time system. Particularly, in the idealized continuous-time CSMA network, we proposed a dual-band approach to facilitate the message passing on interference tolerance levels, and showed that the CSMA scheduling algorithm can achieve throughput optimality under the SINR model. For the more difficult discrete-time case, we developed a “conservative” scheduling algorithm in which a more stringent SINR constraint is imposed based on the MIMO-pipe model. We showed that an efficiency ratio bounded below can be achieved by our distributed scheduling algorithm.

In Chapter 7, we studied the dynamics of complex networks. To that end, we studied the network dynamics from an entropic and Markovian perspective. We first took a network entropy maximization (NEM) view to examine network steady state characteristics, in terms of degree distributions, and explored the underlying rationale connecting network entropy and widely observed phenomena, such as power law degree distributions, exponential degree distributions and Weibull degree distributions. Next, to capture the microscopic behaviors of network dynamics, we developed a two timescale Markov model. This two timescale model offers a general framework to study transient behaviors of network dynamics, and particularly it provides a natural explanation of the surprising network densification phenomenon. Further, it also leads to the NEM degree distributions at the steady state.
8.2 Future Work

We have studied two fundamental issues in cyber-physical systems: network interdependence and information dynamics. There are a number of directions deserving further investigations.

8.2.1 Network Interdependence in Cyber-Physical Systems

In Chapter 2 and Chapter 3, we have studied the robustness of a cyber-physical system in which a cyber-network overlays a physical-network. We modeled both cyber and physical networks as random graphs and assumed that the realization of the graph topology is unknown when allocating the inter-edges between the two networks. However, in many practical cyber-physical systems, such as smart-grids, the network topology structures are usually fixed. Furthermore, the designer could obtain the whole picture of the system before allocating the inter-edges. It is of great interest to consider the case when topology information of two networks is given. Clearly, we can better design the interdependence structure between two networks leading to higher system robustness if we know the intra-edge topology of each network. However, our previous analytical tools based on random graph theory are no longer applicable to these deterministic networks. Therefore, there is a need to develop a new theoretical framework on analyzing and optimizing the system robustness when the network topology information is given.

8.2.2 Information Dynamics in Cyber-Physical Systems

In Chapter 4 and Chapter 5, we have considered information diffusion in an overlaying social-physical network. One important assumption of the system model is that each node independently chooses its neighbors according to
the given degree distribution. Note that in practical scenarios, there could be certain degree of correlations between the connectivity in both networks. Specifically, if two persons are friends, they are more likely to become online friends in Facebook. It is of equal importance to consider the impact of such correlation on information diffusion process. We expect that such studies can be carried out in the same framework based on random graph theory.

Another interesting issue is the joint impact of the network coupling and the hierarchical structure. The hierarchical structure widely exists in Internet [14,77,78] as well as in social organizations [79]. Although the hierarchical structure has been studied in single networks, it still remains open to investigate its impact in coupled networks. For example, in Fig. 8.1, we assume that the information can spread between two networks either across the same layers or different layers, following different manners. Clearly, the hierarchical structure makes the information propagate in a more complicated way and hence the analysis of information epidemic could be particularly challenging.

Up till now, we have focused on network interdependence from two perspectives: system robustness and information diffusion. Another interesting issue is the interactions between the social network and the underlaying
Internet. In the recent ten years, the booming of Web 2.0 technology and mobile social networks has been greatly facilitating the information propagation among online users and also results in a surge of traffic demand on Internet. Clearly, it brings in more stringent requirement on Internet QoS, including new challenges in congestion control, routing, scheduling and cross-layer optimization. On the other hand, the QoS offered by Internet could in turn impact the information diffusion as well as the personal interactions over social network [80]. There is thus a need to develop a new framework for modeling the interdependence between social networks and the underlying Internet, and to develop network management algorithms that improve the QoS of social network services.

In Chapter 6, we studied SINR-based distributed scheduling. Note that there are still many open issues in the MIMO network scheduling. One interesting issue is how to generalize the MIMO-pipe model into different types of channel fading scenarios. In addition to the SINR level, it is also intriguing to consider other parameters in a realistic MIMO scenario to evaluate the QoS of MIMO communication. It is worth studying the joint design of link scheduling and dynamic power control to better leverage the interference among MIMO links. For the more practical discrete-time case, it remains open to develop a CSMA scheduling algorithm with throughput-optimality under the SINR interference model.
REFERENCES


