

Inefficiency on Networks

Alex Rose, MMath.

Thesis submitted to The University of Nottingham
for the degree of Doctor of Philosophy

January 2018

Glory be to the Father, and to the Son: and to the Holy Ghost;
As it was in the beginning, is now, and ever shall be: world without end. Amen.

Abstract

AN ECONOMY IN which individual agents each choose their optimum strategy does not necessarily optimise the average pay-off per individual. One measure of the global inefficiency of such an economy is known as the Price of Anarchy (PoA). Many studies have considered the PoA in the context of traffic networks, in which traffic is routed along paths between origin-destination nodes and a cost or “latency” is incurred for traversing the edges of the network. Costs associated to edges are increasing functions of the traffic routed to the edge, creating the effect of congestion.

The majority of these studies have focussed on deriving upper bounds on the PoA under various conditions, such as polynomial edge cost functions. While these studies are useful for identifying the maximum potential inefficiency, they neglect to consider the global causes for inefficiency or how inefficiency can be reduced. Moreover, they do not consider the PoA outside of the context of traffic networks.

This thesis considers the variation of the PoA with the total traffic volume and the potential for congestion, with the results displaying new and non-trivial scaling effects. An investigative study determines the causes of the scaling effects and identifies conditions for which high levels of inefficiency occur. A method for reducing the PoA by restricting the amount of traffic on certain links is presented and analysed.

A novel context for the PoA involving the spatial, pairwise matching of agents is also considered. Motivated by the recent emergence of ride-sharing companies, this model considers the inefficiency when agents seek to pair with their nearest neighbour, compared to the optimum matching which minimises the distance between paired agents. The PoA is again analysed as a function of various system parameters. When agents enter the system at different times, it is shown that it may be more efficient to delay a matching until multiple pairs of agents have entered, and the “Price of Impatience” is introduced to measure the inefficiency of matching agents at their entry times.

Acknowledgements

I would like to thank the EPSRC for providing my funding, my supervisors for their enormous contribution to the research and writing of this thesis, and my parents for all of their love and support.

Contents

1	Motivation and Literature Review	1
1.1	Justification for the Research	1
1.2	Introduction to network theory	3
1.3	Analysis of network structure	4
1.4	Network generation models	6
1.4.1	Erdős–Rényi random graphs	6
1.4.2	Watts–Strogatz model	7
1.4.3	Barabási–Albert model	9
1.4.4	Jackson–Rogers hybrid model	10
1.5	Traffic routing on networks	11
1.5.1	Notation	11
1.5.2	User Equilibrium and System Optimum traffic flows	12
1.6	The price of anarchy	14
1.7	Bipartite graph matchings	19
1.8	Summary	21
2	Phenomenology of the lattice network model	23
2.1	Introduction	23
2.2	Method for computing the traffic flows	24
2.3	Critique of original study	26
2.4	Results	29
2.5	Conclusion	34

3	The cause of inefficiency in the lattice network model	37
3.1	Introduction	37
3.2	The simplified model	37
3.3	Notation	38
3.4	Identifying local maxima of $\mathcal{P}$	40
3.4.1	Derivation of route transition points	40
3.4.2	Variation of $\mathcal{P}$ with respect to T	42
3.4.3	Variation of $\bar{\mathcal{P}}$ with respect to p	45
3.5	Application to the lattice network model	47
3.5.1	Variation of $\mathcal{P}$ with T	49
3.5.2	Variation of $\bar{\mathcal{P}}$ with p	50
3.5.3	Dependence of $\bar{\mathcal{P}}(p)$ on lattice aspect ratio	52
3.6	Summary	54
4	Restricting edge capacities: a method for reducing inefficiency	55
4.1	Introduction	55
4.2	Methodology	56
4.3	Results	56
4.4	Discussion	58
4.5	Testing the theory using the simplified model	63
4.6	Effect of the cost of cap implementation	65
4.7	Summary	66
5	Inefficiency for bipartite matching of agents	67
5.1	Introduction	67
5.2	Methodology	68
5.3	Matchings on a square domain	69
5.3.1	Variation with the size of the domain	71
5.3.2	Variation with the number of pairs of agents	71
5.3.3	Variation with aspect ratio of domain	74

CONTENTS

5.3.4	Variation with the ratio of riders to drivers	74
5.3.5	Summary	76
5.4	Matchings on networks	78
5.4.1	Diamond lattice network	79
5.4.2	Random graphs	79
5.4.3	Small-world networks	84
5.4.4	Scale-free networks	89
5.4.5	Summary	91
5.5	The price of impatience	93
5.5.1	Introduction	93
5.5.2	Methodology	94
5.5.3	Results	97
5.6	Summary	97
6	Conclusions and Further Work	102
6.1	Traffic network model	102
6.1.1	Summary	102
6.1.2	Further research	103
6.2	Bipartite matching model	105
6.2.1	Summary	105
6.2.2	Further research	106
	Appendix	108
A.1	Variation of $\mathcal{P}$ with p	108
A.1.1	$\mathcal{P}(N_{k^*})$ increasing	109
A.1.2	$\mathcal{P}(N_{k^*})$ decreasing	111
A.2	Solving for $\mathcal{P}$ for the simplified model	112
A.3	Variation with T	114
A.3.1	$\mathcal{P}(T)$ constant or decreasing	114
	References	116

CONTENTS

List of Symbols and Abbreviations	119
--	------------

Motivation and Literature Review

1.1 Justification for the Research

The efficiency of the allocation of scarce resources is fundamental in any economy. In the absence of any external force, the agents within an economy will generally choose to optimise their individual pay-offs. Such an economy, however, does not necessarily maximise global efficiency. In game theory, the shortcoming of this state — known as the Nash Equilibrium — has been demonstrated by the well-known Prisoner's Dilemma [1] (see figure 1.1.1), which proves that when individuals pursue their own interests without cooperation, the outcome can be sub-optimal for all actors involved.

Prisoner A \ Prisoner B	Prisoner B stays silent (<i>cooperates</i>)	Prisoner B betrays (<i>defects</i>)
	Prisoner A stays silent (<i>cooperates</i>)	Prisoner A betrays (<i>defects</i>)
Prisoner A stays silent (<i>cooperates</i>)	Each serves 1 year	Prisoner A: 3 years Prisoner B: goes free
Prisoner A betrays (<i>defects</i>)	Prisoner A: goes free Prisoner B: 3 years	Each serves 2 years

FIG. 1.1.1: The Prisoner's Dilemma. ¹

Another inherent characteristic of an economy is the interconnectedness of its components, whereby local events can have multiple knock-on effects. For example, if a business owner discovers a cheaper method for producing a product, then they can

¹Image downloaded from https://en.wikipedia.org/wiki/Prisoner's_dilemma

reduce the price they charge for the product while maintaining profits. This in turn forces every other business within that industry to lower their prices or risk losing customers and going out of business. Purchasers of the product benefit from the lower prices by saving some of their money, which they can then spend on other products that they otherwise would not have purchased. An economy is therefore analogous to a complex network [2]. This thesis aims to draw together the above two fundamental features of an economy — the self-seeking nature of agents and the underlying complex network structure — with the goal of identifying the causes of and conditions for high inefficiency.

In this thesis, two models of economic systems are considered and a measure known as the Price of Anarchy (PoA) [3] is used as a proxy for inefficiency. The main interest of this thesis is to identify the behaviour and scaling of the PoA with certain parameters of the system, and the causes of such behaviour.

The first model consists of a traffic network with a diamond lattice structure, originally considered in [4]. A more detailed review of [4] will be provided later in the chapter, but a key feature of this model is that the network is subject to congestion, so that inefficiency occurs when too much traffic is routed to some sections of the network and too little traffic is routed to other sections of the network. Applications of this model include road and computer networks [5–8], the use of public services [9], sports strategies [10, 11] and supply chains [12]. This thesis considers several extensions to the model of [4] involving varying the total traffic volume and altering the large-scale geometry of the lattice, producing different results and conclusions.

The second model considered is the spatial, pairwise matching of two sets of agents, who can be thought of as providers and consumers of a service. Here, the matching produced when all agents seek to match with their nearest available neighbour does not necessarily minimise the distance between matched agents. While the Euclidean metric is used here to measure the distance between agents, the distance can be considered more generally as the “cost” of matching two agents. Later on, a notion of time is included in the model such that agents are added into the system according to a random Poisson process. It is shown that delaying the matching to allow further agents into the system can increase efficiency. The inefficiency of matching agents at the moment they enter the system is quantified by an adaptation of the PoA, termed the “Price of Impatience”. The optimum delay time that produces the most efficient sequence of matchings is analysed. The motivation for this model is the recent success

of car-sharing companies that assign drivers to riders via mobile software applications. However, the model has a more general purview to any system in which a collection of agents are matched, such as hospitals and medical students [13], Internet users and servers [14], men and women, and universities and applicants [15].

1.2 Introduction to network theory

The characteristic of complexity can be found in a range of systems in nature and society, including ecological networks of predator-prey relationships, the World Wide Web (webpages connected by hyperlinks), the Internet (computers connected by data links), and social networks of relationships between people [16]. A common feature of these systems is that their emergent properties are not governed solely by the characteristics of their components, but also by the structure of the underlying network in which the components are embedded. Therefore, the emergence of behavioural properties on the global level may differ from those on the local level in a way that is not easily perceptible. It is generally regarded that a system is “complex” if the emergent properties of the system cannot be traced to any particular component or connection taken in isolation, but are the result of the topology of the underlying network as a whole. Moreover, the topology of the network need not be consciously designed, but is created through the many individual interactions between its components. Network Theory aims to provide insights into such complex systems through the study and analysis of networks. Broadly speaking, this is achieved in the following ways: 1) characterising the structure of the networks of real-world systems; 2) creating network-generating models that capture the key structural properties of real networks; and 3) analysing the effect of network structure on the function and performance of the system.

A network is a set of nodes or vertices V and a set of edges E that connect pairs of nodes. A network is *directed* if the edges function only in one direction, otherwise the network is *undirected*. The edges of a network may be assigned values, typically denoting distance or strength. A network is *weighted* if these values vary between edges, otherwise the network is *unweighted*. Two nodes are said to be *adjacent* if there is an edge between them. The *degree* d_i of a node i is the number of nodes that are adjacent to i and the *neighbourhood* N_i is the set of nodes that are adjacent to i . A *loop* is an edge that is connected to the same node on both ends, denoting self-interaction. *Parallel edges* are edges that connect the same pair of nodes. A network is *simple* if it does not contain any loops or parallel edges. A *path* is an ordered sequence of alternating nodes and edges that begins and ends with a node and whose terms are unique within the

sequence. A *cycle* is a path that begins and ends with the same node. A network is *connected* if there exists a path between every pair of nodes. As well as visually, a network can also be represented by an *adjacency matrix* – a square matrix whose dimension is equal to the number of nodes and whose entries a_{ij} denote the connection between nodes i and j . For a simple, unweighted network these are given by $a_{ij} = 1$ if nodes i and j are adjacent and $a_{ij} = 0$ otherwise, with $a_{ii} = 0 \forall i \in V$. For a weighted network, the a_{ij} 's can take a range of values denoting the strength or distance of the connections.

1.3 Analysis of network structure

One method of analysing the structure of a network is to characterise the structural properties of individual nodes and edges. Of particular interest is the influence of a given node within the network and a number of measures called *centralities* have been proposed in order to quantify this. (Similar centralities have been proposed with regards to the influence of edges but only node centralities shall be reviewed here.) One such centrality is the *degree centrality* [18]. The degree centrality is given by $d_i / (N - 1)$, where $N := |V|$ is the number of nodes in the network, and works under the assumption that a higher degree gives a node greater influence. Another centrality is the *closeness centrality* [19] which considers the lengths of the shortest paths, called *geodesics*, between a node and all other nodes. With the assumption that the ability to reach a high number of nodes within short path distances makes a node influential, the closeness centrality of node i is given by

$$C_L(i) := \frac{N - 1}{\sum_{j \in V} l(i, j)},$$

where $l(i, j)$ denotes the length, given by the number of edges, of a geodesic between nodes i and j . Similarly, the *decay centrality* [20] for node i is given by

$$C_D(i) := \sum_{j \in V} \delta^{l(i, j)},$$

where $\delta \in (0, 1)$ is a decay parameter. For small δ , only nodes that are adjacent to i have significant weight, in which case the decay centrality is proportional to the degree centrality. For larger δ , nodes that are more indirectly connected to i via other nodes have significant weight. The *betweenness centrality* [21] for node i considers the proportion of geodesics between every pair of nodes that contain i , given by

$$C_B(i) := \sum_{j, k \neq i} \frac{P_i(j, k)}{P(j, k)} \bigg/ \binom{N - 1}{2},$$

where $P(j, k)$ is the number of geodesics between nodes j and k , $P_i(j, k)$ is the number of these geodesics that contain node i , and the denominator is the binomial coefficient. Many more centralities have been proposed and no centrality is regarded as superior to all others. Rather, they each provide different ways of measuring the influence of a node and the extent to which they achieve this is subjective and context-dependent. In addition to influence within the network, another interesting property of a node is the level of connectivity within its neighbourhood. The neighbourhood N_i of a node i is the set of nodes that are adjacent to i . The *local clustering coefficient* [24] of node i measures the proportion of potential connections within N_i that are realised, given by

$$C_O(i) := \frac{1}{d_i(d_i - 1)} \sum_{\substack{j, k \in N_i \\ j, k \neq i}} a_{jk}.$$

A *community* is a subset of nodes in which edges between nodes within the community are more common than edges connecting the community to outside nodes.

Another method of analysing the structure of a network is to characterise the global properties of the network as a whole. These include:

- Network *diameter*:

$$\max\{l(i, j) : i, j \in V\}.$$

- Average *path length*:

$$\langle l \rangle := \frac{1}{N(N-1)} \sum_{i \neq j} l(i, j).$$

- Average *degree*:

$$\langle d \rangle := \frac{1}{N} \sum_{i \in V} d_i \equiv \frac{2|E|}{|V|}.$$

- Global *clustering coefficient*:

$$\langle C \rangle := \frac{1}{N} \sum_{i \in V} C_O(i).$$

- Degree *distribution* $P(d)$, which is the probability distribution of a given node having degree d .

- And *assortativity coefficient*:

$$\Gamma := \frac{c \sum_{i \in E} d_{i,1} d_{i,2} - \left[c \sum_{i \in E} \frac{1}{2} (d_{i,1} + d_{i,2}) \right]^2}{c \sum_{i \in E} \frac{1}{2} (d_{i,1}^2 + d_{i,2}^2) - \left[c \sum_{i \in E} \frac{1}{2} (d_{i,1} + d_{i,2}) \right]^2},$$

where $d_{i,1}$ and $d_{i,2}$ are the degrees of the nodes that are connected by edge i and $c = \frac{1}{|E|}$. This measures the extent to which nodes with similar degree are adjacent.

1.4 Network generation models

1.4.1 Erdős–Rényi random graphs

The earliest prominent network-generating model is the Erdős–Rényi random graph [22] in which every pair of nodes is connected independently and uniformly with some probability p (see figure 1.4.1). Random-graph theory studies the properties of the probability space associated with random graphs. In general, the limit $N \rightarrow \infty$ is considered so that the law of large numbers provides probabilistic accuracy. If $\langle d(N) \rangle$ is the average degree of a network with N nodes, then the network is almost certainly connected if $\langle d(N) \rangle \geq \log N$ [16]. However, networks that are highly dense with edges are generally uninteresting and unrepresentative of real networks, so the restriction $\langle d(N) \rangle / N \rightarrow 0$ as $N \rightarrow \infty$ is applied. When both these conditions are satisfied, it has been shown for Erdős–Rényi random graphs (as well as a number of general network models) that the average path length $\langle l \rangle$ and diameter are approximately proportional to $\log N / \log \langle d(N) \rangle$ as $N \rightarrow \infty$ [17]. Since the average path length scales logarithmically with the number of nodes in the network, it tends to be small in comparison to the overall size of the network, a property known as “small-world”. A more intuitive example of the small-world property was displayed in the experiments of [23], in which random individuals in Nebraska and Kansas, USA, were asked to try to get a letter to a specific individual in Boston. If a participant did not know the target individual, they were instructed to send the letter to someone they knew who they thought would be more likely to know the person in Boston. Surprisingly, it was found that the letter arrived after an average of just six steps.

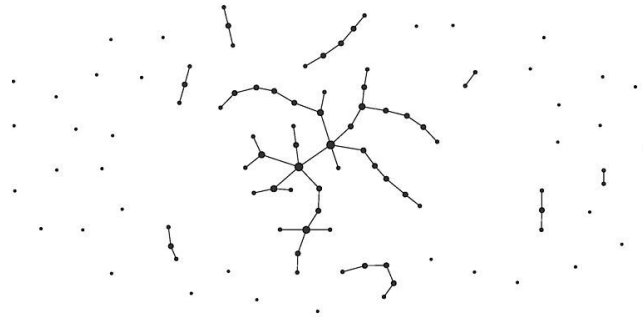


FIG. 1.4.1: Example of an Erdős–Rényi random graph with $p = 0.01$.²

²Image downloaded from https://en.wikipedia.org/wiki/Erdős–Rényi_model

The degree distribution of Erdős–Rényi random graphs is binomial:

$$P(d) = \binom{N-1}{d} p^d (1-p)^{(N-1)-d},$$

which for large N and small p can be approximated by a Poisson distribution:

$$P(d) \approx \frac{N^d}{d!} p^d e^{-Np}.$$

In the limit $N \rightarrow \infty$, there are certain thresholds for the value of p above which a property is almost certainly present in the network and below which it is almost certainly absent. Some of these thresholds are:

- $1/N^2$ - the network contains at least one edge.
- $1/N$ - the network contains a *unique giant component*. This is a subset of nodes that is connected and whose size scales with N .
- $\log N/N$ - the network is connected.

Since connections between nodes are formed independently the global clustering coefficient is simply $\langle C \rangle = p$.

Before the development of powerful computers to collect, store and analyse large data sets, it was assumed that the structure of real-world networks closely followed that of the Erdős–Rényi model. However, while most real networks exhibit the small-world property, other structural properties of many large networks are significantly different from those found in random graphs. Despite this, the Erdős–Rényi model remains a useful benchmark for comparison of real networks to random graphs with identical size and average degree.

1.4.2 Watts–Strogatz model

One of the structural properties for which real networks deviate from random graphs is the global clustering coefficient, which tends to be much higher in real networks such as the World Wide Web, the Internet, social and ecological networks, etc. In the case of social networks, it is intuitive that two people are more likely to be friends if they have a common friend than if they do not. New network-generating models were developed that aimed to recreate the structural properties of real networks. One such model is the Watts–Strogatz model [24], which is able to generate networks with both high clustering and small-world characteristics. The generation of this model is as follows (see figure 1.4.2a):

1. Nodes are positioned in the shape of a ring and each node is connected to its d nearest neighbours ($d/2$ on each side).
2. For each node i , every edge that connects i to some $j \in N_i$ is “rewired” with some probability p . If the edge is rewired, it detaches from node j and attaches to some node chosen uniformly at random such that the edge does not form a loop or a parallel edge (rewired edges are then excluded from future rewiring).

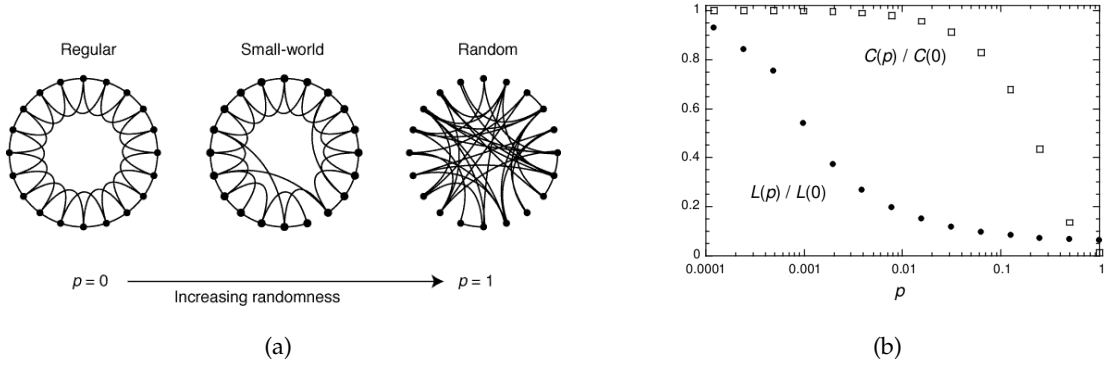


FIG. 1.4.2: (a) Watts–Strogatz network (Figure 1 of [24]). (b) The average path length and the global clustering coefficient as functions of the rewiring probability (Figure 2 of [24]).

The properties of the networks generated through this mechanism are dependent on the rewiring probability p . For $p = 0$, the network is a regular ring lattice with a high global clustering coefficient

$$\langle C(p = 0) \rangle = \frac{3(d - 2)}{4(d - 1)},$$

since a large proportion of each node’s neighbours are connected. It also has a high average path length $\langle l(p = 0) \rangle = N/2d$, because the geodesics between nodes on opposite sides of the ring have a high length. For $p = 1$, the network is a random graph which, as discussed above, has a small global clustering coefficient and a small average path length. For a small value of $p > 0$, the network exhibits both a high global clustering coefficient and a small average path length. This is achieved because rewired edges create shortcuts across the ring. Only a few shortcuts are sufficient for significantly lowering the average path length, while the network retains most of the lattice ring structure that produces high clustering. Figure 1.4.2b shows that these properties persist for a range of p [24].

1.4.3 Barabási–Albert model

Another structural property for which real networks deviate from random graphs is the degree distribution. Random graphs have a Poisson degree distribution whereas a number of real networks were found to exhibit a power law degree distribution $P(d) \sim d^{-\gamma}$ [16], a property known as *scale-free*. Although that the Watts–Strogatz model is able to recreate the clustering and small-world properties, it produces networks with a degree distribution similar to that of a random graph. Furthermore, it was noted that many real networks (such as the World Wide Web) have formed by growing over time — typically increasing in size as new nodes join the network — and the growth process of a network is likely to have an effect on its structure. However, the generations of both the Erdős–Rényi model and the Watts–Strogatz model assume a fixed network size and therefore neglect growth. These problems were addressed in a seminal paper that presented the preferential attachment model [25]. The generation procedure of this model begins with m_0 nodes and then new nodes are created one at a time. Each new node forms $m \leq m_0$ connections with the existing nodes, connecting to any given node with probability proportional to its degree. In other words, the probability Π that the new node connects to some node i is given by

$$\Pi(d_i) = \frac{d_i}{\sum_{j \in V} d_j}.$$

This process of *preferential attachment* or “the rich get richer” leads to the formation of nodes with a very high degree called “hubs”, because nodes that attain connections by chance near the beginning of the network growth are more likely to gain further connections from future nodes. As the network grows large, it contains many nodes with a relatively small degree as well as several hubs (see figure 1.4.3). In this way, the model generates scale-free networks with a small average path length and a power law degree distribution $P(d) \sim d^{-\gamma}$ with exponent $\gamma = 3$. Real networks that have been found to exhibit a power law degree distribution with an exponent of approximately 3 are the World Wide Web, sexual contact networks, airport/airline networks, and linguistics networks (in which words are nodes and two words are connected if they appear next to each other in sentences) [16].

It was found that both growth and preferential attachment are necessary in the generation mechanism for producing networks with a power law degree distribution [25], suggesting that the formation of real networks that exhibit power law degree distributions are also dependent on growth and preferential attachment. This seems likely in the case of the World Wide Web, for example, which is a constantly growing net-

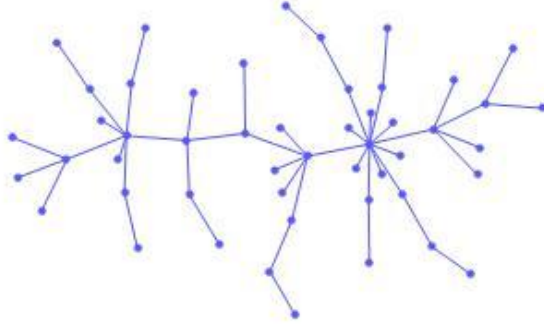


FIG. 1.4.3: A scale-free network generated by the Barabási–Albert model with $N = 50$ and $m = m_0 = 1$.³

work where popular webpages tend to gain more connections from new webpages. The Barabási–Albert model was therefore considered a major breakthrough at the time it was introduced, since it was assumed that real-world networks were typically scale-free. This assumption has been challenged by a recent study into real-world network structures, which found that of the 1000 network data sets examined, only 4% of them were likely to be scale-free and 52% were highly unlikely to be scale-free [26]. However, the conceptual and technical validity of this study has been criticised by Barabási himself [27].

One shortcoming of the Barabási–Albert model is that it gives a power-law degree distribution with a fixed exponent, while the exponents of real networks that exhibit power law degree distributions range between 1 and 3 [16]. Moreover, although its global clustering coefficient is higher than that of a random graph, it is smaller than those found in real networks. These issues were addressed by the Jackson–Rogers hybrid model [28], which also generates networks through a growth process.

1.4.4 Jackson–Rogers hybrid model

At each time step, a new node is created and some fraction a of its edges are connected to nodes uniformly at random, and the remaining fraction $(1 - a)$ are connected uniformly at random to nodes within the neighbourhoods of those nodes. The parameter a allows for a variety of degree distributions, with $a \approx 1$ giving a negative exponential and $a \approx 0$ giving a power law similar to that of the Barabási–Albert model. This is

³Image downloaded from https://en.wikipedia.org/wiki/Barabási–Albert_model

useful for making comparisons between the model and real networks, since the value of a can be tuned so that the model mimics the properties of the network of interest, which can then provide information about the formation process of the network. For example, the best fit for the World Wide Web degree distribution is $a = 0.43$ [28]. This shows that while preferential attachment is a strong factor, there is also a significant tendency for connections to be formed uniformly at random. This model also exhibits the small-world property and, if a is not too small, clustering.

1.5 Traffic routing on networks

Network theory also considers dynamical processes acting on networks, examples of which include the transmission of: data packets through a computer network, traffic through a transportation network, and an infectious disease through a society. In this regard, network-generating models are used not only to provide insights into how real networks are formed, but also to generate proxy networks on which model dynamical processes can operate. The particular dynamical process that will be considered in the research of this thesis is traffic routing, the properties of which will now be described.

1.5.1 Notation

Traffic is routed from source nodes to corresponding destination nodes via paths of edges, where each edge has an associated cost function denoting the cost or “latency” of traversing the edge. The costs of edges are nonnegative, nondecreasing and continuously differentiable functions of the quantity of traffic routed to the edge. Edges with strictly increasing cost functions create the effect of congestion. This model has obvious applications to road and computer networks, which respectively carry traffic in the form of motor vehicles and data packets. However, its applicability extends to any system in which the pay-off to the users of a resource declines with increasing usage. Further examples of such systems include public health services, sports strategies and supply chains, each of which are discussed below.

Let $(s_1, d_1), (s_2, d_2), \dots, (s_k, d_k)$ be the k source-destination pairs, $\mathcal{Q}_i$ the set of paths between nodes s_i and d_i and define $\mathcal{Q} = \bigcup_i \mathcal{Q}_i$. A *flow* is a function $\tau : \mathcal{Q} \rightarrow \mathbb{R}^+$ where $\tau_q := \tau(q)$ is the quantity of traffic routed to path $q \in \mathcal{Q}$. Denote by $\mathbb{E}$ the set of all edges of the network, then the traffic routed to any edge $e \in \mathbb{E}$ is given by:

$$\tau_e := \sum_{q \in \mathcal{Q} : e \in q} \tau_q.$$

A finite and positive *traffic volume* T_i is associated to each (s_i, d_i) -pair, such that

$$\sum_{q \in \mathcal{Q}_i} \tau_q = T_i.$$

In non-atomic models of congestion, the traffic volume represents an infinite number of users (or agents) each with an infinitesimally small weight of traffic. (This is in contrast to the atomic case in which users are assigned a finite weight, which will not be discussed here.) Finally, the cost function associated to a given edge e is denoted by $c_e(\tau_e)$. As noted above, the cost functions of edges are assumed to be nonnegative, differentiable and nondecreasing, but are otherwise arbitrary. The cost of traversing path $q \in \mathcal{Q}$ is

$$c_q(\tau) = \sum_{e \in q} c_e(\tau_e),$$

and the *total cost* of the flow τ over the whole network is

$$C(\tau) := \sum_{q \in \mathcal{Q}} \tau_q c_q(\tau).$$

The total cost can also be given by summing over the edges:

$$C(\tau) := \sum_{e \in \mathbb{E}} \tau_e c_e(\tau_e).$$

1.5.2 User Equilibrium and System Optimum traffic flows

A traffic flow of particular interest is the *User Equilibrium* (UE), whose conditions are the same as those of pure Nash Equilibrium in game theory [1]. The UE follows Wardrop's first principle of traffic flow [30], which states that individual agents each choose their routing strategy so as to minimise their own travel costs. It follows that all paths that carry traffic between a given (s_i, d_i) -pair have equal cost under the UE, and this cost is less than or equal to the cost of any path that does not carry traffic. Hence a flow $\tau = \tau^{\text{ue}}$ is the UE if and only if for any $i \in \{1, 2, \dots, k\}$ and any $q_1, q_2 \in \mathcal{Q}_i$,

$$\tau_{q_1}^{\text{ue}} > 0 \implies c_{q_1}(\tau^{\text{ue}}) \leq c_{q_2}(\tau^{\text{ue}}). \quad (1.5.1)$$

Denoting the cost incurred by traffic on paths between (s_i, d_i) under the UE by $C_i(\tau^{\text{ue}})$, the total cost of a UE flow is given by

$$C(\tau^{\text{ue}}) = \sum_{i=1}^k C_i(\tau^{\text{ue}}) T_i.$$

Several studies have considered the total congestion of the UE traffic flow as a function of the total traffic load in the network [31–34]. This was analysed for various network

structures, including regular, random, small-world, and scale-free networks. Of these, scale-free networks generally performed the best (*e.g.* exhibited the lowest UE total cost), followed by small-world networks, random networks, and regular networks.

Another traffic flow of interest is the *System Optimum* (SO). The SO follows Wardrop's second principle of traffic flow [30] in which the total cost $C(\tau)$ is minimised.

The problem of finding a SO flow can be summarised by the following program:

$$\text{minimise } \sum_{e \in \mathbb{E}} \tau_e c_e(\tau_e) \quad (1.5.2)$$

subject to:

$$\sum_{q \in \mathcal{Q}_i} \tau_q = T_i, \quad \forall i \in \{1, 2, \dots, k\},$$

$$\tau_e = \sum_{q \in \mathcal{Q}: e \in q} \tau_q, \quad \forall e \in \mathbb{E},$$

$$\tau_q \geq 0, \quad \forall q \in \mathcal{Q}.$$

Intuitively, a sufficient condition for a flow to be *locally* optimal is that for all $i \in \{1, 2, \dots, k\}$ any transfer of traffic between (s_i, d_i) paths does not decrease $C(\tau)$, *i.e.* the marginal fall in cost by removing traffic from one path is less than or equal to the marginal rise in cost by routing that traffic to another path. Since the local minimum and global minimum of a convex function on a convex set coincide [35], and the objective function of (1.5.2) is convex, this condition is also sufficient for a flow to be the SO. For any $q \in \mathcal{Q}$, define

$$c_q'(\tau) := \sum_{e \in q} \frac{d}{d\tau_e} (\tau_e c_e(\tau_e)) = \sum_{e \in q} (c_e(\tau_e) + \tau_e c_e'(\tau_e)).$$

A flow $\tau = \tau^{\text{so}}$ is the SO if and only if, for any $i \in \{1, 2, \dots, k\}$ and any $q_1, q_2 \in \mathcal{Q}_i$,

$$\tau_{q_1}^{\text{so}} > 0 \implies c_{q_1}'(\tau^{\text{so}}) \leq c_{q_2}'(\tau^{\text{so}}). \quad (1.5.3)$$

Note the similarity of conditions (1.5.1) and (1.5.3). It follows that a SO flow can be interpreted as a UE flow with respect to a different set of edge cost functions. Namely, if τ satisfies the SO for edge cost functions $c_e(\tau_e)$, then τ satisfies the UE for edge cost functions given by $c_e(\tau_e) + \tau_e c_e'(\tau_e)$. Moreover, finding a UE flow for a network with cost functions $c_e(\tau_e)$ can be summarised by the following optimisation problem:

$$\text{minimise } \sum_{e \in \mathbb{E}} \int_0^{\tau_e} c_e(x) dx \quad (1.5.4)$$

subject to:

$$\sum_{q \in \mathcal{Q}_i} \tau_q = T_i, \quad \forall i \in \{1, 2, \dots, k\},$$

$$\tau_e = \sum_{q \in \mathcal{Q}: e \in q} \tau_q, \quad \forall e \in \mathbb{E},$$

$$\tau_q \geq 0, \quad \forall q \in \mathcal{Q}.$$

Since the above convex program has a continuous objective function and a compact feasible region, the existence of a UE flow follows [7]. Moreover, the UE flow is unique in the sense that all flows that satisfy the UE have equal total cost $C(\tau)$ [7, 36]. Both the UE and the SO flows can therefore be solved using standard convex minimum cost flow algorithms [5].

In the context of real road networks, one would assume that individual drivers each choose the quickest route to their destination, and that the UE therefore approximates the traffic flow. However, with technological advances — such as the development of satellite navigation systems and driverless cars — implementing the SO traffic flow could be a realistic prospect in the near future.

1.6 The price of anarchy

When agents choose “selfish” routing strategies, the total cost of the flow may be sub-optimal. An interesting example of this is *Braess’s paradox* [37] shown in Figure 1.6.1. When a unit of traffic is routed from node s to node d under the UE, half the unit is routed to the path $s \rightarrow v \rightarrow d$ and the other half is routed to $s \rightarrow w \rightarrow d$, giving a travel cost of $3/2$. With the addition of an edge with zero cost, all of the traffic is routed to the newly-formed path $s \rightarrow v \rightarrow w \rightarrow d$ resulting in the travel cost rising to 2. Counter-intuitively, the addition of an extra edge with zero cost to the network results in a more costly UE flow. The level of inefficiency created by the selfish behaviour of network users is the subject of interest of this thesis, and can be measured by the *price of anarchy* $\mathcal{P}$ [3] given by:

$$\mathcal{P} := \frac{C(\tau^{\text{ue}})}{C(\tau^{\text{so}})}.$$

Before the addition of the new edge in Braess’s paradox, the UE and the SO are the same, so $\mathcal{P} = 1$. This flow remains the SO after the addition of the new edge but since the UE has changed, $\mathcal{P} = 4/3$. Much research has focussed on deriving upper bounds

on $\mathcal{P}$ for various families of networks. The first paper in this area showed that a tight upper bound on networks with affine cost functions is $4/3$ [7]. An alternative geometric proof of this result was derived in [38]. Subsequent work [8] showed that an upper bound for networks with separable, polynomial cost functions of degree $\leq k$ is

$$\left[1 - k(k+1)^{-\frac{k+1}{k}}\right]^{-1}.$$

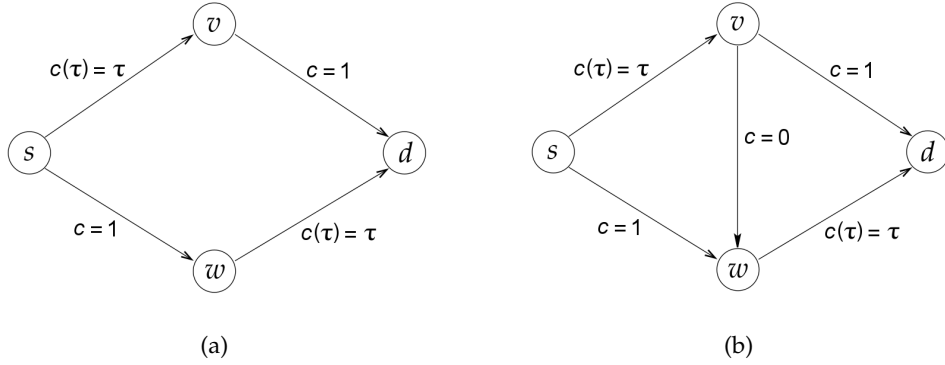


FIG. 1.6.1: Braess's paradox [37]. (a) When a unit of traffic is routed under the UE from node s to node d half a unit is routed to path $s \rightarrow v \rightarrow d$ and the other half to path $s \rightarrow w \rightarrow d$, giving a total cost $3/2$. (b) With the addition of a cost-free edge connecting nodes v and w all traffic under the UE is routed to path $s \rightarrow v \rightarrow w \rightarrow d$, giving a total cost $2 > 3/2$.

In deriving upper bounds on the price of anarchy the above research provides the theoretical maximum inefficiency across broad families of networks. However, it does not provide any general rules regarding when this maximum inefficiency will be realised and it neglects the variation in the price of anarchy with respect to properties such as network structure and traffic volume. These concerns were addressed in [6] which considered real road networks in Boston, London and New York City, where the edges of the network are the roads and the nodes are intersections of roads. Traffic was routed between a chosen source-destination pair, and edge costs followed the Bureau of Public Roads function [39] widely used in civil engineering given by

$$c_e(\tau_e) = \frac{d_e}{v_e} \left[1 + \alpha \left(\frac{\tau_e}{p_e} \right)^\beta \right],$$

where d_e is the length, v_e is the speed limit (which was constant over all edges for simplicity) and p_e is the capacity (number of lanes) of edge e . The parameters α and β were fitted to empirical data and given by $\alpha = 0.2$ and $\beta = 10$. The price of anarchy was considered as a function of traffic volume (vehicles per hour), with the results displayed in Figure 1.6.2a. We observe that $\mathcal{P} = 1$ for small traffic volume values. At intermediate values, $\mathcal{P}$ undergoes a series of fluctuations before reaching a principle peak and then falling to unity at large traffic volume. Intuitively, it is expected that the price of anarchy is low at small traffic volumes due to trivial congestion effects. Similarly, it is expected that the price of anarchy is low at very high traffic volumes, since the majority of traffic is forced to take costly routes under both traffic flows. However, it is not obvious why the price of anarchy is non-negligible at intermediate traffic volumes. Numerical simulations were also produced for synthetic networks with structures given by a regular lattice and the Erdős–Rényi, Watts–Strogatz and Barabási–Albert generation models. In this case the edge costs were affine functions $c_e(\tau_e) = a_e\tau_e + b_e$ where a_e and b_e were random integers, and $\mathcal{P}$ was calculated by routing traffic between each pair of nodes separately and taking the average, $\langle \mathcal{P} \rangle$. The results are displayed in Figure 1.6.2b, and show that the price of anarchy is again maximised at an intermediate value of traffic volume but that the detailed structure observed in the real road networks are absent. It was proposed that the maximum price of anarchy occurred at traffic volume values such that $\langle \tau_e \rangle \approx \langle b_e \rangle / \langle a_e \rangle$, where the average is taken over edges with positive flow. The absence of local maxima was explained by the averaging over many s - d pairs, but no explanation was given for the presence of local maxima in the analysis of real road networks. Work presented in the later chapters of this thesis will identify the causes of these local maxima.

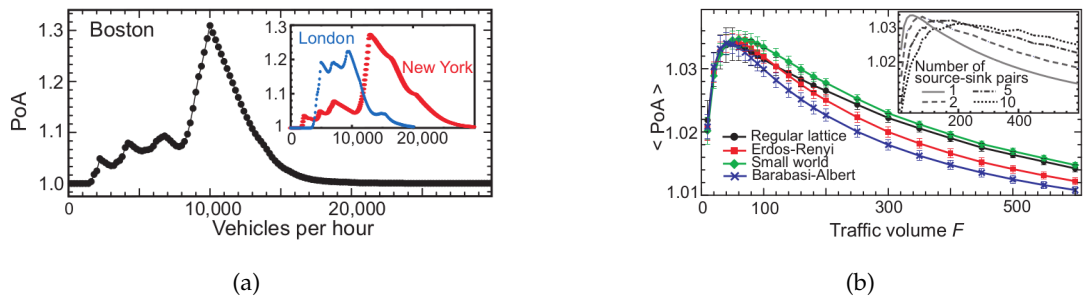


FIG. 1.6.2: The variation in the price of anarchy with traffic volume on (a) road networks in Boston, London and New York (figure 3a of [6]); (b) synthetic networks of standard network generating models (figure 3b of [6]).

The variation in the price of anarchy with respect to traffic volume in the context of public service systems was considered in [9], where agents could choose either to use a public service with an increasing cost function or “balk” and pay a fixed cost. It was proved analytically that the price of anarchy is at unity in the limit of infinite traffic volume. It was also shown for a simple model corresponding to a single s - d pair, that the price of anarchy is maximised when the cost of using the public service is equal to the cost of balking, *i.e.* when the capacity of the system is equal to the perceived worth of service (see figure 1.6.3a). The price of anarchy was also considered for a case study using real hospital data, where patients for elective knee replacement surgery across Wales could choose which public health centre to use, or to balk and pay a fixed cost β . Figure 1.6.3b shows that the numerical simulations display similar behaviour to the theoretical analysis.

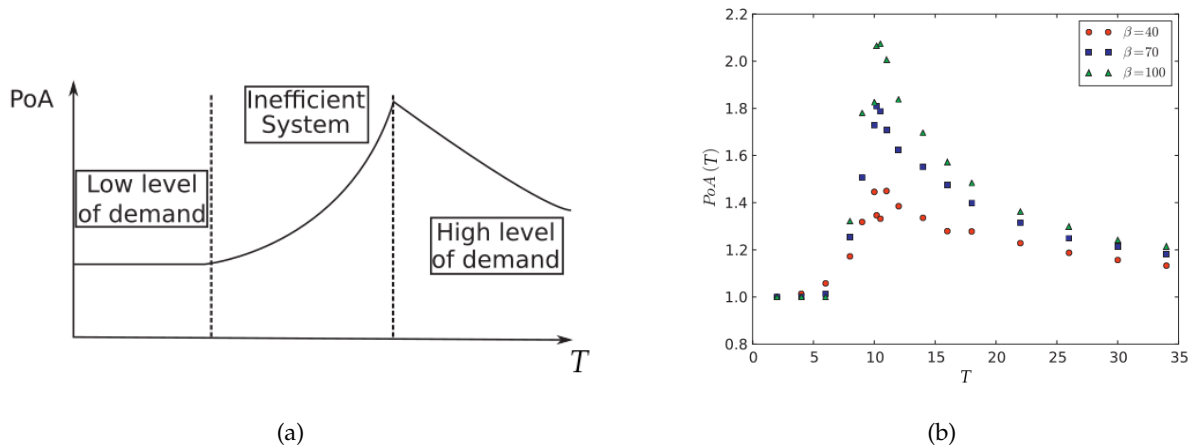


FIG. 1.6.3: (a) Schematic representation of the price of anarchy as a function of traffic volume (figure 6 of [9]). (b) Results from a case study of the choice of facility for elective knee replacement surgery across Wales (figure 11 of [9]).

More recently the issue of fluctuations in the price of anarchy was addressed in [40]. It was shown for several network examples that as the traffic volume is varied, the set of paths that carry traffic undergoes expansions and contractions. The thresholds for which these occur were called *route transition points*. As traffic volume increases, paths that carry traffic become more congested and more expensive to traverse, whereas the costs of paths that carry no traffic remain the same. Increasing the traffic volume gener-

ally expands the set of traffic-carrying paths. However, a network example was given in [40] showing that an increase in the traffic volume of one s - d pair can result in a *contraction* to the set of traffic-carrying paths in another s - d pair, and therefore a contraction to the set of traffic-carrying paths in the network. Numerical results on several network examples showed that the derivative of the price of anarchy with respect to the overall traffic volume increased with the SO expansions and the UE contractions, and decreased with the UE expansions and the SO contractions. Denoting corresponding route transition points under the UE and the SO by $\mathcal{T}^{\text{ue}}$ and $\mathcal{T}^{\text{so}}$, respectively, it was shown that for networks with edge cost functions of the form $c_e(\tau_e) = a_e + b_e \tau_e^\beta$ (with $a_e, b_e, \beta > 0$):

$$\mathcal{T}^{\text{so}} = \frac{1}{\sqrt[\beta]{\beta + 1}} \mathcal{T}^{\text{ue}},$$

and it was observed that the price of anarchy is maximised (and non-differentiable, *i.e.* a cusp) at the UE expansions.

Another study [4] considered the variation in the price of anarchy with respect to the potential level of congestion that can occur. This was analysed on a diamond lattice network with aspect ratio 1:1, as shown in figure 1.6.4a. The network had a heterogeneous edge structure comprising two types of edges: those with a variable cost function $c_e(\tau_e) = \tau_e$ and those with a fixed cost $c_e = 1$. In the context of road networks, those edges with a variable cost function could represent roads that provide a direct route but have few lanes (and are therefore susceptible to congestion), while edges with a fixed cost could represent roads that provide a longer route but have many lanes. A proportion p of edges in the network — selected uniformly at random — were of the former type, with the rest being of the latter type. The potential level of congestion is therefore parameterised by p . Traffic was routed horizontally from the left side of the lattice to the right side, with periodic boundary conditions applied in the vertical direction. The network was therefore directed, in the sense that edges could route traffic in only one direction (left to right). The price of anarchy was considered with varying $p \in [0, 1]$, with the traffic volume fixed at one unit, averaging over multiple network realisations for a given value of p .

The results displayed in figure 1.6.4b show that the price of anarchy is small for both small and large p , and maximised at intermediate p . Recall that a similar result was observed in several of the studies that considered the variation of the price of anarchy with traffic volume T . This makes sense since p and T have inverse effects on the system: p parameterises the network's potential for congestion whereas T governs how much of this potential is realised. When p is relatively small, the price of anarchy is

small because there is little potential for congestion. When p is relatively large, the price of anarchy is again small because there is too little traffic to achieve the congestion. The price of anarchy is non-negligible when the potential for congestion and the traffic volume are corresponding. The precise value of p at which the price of anarchy was maximised appeared to tend to the percolation threshold value p_c of the lattice with increasing lattice-size. This is a critical value such that in the limit $N \rightarrow \infty$, a path across the lattice composed entirely of variable-cost edges (for simplicity such a path shall be called a *variable path*) almost certainly does not exist for $p < p_c$ and almost certainly does exist for $p > p_c$. Hence the following explanation for the maximisation of the price of anarchy was given in [4]:

1. For $p < p_c$, there does not exist a variable path so traffic under the UE is “*naturally distributed over many paths with relatively low susceptibility to congestion*” and the price of anarchy is not too large.
2. For $p \approx p_c$, there exists exactly one variable path and, since this is theoretically the cheapest path in the network to traverse, it is heavily overused at the UE and the price of anarchy is large.
3. For $p > p_c$, there exist multiple variable paths over which traffic under the UE is dispersed, resulting in less congestion and a smaller price of anarchy.

This model will be revisited, generalised and studied in depth in Chapters 2-4. We will show that the conclusions above are only true in the large system-size limit and that the behaviours can be explained by more general principles.

1.7 Bipartite graph matchings

The other action on networks that will be considered within the research of this thesis is the pairwise matching of agents, which can be represented by bipartite graphs. A bipartite graph is a graph whose nodes can be partitioned into two disjoint sets such that no edge links two nodes of the same set. The two sets of nodes often represent two different types of agents, such as hospitals and interns [13], workers and firms [41], universities and applicants [15], parents and children etc. The research of this thesis will specifically focus on one-to-one matchings, where each agent is adjacent to at most one agent from the opposite set. Each agent ranks the agents from the opposite set in the order of their preference. A matching is stable if there does not exist two agents that

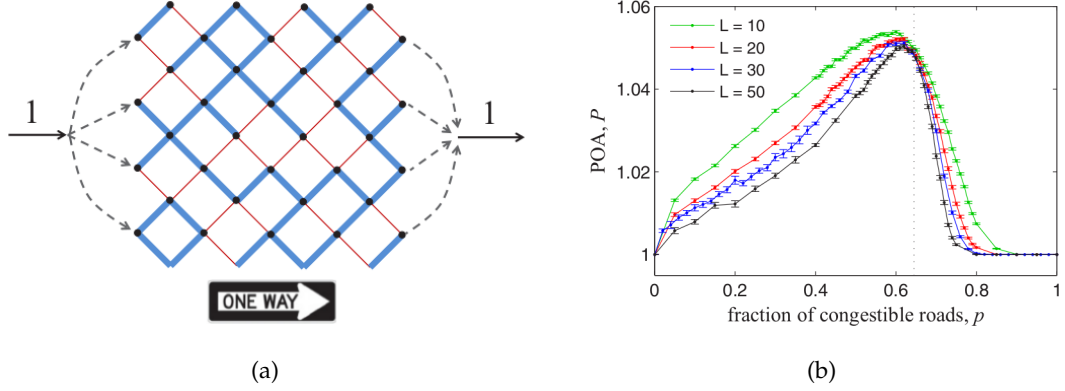


FIG. 1.6.4: (a) The diamond-lattice network model considered in [4] (figure 1b). A unit of traffic is routed from the nodes on the left side of the lattice to the nodes on the right side. The edges have either a fixed cost function $c_e = 1$ (red) or a variable cost function $c_e(\tau_e) = \tau_e$ (blue). (b) Numerical results of [4] (figure 3) for the variation of the price of anarchy with the proportion p of variable-cost edges. L denotes network size and the vertical dotted line indicates the percolation threshold $p_c \approx 0.6447$.)

would both prefer to be paired with each other than with their actual partners. Given two sets of agents and their preference lists, a seminal paper on stable matchings [15] described a procedure for obtaining a stable matching, thereby proving that a stable matching always exists. This procedure was also extended to many-to-one matchings, where the agents of one set form multiple partners each while the agents of the other set form just one partner each.

When matchings are represented as bipartite graphs, the weights of the edges can denote the pay-offs that the agents receive from their pairing. Alternatively, edge weights can denote the costs that the agents incur from their pairing. In the latter case, of particular interest is the matching for which the sum of the edge weights is minimised. An algorithm for computing such a matching can be found in [42]. As it will be seen in Chapter 6, minimum-weight matchings are not necessarily stable.

Since stable matchings are the result of agents acting selfishly, bipartite graph matchings represent another context where the price of anarchy exists, and can be quantified by the ratio of the weights of the stable matching and the minimum-weight matching. Similarly to the case of the traffic network model, there have been a number of studies concerning upper bounds on the PoA for graph matchings [44–46]. However, there

have been no studies regarding the issue of when a high PoA is to be expected. This will be the subject of Chapter 5, together with the impact of the environment in which the matchings occur and the time it takes to implement them.

1.8 Summary

This chapter has described the motivation for the research undertaken in this thesis and provided an overview of the relevant research to date. After reviewing the literature, the following issues have been identified as deficiencies within the research area:

1. The objective of most studies concerning inefficiency in traffic networks has been to derive upper bounds on the PoA. Few studies have considered how the PoA varies with respect to certain parameters of the system.
2. There has been no thorough investigation into a practical method for reducing the PoA.
3. The behaviour of the PoA has not been considered within the context of bipartite graph matchings.

The objective of this thesis is to advance progress in each of the above areas, with Chapters 2 and 3 addressing 1); Chapter 4 addressing 2); and Chapter 5 addressing 3). A more detailed description of the thesis structure is as follows.

Chapter 2 discusses the theoretical shortcomings of the conclusions drawn in [4]. The results of [4] are replicated and analysed in a more generalised model, and a new feature is observed. A more thorough survey into the variation of the PoA is presented, considering the impact of the traffic volume and the aspect ratio of the lattice network. The novel behaviours displayed by the numerical results suggest that a new theory for the cause of such behaviours is required.

Chapter 3 introduces a simplified model that produces the essential results of the general model and allows for analysis of the traffic flows, as well as interpretation of the rich phenomenology of the computer simulation results. Critical thresholds for classes of paths to carry traffic are derived for both the UE and SO traffic flows, and it is shown that the PoA is maximised at these thresholds for the UE. The cause for this correspondence is provided and a new theory for the behaviour of the PoA is discussed. Numerical simulations show that this theory holds both for the simplified model and the

original model.

Chapter 4 considers a method for reducing inefficiency by implementing “caps” that restrict the amount of traffic certain edges can carry. This forces traffic to be redistributed from overly congested paths to less congested paths. An application of this could be to restrict the number of vehicles on a particularly busy road, thereby allowing traffic to move more freely. This restriction could be enforced implicitly via variable speed limits, which are already widely used on congested sections of the British motorway network. The optimum number of caps is considered as a function of cost per cap implementation. It is shown that implementing caps on only a small percentage of edges in the network is sufficient for completely reducing inefficiency.

Chapter 5 introduces the bipartite graph matching model, in which two sets of agents with randomly distributed spatial locations are matched into pairs. The variation of the PoA and the weights of the matchings is examined with respect to various system parameters, including the size and shape of the domain that contains the agents, the number of agents in the system, and the ratio of the sizes of the two sets of agents. Agent locations are considered within a rectangular plane as well as on the nodes of networks, including the lattice structure used in [4] and the random networks described in Section 1.4. A notion of time is then added to the model, such that agents enter the system at different times determined by a random Poisson process. In this case, it is shown that optimum efficiency is obtained by delaying the matching until multiple agents have entered the system. The inefficiency of matching agents at their entry times is analysed using an adaptation of the PoA, termed the “Price of Impatience”.

Chapter 6 provides a summary of the research undertaken in this thesis and discusses the potential for further work. Finally, the Appendix provides further analytical results for the simplified traffic model that support the conclusions drawn in Chapter 3, and formulas for the total costs of the traffic flows, and hence the PoA, are derived. All numerical results in this thesis were generated by code written in C++, which can be located with the electronic submission of the thesis.

Phenomenology of the lattice network model

2.1 Introduction

This chapter will consider several extensions to the model of [4], which was described in Section 1.6. The numerical results obtained in this chapter will display a richness that cannot be accounted for by the conclusions drawn in [4]. Extensions that will be considered are variations to the aspect ratio of the lattice and the total traffic routed in the network. Note the lattice in [4] has an aspect ratio of 1, i.e. the width of the lattice is equal to its length. Given that traffic in this model is routed from one side of the lattice to the opposite side, it might be expected that a lattice with unequal width and length would display different results, and the results presented in this chapter will show that this is indeed the case. The results of [4] considered the variation of $\mathcal{P}$ with the potential congestion parameter p for a fixed traffic load T of one unit. In chapter 1, several studies were reviewed showing that $\mathcal{P}$ is highly dependent on T , and hence a traffic load of $T = 1$ is too restrictive and unrepresentative of real systems. Some of the numerical results presented in these studies applied to individual networks, whereas others displayed $\mathcal{P}$ as an average over multiple network realisations. For example, the numerical results of [6] for real road networks displayed values of $\mathcal{P}$ for each network, whereas the results in the same study for synthetic networks were averages over ensemble realisations. This chapter will provide numerical results for both individual and ensemble network realisations, and quantify in detail the dependence of $\mathcal{P}$ on p , T and lattice aspect ratio. For the remainder of this thesis, the price of anarchy for an individual realisation will be denoted $\mathcal{P}$, and the average over ensemble realisations will be denoted $\bar{\mathcal{P}}$.

2.2 Method for computing the traffic flows

In order to compute the traffic flows for a given network realisation, a numerical method from [47] (an earlier version of [4]) is employed, and works as follows. Denote by $\mathbb{E}$ the set of all edges of the network and by $\mathcal{Q}$ the set of all paths across the lattice (*i.e.* from source to destination). For an edge $i \in \mathbb{E}$, define the following indicator function:

$$\chi_i := \begin{cases} 1, & \text{if } i \text{ is a variable edge;} \\ 0, & \text{if } i \text{ is a fixed edge.} \end{cases}$$

The cost for a user to traverse a path $q \in \mathcal{Q}$ is

$$c_q = \sum_{i \in q} (\chi_i \tau_i + (1 - \chi_i)),$$

where τ_i is the traffic on edge i . The total cost of all traffic on path q is given by:

$$C_q = \sum_{i \in q} (\chi_i \tau_i^2 + (1 - \chi_i) \tau_i).$$

The algorithm for computing the traffic flows begins with the traffic uniformly distributed over all the paths. An iterative procedure is then implemented whereby two paths A and B are chosen at random and some traffic $\delta\tau \in \mathbb{R}$ is transferred from each edge of B to each edge of A .

The calculation of $\delta\tau$ is dependent on whether the UE or the SO is being computed. Recall from Section 1.5.2 that under the UE, all paths that carry traffic incur the same cost to traverse. Therefore, to compute the UE, $\delta\tau$ is calculated such that a transfer of $\delta\tau$ from path B to path A results in $c_A = c_B$. Under the SO, the total cost of the traffic flow is minimised, so when computing the SO $\delta\tau$ is calculated such that $(C_A + C_B)$ is minimised (detailed derivations of $\delta\tau$ in both cases will be provided below). To improve the convergence of the algorithm, paths are generated via a Monte Carlo sampling whereby each edge is chosen with probability proportional to the amount of traffic the edge carries at the current iteration. Once the algorithm has terminated (see below for criterion), the total cost of the flow over the entire network is calculated by:

$$C = \sum_{i \in \mathbb{E}} (\chi_i \tau_i^2 + (1 - \chi_i) \tau_i).$$

Note that this simple numerical method is not guaranteed to converge to the exact the UE or the SO flow. In order to minimise the effect of errors, results are produced by iterating over many pairs of paths per network realisation and taking the average $\bar{P}$ over multiple realisations for each choice of parameters (p, T) . These choices are based

on a convergence criterion given in [47].

For the computation of the UE, $\delta\tau$ is transferred from each edge of B to each edge of A to equalise the costs of the two paths:

$$\sum_{i \in A} (\chi_i(\tau_i + \delta\tau) + (1 - \chi_i)) = \sum_{i \in B} (\chi_i(\tau_i - \delta\tau) + (1 - \chi_i)).$$

Solving for $\delta\tau$ gives

$$\delta\tau = \frac{\sum_{i \in B} (\chi_i\tau_i + (1 - \chi_i)) - \sum_{i \in A} (\chi_i\tau_i + (1 - \chi_i))}{\sum_{i \in A} \chi_i + \sum_{i \in B} \chi_i}. \quad (2.2.1)$$

Note that $\delta\tau$ may take a negative value here, which denotes a transfer of traffic from A to B . However, $\delta\tau$ may not leave a negative amount of traffic on any of the edges. To this end, $\delta\tau^{\text{ue}}$ is assigned as follows:

$$\delta\tau^{\text{ue}} = \begin{cases} \min_{i \in B} \tau_i, & \text{if } \delta\tau > \min_{i \in B} \tau_i \\ -\min_{i \in A} \tau_i, & \text{if } -\delta\tau > \min_{i \in A} \tau_i \\ \delta\tau, & \text{otherwise,} \end{cases}$$

where $\delta\tau$ is given by (2.2.1). Finally, a damping parameter $\lambda = 0.5$ is applied: $\delta\tau^{\text{ue}} \mapsto \lambda\delta\tau^{\text{ue}}$. This is to avoid oscillatory behaviour in the numerical computation, which slows down convergence and produces inaccuracy (see figure 2.2.1).

For the computation of the SO, $\delta\tau$ minimises

$$C_A + C_B = \sum_{i \in A} (\chi_i(\tau_i + \delta\tau)^2 + (1 - \chi_i)(\tau_i + \delta\tau)) + \sum_{i \in B} (\chi_i(\tau_i - \delta\tau)^2 + (1 - \chi_i)(\tau_i - \delta\tau)),$$

so

$$\frac{d}{d(\delta\tau)}(C_A + C_B) = 0. \quad (2.2.2)$$

Solving (2.2.2) for $\delta\tau$ gives

$$\delta\tau = \frac{\sum_{i \in B} (2\chi_i\tau_i + (1 - \chi_i)) - \sum_{i \in A} (2\chi_i\tau_i + (1 - \chi_i))}{2 \left(\sum_{i \in A} \chi_i + \sum_{i \in B} \chi_i \right)}. \quad (2.2.3)$$

As before, $\delta\tau^{\text{so}}$ is given by

$$\delta\tau^{\text{so}} = \begin{cases} \min_{i \in B} \tau_i, & \text{if } \delta\tau > \min_{i \in B} \tau_i \\ -\min_{i \in A} \tau_i, & \text{if } -\delta\tau > \min_{i \in A} \tau_i \\ \delta\tau, & \text{otherwise,} \end{cases}$$

where $\delta\tau$ is given by (2.2.3), and a damping parameter $\lambda = 0.5$ is again applied: $\delta\tau^{\text{so}} \mapsto \lambda\delta\tau^{\text{so}}$.

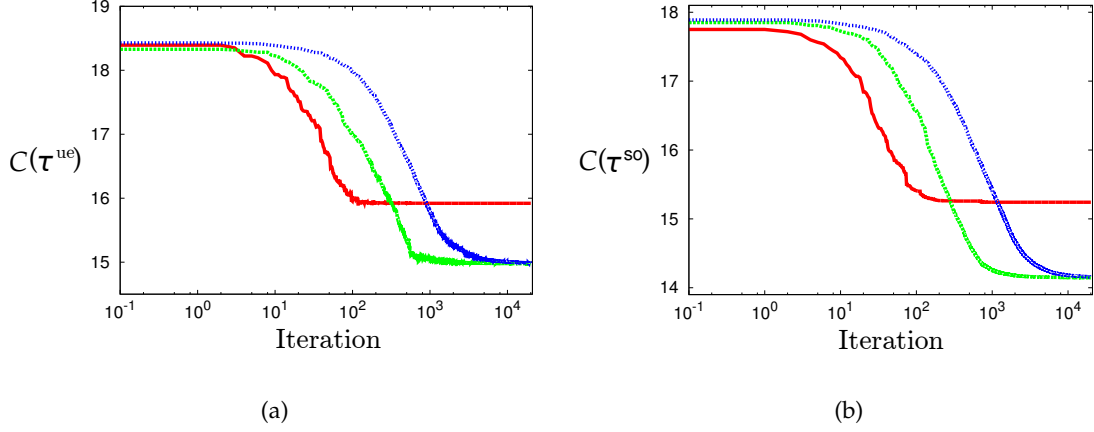


FIG. 2.2.1: Example of the evolution of the total cost under the UE (a) and the SO (b) with the number of iterations of transferring traffic between two paths chosen at random. Plots are given for $\lambda = 0.1$ (blue), 0.5 (green) and 1 (red). The total cost obtained with $\lambda = 1$ is larger than for smaller values of λ , and therefore does not represent the true total cost. Moreover, while the total cost converges to the same value for both $\lambda = 0.1$ and 0.5 , convergence is obtained faster for $\lambda = 0.5$, which is the value used in subsequent calculations.

2.3 Critique of original study

Recall from Section 1.6 the principal conclusion of [4] that $\bar{\mathcal{P}}$ as a function of the congestion potential p is maximised at the percolation threshold p_c of the lattice. The percolation threshold is a critical value at which there emerges a path across the lattice composed only of variable-cost edges (termed “variable paths”). Since such a path is theoretically the cheapest path to traverse, it is heavily overused at the UE flow which produces a high $\mathcal{P}$. The $\mathcal{P}$ then decreases with the emergence of more variable paths as traffic is then dispersed over these paths. If this is true then why does this phenomenon only occur for variable paths? At smaller values of p , the cheapest path will comprise a combination of both variable-cost and fixed-cost edges, and it might be expected that $\mathcal{P}$ would also decrease with the emergence of multiple cheapest paths of this kind as well.

Figure 2.3.1a shows a recreation of the numerical results of [4] for a lattice with size $L = 10$, where L is the number of source/destination nodes and $2L$ is the path length across the lattice. Whilst there is excellent agreement between this recreation and the original results of [4], the increased resolution used in obtaining the recreation has revealed a local maximum at small p . This feature is not present in the original results, because these were produced using a low density of points at lower values of p . As stated above, if the explanation given in [4] for the maximisation of $\bar{\mathcal{P}}$ is correct, then the new feature observed in figure 2.3.1 can be explained as follows:

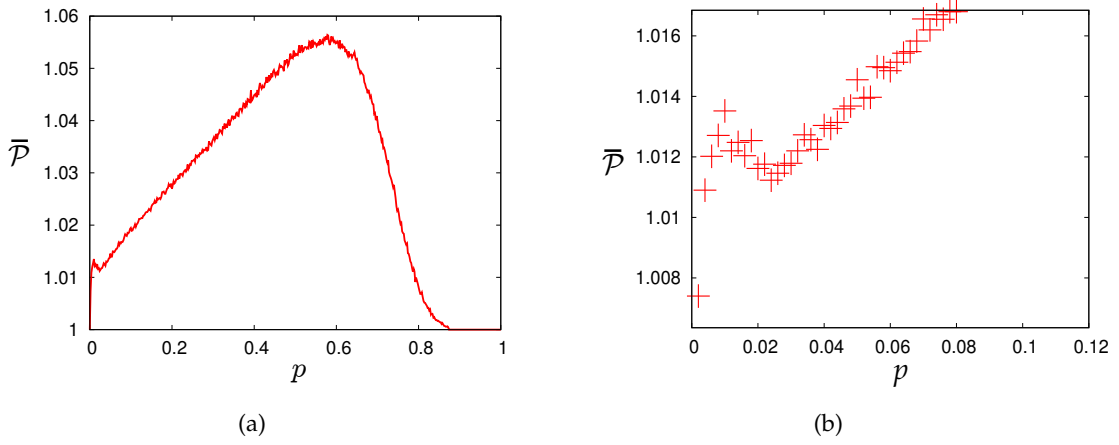


FIG. 2.3.1: (a) Recreation of the numerical results of [4] using the computational method outlined in Section 2.2. The congestion parameter p denotes the proportion of edges in the network with variable-cost function $c_i(\tau_i) = \tau_i$, with the remaining proportion having a fixed cost $c_i = 1$. The average price of anarchy $\bar{\mathcal{P}}$ over ensemble realisations of the diamond lattice network model is plotted as a function of p . The lattice considered here has $L = 10$ source/destination nodes and a path length of 20. (b) Highlight of the local maximum at small p , a feature that was absent in [4].

1. At very small p , the majority of network realisations contain either no variable-cost edges, in which case $\mathcal{P} = \mathcal{P}_0 = 1$, or exactly one variable-cost edge, in which case $\mathcal{P} = \mathcal{P}_1 > 1$. Consider the latter case, and recall that the cost of traversing a fixed-cost edge is $c_e = 1$, the cost of traversing a variable cost edge is $c_e(\tau_e) = \tau_e$, and the total traffic volume is one unit. It follows that under the UE, all traffic is routed to paths that contain the variable-cost edge. This can be

understood by considering a flow in which some traffic $\tau_f > 0$ is routed to paths that contain only fixed-cost edges, with $\tau_v = 1 - \tau_f$ routed to paths that contain the variable-cost edge. The cost of traversing the variable-cost edge is then $c_v(\tau_v) = \tau_v < 1$, and the cost of traversing a path that contains the variable-cost edge is $c_{q_v} = (2L - 1) + \tau_v < 2L$. The cost of traversing a path that contains only fixed-cost edges is $c_{q_f} = 2L > c_{q_v}$ and since some traffic is routed to some of these paths, the flow does not satisfy the UE. Hence under the UE, all traffic is routed to paths that contain the variable-cost edge, i.e. $\tau_v = 1$, and the total cost of the flow is $C(\tau^{ue}) = 2L$.

A flow under the SO minimises the total cost

$$\begin{aligned} C(\tau) &= \tau_v c_{q_v}(\tau_v) + \tau_f c_{q_f} \\ &= \tau_v(2L - 1 + \tau_v) + (1 - \tau_v) \cdot 2L, \end{aligned}$$

which gives $\tau_v = \tau_f = 1/2$. Hence $C(\tau^{so}) = 2L - 1/4$ and the price of anarchy is

$$\mathcal{P}_1 = \frac{2L}{2L - \frac{1}{4}}.$$

2. As p increases, the proportion of realisations that contain no variable-cost edges decreases and the proportion of realisations that contain exactly one variable-cost edge increases. Since $\mathcal{P}_1 > \mathcal{P}_0$, the average price of anarchy $\bar{\mathcal{P}}$ increases.
3. As p increases further, a proportion of realisations will contain exactly two variable-cost edges. Realisations of this type can be separated into two relevant groups: those in which there exists at least one path that contains both variable-cost edges (Group 1), and those in which no such path exists (Group 2). Group 1 is similar to the case of a network realisation with exactly one variable-cost edge: under the UE all traffic is routed to paths that contain both variable-cost edges giving $C(\tau^{ue}) = 2L$, and under the SO half of the traffic is routed to entirely fixed-cost paths giving $C(\tau^{so}) = 2L - 1/2$. The price of anarchy is therefore

$$\mathcal{P}_2 = \frac{2L}{2L - \frac{1}{2}}.$$

In the case of Group 2 there are two sets of theoretically cheapest paths: those paths that contain one of the variable-cost edges and those that contain the other variable-cost edge. Under both the UE and the SO, half the traffic is routed to each set and hence $\mathcal{P}_3 = 1$. Note that this is a case in which there are multiple cheapest paths (or multiple sets of cheapest paths), which, as noted in [4], results in traffic being dispersed over multiple paths under both traffic flows and a low

price of anarchy. If the ensemble of realisations is dominated by Group 2, the average price of anarchy $\bar{\mathcal{P}}$ would decrease.

The above would account for the presence of the local maximum at small p , but then raises the question of why further local maxima corresponding to cheapest paths with $2, 3, \dots, 2L - 1$ variable-cost edges are absent. The next section will show that these local maxima are indeed present but that the effects are negligible for square lattices (*i.e.* with aspect ratio 1:1). However, these effects become prominent and the local maxima visible once an adjustment is made to the aspect ratio of the lattice.

2.4 Results

As noted in Section 2.1, the lattice considered in [4] has aspect ratio 1 so that its vertical dimension is equal to its horizontal dimension. Denote by I the number of source nodes of the lattice, and by J the path length across the lattice. The aspect ratio is then given by $2I/J$ (see figure 2.4.1a), *e.g.* $J = 2I$ for lattices with aspect ratio 1. Figure 2.4.2 shows the variation in $\bar{\mathcal{P}}$ with respect to p for lattices with (a) $I = 80, J = 3$; (b) $I = 80, J = 4$; (c) $I = 80, J = 5$; and (d) $I = 80, J = 6$. For all lattices considered, $2I$ is much greater than J , meaning that the vertical dimension is much larger than the horizontal dimension. It can be observed that the variation in $\bar{\mathcal{P}}$ exhibits “ripples”, and that the number of such ripples is equal to the path length of the lattice (the inset of figure 2.4.2d shows the first peak at $p \approx 0$, which is not easily visible in any of the plots). The cause of these ripples will be shown in Chapter 3.

As discussed in Section 1.6, previous research into traffic routing has also considered the variation in $\mathcal{P}$ with respect to the traffic volume. Recall that [4] considered only a traffic volume of $T = 1$. Figure 2.4.3a-c shows the variation in $\mathcal{P}$ with T for several individual network realisations of a lattice with aspect ratio 1, for $p = 0.3$, $p = 0.6$ and $p = 0.8$. It can be observed that $\mathcal{P}$ exhibits irregular fluctuations for both $p = 0.3$ and $p = 0.6$, with the number of local maxima and their locations in T varying across network realisations. By contrast, for $p = 0.8$ there is a single maximum located at approximately the same value of T for each realisation. Consequently, if an ensemble of realisations were averaged for $p = 0.8$, this prominent peak would remain largely invariant in location and size. Figure 2.4.3d-f displays similar results for a lattice with $2I > J$. These results show that fluctuations with respect to traffic volume that have been observed in previous research studies [6, 9, 40] also occur for the diamond lattice network, for both $2I = J$ and $2I > J$. This would suggest that there exist route transi-

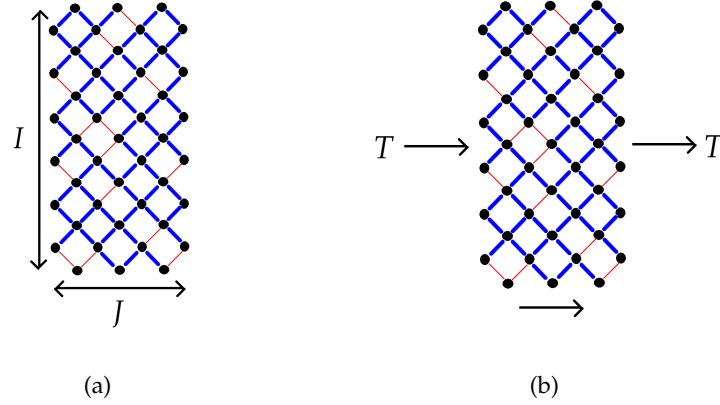


FIG. 2.4.1: (a) The aspect ratio of the lattice is determined by I and J . (b) The total traffic routed across the lattice is given by T . Thick blue links denote edges with a fixed cost, while thin red links denote edges with a variable cost function.

tion points in the manner of [40] (see section 1.6), and this will be addressed in Chapter 3.

Figure 2.4.4a-c shows the variation in $\bar{\mathcal{P}}$ averaged over multiple realisations, for a lattice with $I = 10$ and $J = 20$ (corresponding to the individual realisations of Figure 2.4.3a-c). For $p = 0.3$ and $p = 0.6$, it can be observed that the local maxima at intermediate T for individual realisations turn into a plateau in the average, while the final peak at high T is retained. For $p = 0.8$, the plateau is absent and only the final peak is present. Figure 2.4.4d-e shows the variation in $\bar{\mathcal{P}}$ for a lattice with $I = 80$ and $J = 6$ (corresponding to the individual realisations of Figure 2.4.3d-e). It can be observed for $p = 0.2$ and $p = 0.4$ that multiple local maxima are retained in the average. For $p = 0.7$, individual realisations exhibit similar behaviour – similarly to the case of $I = 10$ and $J = 20$ for $p = 0.8$ – and the average is therefore a replicate of them.

Finally, Figure 2.4.5 displays a heat-map showing the variation in $\bar{\mathcal{P}}$ with respect to both p and T . This is given for lattices with $2I = J$ (top row) and $2I > J$ (bottom row), and with network size doubling from left to right. There are two principal features in $\bar{\mathcal{P}}(p, T)$. First, a large thin region of increased $\bar{\mathcal{P}}$ localised in T but quasi-invariant with p (extending for $0.2 \lesssim p \lesssim 0.8$) is present in *all* networks. Its precise location is dependent on aspect ratio (for $2I = J$ it occurs at $T \sim 10$, whereas for $2I > J$ its location moves to $T \sim 100$). This feature was observed in figure 2.4.4, where the final peaks in $\bar{\mathcal{P}}(T)$ are preserved for each value of p considered. The second more complex

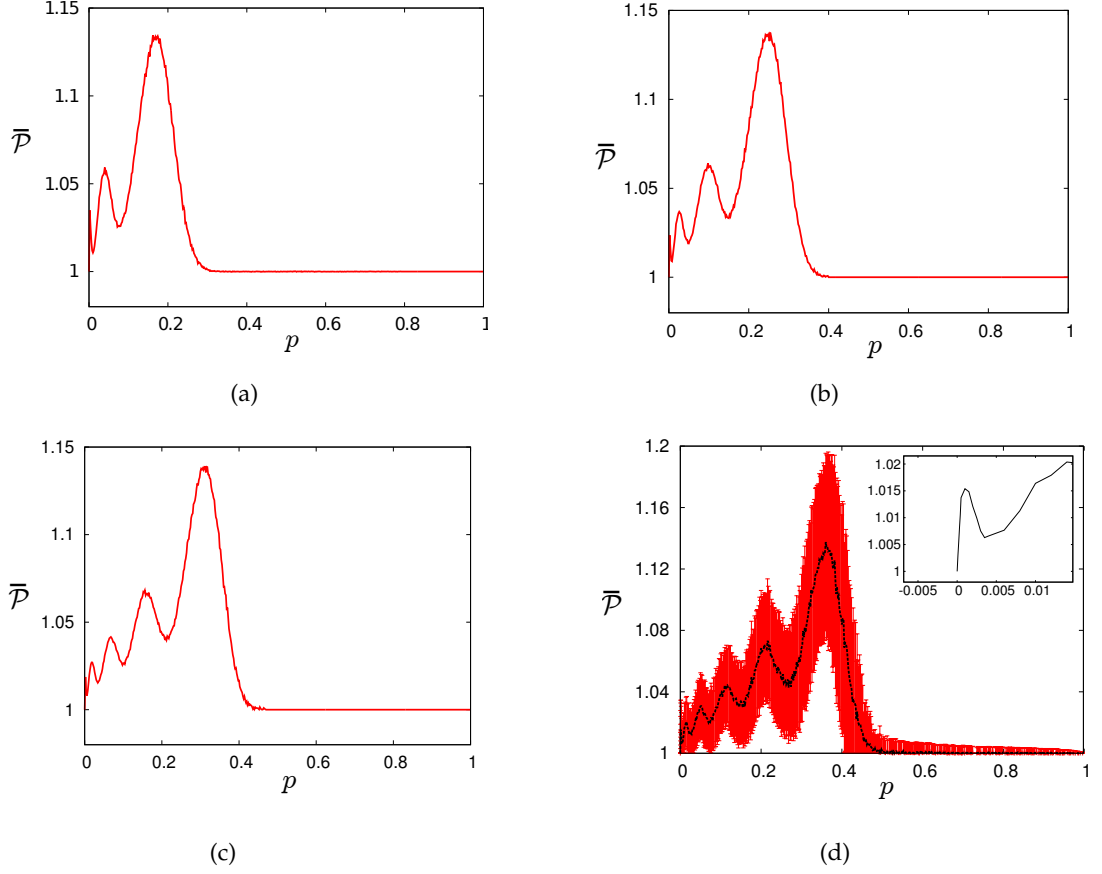


FIG. 2.4.2: The variation of $\bar{P}$ with respect to p exhibits ripples for lattices with I much greater than J , with the number of ripples being equal to J . Lattices considered here have dimensions: (a) $I = 80, J = 3$; (b) $I = 80, J = 4$; (c) $I = 80, J = 5$; and (d) $I = 80, J = 6$. The red shaded area in (d) represents one standard deviation from the mean, with the inset displaying the first ripple occurring at small p .

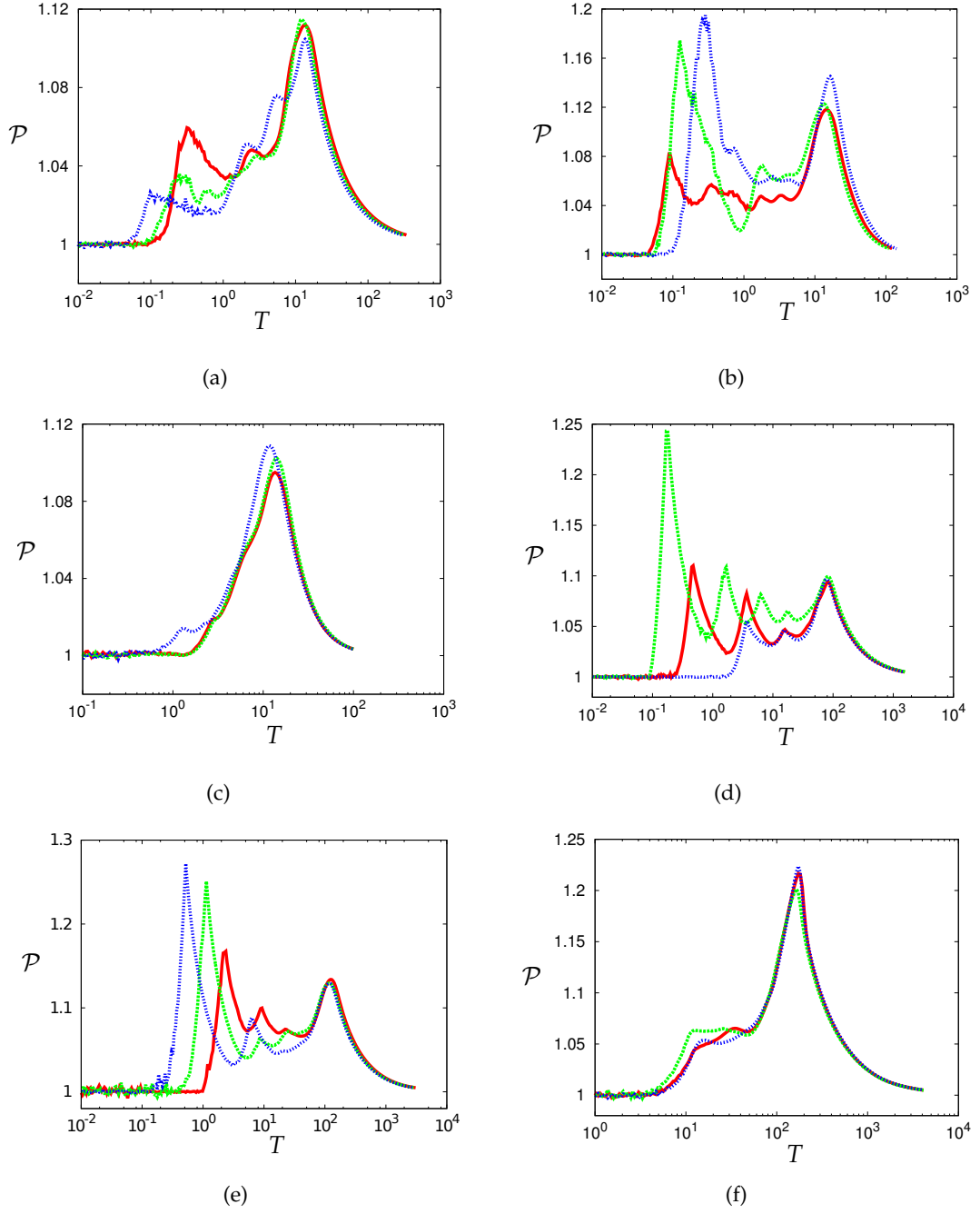


FIG. 2.4.3: (a)-(c) The variation of $\mathcal{P}$ with respect to T for three individual realisations of a lattice with $I = 10$ and $J = 20$, for (a) $p = 0.3$, (b) $p = 0.6$ and (c) $p = 0.8$. (d)-(f) The variation of $\mathcal{P}$ with respect to T for three individual realisations of a lattice with $I = 80$ and $J = 6$, for (d) $p = 0.2$, (e) $p = 0.4$ and (f) $p = 0.7$.

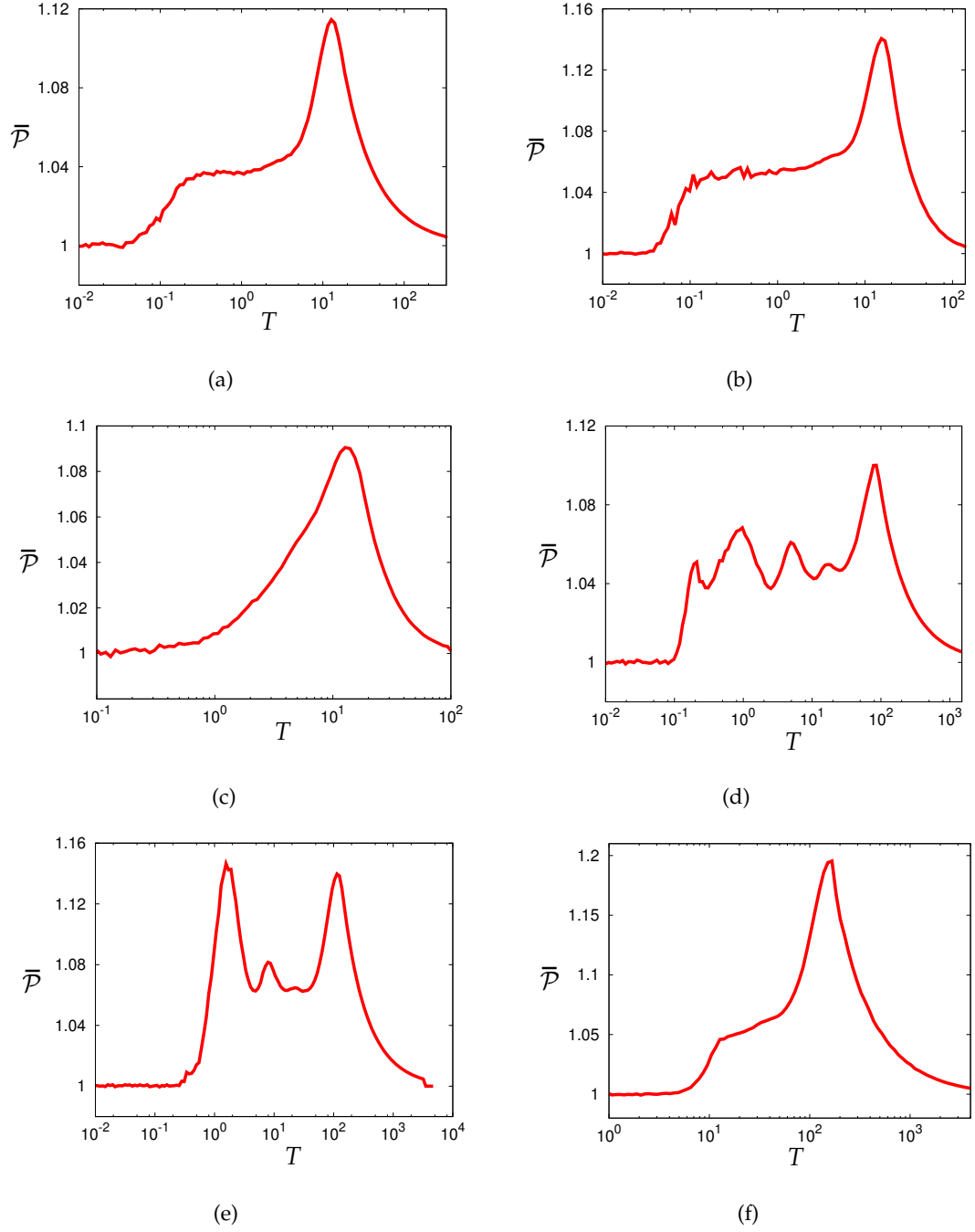


FIG. 2.4.4: (a)-(c) The variation of $\bar{P}$ with respect to T for a lattice with $I = 10$ and $J = 20$, for (a) $p = 0.3$, (b) $p = 0.6$ and (c) $p = 0.8$. (d)-(f) The variation of $\bar{P}$ with respect to T for a lattice with $I = 80$ and $J = 6$, for (d) $p = 0.2$, (e) $p = 0.4$ and (f) $p = 0.7$.

feature is more strongly dependent on network characteristics, in particular the aspect ratio. For $2I > J$ there are striations through the (p, T) -landscape, which reflect the ripples present in figure 2.4.2. For $2I = J$, there is a localised and globular morphology concentrated around $p = 0.6$, $T = 0.1$, which reflects the global maximum observed in figure 2.3.1a. The results of Figures 2.4.5(c,d) highlight that for fixed aspect ratio, the morphology of the landscapes of $\mathcal{P}$ are not strongly dependent on network size. However, the relative importance of the two features described is reversed as the network size increases: in small networks, the dominant peak in $\bar{\mathcal{P}}$ occurs for large T ; for larger networks, the globular or striated regions dominate. This is maintained for further increases in network size. For $2I < J$, additional simulations that are not illustrated here show that the locations and relative strength of the prominent features in the landscape of $\bar{\mathcal{P}}$ for square networks prevail. Since [4] considered the variation of $\bar{\mathcal{P}}$ with p only for $T = 1$, those results correspond to a single slice through the landscape of $\bar{\mathcal{P}}$ (and only for square lattices) and therefore miss much of the complexity.

2.5 Conclusion

This chapter has provided a critique of [4] and considered several extensions to the diamond lattice model that reveal novel behaviours that warrant detailed explanation. A higher resolution recreation of the results of [4] displayed a new feature at $p \approx 0$ that was absent in [4] and could not be explained by the conclusions therein. An alteration to the aspect ratio of the lattice was considered such that the height of the lattice is greater than the width, and displayed significantly different results to the original model. The emergence of ripples was observed in the behaviour of $\bar{\mathcal{P}}$, with the number of ripples being equal to the path length across the lattice. Numerical results were then presented for the variation of $\mathcal{P}$ with the total traffic load T . These were given both for individual and ensemble realisations, and for lattices of both types of aspect ratio. The variation of $\mathcal{P}$ again displayed local maxima, although the fluctuations were more irregular between realisations than the case of the variation in p . Finally, 3D plots displayed the rich structure of the variation of $\mathcal{P}$ with both p and T , as well as its dependence on the size and aspect ratio of the lattice. This chapter has therefore provided a thorough survey of the variation of $\mathcal{P}$, and revealed a rich phenomenology that the explanation offered in [4] cannot explain.

Since $\mathcal{P}$ is a proxy for inefficiency, it is clearly desirable to discover the causes for the

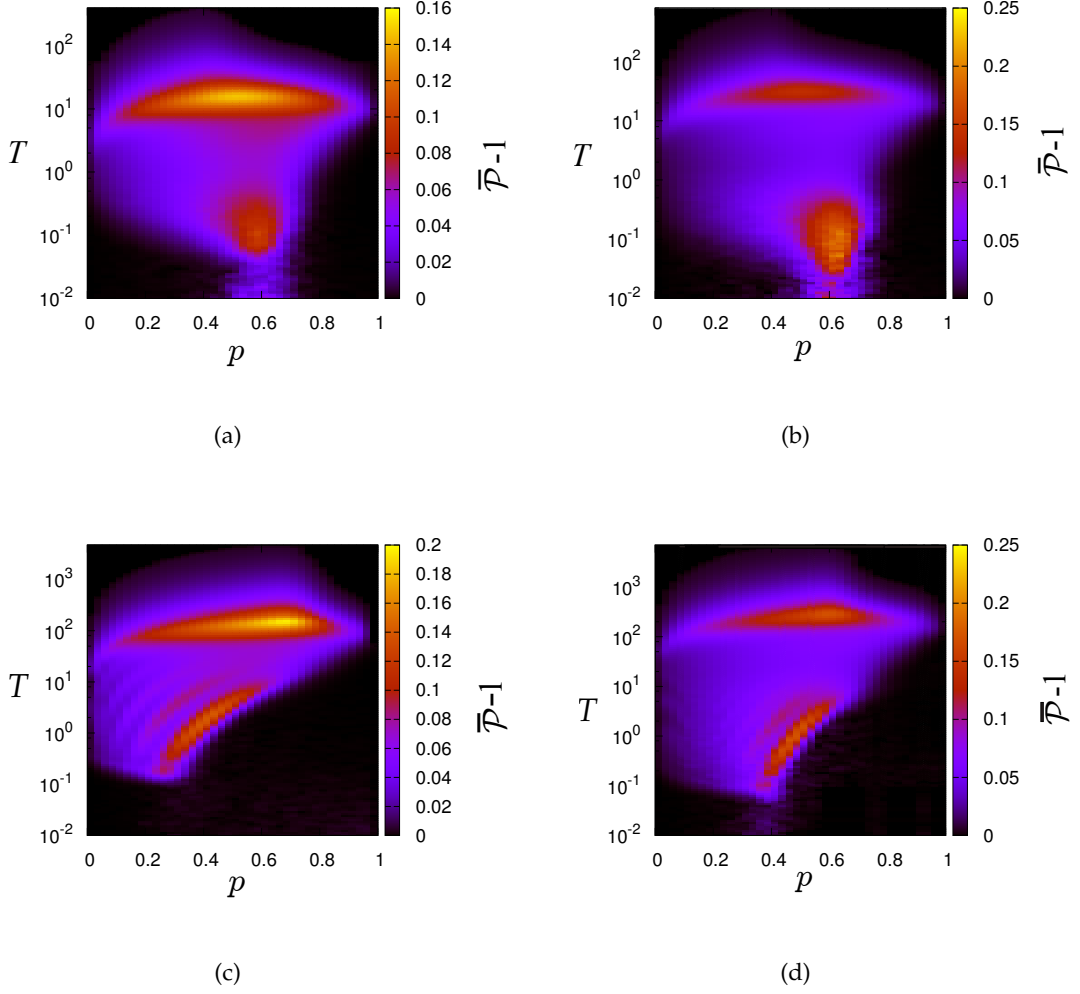


FIG. 2.4.5: Heat maps of $\bar{P}$ with respect to both p and T for lattices with (a) $I = 10$, $J = 20$; (b) $I = 20$, $J = 40$; (c) $I = 80$, $J = 6$; and (d) $I = 160$, $J = 12$. $(\bar{P} - 1)$ is displayed here to enhance contrast of colour.

variation of $\mathcal{P}$. Attaining knowledge of the conditions that create high inefficiency would be of fundamental importance in the design of networks, so that such conditions could be avoided and efficiency maximised. A network designer may want to minimise $\mathcal{P}$ over certain regions of p and T . The ripples of $\bar{\mathcal{P}}$ have an interesting implication: an alteration to the value of a parameter (whether it's p or T) can either increase or decrease inefficiency depending on the initial value of the parameter relative to the locations of the ripples. In this regard, the non-monotonic behaviour of $\mathcal{P}$ is of particular interest, as it shows that the parameters p and T could (in theory) be fine-tuned so that $\mathcal{P}$ attains a local minimum. This, however, can only be achieved if the rules for the variation of $\mathcal{P}$ are understood. The next chapter will address this issue by considering a simpler model of traffic flow that nevertheless captures the essential features embodied by the full model.

The cause of inefficiency in the lattice network model

3.1 Introduction

Attempting to understand the behaviour of the PoA observed in chapter 2 is complicated due to the computational complexity of the traffic flows. This complexity arises from the lattice network structure in which traffic-carrying paths can overlap one another, making analysis of the traffic flows very difficult. This chapter therefore introduces a simplified version of the model in which each source node is connected to a destination node by a single path containing no shared links. It will be shown that this model produces similar numerical results to the original model and, crucially, allows for analytical tractability of the traffic flows. Route transition points are derived and a detailed explanation of the variation of the PoA is provided. The conclusions drawn from the simplified model are then demonstrated to capture the essential behaviours exhibited by the full model, thereby providing a full explanation of the phenomena observed in Chapter 2.

3.2 The simplified model

Note that each edge of the diamond lattice network composes multiple paths across the lattice, so that paths can overlap one another by containing one or more common edges. As a result of this, calculating the UE and the SO traffic flows is a complicated problem, and it is difficult to infer any properties about them with respect to congestion potential and traffic volume. This problem can be simplified by considering a set of non-overlapping paths, classified such that each are composed of the same number

of edges and connect a source node to a destination node. All other features of the original model considered in Chapter 2 are retained: a traffic volume T is routed from the source nodes to the destination nodes, with no restrictions on the amount of traffic any given path can carry, and a proportion p of the edges have variable cost and $(1 - p)$ have fixed cost. A result of this simplification is that the following two conditions are necessary for a traffic flow to satisfy both the UE and the SO (which were not necessary under the original lattice network model):

1. *For any given path, each edge of the path carries identical traffic.*
2. *All paths with an equal number of variable edges carry an equal amount of traffic.*

Condition 1 is a trivial consequence of the paths not overlapping. To understand why condition 2 is necessary, consider a traffic flow in which this condition does not hold, i.e. there exist two paths P_1 and P_2 with an equal number of variable edges and that P_1 carries more traffic than P_2 . In this case, the variable edges of P_1 are more expensive to traverse than the variable edges of P_2 , hence P_1 is more expensive to traverse than P_2 and the flow does not satisfy the UE. Moreover, since P_1 carries more traffic and is more expensive to traverse, the total cost of the traffic on the two paths would clearly be smaller if the paths carried equal traffic, so the flow does not satisfy the SO either. Hence the initial premise is false and P_1 and P_2 must carry the same quantity of traffic. (The one exception to Condition 2 is paths with no variable edges, since the cost of these paths is the same regardless of the amount of traffic carried by them.)

The traffic flows within this simplified model are now analytically tractable. This will allow for the derivation of route transition points and the locations of local maxima in $\mathcal{P}$. Numerical results will be shown to display very similar behaviour to the original lattice model and the causes for these behaviours will be provided. In the Appendix, the simplified model will be used to support this theory and the total costs under both the UE and the SO will be derived.

3.3 Notation

As in Chapter 2, for any given realisation of the simplified model, the number of source/destination nodes will be denoted by I and the path length from a source node to a destination node will be denoted by J . Denote by N_k the number of paths that are composed of k variable edges — such paths shall be referred to as “ k -paths” — and by τ_k the traffic flow on each k -path. Denote by k^* the number of variable edges within

each path containing the greatest number of variable edges out of all of paths in the network, *i.e.*

$$k^* := \max\{k : N_k > 0\}.$$

The traffic volume T then satisfies:

$$T = \sum_{k=0}^{k^*} \tau_k N_k.$$

The cost of traversing a k -path is

$$c_k(\tau_k) = k\tau_k + (J - k). \quad (3.3.1)$$

The total cost of the traffic on each k -path is

$$C_k(\tau_k) = \tau_k c_k(\tau_k) = k\tau_k^2 + (J - k)\tau_k, \quad (3.3.2)$$

and the total cost of the flow is

$$C(\tau) = \sum_{k=0}^{k^*} C_k(\tau_k) N_k.$$

Recall that in Chapter 1, sufficient conditions (1.5.1) and (1.5.3) were provided for a flow to satisfy the UE and the SO, respectively. When applied to the current model under consideration, condition (1.5.1) can be simplified to the following: a traffic flow satisfies the UE if and only if for any i such that $N_i > 0$,

$$\tau_i > 0 \implies c_i \leq c_j \quad (3.3.3)$$

for all j such that $N_j > 0$. It follows that for any i and j with $N_i, N_j > 0$,

$$\tau_i, \tau_j > 0 \implies c_i = c_j. \quad (3.3.4)$$

Condition (1.5.3) can also be simplified to the following: a traffic flow satisfies the SO if and only if for any i such that $N_i > 0$,

$$\tau_i > 0 \implies C_i'(\tau_i) \leq C_j'(\tau_j) \quad (3.3.5)$$

for all j such that $N_j > 0$. It follows that for any i and j with $N_i, N_j > 0$,

$$\tau_i, \tau_j > 0 \implies C_i'(\tau_i) = C_j'(\tau_j). \quad (3.3.6)$$

Also note that under both the UE and the SO, for any i with $N_i > 0$,

$$\tau_i = 0 \implies \tau_j = 0 \quad \forall j \leq i. \quad (3.3.7)$$

These results will be useful for the derivation of route transition points, which is provided in the next section.

3.4 Identifying local maxima of $\mathcal{P}$

3.4.1 Derivation of route transition points

Another useful feature of the simplified model is that any realisation of non-overlapping paths can be partitioned into *classes* of paths, such that all paths within a given class contain the same number of variable edges and therefore carry the same amount of traffic. This is displayed in figure 3.4.1a, where red thin edges have a variable cost and blue thick edges have a fixed cost. The amount of traffic routed to each path within a given class is determined by the traffic volume T and the sizes of the classes of paths. For example, for any given j , with $1 \leq j \leq k^*$, the amount of traffic routed to each j -path is inversely proportional to the number of j -paths. The following proposition states sufficient conditions for which a given class of paths — and therefore all classes of paths with fewer variable edges — carries no traffic.

Proposition

For any i such that $N_i > 0$:

- (i) Under the UE, $\tau_k = 0$ for all $k \leq i$ if and only if

$$T \leq \sum_{j=i+1}^{k^*} \frac{j-i}{j} N_j. \quad (3.4.1)$$

- (ii) Under the SO, $\tau_k = 0$ for all $k \leq i$ if and only if

$$T \leq \frac{1}{2} \sum_{j=i+1}^{k^*} \frac{j-i}{j} N_j. \quad (3.4.2)$$

Proof

- (i) **First consider the “if” part.** If $\tau_k = 0$ for all $k \leq i$, then $\tau_i = 0$ and $c_i = J - i$. For any j with $i < j \leq k^*$, if $\tau_j = 0$ then $c_j = J - j < J - i$; if $\tau_j > 0$ then by (3.3.3) it follows that $c_j \leq c_i = J - i$. Hence

$$\begin{aligned} c_j &\leq J - i \quad \text{for } j = i + 1, i + 2, \dots, k^*. \\ \text{(using (3.3.1))} \implies \tau_j &\leq \frac{j-i}{j} \quad \text{for } j = i + 1, i + 2, \dots, k^*. \end{aligned}$$

So

$$T = \sum_{j=0}^{k^*} \tau_j N_j = \sum_{j=i+1}^{k^*} \tau_j N_j \leq \sum_{j=i+1}^{k^*} \frac{j-i}{j} N_j,$$

as required.

Now consider the “only if” part. Suppose there exists a $k \leq i$ such that $\tau_k > 0$. Then by the contrapositive of (3.3.7) it follows that $\tau_j > 0$ for all $j \geq k$. In particular, $\tau_i > 0$ and

$$c_i = i\tau_i + (J - i) \geq J - i.$$

Since $\tau_j > 0$ for all $j > i$, by (3.3.4) it follows that

$$c_j = c_i \quad \text{for } j = i + 1, i + 2, \dots, k^*.$$

$$\text{(using (3.3.1))} \implies \tau_j \geq \frac{j-i}{j} \quad \text{for } j = i + 1, i + 2, \dots, k^*.$$

So

$$T = \sum_{j=0}^{k^*} \tau_j N_j > \sum_{j=i+1}^{k^*} \tau_j N_j \geq \sum_{j=i+1}^{k^*} \frac{j-i}{j} N_j.$$

By the contrapositive, it follows that

$$T \leq \sum_{j=i+1}^{k^*} \frac{j-i}{j} N_j \implies \tau_k = 0 \quad \forall k \leq i,$$

as required.

(ii) Note from (3.3.2) that

$$C'_k(\tau_k) = 2k\tau_k + (J - k), \tag{3.4.3}$$

and **consider again the “if” part.** If $\tau_k = 0$ for all $k \leq i$, then $\tau_i = 0$ and $C'_i(\tau_i = 0) = J - i$. For any j with $i < j \leq k^*$, if $\tau_j = 0$ then $C'_j(\tau_j = 0) = J - j < J - i$; if $\tau_j > 0$ then by (3.3.5) it follows that $C'_j(\tau_j) \leq C'_i(\tau_i = 0) = J - i$. Hence

$$C'_j(\tau_j) \leq J - i \quad \text{for } j = i + 1, i + 2, \dots, k^*.$$

$$\text{(using (3.4.3))} \implies \tau_j \leq \frac{j-i}{2j} \quad \text{for } j = i + 1, i + 2, \dots, k^*.$$

So

$$T = \sum_{j=0}^{k^*} \tau_j N_j = \sum_{j=i+1}^{k^*} \tau_j N_j \leq \frac{1}{2} \sum_{j=i+1}^{k^*} \frac{j-i}{j} N_j,$$

as required.

Now consider the “only if” part. Suppose there exists a $k \leq i$ such that $\tau_k > 0$. Then by the contrapositive of (3.3.7) it follows that $\tau_j > 0$ for all $j \geq k$. In particular, $\tau_i > 0$ and

$$C_i'(\tau_i) = 2i\tau_i + (J - i) \geq J - i.$$

Since $\tau_j > 0$ for all $j > i$, it follows from (3.3.6) that

$$C_j'(\tau_j) = C_i'(\tau_i) \quad \text{for } j = i + 1, i + 2, \dots, k^*.$$

$$\text{(using (3.4.3))} \implies \tau_j \geq \frac{j-i}{2j} \quad \text{for } j = i + 1, i + 2, \dots, k^*.$$

So

$$T = \sum_{j=0}^{k^*} \tau_j N_j > \sum_{j=i+1}^{k^*} \tau_j N_j \geq \frac{1}{2} \sum_{j=i+1}^{k^*} \frac{j-i}{j} N_j.$$

By the contrapositive, it follows that

$$T \leq \frac{1}{2} \sum_{j=i+1}^{k^*} \frac{j-i}{j} N_j \implies \tau_k = 0 \quad \forall k \leq i,$$

as required.

3.4.2 Variation of $\mathcal{P}$ with respect to T

The above proposition will be used to derive insight into the variation of the price of anarchy with respect to both p and T . Consider first the variation of the UE and the SO traffic flows with respect to T . Recall from Chapter 1 that a *route transition point* is a critical threshold for which the set of traffic-carrying paths changes. For the model of this chapter, a route transition point is a traffic threshold for a class of paths to carry traffic. These thresholds under the UE and the SO shall be denoted by $\mathcal{T}_i^{\text{ue}}$ and $\mathcal{T}_i^{\text{so}}$, respectively. The above proposition provides route transition points with respect to T for i -paths to carry traffic, given by

$$\mathcal{T}_i^{\text{ue}} := \sum_{j=i+1}^{k^*} \frac{j-i}{j} N_j \tag{3.4.4}$$

under the UE, and

$$\mathcal{T}_i^{\text{so}} := \frac{1}{2} \sum_{j=i+1}^{k^*} \frac{j-i}{j} N_j \tag{3.4.5}$$

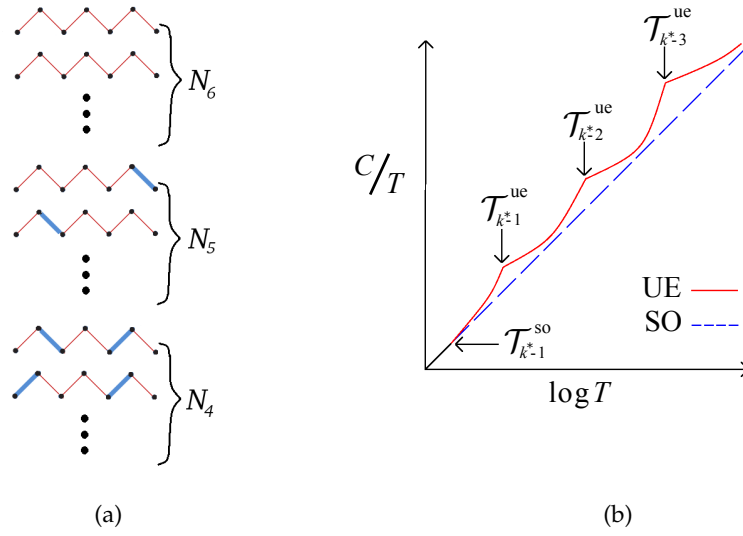


FIG. 3.4.1: (a) The simple model of non-overlapping paths, which can be partitioned into classes of paths composed of the same number of variable edges. (b) A schematic diagram showing the variation of the total costs (per unit traffic) under the UE (red/solid line) and the SO (blue/dashed line) with the traffic volume T . The cusps of the UE curve correspond to the route transition points $\mathcal{T}_i^{\text{ue}}$ where $\mathcal{P}$ exhibits local maxima.

under the SO. The route transition points under the UE and the SO will be referred to as the UE-RTP and the SO-RTP, respectively. Note that $\mathcal{T}_i^{\text{so}} = \mathcal{T}_i^{\text{ue}}/2$, i.e. the threshold for i -paths to carry traffic under the SO is precisely half the threshold for i -paths to carry traffic under the UE. It follows that there exist intervals in T for which the SO utilises a greater number of classes of paths than does the UE. Since traffic is more dispersed when there is a greater number of traffic-carrying paths, the level of congestion on these paths increases with T at a lower rate. Therefore, it is expected that when the SO utilises a greater number of classes of paths than the UE, $C(\tau^{\text{ue}})$ increases with T at a *higher* rate than does $C(\tau^{\text{so}})$ and the price of anarchy increases with T . By contrast, if the UE and the SO are utilising exactly the same classes of paths, then $C(\tau^{\text{ue}})$ and $C(\tau^{\text{so}})$ will increase with T at the *same* rate. To understand the behaviour of $\mathcal{P}$ in this case, consider the function $f(t) := g(t)/h(t)$ (i.e. a ratio of costs) for some variable t and suppose there exists an interval $[a, b]$ such that $g'(t) = h'(t) > 0$ for all $t \in [a, b]$. It follows that if $g(a) = h(a)$, then $f(t) = 1$ for all $t \in [a, b]$; if $g(a) > h(a)$, then $f(t)$ is decreasing over $t \in [a, b]$. By the same token, if the UE and the SO are utilising exactly the same classes of paths for some interval $T \in [T_a, T_b]$, then $\mathcal{P}$ is constant for $T \in [T_a, T_b]$ if $C(\tau^{\text{ue}}, T = T_a) = C(\tau^{\text{so}}, T = T_a)$, and $\mathcal{P}$ is decreasing for $T \in [T_a, T_b]$ if $C(\tau^{\text{ue}}, T = T_a) > C(\tau^{\text{so}}, T = T_a)$.

For a given realisation of paths and a given traffic volume T , denote by $\hat{k}^{\text{ue}}$ and $\hat{k}^{\text{so}}$ the number of variable edges of the traffic-carrying paths with the least number of variable edges under the UE and the SO, respectively, given by

$$\hat{k} := \min\{k : \tau_k > 0\}.$$

When $\hat{k}^{\text{so}} < \hat{k}^{\text{ue}}$, the SO is utilising more classes of paths than the UE, and it is expected that the price of anarchy is increasing with T . When $\hat{k}^{\text{so}} = \hat{k}^{\text{ue}}$, the UE and the SO are utilising the same classes of paths and it is expected that the price of anarchy is either constant or decreasing. This implies that local maxima in the price of anarchy occur at the UE-RTP's given by (3.4.4) for $i = \hat{k}^{\text{so}}$, since

$$\begin{cases} T \leq \mathcal{T}_{\hat{k}^{\text{so}}}^{\text{ue}} & \implies \hat{k}^{\text{ue}} > \hat{k}^{\text{so}}, \\ T > \mathcal{T}_{\hat{k}^{\text{so}}}^{\text{ue}} & \implies \hat{k}^{\text{ue}} = \hat{k}^{\text{so}}. \end{cases}$$

However, since $\hat{k}^{\text{so}}$ varies with T , it follows that $\mathcal{P}$ does not necessarily attain a local maximum at every UE-RTP. To understand why this is true, consider a traffic load T such that $\hat{k}^{\text{ue}} = \tilde{k}$ and $\hat{k}^{\text{so}} = \tilde{k} - 2$, for some $\tilde{k}$. It follows that the price of anarchy would not attain a local maximum at $\mathcal{T}_{\tilde{k}-1}^{\text{ue}}$. This is because when T exceeds $\mathcal{T}_{\tilde{k}-1}^{\text{ue}}$,

	the UE	the SO	$\mathcal{P}$
$\frac{1}{2}\mathcal{T}_i^{\text{ue}} < T \leq \mathcal{T}_i^{\text{ue}}$	$\hat{k}^{\text{ue}} = i + 1.$	$\hat{k}^{\text{so}} = i.$	Increasing
	C^{ue} increases with T at a higher rate than C^{so} .		
$\mathcal{T}_i^{\text{ue}} < T \leq \frac{1}{2}\mathcal{T}_{i-1}^{\text{ue}}$	$\hat{k}^{\text{ue}} = i.$	$\hat{k}^{\text{so}} = i.$	Decreasing
	C^{ue} increases with T at the same rate as C^{so} .		

Table 3.1: Description of the variation of $\mathcal{P}$ with T around a route transition point $\mathcal{T}_i^{\text{ue}}$ such that $\mathcal{T}_{i-1}^{\text{ue}}/2 > \mathcal{T}_i^{\text{ue}}$.

$\hat{k}^{\text{ue}} = \tilde{k} - 1 > \tilde{k} - 2 = \hat{k}^{\text{so}}$, so the price of anarchy continues to increase with T . Moreover, the price of anarchy does not attain local maxima at $\mathcal{T}_i^{\text{ue}}$ for any i such that $N_i = 0$, since a route transition point for i -paths does not exist if there are no i -paths. On the other hand, if $N_i > 0$ and $\mathcal{T}_{i-1}^{\text{ue}}/2 > \mathcal{T}_i^{\text{ue}}$ for some $i \leq J$, then $\mathcal{P}$ will certainly attain a local maximum at $\mathcal{T}_i^{\text{ue}}$, as outlined in Table 3.1.

The variation of the price of anarchy with T can therefore be summarised as follows. For small T , $\mathcal{P}$ is at unity since all traffic is routed to the class of k^* -paths under both the UE and the SO. When T exceeds a SO-RTP, the SO utilises more classes of paths than the UE, *i.e.* $\hat{k}^{\text{so}} < \hat{k}^{\text{ue}}$, and $\mathcal{P}$ increases with T . When T exceeds a UE-RTP such that the UE and the SO utilise the same classes of paths, *i.e.* $\hat{k}^{\text{ue}} = \hat{k}^{\text{so}}$, $\mathcal{P}$ decreases with T . This cycle is repeated as T continues to exceed further the UE-RTP's and the SO-RTP's and more classes of paths are utilised, until the class of paths with the fewest variable edges is utilised under both traffic flows and $\mathcal{P}$ saturates to unity. The variation of the total costs under the UE and the SO with T is illustrated by figure 3.4.1b. Figure 3.4.2 displays numerical results for two individual realisations of the simplified model. It can be observed that the simplified model displays similar results to the original lattice model, with $\mathcal{P}$ exhibiting irregular fluctuations both for $2I > J$ and $2I = J$. Moreover, local maxima occur at vertical dotted lines denoting the UE-RTP's given by (3.4.4), suggesting that the above description for the variation of $\mathcal{P}$ is accurate.

3.4.3 Variation of $\bar{\mathcal{P}}$ with respect to p

Now consider the variation in the price of anarchy with respect to p . Recall from Chapter 1 that the parameters p and T have opposite effects on the system, since p can be

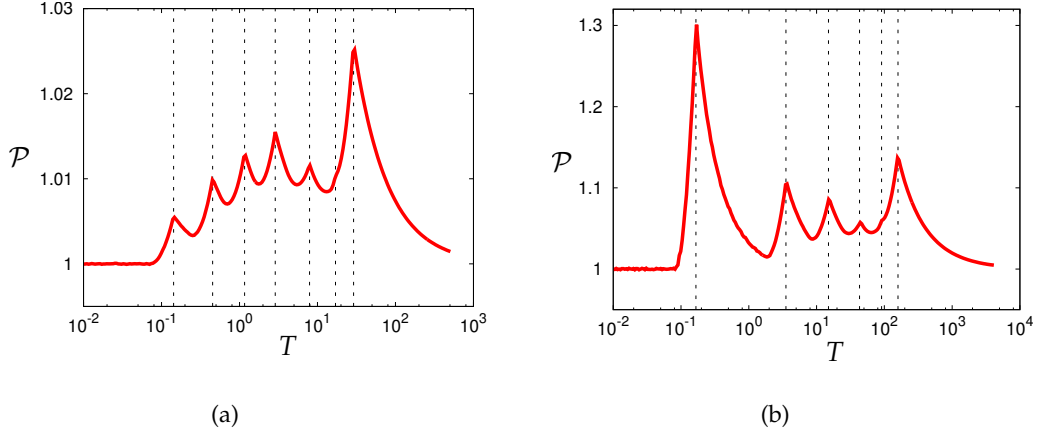


FIG. 3.4.2: The variation in the price of anarchy with respect to T for two individual realisations of the simplified model of non-overlapping paths, with (a) $I = 160$, $J = 6$ and $p = 0.5$; (b) $I = 30$, $J = 30$ and $p = 0.1$. Vertical dotted lines denote route transition points under the UE given by (3.4.4).

interpreted as the congestion potential, whereas T determines the extent to which this potential is realised. While the total costs under the UE and the SO increase with T , they decrease with p . This is because — assuming T is not too large¹ — variable edges are cheaper to traverse than fixed edges. It was shown in Chapter 2 that the variation in the price of anarchy exhibits ripples for lattices whose vertical dimension is larger than their horizontal dimension. Recall that in this case, the average price of anarchy $\bar{P}$ over multiple network realisations is considered for each value of p . Although p is not explicitly present in (3.4.1) or (3.4.2), the set of values $\{N_j\}_{j=0}^{j=J}$ and k^* for a given network realisation are random variables that depend on p (except in the extreme cases of $p = 0$ and $p = 1$). Taking $i = (k^* - 1)$ in (3.4.1) gives the following:

$$\text{Under the UE, } \tau_k = 0 \quad \forall \quad k \leq (k^* - 1) \quad \Longleftrightarrow \quad T \leq \frac{N_{k^*}}{k^*}. \quad (3.4.6)$$

Since T is fixed, property (3.4.6) is dependent only on N_{k^*} and k^* . Figure 3.4.3 displays a heat map showing the pdf of k^* within the ensemble of realisations, and over a range of p values. Although k^* is a random variable, it can be observed that there exist intervals in p for which the majority of realisations exhibit $k^* = i$ for each $i \leq J$. Intuitively, for a given k^* , the average number of k^* -paths will increase with the average number of variable edges, *i.e.* $\langle N_{k^*} \rangle$ will increase with p . Hence, the UE-RTP with respect to p for

¹It is assumed that T is small enough so that the traffic load on variable edges is ≤ 1 . This is because a central feature of the model is that variable edges are cheaper to traverse than fixed edges. At large values of T such that this assumption does not hold, the price of anarchy is trivially 1.

$(k^* - 1)$ -paths is given by:

$$N_{k^*}^{\text{ue}} = Tk^*. \quad (3.4.7)$$

That is, $(k^* - 1)$ -paths carry traffic if $N_{k^*} < Tk^*$, and carry no traffic if $N_{k^*} \geq Tk^*$. Taking $i = (k^* - 1)$ in (3.4.2) gives the corresponding SO-RTP:

$$N_{k^*}^{\text{so}} = 2Tk^*. \quad (3.4.8)$$

So when $N_{k^*} \geq 2Tk^*$, all traffic is routed to the k^* -paths under both the UE and the SO, and $\mathcal{P} = 1$.

Since the variation of $\mathcal{P}$ with T exhibits local maxima at the UE-RTP's, it is plausible that this is also true for the behaviour with p . It follows that J local maxima are formed over each interval in p in which a majority of network realisations exhibit $k^* = i$, for $i = 1, 2, \dots, J$. If T is not an integer, property (3.4.7) will not be possible to realise, in which case $\mathcal{P}$ is maximised when $|N_{k^*} - Tk^*|$ is minimised. When considering the average over an ensemble of realisations, it is therefore expected that local maxima in $\bar{\mathcal{P}}$ coincide with local minima in $\langle \mathcal{G} \rangle$, where

$$\mathcal{G} := \begin{cases} \left| \frac{N_{k^*}}{Tk^*} - 1 \right|, & \text{if } 1 \leq N_{k^*} \leq 2Tk^* \\ 1, & \text{if } N_{k^*} \geq 2Tk^*. \end{cases} \quad (3.4.9)$$

In the above definition, Tk^* is used as a normalising factor. Figure 3.4.4 displays results for two example cases of the simplified model, and shows that local maxima in $\bar{\mathcal{P}}$ do indeed coincide with local minima in $\langle \mathcal{G} \rangle$. The maxima of $\bar{\mathcal{P}}$ have therefore been located, and the behaviour of $\bar{\mathcal{P}}$ with varying p is attributed to the route transition points for $(k^* - 1)$ -paths. In theory, route transition points for i -paths with $i < (k^* - 1)$ also exist, but these become irrelevant when averaging over multiple network realisations.

3.5 Application to the lattice network model

Although the above explanation for the variation of the PoA for the simplified model provides plausible interpretations for the behaviours observed, this chapter has yet to provide a corresponding explanation for the variation of the PoA for the original lattice network model. Figures 3.4.2 and 3.4.4 show that the simplified model displays

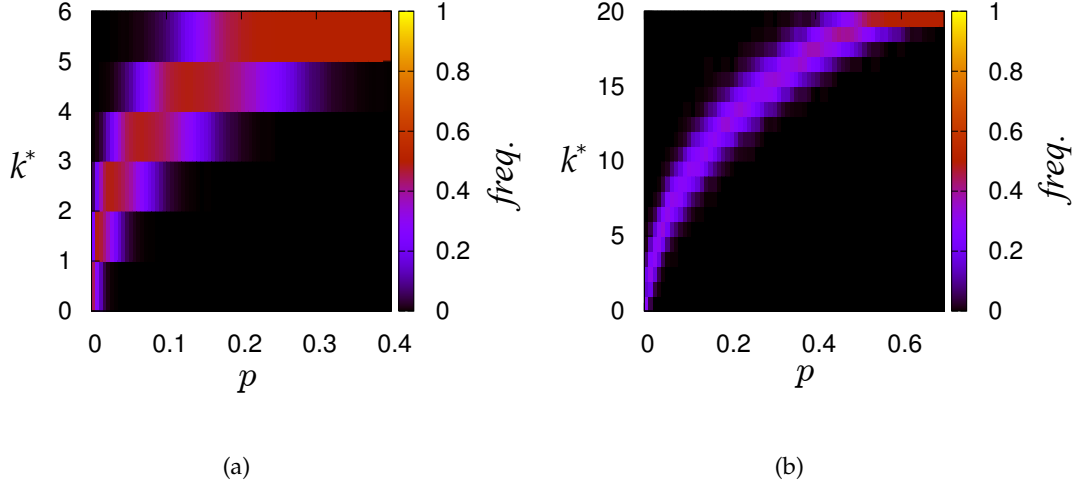


FIG. 3.4.3: 3D heat maps showing the pdf of k^* with varying p for the original lattice network model, with (a) $I = 80, J = 6$; (b) $I = 10, J = 20$. Colours indicate the proportional frequency of a given value of k^* .

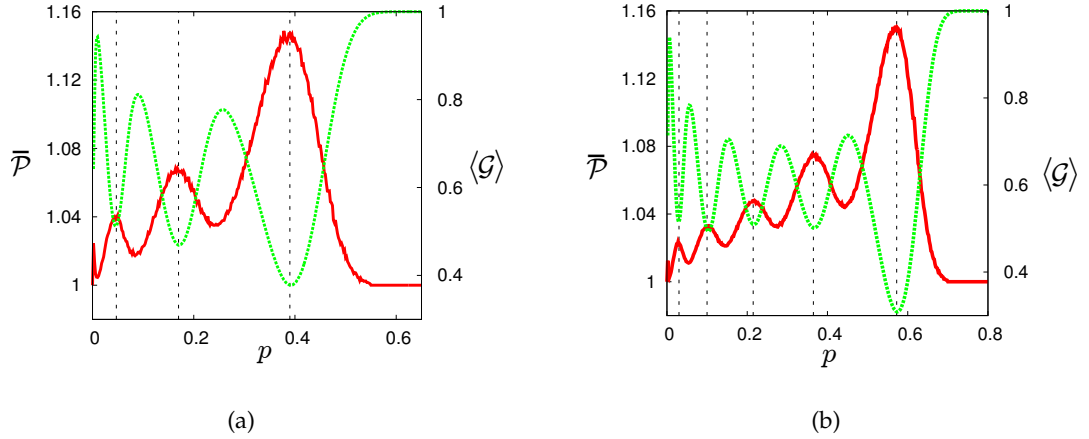


FIG. 3.4.4: Variation of $\bar{P}$ (red) and $\langle \mathcal{G} \rangle$ (green) with respect to p for the simplified model, with fixed $T = 1$ and (a) $I = 160, J = 4$; (b) $I = 160, J = 6$. Vertical dotted lines are displayed to show correspondance between local maxima in $\bar{P}$ and local minima in $\langle \mathcal{G} \rangle$.

similar numerical results to the lattice model, suggesting that the behaviour of the two models have the same underlying cause. This raises the question of whether the above findings can be directly applied to the lattice network model. Recall that in this case, each edge of the network belongs to many paths across the lattice. A categorisation of the paths into classes would not form a partition of the edges of the network, since each edge would be counted multiple times. Variable edges are subject to congestion by traffic from each of the paths that they compose, making the traffic flows for the lattice network highly complex. For example, if two k^* -paths overlap one another by having one or more common variable edges, then these variable edges will exhibit congestion from traffic routed to both paths. It follows that the efficacy of these two overlapping k^* -paths — while perhaps greater than that of a single k^* -path — is lower than that of two non-overlapping k^* -paths. It would be inaccurate to calculate N_j , for any $0 \leq j \leq k^*$, simply by counting the number of j -paths. Therefore, identifying route transition points for the lattice network model requires a different method, which will now be considered.

3.5.1 Variation of $\mathcal{P}$ with T

Recall that the UE follows Wardrop's first principle of traffic flow: all traffic-carrying paths charge the same cost, which is less than or equal to the cost of any path not carrying traffic. It follows that the UE-RTP's can be derived using the total cost $C(\tau^{\text{ue}})$. Denoting the cost of traffic-carrying paths in the lattice network by c , the total cost is simply given by $C(\tau^{\text{ue}}) = Tc$. It follows from (3.3.1) that when an i -path does not carry traffic, its base cost is $c_i = J - i$. This cost is simply equal to the number of fixed edges of the path, which is an integer value. Under the UE, i -paths do not carry traffic when $c \leq c_i$, and i -paths carry traffic if $c > c_i$, hence a UE-RTP occurs at $c = c_i$. Using the above formula for $C(\tau^{\text{ue}})$ and c_i , this equation is equivalent to:

$$\frac{C(\tau^{\text{ue}})}{T} = J - i,$$

where $J - i$ is an integer value. This observation is useful because it can be tested for the lattice model as well as the simplified model. Figure 3.5.1 shows that $\mathcal{P}$ exhibits local maxima at integer values of $C(\tau^{\text{ue}})/T$ for both models, suggesting that the cause of the behaviour is the same for both models. Here, the values of I and J for the two models were chosen so that the examples contained the same number of edges, and the values of p were chosen so that $\mathcal{P}$ exhibited pronounced local maxima.

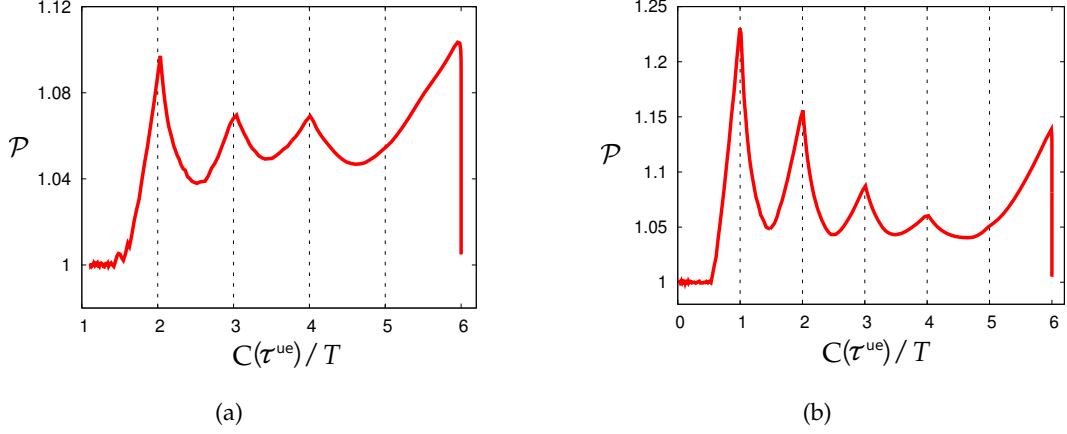


FIG. 3.5.1: The variation in the price of anarchy with respect to $C(\tau^{\text{ue}})/T$ for an individual realisation of (a) the diamond lattice network with $I = 80$, $J = 6$ and $p = 0.2$; (b) the simplified model with $I = 160$, $J = 6$ and $p = 0.5$. Vertical dotted lines denote integer values of $C(\tau^{\text{ue}})/T$.

3.5.2 Variation of $\bar{\mathcal{P}}$ with p

It was shown for the simplified model that the dominant property for the maximisation of $\bar{\mathcal{P}}$ with respect to p was given by (3.4.7). It is plausible that this also holds for the lattice model so long as N_{k^*} is considered as the *effective number* of k^* -paths. To return to the example given above, if a lattice realisation contains two k^* -paths such that at least one variable edge belongs to both of these paths, then N_{k^*} is such that $1 \leq N_{k^*} < 2$ (and equal to 1 if the two paths share all k^* of their variable edges). Hence a more sophisticated method for calculating N_{k^*} is required, and this can be derived using (3.4.6).

Denote by $\mathcal{Q}_k$ the set of k -paths and denote by T_k the total traffic on all k -paths, given by:

$$T_k := \sum_{q \in \mathcal{Q}_k} \tau_q,$$

where τ_q is the traffic routed to path q . Then (3.4.6) can be restated as:

$$\text{Under the UE, } T_k = 0 \quad \forall \quad k \leq (k^* - 1) \iff T \leq \frac{N_{k^*}}{k^*}. \quad (3.5.1)$$

Denote by c_q the cost of traversing path q and let $\hat{q} \in \mathcal{Q}_{k^*-1}$ be a $(k^* - 1)$ -path that does not overlap with any k^* -paths, then

$$c_{\hat{q}} \begin{cases} = J - (k^* - 1), & \text{if } T_{k^*-1} = 0; \\ > J - (k^* - 1), & \text{if } T_{k^*-1} > 0. \end{cases} \quad (3.5.2)$$

Although $\hat{q}$ carries no traffic when $T_{k^*-1} = 0$, the theoretical cost $c_{\hat{q}}$ is still relevant to the following discussion. Recall that under the UE all traffic incurs the same cost c , and the cost of traversing any unused path is greater than or equal to c . It follows that:

$$c_{\hat{q}} \begin{cases} \geq c, & \text{if } T_{k^*-1} = 0; \\ = c, & \text{if } T_{k^*-1} > 0. \end{cases} \quad (3.5.3)$$

Inequalities for c can be obtained from (3.5.2) and (3.5.3):

$$c \begin{cases} \leq J - (k^* - 1), & \text{if } T_{k^*-1} = 0; \\ > J - (k^* - 1), & \text{if } T_{k^*-1} > 0. \end{cases} \quad (3.5.4)$$

The conditions in the above can be given in terms of T using (3.5.1):

$$c \begin{cases} \leq J - (k^* - 1), & \text{if } T \leq N_{k^*}/k^*; \\ > J - (k^* - 1), & \text{if } T > N_{k^*}/k^*. \end{cases}$$

The total cost under the UE is given by $C(\tau^{\text{ue}}) = Tc$, which substituting into the above provides

$$\frac{C(\tau^{\text{ue}})}{T} \begin{cases} \leq J - (k^* - 1), & \text{if } T \leq N_{k^*}/k^*; \\ > J - (k^* - 1), & \text{if } T > N_{k^*}/k^*. \end{cases} \quad (3.5.5)$$

This result can be used to derive the following method for computing N_{k^*} . First, note that for any given network realisation, at least one k^* -path exists (by the definition of k^*), hence $N_{k^*} \geq 1$. The method for computing N_{k^*} begins by computing the UE for $T = 1/k^*$. This value of T is chosen because it satisfies $T \leq N_{k^*}/k^*$ and hence by (3.5.5), $C(\tau^{\text{ue}})/T \leq J - (k^* - 1)$. The value of T is then increased incrementally until $C(\tau^{\text{ue}})/T = J - (k^* - 1)$, at which point it is known that $T = N_{k^*}/k^*$. Since T is known and k^* can be easily observed from any given network realisation, N_{k^*} is simply found using $N_{k^*} = Tk^*$. Pseudocode for computing N_{k^*} is given in Algorithm 1. Figure 3.5.2a displays the variation of $\bar{\mathcal{P}}$ and $\langle \mathcal{G} \rangle$ with respect to p for the lattice model, where N_{k^*} is computed using the above method, and shows that local maxima in $\bar{\mathcal{P}}$ do indeed coincide with local minima in $\langle \mathcal{G} \rangle$. Hence the maxima of $\bar{\mathcal{P}}$ can be located.

Figure 3.5.2b displays a survey of the values p_{max} of p at which $\bar{\mathcal{P}}$ is maximised and $\langle \mathcal{G} \rangle$ is minimised (for $k^* = J$). This is plotted against lattice size, for fixed aspect ratio $2I/J = 80/3$ and for $T = 0.2$ (blue), $T = 1$ (red) and $T = 5$ (green). Correspondence

Algorithm 1 Algorithm for computing N_{k^*} , where $\mathbb{E}$ denotes the set of all the edges of the network, τ^i the traffic routed to edge $i \in \mathbb{E}$, and $\epsilon = 0.0001$ a small constant.

```

    Compute the UE for  $T = 1/k^*$ ;
    while (  $C(\tau^{\text{ue}}) < J - (k^* - 1)$  ) {
        for (  $i \in \mathbb{E}$  ) {
             $\tau^i = \tau^i + \tau^i \epsilon / T$ ;
        }
         $T = T + \epsilon$ ;
    }
    Set  $N_k^* = Tk^*$ ;
    
```

between the maximum of $\bar{\mathcal{P}}$ and the minimum of $\langle \mathcal{G} \rangle$ holds for all lattice sizes and values of T . It can also be observed that $p_{\max}$ increases with lattice size. This is because J , and hence the value of N_{k^*} that minimises $\langle \mathcal{G} \rangle$ for $k^* = J$, increases with lattice size. For the same reason, $p_{\max}$ also increases with T , and tends to the lattice percolation threshold only as T becomes small. Hence the theory proposed in [4] — that the price of anarchy is maximised at the percolation threshold — is not universal and only holds for particular systems.

Recall from Chapter 2 that while the variation of $\mathcal{P}$ with T displayed similar behaviour for both $2I > J$ and $2I = J$, the variation of $\bar{\mathcal{P}}$ with p displayed markedly different behaviour between $2I > J$ and $2I = J$. The effect of the lattice aspect ratio on the behaviour of $\bar{\mathcal{P}}$ as a function of p will therefore be discussed in the next subsection.

3.5.3 Dependence of $\bar{\mathcal{P}}(p)$ on lattice aspect ratio

In Chapter 2, it was shown that $\bar{\mathcal{P}}(p)$ exhibited ripples only when the height of the lattice was significantly greater than its width. It was noted from figure 3.4.3 that for each i , $1 \leq i \leq J$, there exist intervals of p where the majority of realisations exhibit $k^* = i$, so that these realisations are dominant in determining the value of $\bar{\mathcal{P}}$. When p is at the lower end of any given interval, the value of $\langle N_{k^*} \rangle$ is small. As p increases within the interval, so does $\langle N_{k^*} \rangle$. The cause of the ripples for lattices with $2I/J > 1$ can therefore be explained as follows. When $\bar{\mathcal{P}}(p)$ is increasing, fewer network realisations are exhibiting $N_{k^*} > Tk^*$ for $k^* = i - 1$, which gives a small $\mathcal{P}$, and more network realisations are exhibiting $N_{k^*} \sim Tk^*$ for $k^* = i$, which gives a higher $\mathcal{P}$. When $\bar{\mathcal{P}}(p)$ is decreasing, $\langle N_{k^*} \rangle$ is increasing with p and more network realisations are exhibiting

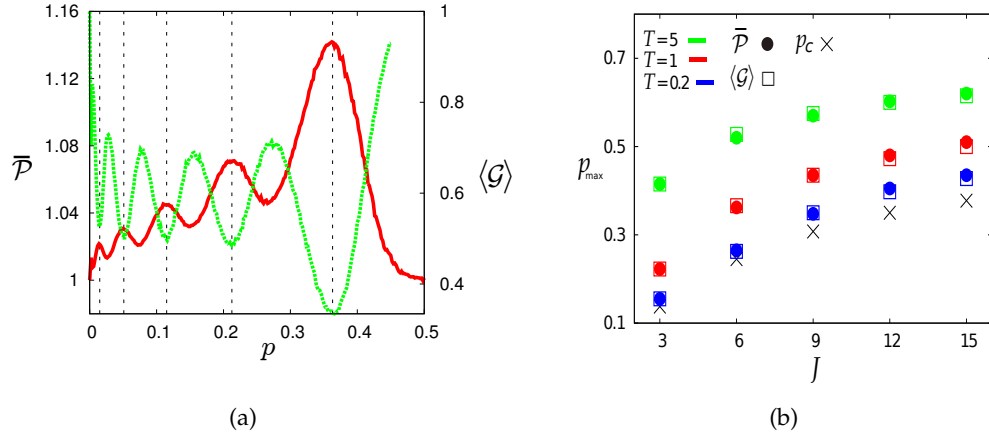


FIG. 3.5.2: (a) The variation in $\bar{P}$ (red) and in $\langle \mathcal{G} \rangle$ (green) with respect to p for the diamond lattice network with $I = 80$ and $J = 6$, with fixed $T = 1$. (b) The principal peak locations $p_{\max}$ of $\bar{P}(p, T = \text{constant})$ (circles) for the lattice network and the corresponding minimum of $\langle \mathcal{G} \rangle$ (squares) for $T = 0.2$ (blue), $T = 1$ (red) and $T = 5$ (green). This is plotted against network size denoted by J , where all lattices considered have the same aspect ratio $2I/J = 80/3$. The percolation threshold p_c (black crosses) is estimated as the value of p in which precisely half of the network realisations contain a J -path.

$N_{k^*} > Tk^*$ for $k^* = i$, giving a smaller $\mathcal{P}$.

So why are ripples absent for lattices with $2I/J = 1$? Note from figure 3.4.3 that the intervals of p are larger for lattices with $2I/J > 1$ than with $2I/J = 1$. This means that for $2I/J = 1$, $\langle N_{k^*} \rangle$ for some $k^* = i$ remains relatively small before k^* increments to $i + 1$, which does not allow for $\bar{\mathcal{P}}(p)$ to start decreasing. The two exceptions to this rule are when $k^* = 1$ and $k^* = J$, which is why figure 2.3.1 shows that $\bar{\mathcal{P}}(p)$ attains a local maximum at very small p and a global maximum at around the percolation threshold.

3.6 Summary

This chapter has considered the underlying causes for the variation of the PoA observed in chapter 2. It was recognised that identifying these causes in the original model analytically would be too difficult due to the complexity of the traffic flows. Hence a new model consisting of non-overlapping paths between source nodes and destination nodes was introduced. This simplified model produced similar numerical results to the original model but also allowed for analytical investigation of the traffic flows. From these analytical findings, it was deduced that the UE and the SO utilise different classes of paths under different conditions. The variation of the UE and the SO costs, and hence the PoA, with p and T is governed by the classes of paths being utilised under the two traffic flows. Route transition points under the UE and the SO were derived. A new theory was provided proposing that the PoA is expected to be maximised at route transition points under the UE, which numerical results confirmed. Finally, the conclusions drawn from the analysis of the simplified model were tested against the results observed in Chapter 2, and it was confirmed that the behaviour of the PoA for the original lattice model was governed by the same underlying causes as that of the simplified model. The results of this chapter and Chapter 2 were published in [48].

This chapter has developed a theory for the behaviour of the PoA based on the route transition points of the traffic flows, and has provided significant numerical evidence to support this theory. The next chapter will consider a method for reducing the PoA by restricting the traffic capacity on certain edges of the network. Analysis will consider the effect of the cost of “cap implementation” on: the optimum number of caps, the efficiency of cap implementation, and the reduction to the UE total cost and the PoA.

Restricting edge capacities: a method for reducing inefficiency

4.1 Introduction

This chapter builds on and finds a way of exploiting the investigation of the previous chapters into the behaviour of the POA. Recalling that the POA is a proxy for inefficiency, it would clearly be desirable to take actions that have the effect of lowering the POA. This chapter therefore considers a method for achieving this goal via an alteration to the functioning of the network without affecting its architecture. Namely, a “cap” is placed on one or more of the most congested edges of the network restricting the amount of traffic the edge can carry. On real road networks, such a cap would represent a restriction to the flux of traffic on a particular road, which could be enforced via a traffic light system or by a traffic code in which vehicles must keep a certain distance from the vehicle directly in front. It will be shown that the total cost of the UE traffic flow typically *decreases* as a result of capping edges, since traffic that was previously using a capped edge is redistributed to less congested edges. This phenomenon is similar to the well-known Braess’s paradox in which the removal of an edge results in a more efficient UE flow (see Chapter 1). For any given network realisation, the optimum number of caps for reducing inefficiency is computed. If there is no cost associated with cap implementation, then clearly the optimum strategy is to place caps on all edges such that the resulting flow is the SO and inefficiency is reduced altogether. Since this case is trivial, a positive cost per cap will be considered. A rationale for the optimum number of caps will be provided, and the reduction to the UE cost and the price of anarchy will be analysed with respect to the cost of cap implementation. The dependence of the average optimum number of caps on the cost of cap implementation

will also be analysed.

4.2 Methodology

Caps are applied over N_{cap} iterations, with the most congested edge being capped at each iteration. The value of the cap applied to a given edge is set to the amount of traffic routed to that edge under the SO traffic flow. Denote the cost of implementing each cap on an edge by μ . We are interested in finding the number of caps that optimises efficiency. If implementing x caps results in a reduction to the total cost of the UE flow by greater than μx , then that represents a cost-effective strategy. However, it is not necessarily the *most* cost-effective strategy. Consider the example given in Table 4.1, which shows the total cost decreasing as caps are applied. When the cost per cap is $\mu = 0.01$, the first three caps are cost-effective, since they each reduce the total cost by > 0.01 . Since the 4th cap decreases the cost by < 0.01 , the optimum number of caps in this case is 3. Note that the application of 4 caps is not the most cost-effective strategy, even though it reduces the initial cost by more than 4μ .

Denoting the total cost of the UE flow at the i th iteration by $C^i(\tau^{\text{ue}})$, the number of caps that optimises efficiency is given by:

$$i^* := \max\{i : C^j(\tau^{\text{ue}}) - C^i(\tau^{\text{ue}}) > (i - j)\mu \quad \forall j < i\}.$$

For the purpose of speeding up the computational results, the optimum number of caps will be bounded by

$$\mathcal{N} := \min\{i^*, N_{\text{cap}}\}.$$

It will be seen that this does not compromise accuracy, however, since the value of N_{cap} will be chosen such that virtually all network realisations exhibit $\mathcal{N} < N_{\text{cap}}$ whenever $\mu > 0$. For any given realisation, the value of $\mathcal{N}$ depends on μ , since it is clearly an unwise policy to apply caps when the cost of implementation is greater than the reduction in the cost of the traffic flow.

4.3 Results

Figure 4.3.1 considers the reduction δC in the total cost after $\mathcal{N}$ caps are applied to the network, where

$$\delta C := \frac{C^0(\tau^{\text{ue}}) - C^{\mathcal{N}}(\tau^{\text{ue}})}{C^0(\tau^{\text{ue}})}.$$

i	0	1	2	3	4
$C^i(\tau^{\text{ue}})$	1.2	1.1	1.05	1.01	1.001

Table 4.1: A theoretical example of the total UE cost decreasing as caps are applied.

When the cost per cap is $\mu = 0.01$, the first three caps are cost-effective, since they each reduce the total cost by > 0.01 . Since the 4th cap decreases the cost by < 0.01 , the optimum number of caps in this case is 3.

Each point on the graph represents a realisation of a lattice network with $I = 80$, $J = 6$, $p = 0.36$, $T = 1$, and $N_{\text{cap}} = 10$. The value of p here was chosen because it maximises $\langle \mathcal{P} \rangle$. As a result, almost all network realisations exhibit $k^* = J = 6$, where k^* is the number of variable-cost edges within paths that contain the greatest number of variable-cost edges. The value of N_{cap} was chosen such that $\mathcal{N} < N_{\text{cap}}$ for almost all network realisations when μ is not negligibly small. The values of p and N_{cap} for other sets of parameter values considered in this chapter shall also be chosen in this way. Plots are given for $\mu = 0, 0.001, 0.01, 0.03$, and 0.04 . To put these values into perspective, the value of $C^0(\tau^{\text{ue}})$ for the set of parameters considered ranges from 0.4 to 1.3 and has a mean value of ≈ 0.98 . The colour and shape of a point represents the value of $\mathcal{N}$ for that realisation, with the symbol key displayed in figure 4.3.1f. The first thing of note is that the points lie in an inverted ‘V’ shape for all of the plots. Figure 4.3.1a displays results for $\mu = 0$, and shows that for many of the points, $\mathcal{N}$ takes its maximum possible value: $\mathcal{N} = N_{\text{cap}} = 10$. This value of μ is trivial, since if the implementation of each cap is cost-free then there is no reason not to apply caps to all edges of the network. Consequently, the points are unordered.

Figure 4.3.1b displays results for $\mu = 0.001$ and shows that while the points remain in an inverted ‘V’ shape, the colouring of the points has drastically changed. Firstly, $\mathcal{N} < N_{\text{cap}}$ for almost all realisations. This is because the cost of cap implementation is now taking effect, even for a small value of μ . Moreover, realisations of a given value of $\mathcal{N}$ now exhibit ordering, tending to congregate at a similar location on the graph. The relative frequency histogram of $\mathcal{N}$ for $\mu = 0.001$ is displayed in figure 4.3.4a. Figures 4.3.1(b-e) show that as μ increases, realisations that exhibit $\mathcal{N} \geq 5$ and $\delta C > 0$ for small μ exhibit $\mathcal{N} = 0$ and $\delta C = 0$ at larger μ . Moreover, many realisations that exhibit $\mathcal{N} < 5$ at small μ exhibit a smaller $\mathcal{N}$ and smaller δC at larger μ .

Figure 4.3.2 shows the reduction $\delta \mathcal{P} := \mathcal{P}^0 - \mathcal{P}^{\mathcal{N}}$ in the price of anarchy, where $\mathcal{P}^i$

denotes the price of anarchy at the i th iteration. Plots are displayed of $\delta\mathcal{P}$ against $\mathcal{P}^0$ for the same values of μ considered in figure 4.3.1. It can be observed that many realisations lie on the line $\delta\mathcal{P} = \mathcal{P}^0 - 1$. This rearranges to $\mathcal{P}^0 - \delta\mathcal{P} = 1$, indicating that inefficiency is completely reduced after the caps are implemented. Figures 4.3.2b and 4.3.2c show that realisations that satisfy $\mathcal{P}^0 - \delta\mathcal{P} = 1$ exhibit $\mathcal{N} \leq 5$, whereas realisations that do not satisfy $\mathcal{P}^0 - \delta\mathcal{P} = 1$ exhibit $\mathcal{N} > 5$.

Figure 4.3.3 displays results for a lattice network with $I = 80$, $J = 6$, $p = 0.42$, $T = 2$, $N_{\text{cap}} = 15$, and $\mu = 0.005$, with the corresponding relative frequency histogram of $\mathcal{N}$ displayed in figure 4.3.4b. In this case, the value of $C^0(\tau^{\text{ue}})$ ranges from 1 to 2.5 and has a mean value of ≈ 2.01 . The results are similar to the previous case but with two noteworthy differences:

1. Realisations where δC and $\delta\mathcal{P}$ are maximised exhibit $\mathcal{N} \approx 10$, as opposed to $\mathcal{N} \approx 5$ in the previous case.
2. The initial cost corresponding to the maximum δC is given by $C^0(\tau^{\text{ue}}) = 2$, as opposed to $C^0(\tau^{\text{ue}}) = 1$ in the previous case.

The next section will discuss the causes for the above observations.

4.4 Discussion

Clearly, when caps are applied to the edges of the most congested paths, the traffic on these paths is redistributed to paths with fewer variable edges, thereby reducing the cost of the traffic flow. For any given network realisation, the paths with the greatest number of variable edges (*i.e.* the k^* -paths) will carry the most traffic under the UE flow. Hence edges that belong to these paths are the first to be capped. Note that a cap on just one edge of a path is sufficient for reducing the traffic on all edges of the path, and hence has a global impact on the traffic flow. Once an edge of each k^* -path has been capped, any further caps to edges of k^* -paths would have only a negligible effect on the UE traffic flow and cost. It is plausible, then, that the optimum number of caps is the minimum number required to cap all k^* -paths. It would be expected that realisations with the same N_{k^*} and $\mathcal{N}$ will have similar values of $C^0(\tau^{\text{ue}})$ and δC , which explains why points of a given colour exhibit clustering in the plots.

Note from figure 4.3.1 that (except in the trivial case of $\mu = 0$) $\mathcal{N}$ generally decreases with increasing $C^0(\tau^{\text{ue}})$. This is because the cost of the flow decreases with the number

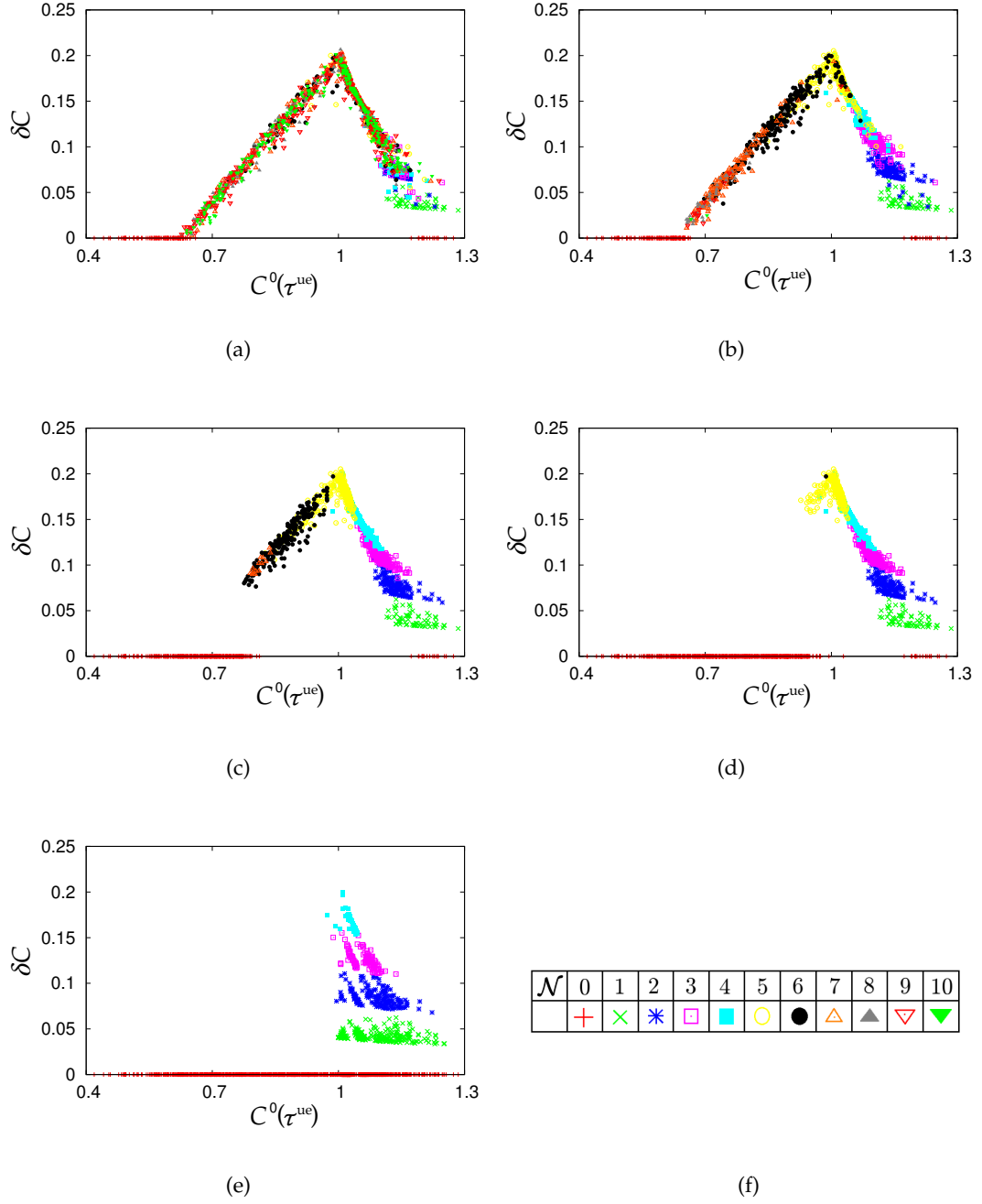


FIG. 4.3.1: (a)-(e) Plots of δC against $C^0(\tau^{\text{ue}})$ for $I = 80$, $J = 6$, $p = 0.36$, $T = 1$, $N_{\text{cap}} = 10$, and (a) $\mu = 0$, (b) $\mu = 0.001$, (c) $\mu = 0.01$, (d) $\mu = 0.03$, (e) $\mu = 0.04$. (f) Key displaying the correspondence between the shape and colour of a given point and the value of $\mathcal{N}$ for that realisation.

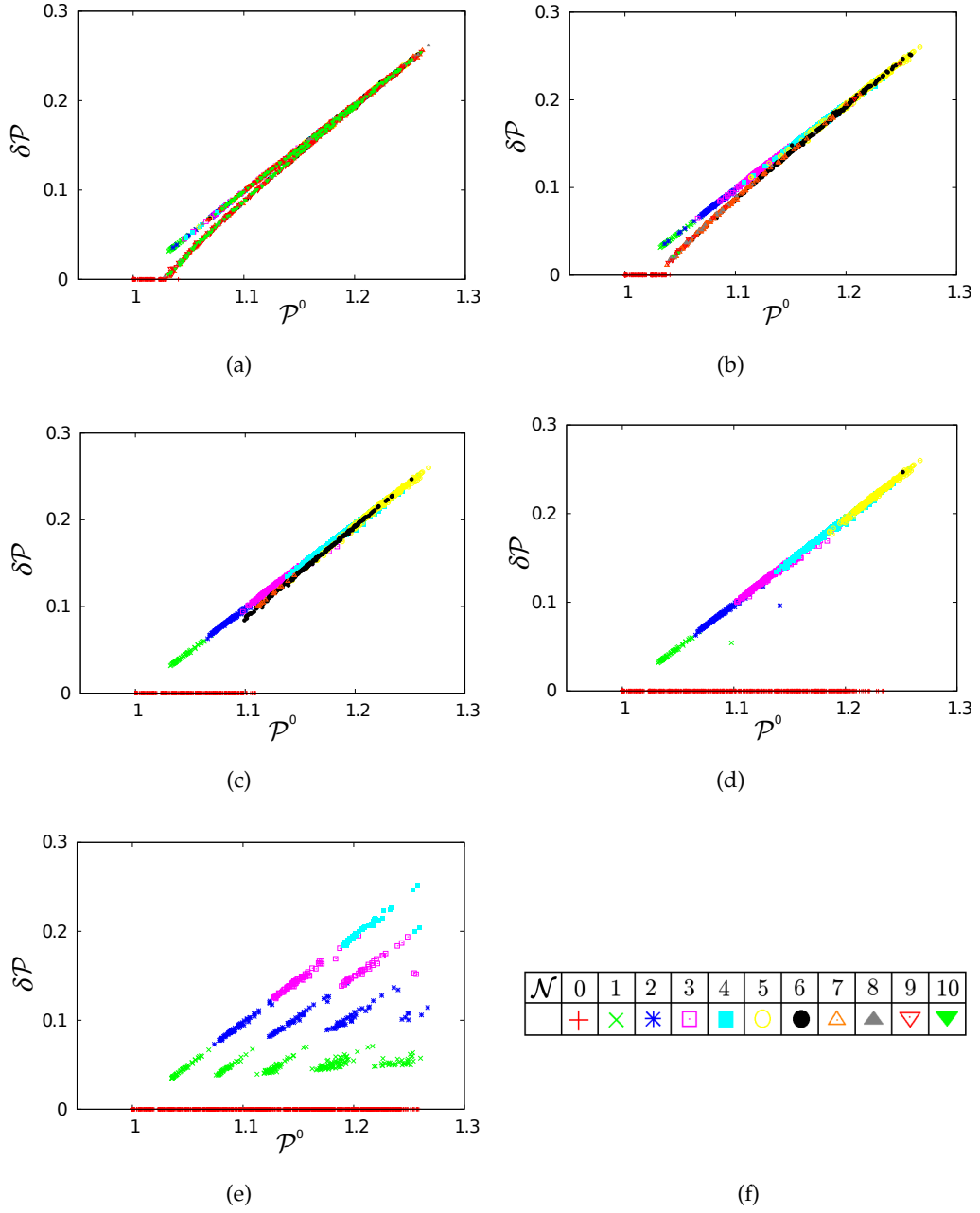


FIG. 4.3.2: (a)-(e) Plots of $\delta\mathcal{P}$ against $\mathcal{P}^0$ for $I = 80$, $J = 6$, $p = 0.36$, $T = 1$, $N_{\text{cap}} = 10$, and (a) $\mu = 0$, (b) $\mu = 0.001$, (c) $\mu = 0.01$, (d) $\mu = 0.03$, (e) $\mu = 0.04$. (f) Key displaying the correspondence between the shape and colour of a given point and the value of $\mathcal{N}$ for that realisation.

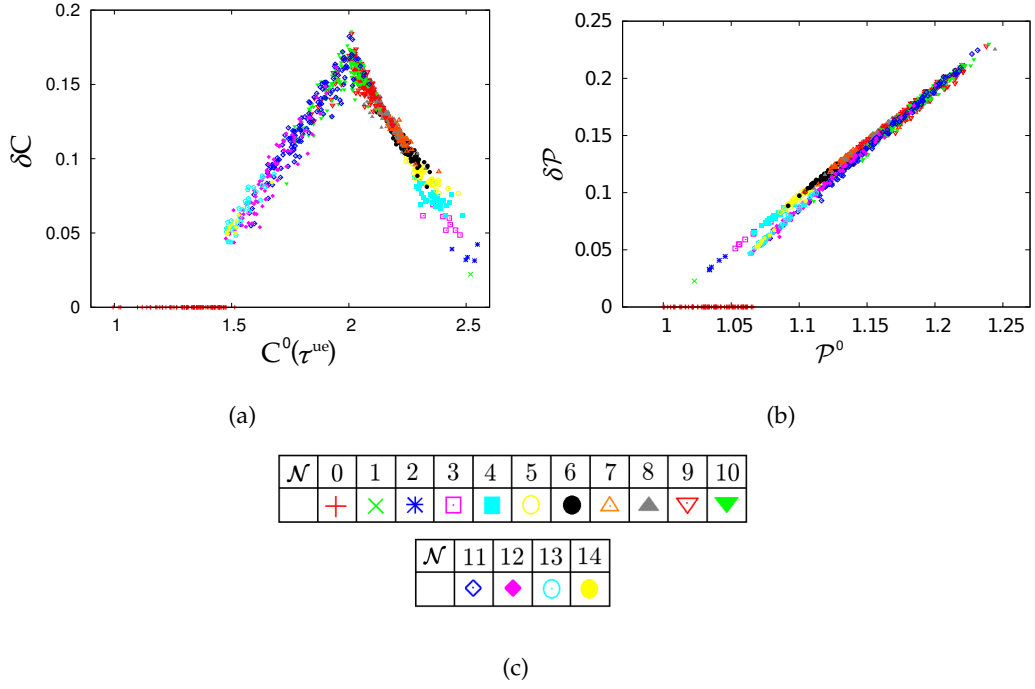


FIG. 4.3.3: Plots of (a) δC against $C^0(\tau^{ue})$ and (b) δP against P^0 for $I = 80$, $J = 6$, $p = 0.42$, $T = 2$, $N_{\text{cap}} = 15$, and $\mu = 0.005$. (c) Key displaying the correspondence between the shape and colour of a given point and the value of $\mathcal{N}$ for that realisation.

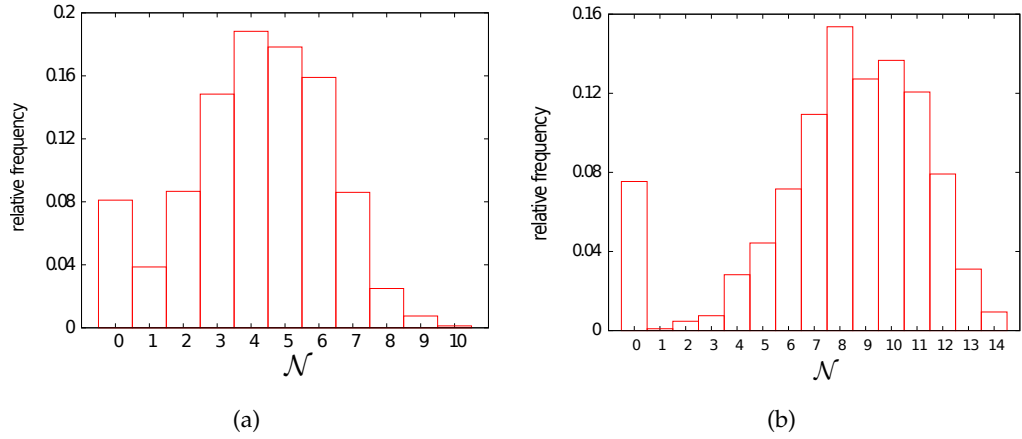


FIG. 4.3.4: Relative frequency histograms of the values of $\mathcal{N}$ for (a) $I = 80$, $J = 6$, $p = 0.36$, $T = 1$, $N_{\text{cap}} = 10$, and $\mu = 0.001$. (b) $I = 80$, $J = 6$, $p = 0.42$, $T = 2$, $N_{\text{cap}} = 15$, and $\mu = 0.005$.

of k^* -paths, *i.e.* $C^0(\tau^{\text{ue}})$ decreases with increasing N_{k^*} , and $\mathcal{N}$ scales with N_{k^*} . An exception to this trend are realisations with $\mathcal{N} = 0$, which exhibit the lowest and also the highest $C^0(\tau^{\text{ue}})$. Those with a low value of $C^0(\tau^{\text{ue}})$ correspond to network realisations that contain many k^* -paths. For such realisations, all traffic is routed to the k^* -paths, and initial caps shift traffic away from the k^* -paths that are capped and onto k^* -paths that are not capped. Since the total cost of the traffic over a given class of paths is minimised when the traffic is uniformly distributed on the paths, this shift of traffic results in an *increase* to the total cost of the flow. The cost starts to decrease only when enough k^* -paths are capped such that some of the traffic is routed to $(k^* - 1)$ -paths. For the case of $\mu = 0$, these realisations would have a large $\mathcal{N}$ but for the maximum possible number of caps considered here being $N_{\text{cap}} = 10$. When μ is non-negligible, the cost of implementing the number of caps required to reduce the cost of the flow is not cost-effective, *i.e.* the cost of the caps outweighs the reduction in the cost of the flow, hence $\mathcal{N} = 0$. As μ increases, a greater number of realisations fall into this category and exhibit $\mathcal{N} = 0$. Those points with $\mathcal{N} = 0$ and a high value of $C^0(\tau^{\text{ue}})$ correspond to network realisations with $k^* = J - 1$ (*i.e.* no J -paths exist). Such realisations also contain an abundance of k^* -paths hence, for the same reason as above, $\mathcal{N} = 0$.

It was observed from figure 4.3.2 that $\delta\mathcal{P}$ is strictly increasing with $\mathcal{P}^0$. This is because a higher $\mathcal{P}^0$ indicates that k^* -paths are overused to a greater degree and hence caps on these paths are more effective. It might have been expected that the reduction in the total cost and in the price of anarchy would monotonically increase with the optimum number of caps, *i.e.* δC and $\delta\mathcal{P}$ would monotonically increase with $\mathcal{N}$. However, figures 4.3.1 and 4.3.2 show that for $\mu > 0$, both δC and $\delta\mathcal{P}$ are maximised for realisations corresponding to $\mathcal{N} = 5$. Recall from Section 3.4.3 that $\mathcal{P}$ is maximised when $N_{k^*} = Tk^*$, such that all traffic under the UE is routed to the k^* -paths. Since $k^* = J = 6$ and $T = 1$, it follows that $\mathcal{P}$ — and hence δC and $\delta\mathcal{P}$ — is maximised when $N_6 = 6$. This raises the question of why the points that maximise the reduction of $C^0(\tau^{\text{ue}})$ and $\mathcal{P}^0$ mostly correspond to realisations with $\mathcal{N} = 5$ rather than $\mathcal{N} = 6$. Recall that for the lattice network, k^* -paths can overlap one another by sharing variable edges. Any variable edge that belongs to two k^* -paths will surely be one of the most congested edges in the network, and a cap applied to this edge will reduce the traffic on both k^* -paths. It is plausible that most realisations that exhibit $N_6 = 6$ contain an edge that belongs to two of the 6-paths, in which case only 5 caps are necessary to cover all of the 6-paths. Similarly, for the case given in figure 4.3.3, caps on 10 of the edges is sufficient for capping all k^* -paths when $N_{k^*} = Tk^* = 12$. Finally, when $N_{k^*} = Tk^*$ the initial cost of the UE is $C^0(\tau^{\text{ue}}) = T$, which explains why δC is maximised at $C^0(\tau^{\text{ue}}) = 1$ in figure

4.3.1 and $C^0(\tau^{\text{ue}}) = 2$ in figure 4.3.3.

4.5 Testing the theory using the simplified model

The previous section provided plausible explanations for the numerical results presented in Section 4.3. This section will confirm these findings by considering the case of the simplified model of non-overlapping paths that was introduced in Chapter 3.

Recall that one of the conjectures of Section 4.4 is that, whenever the cost per cap is non-negligible, the optimum number of caps $\mathcal{N}$ is the minimum number that is necessary to cap every k^* -path. Since there is no overlapping between paths in the simplified model, it would be expected that $\mathcal{N} = N_{k^*}$ for any given network realisation. The numerical results of this section have therefore been computed under this assumption. Figure 4.5.1 presents results for the simplified model with $I = 160$, $J = 6$, $p = 0.55$, and $T = 1$. The values of I , J and T here were chosen because they are the same values considered in the previous chapter, and the value of p was chosen because it again maximises $\langle \mathcal{P} \rangle$. It was described in Section 4.4 that δC and $\delta \mathcal{P}$ are maximised for realisations with $N_{k^*} = Tk^*$. It can be seen in figures 4.5.1a and 4.5.1b that δC and $\delta \mathcal{P}$ are maximised for realisations with $\mathcal{N} = 6$. Since $Tk^* = TJ = 6$, the results of figures 4.5.1a and 4.5.1b therefore support the conjecture that $\mathcal{N}$ is the minimum number of caps that is necessary to cap every k^* -path. Note that the points displayed in figure 4.5.1a are more distinctly categorised by their value of $\mathcal{N}$ than in the lattice model. This is because — as described in Chapter 3 — N_{k^*} always takes an integer value for the simplified model, whereas its effective value for the lattice model need not be an integer.

In the analysis of $\delta \mathcal{P}$ for the lattice model, it was noted that realisations with a smaller value of $\mathcal{N}$ satisfied $\delta \mathcal{P} = \mathcal{P}^0 - 1$, whereas realisations with a larger $\mathcal{N}$ did not. Figure 4.5.1b shows that *all* of the realisations satisfy $\delta \mathcal{P} = \mathcal{P}^0 - 1$ in the simplified model. This suggests that the overlapping of paths is the reason that realisations with a higher $\mathcal{N}$ do not satisfy $\delta \mathcal{P} = \mathcal{P}^0 - 1$ in the lattice model. Finally, figure 4.5.1c shows the value of k^* for each point, and confirms the conjecture that $k^* = J - 1$ for realisations with $\mathcal{N} = 0$ and a high value of $C^0(\tau^{\text{ue}})$.

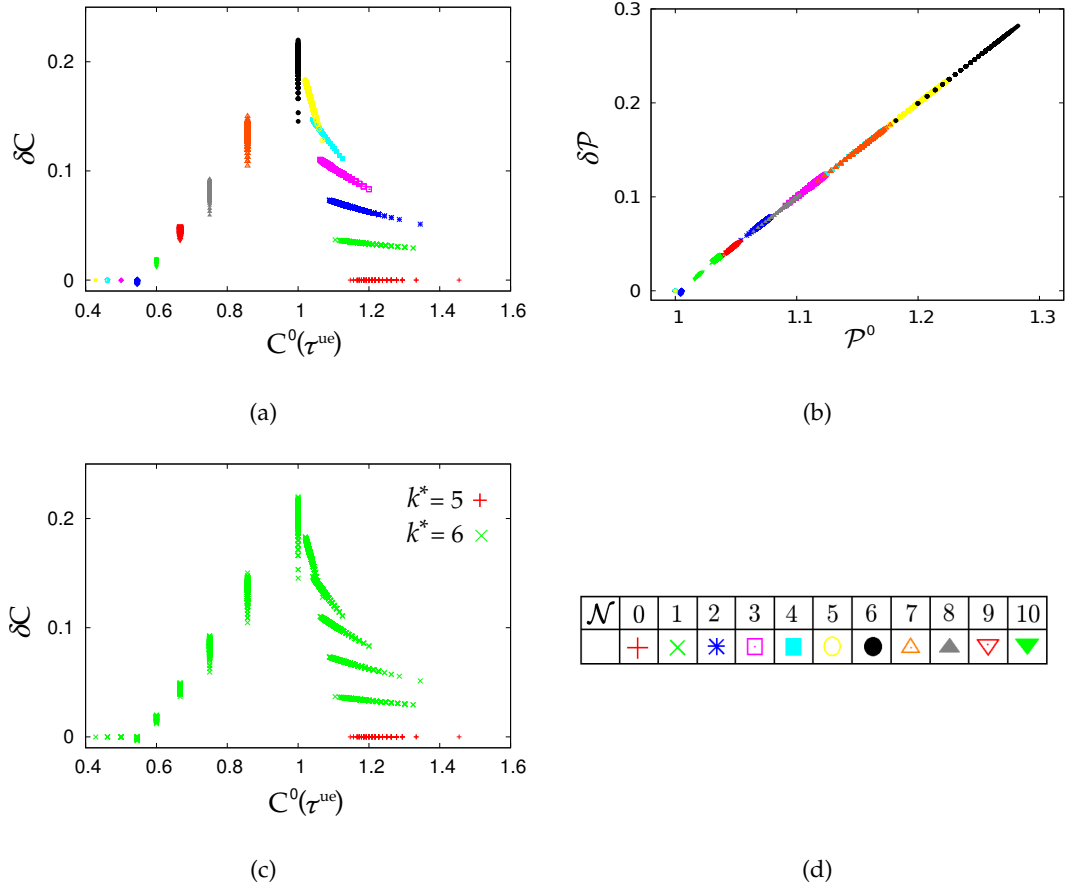


FIG. 4.5.1: Numerical results for the simplified model with $I = 160$, $J = 6$, $p = 0.55$, and $T = 1$. For figures (a) and (b), the colour and shape of a point is given by the key displayed in figure (d). (a) δC against $C^0(\tau^{\text{ue}})$, where the number of caps applied is set to $\mathcal{N} = N_{k^*}$. (b) $\delta \mathcal{P}$ against $\mathcal{P}^0$. (c) δC against $C^0(\tau^{\text{ue}})$ where realisations are coloured by their value of k^* .

4.6 Effect of the cost of cap implementation

Recall that μ can be considered as the cost of implementing a cap, which to a policy-maker could take the form of time and resources associated with the implementation and maintenance of the cap. Caps will be used more sparingly with a higher cost of implementation, which in turn restricts the reduction to the cost of the flow. Figure 4.6.1a shows the variation of the average $\langle \mathcal{N} \rangle$ (red points) with μ for the lattice network with $I = 80$, $J = 6$, $p = 0.36$, $T = 1$, and $N_{\text{cap}} = 10$. As μ increases from 0, there is an initial steep fall in the value of $\langle \mathcal{N} \rangle$ followed by a steady decline and another steep fall to 0. The variation in the average reduction to the cost $\langle C^0(\tau^{\text{ue}}) - C^{\mathcal{N}}(\tau^{\text{ue}}) \rangle$ (green points) displays similar behaviour but without the drop at small μ , which suggests that the initial decline of $\langle \mathcal{N} \rangle$ is due to the disuse of those caps that reduce the cost by an insignificant amount. Also displayed is the average total cost of cap implementation $\mu \langle \mathcal{N} \rangle$ (blue points). When this approaches the average reduction to the cost of the flow, the application of caps ceases to be cost-effective and $\langle \mathcal{N} \rangle$ quickly falls to 0. Figure 4.6.1b shows similar results for the lattice network with $I = 80$, $J = 6$, $p = 0.42$, $T = 2$, and $N_{\text{cap}} = 15$. Since the lattice networks considered here comprise 960 edges, note that even for small μ the value of $\langle \mathcal{N} \rangle$ is a very small fraction of the overall number of edges.

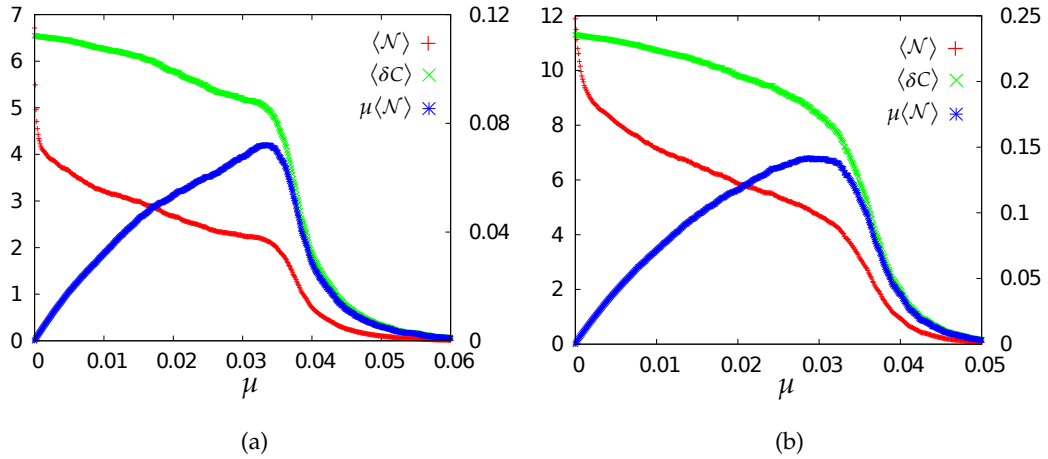


FIG. 4.6.1: The average optimum number of caps $\langle \mathcal{N} \rangle$ (left-hand y-axis), the average reduction in the cost of the UE flow $\langle C^0(\tau^{\text{ue}}) - C^{\mathcal{N}}(\tau^{\text{ue}}) \rangle$ (right-hand y-axis) and the average cost of cap implementation $\mu \langle \mathcal{N} \rangle$ (right-hand y-axis) as functions of μ , for the lattice network with (a) $I = 80$, $J = 6$, $p = 0.36$, $T = 1$, and $N_{\text{cap}} = 10$; (b) $I = 80$, $J = 6$, $p = 0.42$, $T = 2$, and $N_{\text{cap}} = 15$.

4.7 Summary

The findings of this section show that applying caps on several of the most congested edges in the network can significantly reduce the cost of the UE flow and the price of anarchy. For many network realisations, cap implementation eliminates inefficiency altogether. It was also shown that the optimum number of caps is the minimum number required to cap all k^* -paths, and the average optimum number of caps is a small proportion of the overall number of edges in the network. For the management of road traffic networks, for example, this suggests that restricting the traffic flow on a select few roads would be sufficient for reducing inefficiency, and may be a more convenient and cost-effective method than having a central system direct all of the traffic. Finally, the average optimum number of caps was considered as a function of the cost μ per cap. As μ increases and the total cost of cap implementation approaches the reduction to the cost of the UE flow, cap implementation eventually becomes redundant. Although the average value for the initial UE cost $C^0(\tau^{\text{ue}})$ is very different for the two lattice network cases considered (0.98 and 2.01), the value of μ at which cap implementation becomes redundant is ≈ 0.05 for both cases.

This chapter concludes the research of this thesis using the traffic-flow network model, and the next chapter will consider the price of anarchy in a different context. A new model consisting of the pairwise matching of agents within a spatial domain will be introduced. When all agents seek to pair with their nearest available neighbour, the resulting matching may not minimise the distance between paired agents. The price of anarchy is again used as a proxy for inefficiency and analysed as a function of various parameters of the system, including the number of pairs of agents, the size and shape of the domain, and the ratio between the sizes of the two sets of agents.

Inefficiency for bipartite matching of agents

5.1 Introduction

This chapter diverges from the traffic flow model that has been used hitherto, and considers inefficiency in the bipartite matching of agents. While this model has a general purview to any system in which two sets of agents are matched into pairs, the original motivation was the recent popularity of ride-sharing companies such as Uber and Lyft. In this setting, the assignment of drivers to requests from riders resembles a bipartite graph matching problem, where riders and drivers are two disjoint sets of graph nodes and dispatches of drivers to riders denote graph edges. With this in mind, we consider the matchings of two disjoint sets of agents located on a planar domain. Initially, distances between agents will be given by the Euclidean distance in the plane. Then the locations of agents will lie on the nodes of a network embedded in the plane, which should provide a more accurate simulation of the matching of riders and drivers on a road network. While the traffic model of Chapters 2-4 considered a directed network, the matchings and networks considered in this chapter will be undirected. The inefficiency measured using the “price of anarchy” will again be obtained and explored for this system.

In the context of ride-sharing, requests from riders and the availability of drivers vary over time. For this reason, a notion of time shall also be considered within the model, such that agents are introduced into the system via a random Poisson process. Here, the weight of a matching depends not only on the distance between matched agents, but also the time between an agent entering the system and its being matched. The in-

efficiency of matching agents at their entry times — dubbed the “price of impatience” — shall be analysed.

5.2 Methodology

Given two sets of agents A and B , a matching is a set $M \subset A \times B$ such that each agent $i \in A \cup B$ is represented by at most one pair $m \in M$. The weight of a matching M is given by the sum of the distances between paired agents:

$$w(M) := \sum_{m \in M} d(m), \quad (5.2.1)$$

where $d(m)$ denotes the Euclidean distance between pair m .

For any given realisation of agent locations, two classes of matchings corresponding to the User Equilibrium (UE) and the System Optimum (SO) are considered. Under the UE, each agent seeks to pair with its nearest neighbour of the opposite set, where mutual agreement between two agents is required for a pair to form. It may be that some agents are unable to pair with their nearest neighbours, in which case they will seek to pair with their second nearest neighbours, and so on until the greatest number of pairings between agents is achieved. In contrast, the SO matching minimises (5.2.1). Similarly to the traffic flow model, the UE and the SO matchings are not equivalent and therefore do not necessarily have equal weights. Denoting the UE and the SO matchings by M^{ue} and M^{so} , respectively, the price of anarchy is given by:

$$\mathcal{P} := \frac{w(M^{\text{ue}})}{w(M^{\text{so}})}.$$

Consider the example shown in Figure 5.3.1a in which two sets of agents $A = \{a_1, a_2\}$ and $B = \{b_1, b_2\}$ are located on a vertical line. From top to bottom, the positions of the agents are ordered as a_1, b_1, a_2, b_2 . The distance between a_1 and b_1 and between a_2 and b_2 is $y > 0$, while the distance between b_1 and a_2 is $x > 0$. When $x < y$, agents b_1 and a_2 are nearest neighbours and mutually agree to pair up under the UE. The UE matching is therefore given by $M^{\text{ue}} = \{(b_1, a_2), (a_1, b_2)\}$, which has weight $w(M^{\text{ue}}) = 2(x + y)$. The SO matching is given by $M^{\text{so}} = \{(a_1, b_1), (a_2, b_2)\}$, which has weight $w(M^{\text{so}}) = 2y$. The price of anarchy is

$$\mathcal{P} = \frac{2(x + y)}{2y} < 2.$$

The upper bound of 2 in the above is significantly greater than the upper bound of $4/3$ exhibited by traffic networks with linear cost functions. This simple example therefore demonstrates that the inefficiency in bipartite graph matching can potentially reach high levels.

5.3 Matchings on a square domain

Suppose the agents are distributed uniformly at random on a square, planar domain with side length S , as shown in figure 5.3.1b. Denote by N the size of each set of agents, so that a matching consists of N pairs. The distance between agents is simply the Euclidean distance in the plane, and the weight of a matching is given by (5.2.1). It is clear that a SO matching exists for any configuration of agent locations, since it is simply the matching with minimum weight. Moreover, for any two sets of agents A and B , there always exists at least one pair of agents that are nearest neighbours, *e.g.* the shortest edge of the complete bipartite graph of A and B . Hence the procedure that iteratively pairs all nearest neighbours and removes them from the system produces a UE matching. Since the distances between agents are real values, each agent has exactly one nearest neighbour and it follows that the UE matching is unique. Pseudocode for computing the UE matching is given in Algorithm 2. The SO matching can be computed using a standard algorithm for finding a minimum weight perfect bipartite matching, which can be found in [42].

The precise algorithm that Uber uses to assign riders and drivers does not seem to be available to the public. Intuitively, minimising the distances between rider-driver pairs would maximise overall rider and driver satisfaction, hence it would be reasonable for Uber to adopt the SO matching. However, it is possible that factors other than the distances between matched pairs are taken into account in the assignment. For example, there is speculation that new riders are more likely to be paired with drivers with a high rating, and new drivers are more likely to be paired with riders with a high rating [43].

Denote the weights of the UE and the SO matchings by w^{ue} and w^{so} , respectively. Continuing with the theme of this thesis, we are interested in how $\mathcal{P}$, w^{ue} and w^{so} vary with certain parameters of the system, including the size of the domain, the aspect ratio of the domain, the number of pairs of agents being matched, and the ratio of “riders” to “drivers”. Since the space of realisations is infinitely large, the average values $\bar{\mathcal{P}}$, $\bar{w}^{\text{ue}}$

and $\overline{w}^{\text{so}}$ over multiple realisations (1000) will be considered.

Algorithm 2 Psuedocode for computing the UE matching, where $n(i)$ denotes the nearest neighbour of the opposite set to agent $i \in A \cup B$.

```

count = 0
while (count < N) {
    for all ( $a \in A, b \in B$ ) {
        assign  $n(a) \in B$  and  $n(b) \in A$ 
    }
    for all ( $a \in A$ ) {
        if (  $n(n(a)) == a$  ) {
            pair  $a$  and  $n(a)$  and remove them from  $A$  and  $B$ .
            count = count + 1
        }
    }
}

```

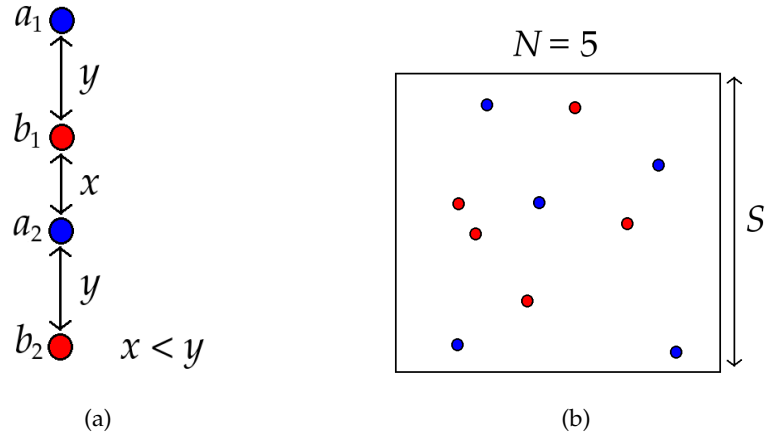


FIG. 5.3.1: (a) A simple example of the price of anarchy in bipartite graph matching. The UE matching is given by $M^{\text{ue}} = \{(b_1, a_2), (a_1, b_2)\}$; the SO matching is given by $M^{\text{so}} = \{(a_1, b_1), (a_2, b_2)\}$; the price of anarchy is $\mathcal{P} = \frac{2(x+y)}{2y}$. (b) Two sets of $N = 5$ agents are distributed uniformly at random in a square, planar domain with side length S .

5.3.1 Variation with the size of the domain

Figure 5.3.2 considers $\bar{\mathcal{P}}$, $\bar{w}^{\text{ue}}$ and $\bar{w}^{\text{so}}$ as functions of the size of the domain, denoted by S . All errorbars in this chapter indicate one standard deviation from the mean. Figure 5.3.2a displays the variation of $\bar{\mathcal{P}}$ with S , with the number of pairs of agents $N = 10$ fixed, and shows that $\bar{\mathcal{P}}$ is invariant with S . Figures 5.3.2b and 5.3.2c display the variations of the two matching weights with S , and show that there is a linear relationship.

To understand these behaviours, suppose that the origin is at the lower left-hand corner of the square domain. For a given configuration of agent locations, consider varying the size of the domain whilst keeping the locations of the agents relative to the domain fixed. So an agent at position $(1, 1)$ for $S = 10$ will be at position $(2, 2)$ for $S = 20$. Although the distance between matched agents varies, the UE and the SO matchings remain the same regardless of the value of S . Therefore, w^{ue} and w^{so} vary at the same rate (with affine scaling) proportional to S , and $\bar{\mathcal{P}}$ remains constant.

5.3.2 Variation with the number of pairs of agents

Figure 5.3.3a is a log-plot of the variation of $\bar{\mathcal{P}}$ with N , for fixed $S = 10$, and shows that $\bar{\mathcal{P}}$ varies logarithmically with N . Figure 5.3.3b shows that the variation of $\bar{w}^{\text{ue}}$ is proportional to $f(N) := \sqrt{N} + \log(N) + \sqrt{N} \log(N)$, i.e. $\bar{w}^{\text{ue}} = O(\sqrt{N} \log(N))$, while figure 5.3.3c shows that the variation of $\bar{w}^{\text{so}}$ is given by $\bar{w}^{\text{so}} = O(\sqrt{N})$. All fittings of functions to data in this chapter were carried out using the “fit” command in *gnuplot*, with the choice of functions being determined via trial and error. Note that, although the average distance between paired agents is decreasing with N , the average weights of the matchings are increasing functions of N . The increasing behaviour of $\bar{\mathcal{P}}$ as a function of N shows that a greater number of pairs of agents results in more inefficient matching sequences under the UE.

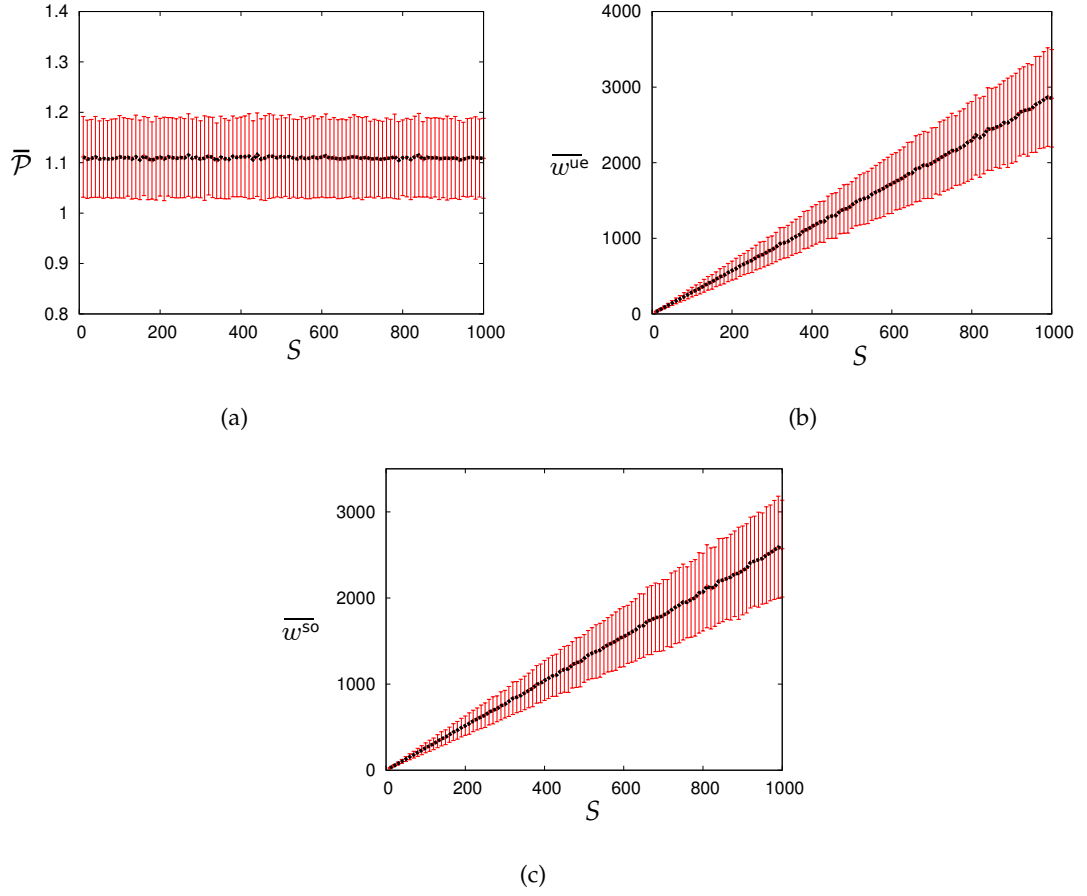


FIG. 5.3.2: The variation of (a) $\bar{\mathcal{P}}$ (b) $\overline{w^{ue}}$ and (c) $\overline{w^{so}}$ with system size S , for fixed $N = 10$.

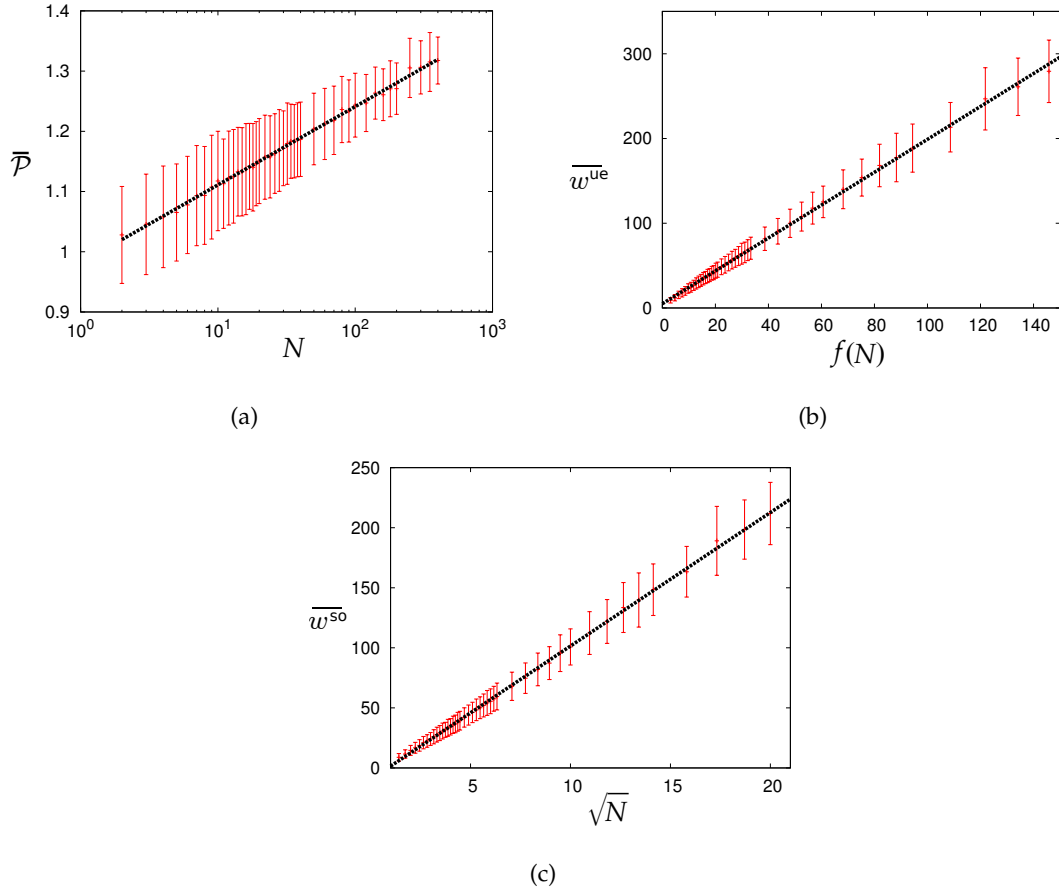


FIG. 5.3.3: The variation of (a) $\bar{\mathcal{P}}$ (b) $\bar{w}^{ue}$ and (c) $\bar{w}^{so}$ with the number of pairs of agents N , for fixed $S = 10$. The function f is given by $f(N) := \sqrt{N} + \log(N) + \sqrt{N} \log(N)$.

5.3.3 Variation with aspect ratio of domain

In Chapters 2 and 3, it was shown that the aspect ratio of the lattice network had a significant effect on the behaviour of $\bar{\mathcal{P}}$. This provides a motivation for an investigation into the effect of the aspect ratio of the domain in the current context of matchings. Suppose that the shape of the domain containing the agents is rectangular, and denote the two side lengths by S_1 and S_2 . Figure 5.3.4 considers the effect of varying the aspect ratio $V = S_1/S_2$, where $S_1 \geq S_2$, while keeping the area $S_1 S_2$ fixed at 100. Figure 5.3.4a displays a log-plot of the variation of $\bar{\mathcal{P}}$ with V ranging from 1 (a 10x10 square domain) to 100 (a 100x1 rectangle), for fixed $N = 10$. It can be observed that $\bar{\mathcal{P}}$ peaks at an intermediate value of V . Figure 5.3.4b shows that the corresponding variation of $\overline{w^{\text{ue}}}$ is given by $\overline{w^{\text{ue}}} = O(\sqrt{V})$. Similar results for the variation of $\overline{w^{\text{so}}}$ can be observed in figure 5.3.4c. It is unsurprising that the matching weights would increase with V , since the maximum distance between any points in the domain (*i.e.* the diagonal of the rectangle) also increases with V . However, the matching weights clearly increase with V at different rates, since $\bar{\mathcal{P}}$ is not constant for varying V . Figure 5.3.4d displays the variation of $\bar{\mathcal{P}}$ for different values of N , and shows that $\bar{\mathcal{P}}$ peaks at $V \approx 7$ for each N .

To give this result some context, it follows that the UE matching is more inefficient at matching riders and drivers in Manhattan, New York City, which has an aspect ratio of about 7, than in Brooklyn, which has more of a square shape. However, it should be noted from figure 5.3.4a that the standard deviation of $\bar{\mathcal{P}}$ is quite large relative to the range of $\bar{\mathcal{P}}$ over $1 \leq V \leq 100$.

5.3.4 Variation with the ratio of riders to drivers

Suppose that the cardinalities of the two sets of agents A and B are not necessarily equal, and define $R := |A|/|B|$. Figure 5.3.5a displays the variation of $\bar{\mathcal{P}}$ with R , for fixed $|B| = 10$ (varying $|A|$) and $S_1 = S_2 = 10$, and shows that $\bar{\mathcal{P}}$ is maximised at $R = 1$, *i.e.* when the two sets contain an equal number of agents. Figures 5.3.5b and 5.3.5c show that $\overline{w^{\text{ue}}}$ and $\overline{w^{\text{so}}}$, respectively, are also maximised at $R = 1$. Figure 5.3.5d displays a heat map of $\bar{\mathcal{P}}$ for varying $|A|$ and $|B|$, confirming that $\bar{\mathcal{P}}$ is maximised at $R = 1$.

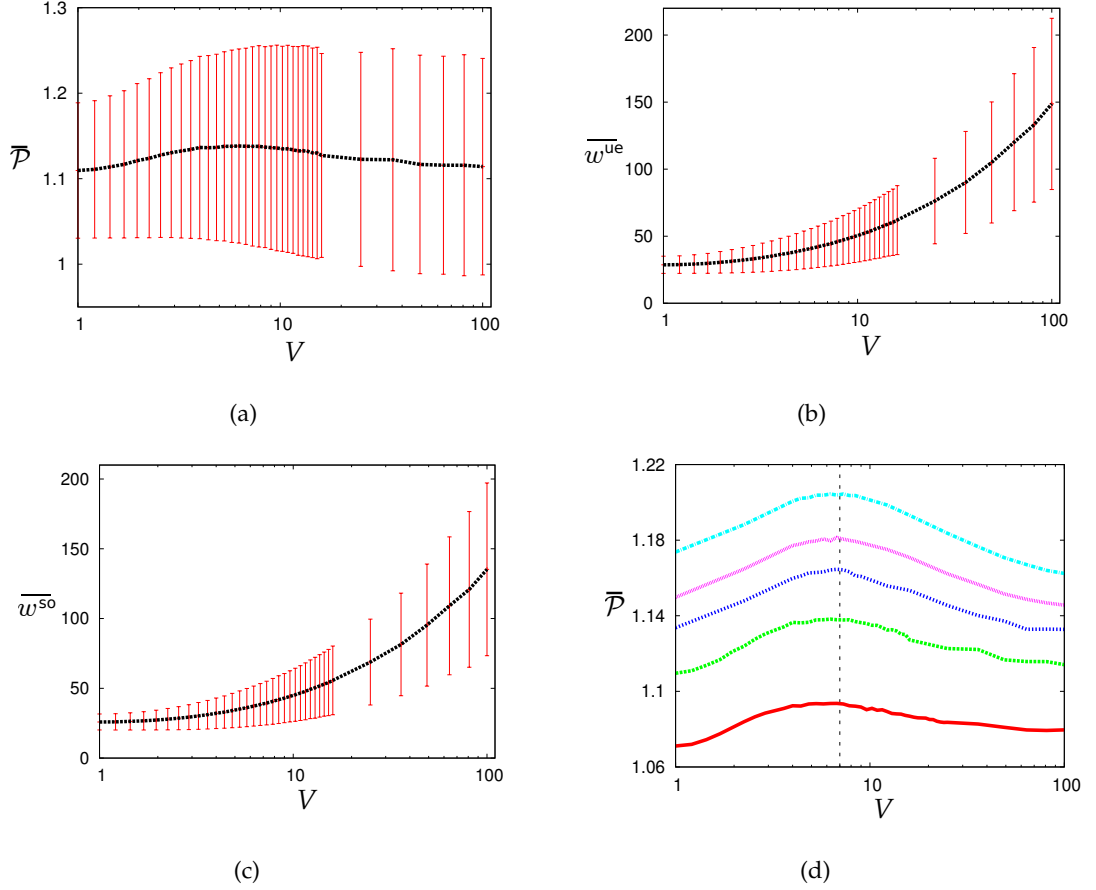


FIG. 5.3.4: The variation of (a) $\bar{\mathcal{P}}$ (b) $\overline{w^{ue}}$ and (c) $\overline{w^{so}}$ with the aspect ratio $V = S_1/S_2$ of the domain, for fixed $N = 10$ and domain area $S_1S_2 = 100$. (d) The variation of $\bar{\mathcal{P}}$ with V for (bottom to top) $N = 5, 10, 15, 20$, and 30 . The vertical dotted line in (d) is at $V = 7$.

These results can be explained as follows. When $|A| \neq |B|$, some of the agents within the larger set of A and B will not be paired, hence the cardinality of the matchings is equal to the cardinality of the smaller set, *i.e.* $|M^{\text{ue}}| = |M^{\text{so}}| = \min(|A|, |B|)$. When $R < 1$, with fixed $|B|$ and varying $|A|$, the number of paired agents is equal to $|A|$, which is increasing with R . The results of figure 5.3.3 show that $\bar{\mathcal{P}}$, $\bar{w}^{\text{ue}}$ and $\bar{w}^{\text{so}}$ increase with the number of paired agents, hence $\bar{\mathcal{P}}$, $\bar{w}^{\text{ue}}$ and $\bar{w}^{\text{so}}$ increase with R for $R < 1$. For $R > 1$, the number of paired agents remains the same (equal to $|B|$, since $|B| < |A|$) but the number of excess agents (*i.e.* agents that will not be paired) in A increases with R . The average distance between paired agents decreases with increasing $|A|$, since the agents of B have a greater “choice” of agents from A , hence $\bar{w}^{\text{ue}}$ and $\bar{w}^{\text{so}}$ decrease with R for $R > 1$.

Since it is less obvious why $\bar{\mathcal{P}}$ also decreases with R for $R > 1$, consider the example shown in figure 5.3.6. This example is similar to the example given in figure 5.3.1a, but contains an extra red agent b_3 with a horizontal distance of $z > 0$ from the other agents and a vertical coordinate identical to a_1 . When $z > y$, the SO matching is the same as before: $M^{\text{so}} = \{(a_1, b_1), (a_2, b_2)\}$ with a weight of $w^{\text{so}} = 2y$. When $z < 2y + x$, however, the UE matching is now given by $M^{\text{ue}} = \{(b_1, a_2), (a_1, b_3)\}$ with a weight of $w^{\text{ue}} = x + z$, and $\mathcal{P} = \frac{x+z}{2y}$. Note that the new agent causes a decrease in w^{ue} but not in w^{so} . This is because when agent a_1 is unable to pair with its initial nearest neighbour (b_1) under the UE matching, there are now two other red agents with which it can pair. For $z < 2y + x$, a_1 chooses to pair with b_3 and the value of w^{ue} is lower than it would be in the absence of b_3 . This example shows how excess agents can lower the weight of the UE matching without lowering the weight of the SO matching, thereby decreasing $\mathcal{P}$.

5.3.5 Summary

This section has considered inefficiency in the pairwise matching of agents located within an empty, planar domain. Explanations have been provided for some — but not all — of the results presented in this section. A simple thought experiment showed that the UE and SO matchings for a given realisation are invariant with the size of the domain, developing a good understanding of the behaviours of $\bar{\mathcal{P}}$, $\bar{w}^{\text{ue}}$ and $\bar{w}^{\text{so}}$ for this case. A phenomenological explanation for the variation with the ratio R of riders to drivers was given. For the case of varying N , it was deduced that a greater number of agents produces higher average matching weights and more inefficient matching

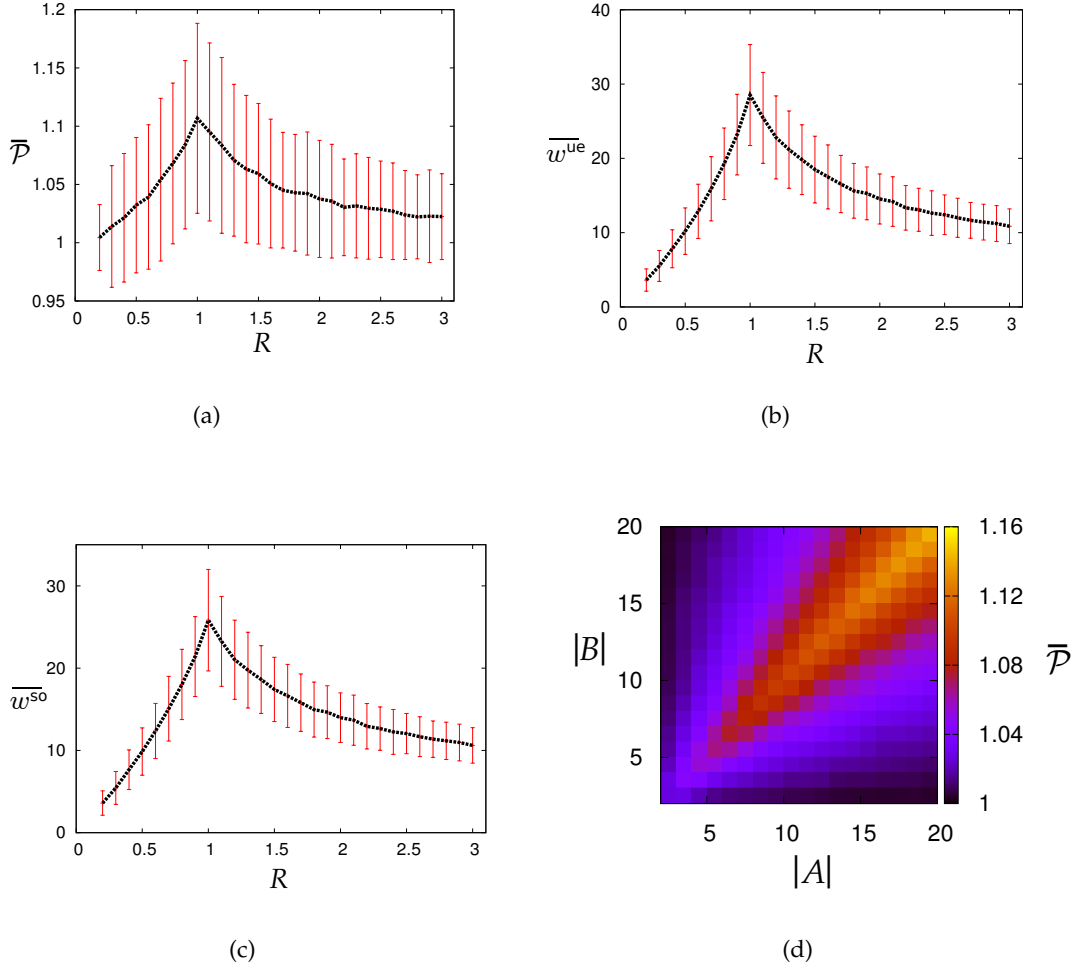


FIG. 5.3.5: The variation of (a) $\bar{\mathcal{P}}$ (b) $\overline{w^{ue}}$ and (c) $\overline{w^{so}}$ with the ratio $R := |A|/|B|$ of the sizes of the two sets of agents, for fixed $|B| = 10$ (varying $|A|$) and $S_1 = S_2 = 10$. (d) 3D heat map of $\bar{\mathcal{P}}$ with $|A|$ and $|B|$, for fixed $S_1 = S_2 = 10$.

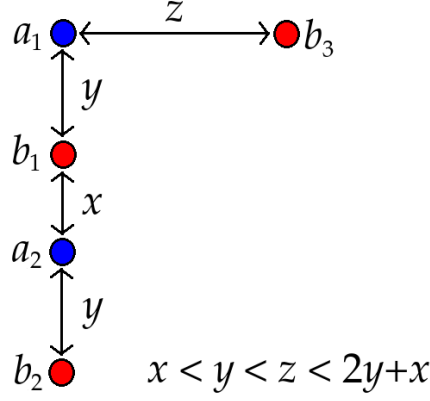


FIG. 5.3.6: The example given in figure 5.3.1a but with an added red agent b_3 . The UE matching is $M^{\text{ue}} = \{(b_1, a_2), (a_1, b_3)\}$ with a weight of $w^{\text{ue}} = x + z$; the SO matching is $M^{\text{so}} = \{(a_1, b_1), (a_2, b_2)\}$ with a weight of $w^{\text{so}} = 2y$; the price of anarchy is $\mathcal{P} = \frac{x+z}{2y}$. This example illustrates that the addition of excess agents can reduce the weight of the UE matching while keeping the SO matching the same, thereby reducing $\mathcal{P}$.

sequences under the UE. It was conjectured that the matching weights increase with the domain aspect ratio V due to the maximum distance between points in the domain increasing with V . Finally, it was shown that $\bar{\mathcal{P}}$ is maximised at $V \approx 7$. This result has an interesting implication to the context of ride-sharing: the inefficiency of matching riders and drivers is higher for rectangular regions (*e.g.* Manhattan) than for square regions (*e.g.* Brooklyn).

5.4 Matchings on networks

In the previous section, agents were situated within an empty, planar domain and the distance between agents was given by the Euclidean metric. An extension to this model will now be considered in which agents are distributed uniformly at random to the nodes of a network with particular topology. The networks considered will be spatial (and weighted), and the distance between agents will be given by the length of the geodesic path between the nodes that they occupy. The addition of a network into the model represents a continuation of the theme of previous chapters — inefficiency on networks — and is pertinent to the context of ride-sharing, since a driver would need to travel along a road network route in order to reach a rider. Network topologies that

will be considered include the diamond lattice structure of Chapters 2-4, as well as random, small-world and scale-free networks. The variation of the price of anarchy will again be considered with respect to the parameters analysed in the previous section, as well as standard parameters relating to the generation of the network structure, such as the number of nodes and the density of edges. Numerical results for the diamond lattice network will again be averaged over 1000 realisations of agent locations. Since the generations of random, small-world and scale-free networks are random, numerical results in these cases will be averaged over multiple realisations of the network structure (500) as well as multiple realisations of agent locations for each network (50).

5.4.1 Diamond lattice network

Since Chapters 2-4 considered a traffic network with a diamond lattice structure, it is fitting to consider the matching of agents on such a network. The results of Section 5.3.1 suggest that the length of each edge of the lattice is arbitrary, and will be taken to be 1. The number of nodes down the vertical side of the lattice is again denoted by I , and the path length spanning the width of the lattice is again denoted by J .

Figure 5.4.1a displays the variation of $\bar{\mathcal{P}}$ with J , for fixed $N = 5$ and aspect ratio $2I/J = 1$ (*i.e.* a square lattice), showing that $\bar{\mathcal{P}}$ is invariant with lattice size. Figure 5.4.1b shows that $\bar{\mathcal{P}}$ increases logarithmically with the number of pairs of agents N . Figure 5.4.1c shows that the variation of $\bar{\mathcal{P}}$ with lattice aspect ratio V is maximised at $V \approx 7$. Figure 5.4.1d shows that the variation of $\bar{\mathcal{P}}$ with the ratio R between the sizes of the two sets of agents is maximised at $R = 1$.

Figure 5.4.1 therefore shows that the variation of $\bar{\mathcal{P}}$ displays very similar behaviour to the empty domain case considered in section 5.3. The results for the variations of $\bar{w}^{\text{ue}}$ and $\bar{w}^{\text{so}}$ also show similar results to the previous section, and hence have been omitted. Indeed, the distance metric on a regular lattice does not significantly differ from that on a plane, and tends to the Euclidean metric as the lattice edge length tends to 0. This raises the question of how corresponding results behave when the network structure is random rather than regular, which will now be considered.

5.4.2 Random graphs

Recall from section 1.4.1 that Erdős–Rényi random graphs are generated such that each pair of nodes is adjacent with some probability p_e , for a fixed number of nodes n . In the

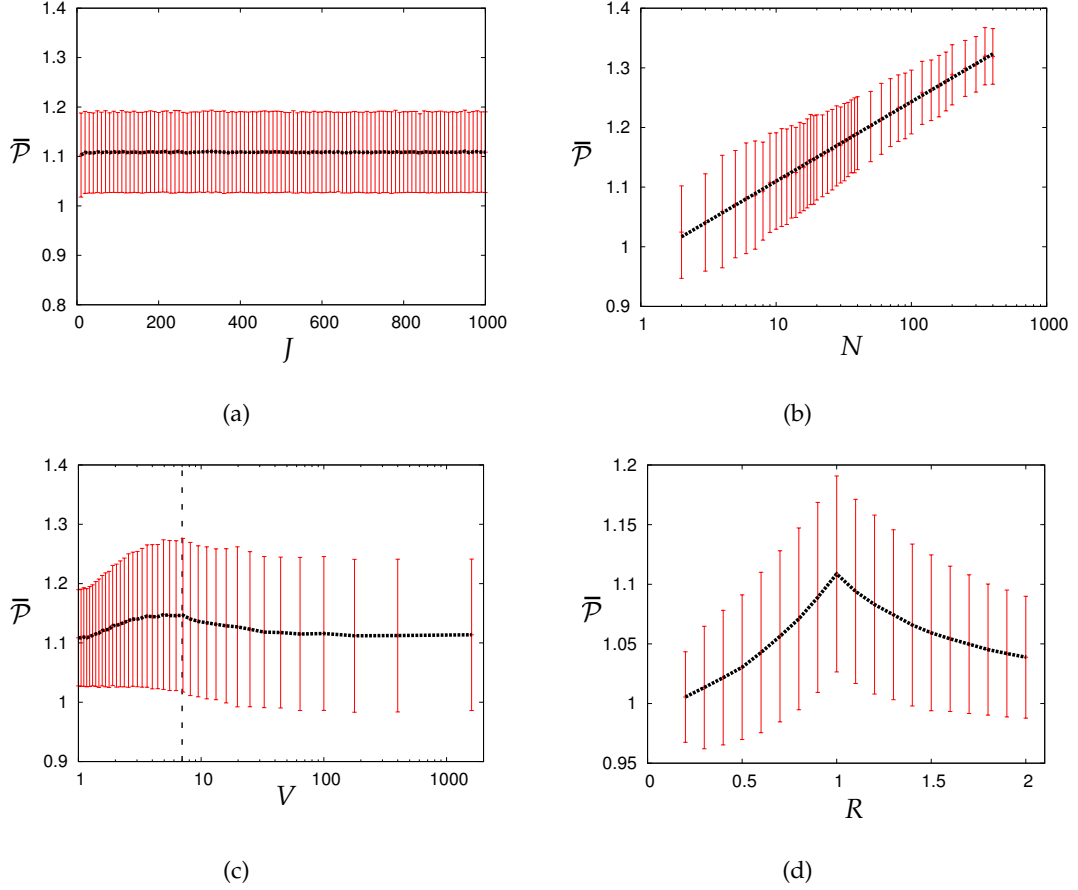


FIG. 5.4.1: Numerical results for matching agents located on the nodes of a diamond lattice network. (a) The variation of $\bar{P}$ with lattice size, given by the path-length J (with $2I = J$), for fixed $N = 5$. (b) The variation of $\bar{P}$ with the number of pairs of agents N , for fixed $I = 500$ and $J = 1000$. (c) The variation of $\bar{P}$ with lattice aspect ratio V , for fixed $N = 5$ and $2IJ = 1,000,000$. The vertical dotted line indicates that $\bar{P}$ is maximised at $V \approx 7$. (d) The variation of $\bar{P}$ with the ratio R of the sizes of the sets of agents, for fixed $|A| = 10$ (varying $|B|$), $I = 50$ and $J = 100$.

context of matchings, we are interested in spatial networks where the nodes are embedded in Euclidean space and the edges denote real physical connections (*e.g.* roads). For this reason, the nodes are distributed uniformly at random within a square, planar domain, and the lengths of the edges are given by the Euclidean metric. The side length of the domain is fixed at 10 without loss of generality, since (as it has already been shown) the price of anarchy is invariant with the relative distance between agents.

Figure 5.4.2a displays the variation of $\bar{\mathcal{P}}$ (red), $\bar{w}^{\text{ue}}$ (green) and $\bar{w}^{\text{so}}$ (blue) with the probability p_e that any given pair of nodes is adjacent, for a fixed number of nodes $n = 100$ and pairs of agents $N = 10$. The variation of $\bar{\mathcal{P}}$ displays several prominent features:

1. a local maximum at $p_e \approx \log(n)/n$.
2. a local minimum at $p_e \approx 0.11$.
3. an increase with p_e to a plateau as $p_e \rightarrow 1$.

Figure 5.4.2b shows that these features hold for $N = 5$ (brown), 10 (red) and 15 (purple), suggesting that they are independent of N . This will be shown to be a recurring result throughout this section. The matching weights also exhibit a local maximum at around the critical threshold for the network to be connected [16]: $\log(n)/n \approx 0.046$. At values of p_e lower than this threshold, the network is likely to be unconnected. In this case, there may be instances where there does not exist a path between one agent and any available agent from the opposite set, so that some agents will be left unpaired and the matching will consist of fewer than N pairings. The matching weights and the price of anarchy increase with the number of pairings, which explains the sharp increase at small p_e . When p_e reaches $\log(n)/n$, network realisations are likely to be connected and all agents are paired. The matching weights then decrease with increasing p_e since the average geodesic length between nodes decreases.

Figure 5.4.3a displays the variation of $\bar{\mathcal{P}}$ (red), $\bar{w}^{\text{ue}}$ (green) and $\bar{w}^{\text{so}}$ (blue) with the number of pairs of agents N , for fixed $p_e = 0.05$ and $n = 100$. Figure 5.4.3b displays the variation of $\bar{\mathcal{P}}$ with N for $p_e = 0.05$ (red), 0.11 (turquoise) and 0.6 (orange). These values of p_e were chosen because they correspond to each of the interesting features listed above for the variation of $\bar{\mathcal{P}}$ with p_e . Unlike the previous models considered (the empty domain and the diamond lattice network), $\bar{\mathcal{P}}$ is not well fitted by a logarithmic function alone but rather $f(N) = a + b \log(N) + c\sqrt{N}$, for some constants a, b and c . This will be a recurring result throughout this section, and the particular values of a, b

and c for all cases considered are given in Table 5.1.

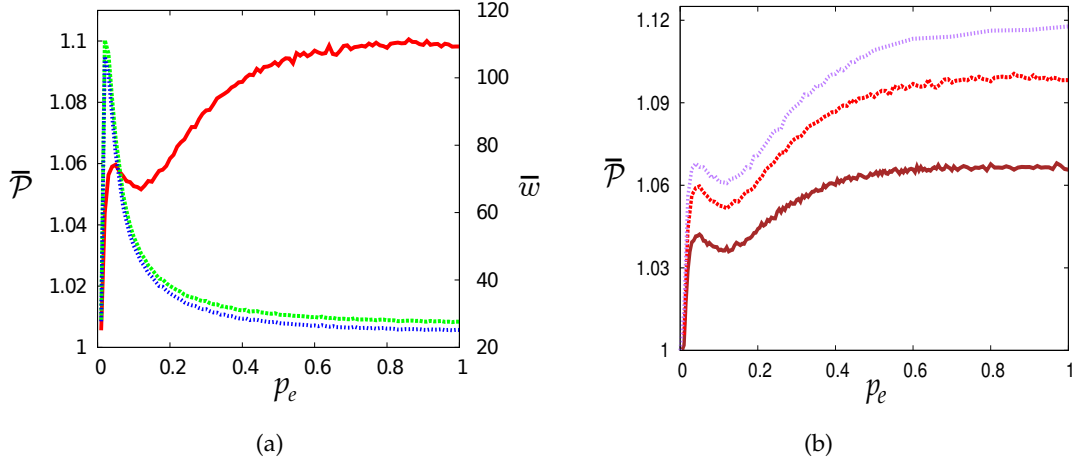


FIG. 5.4.2: Numerical results for matching agents that are located on the nodes of Erdős-Rényi random networks with size $n = 100$. (a) The variation of $\bar{P}$ (red), $\bar{w}^{\text{ue}}$ (green) and $\bar{w}^{\text{so}}$ (blue) with the probability p_e that any given pair of nodes is adjacent, for fixed $N = 10$. (b) The variation of $\bar{P}$ with p_e for $N = 5$ (brown), 10 (red) and 15 (purple).

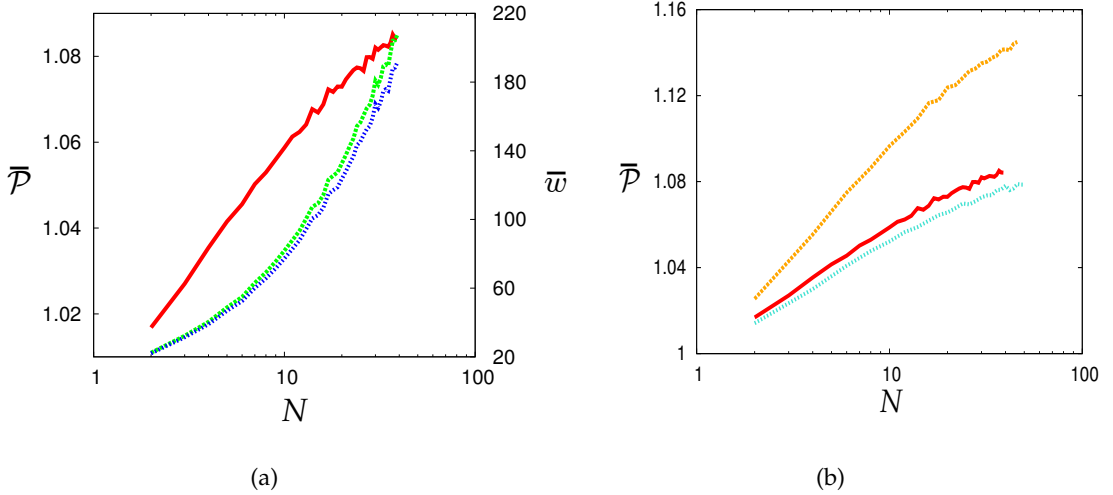


FIG. 5.4.3: Numerical results for matching agents that are located on the nodes of Erdős-Rényi random networks with size $n = 100$. (a) The variation of $\bar{P}$ (red), $\bar{w}^{\text{ue}}$ (green) and $\bar{w}^{\text{so}}$ (blue) with the number of pairs of agents N , for fixed $p_e = 0.05$. (b) The variation of $\bar{P}$ with N for $p_e = 0.05$ (red), 0.11 (turquoise) and 0.6 (orange). Inspection finds that the function $f(N) = a + b \log(N) + c\sqrt{N}$ provides a good fit for $\bar{P}$.

Figure 5.4.4a displays the variation of $\bar{P}$ (red), $\bar{w}^{\text{ue}}$ (green) and $\bar{w}^{\text{so}}$ (blue) with the number of nodes n of each network realisation, for fixed $N = 10$. For each value of n , the connection parameter is set equal to the threshold for the network to be connected: $p_e = \log(n)/n$. It can be seen that the matching weights monotonically increase with n , while the price of anarchy first increases at small n before steadily decreasing for large n . Recall that the threshold $p_e = \log(n)/n$ applies in the limit $n \rightarrow \infty$, so for small n the network is less likely to be connected. As mentioned above, if the network is not connected then some agents may not be matched, which results in lower matching weights and a lower price of anarchy. This would explain the sharp rise in the matching weights and the price of anarchy at small n . When the network is large enough so that all agents are matched (at $n \approx 50$), the price of anarchy monotonically decreases with the network size. Figure 5.4.4b shows that this behaviour holds for $N = 5$ (brown), 10 (red) and 15 (purple).

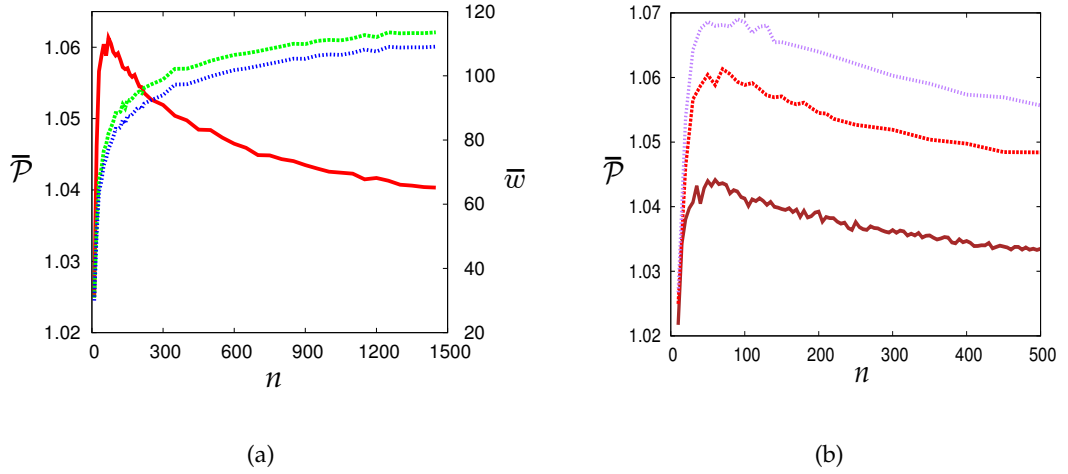


FIG. 5.4.4: Numerical results for matching agents that are located on the nodes of Erdős–Rényi random networks. (a) The variation of $\bar{P}$ (red), $\bar{w}^{\text{ue}}$ (green) and $\bar{w}^{\text{so}}$ (blue) with the number of nodes n of each network realisation, for fixed $N = 10$. For each value of n , the connection parameter is set equal to the threshold for the network to be connected: $p_e = \log(n)/n$. (b) The variation of $\bar{P}$ with n for $N = 5$ (brown), 10 (red) and 15 (purple).

5.4.3 Small-world networks

Watts–Strogatz “small-world” networks are generated by rewiring each edge of a regular ring lattice with probability p_w (see Section 1.4.2 for a description of the model). The original edges of the ring could correspond to short roads in suburban areas, while the rewired “shortcut” edges could correspond to main roads such as motorways. Since we are interested in spatial networks, the nodes are placed equidistant on a circle with its center at the origin and with radius 10 (WLOG). Before rewiring, each node is adjacent to its d nearest neighbours ($d/2$ on each side in the ring), where d parameterises the density of edges in the network.

Figure 5.4.5a displays the variation of $\bar{P}$ (red), $\bar{w}^{ue}$ (green) and $\bar{w}^{so}$ (blue) with the rewiring probability p_w , for fixed $N = 10$, $n = 100$ and $d = 4$. After a small decrease at small p_w , the matching weights increase significantly with p_w , while $\bar{P}$ again displays several prominent features:

1. a significant drop for $0 < p_w < 0.45$.
2. a minimum at $p_w \approx 0.45$.
3. a local maximum at $p_w \approx 0.8$.

Figure 5.4.5b displays the variation of $\bar{P}$ with p_w for $d = 4$ (brown), 6 (orange) and 8 (turquoise), with fixed $N = 5$. It can be observed that the behaviour of $\bar{P}$ varies significantly for different values of d . For $d = 6$, the minimum and local maximum are less pronounced than for $d = 4$, and are located at larger p_w . For $d = 8$, there is a plateau at $0.2 \leq p_w \leq 0.5$, the minimum at ≈ 0.8 is fairly shallow and the local maximum has vanished.

A better understanding of these results can be derived from figure 5.4.6, which displays semi-log plots of $\bar{P}$, $\bar{w}^{ue}$ and $\bar{w}^{so}$ as well as the global clustering coefficient $Cl(p_w)$. Recall from Chapter 1 that the global clustering coefficient measures the extent to which two nodes with a common neighbour are also adjacent to each other. Moreover, it was shown in [24] that the average path length between nodes drops at a lower value of p_w than does $Cl(p_w)$ (see figure 1.4.2). It might be expected that the matching weights would decrease with the average path length. However, figure 5.4.6a shows that the matching weights decrease very little at small p_w , and rapidly increase for small values of $Cl(p_w)$. This suggests that the matchings very rarely pair two agents that are far

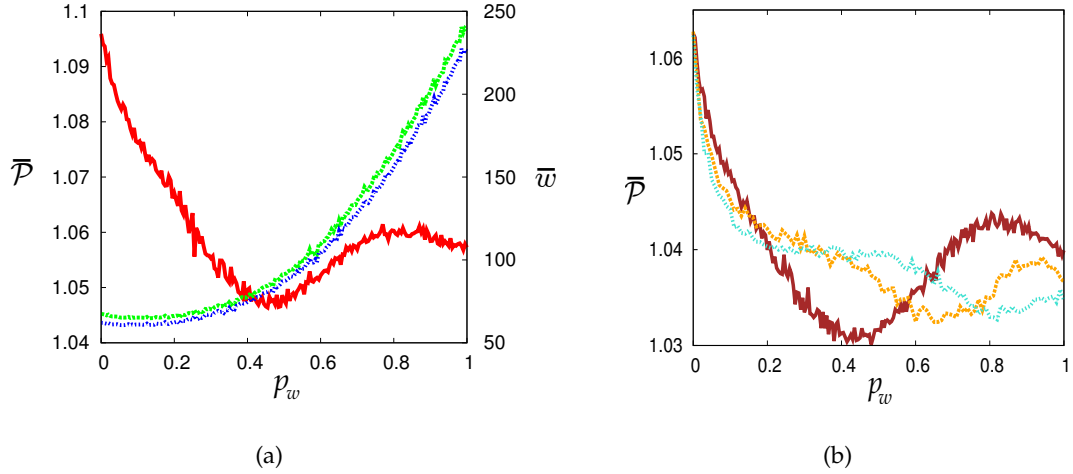


FIG. 5.4.5: Numerical results for matching agents located on the nodes of Watts–Strogatz small-world networks with size $n = 100$. (a) The variation of $\bar{P}$ (red), $\bar{w}^{ue}$ (green) and $\bar{w}^{so}$ (blue) with the rewiring probability p_w , for fixed $d = 4$ and $N = 10$. (b) The variation of $\bar{P}$ with p_w for $d = 4$ (brown), 6 (orange) and 8 (turquoise), with fixed $N = 5$.

apart from each other on the ring, otherwise shortcuts across the ring would significantly decrease the matching weights. If matchings only pair agents that are geometrically close, rewiring edges will actually *increase* the distance between paired agents, which explains the rapid increase in the matching weights at $p_w > 0.2$. Figure 5.4.6b displays the variation of $\bar{P}$ for several values of N , as well as $C(p_w)$. It can be seen that the initial decrease in $\bar{P}$ coincides with the drop in $C(p_w)$. Figures 5.4.6c and 5.4.6d show similar results for $d = 6$ and $d = 8$, respectively. Intuitively, it makes sense that inefficient matching sequences are more likely to occur when there is high clustering between nodes, and it has already been noted that matchings are likely to comprise pairs of agents that are close to each other in the network. However, this does not explain the subsequent behaviour of $\bar{P}$ at larger p_w .

Figure 5.4.7a displays the variation of $\bar{P}$ (red), $\bar{w}^{ue}$ (green) and $\bar{w}^{so}$ (blue) with the number of pairs of agents N , for fixed $d = 4$ and $p_w = 0.05$. The value of p_w here was chosen because the results of [24] indicate that this value generates networks with a small average path length and high global clustering coefficient, which are common features of real world networks [16]. Figure 5.4.7b displays the variation of $\bar{P}$ with N for $p_w = 0.05$ (red), 0.45 (brown) and 0.8 (purple), for fixed $d = 4$. The first value of p_w here was chosen for the same reason as in figure 5.4.7a, while the latter two values were

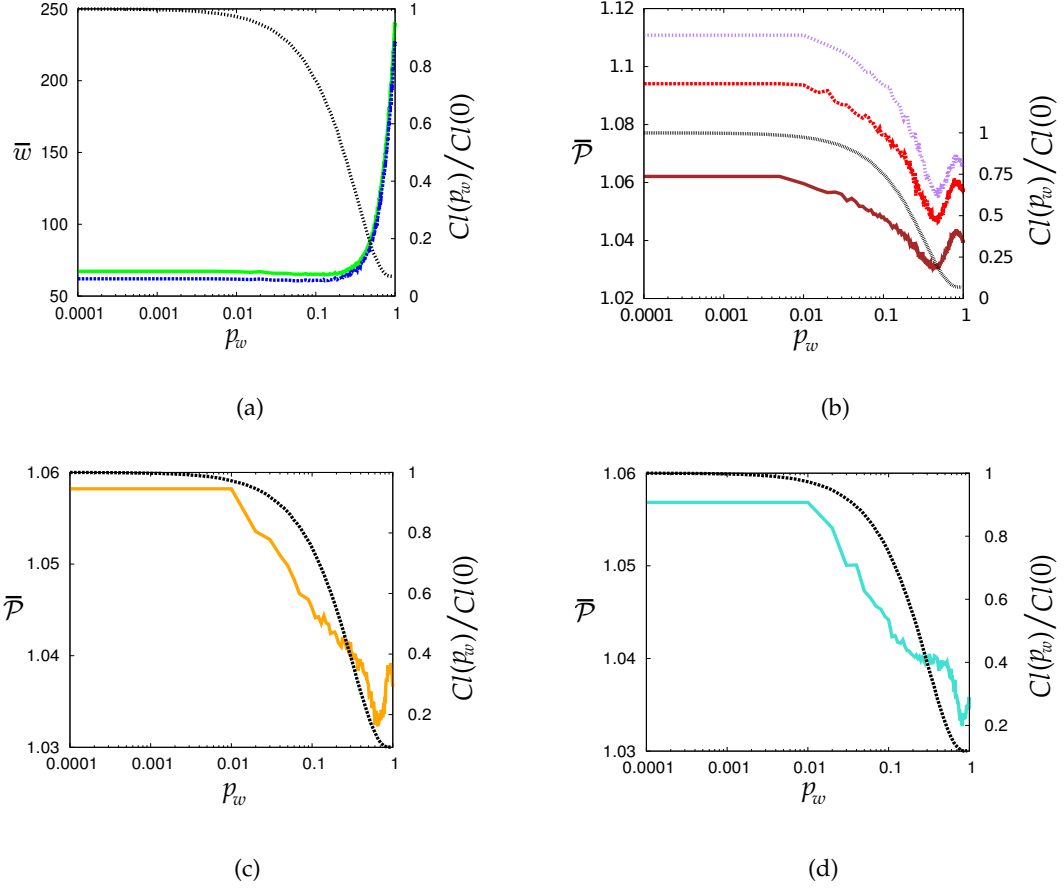


FIG. 5.4.6: Semi-log plots for matching agents located on the nodes of Watts–Strogatz small-world networks with size $n = 100$. (a) The variation of $\bar{w}^{\text{ue}}$ (green), $\bar{w}^{\text{so}}$ (blue) and the (normalised) global clustering coefficient $Cl(p_w)/Cl(0)$ (black) with p_w , for fixed $d = 4$ and $N = 10$. (b) The variation of $\bar{P}$ for $N = 5$ (brown), 10 (red), 15 (purple) and $Cl(p_w)/Cl(0)$ (black) with p_w , for fixed $d = 4$. (c) The variation of $\bar{P}$ (orange) and $Cl(p_w)/Cl(0)$ (black) with p_w , for fixed $d = 6$. (d) The variation of $\bar{P}$ (turquoise) and $Cl(p_w)/Cl(0)$ (black) with p_w , for fixed $d = 8$.

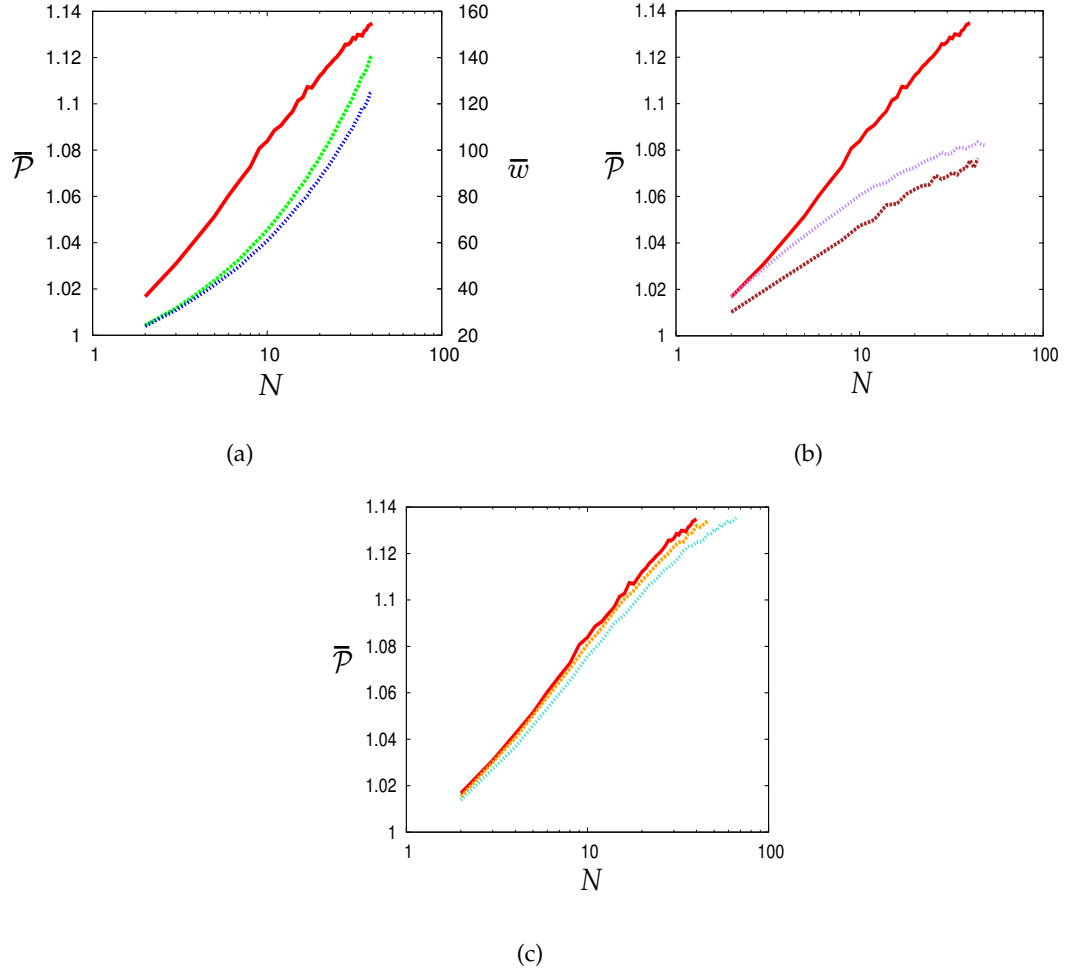


FIG. 5.4.7: Numerical results for matching agents located on the nodes of Watts–Strogatz small-world networks with size $n = 100$. (a) The variation of $\bar{P}$ (red), $\bar{w}^{\text{ue}}$ (green) and $\bar{w}^{\text{so}}$ (blue) with the number of pairs of agents N , for fixed $d = 4$ and $p_w = 0.05$. (b) The variation of $\bar{P}$ with N for $p_w = 0.05$ (red), 0.45 (brown) and 0.8 (purple), for fixed $d = 4$. (c) The variation of $\bar{P}$ with N for $d = 4$ (red), 6 (orange) and 10 (turquoise), for fixed $p_w = 0.05$. Inspection finds that the function $f(N) = a + b \log N + c\sqrt{N}$, for some constants a, b and c , provides a good fit for $\bar{P}$.

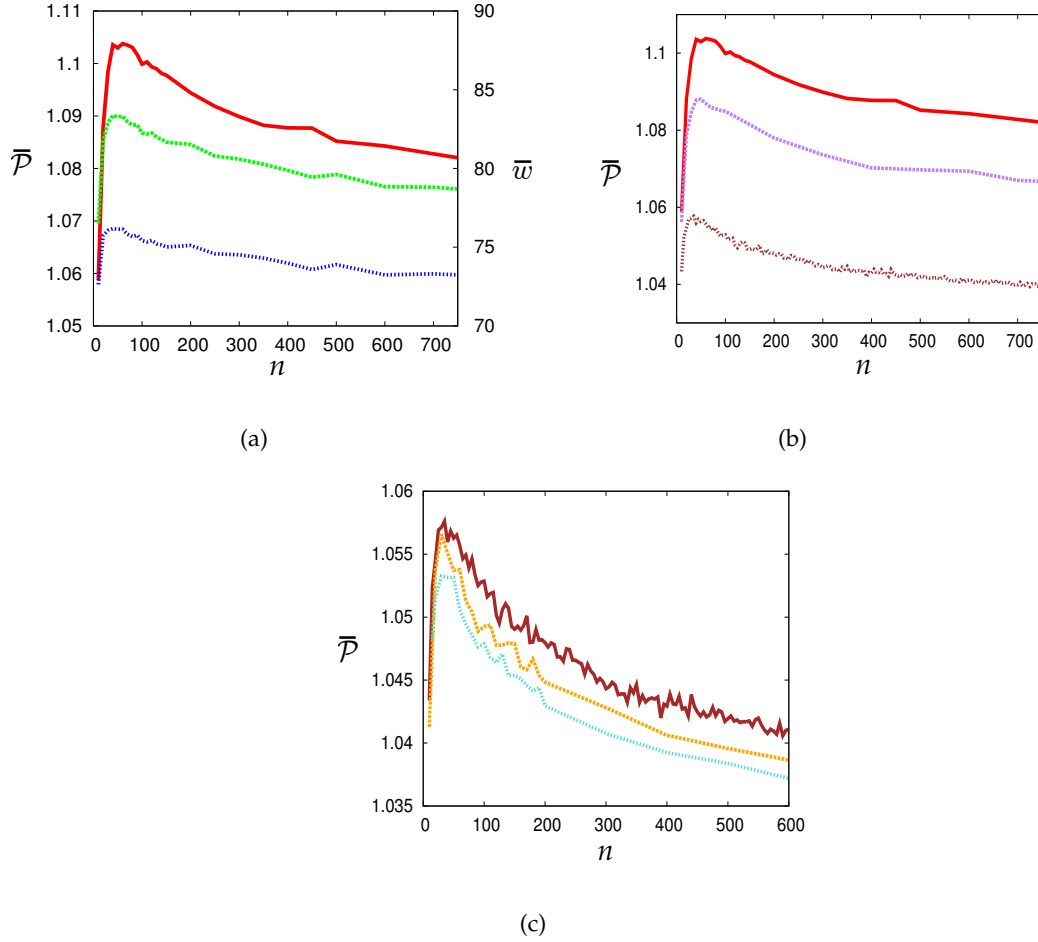


FIG. 5.4.8: Numerical results for matching agents located on the nodes of Watts–Strogatz small-world networks. (a) The variation of $\bar{P}$ (red), $\bar{w}^{ue}$ (green) and $\bar{w}^{so}$ (blue) with the number of nodes n of each network realisation, for fixed $N = 15$, $d = 4$ and $p_w = 0.05$. (b) The variation of $\bar{P}$ with n for $N = 5$ (brown), 10 (purple) and 15 (red), for fixed $d = 4$ and $p_w = 0.05$. (c) The variation of $\bar{P}$ with n for $d = 4$ (brown), 6 (orange) and 10 (turquoise), for fixed $N = 5$ and $p_w = 0.05$.

chosen because figure 5.4.5 indicates that they correspond to the local minimum and local maximum of the variation of $\bar{\mathcal{P}}$ with p_w , respectively. Figure 5.4.7c displays the variation of $\bar{\mathcal{P}}$ with N for $d = 4$ (red), 6 (orange) and 10 (turquoise), for fixed $p_w = 0.05$. Figures 5.4.7b and 5.4.7c show that there is disparity between the variations of $\bar{\mathcal{P}}$ with N for different values of p_w , but only a minor difference for different values of d . Similarly to the matching on Erdős–Rényi graphs, the variation of $\bar{\mathcal{P}}$ with N does not satisfy a log function, but rather the function $f(N) = a + b \log N + c\sqrt{N}$, for some constants a, b and c (see Table 5.1), provides a good fit.

Figure 5.4.8a displays the variation of $\bar{\mathcal{P}}$ (red), $\bar{w}^{ue}$ (green) and $\bar{w}^{so}$ (blue) with the number of nodes n of each network realisation, for fixed $N = 15$, $d = 4$ and $p_w = 0.05$. All three functions display similar behaviour: an initial increase with n to a maximum at $n \approx 30$, followed by a gradual decline thereafter. The initial increase could be attributed to the same cause as for the Erdős–Rényi random networks case: that the network is less likely to be connected at lower values of n . Figures 5.4.8b and 5.4.8c show that the structure of $\bar{\mathcal{P}}$ is the same for different values of N and d , respectively.

5.4.4 Scale-free networks

The generation of Barabási–Albert scale-free networks starts with a complete network of m nodes that are all adjacent to each other. New nodes are then added to the network one by one and each form links with m existing nodes via preferential attachment (see section 1.4.3 for more details). Similarly to the case of random networks, the nodes are distributed uniformly at random within a square, planar domain with side length 10 (WLOG), and the lengths of the edges are given by the Euclidean metric. The generation procedure terminates when the network comprises n nodes. The networks considered for this case will be larger than the random networks and small-world networks considered previously. This is so that network realisations are more likely to exhibit the fundamental structural properties of scale-free networks, such as hubs and a power law degree distribution.

Figure 5.4.9a displays the variation of $\bar{\mathcal{P}}$ (red), $\bar{w}^{ue}$ (green) and $\bar{w}^{so}$ (blue) with the number of nodes m with which each new node connects upon joining the network, for fixed $N = 10$ and $n = 200$. As expected, the matching weights decrease with m since the average geodesic length between nodes decreases. It can be seen that $\bar{\mathcal{P}}$ increases with m , with a particularly large change between $m = 1$ and $m = 2$. Figure 5.4.9b shows that this behaviour holds for $N = 10$ (red), 20 (purple) and 30 (brown).

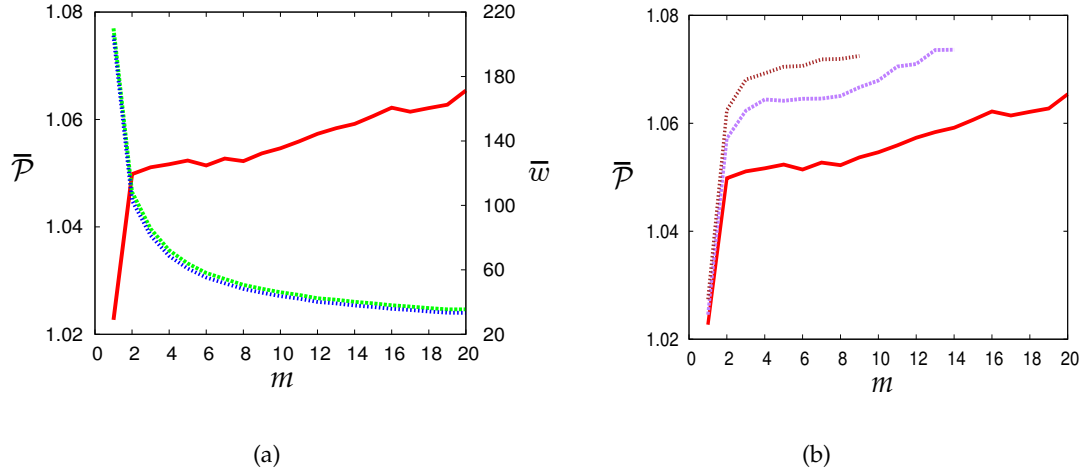


FIG. 5.4.9: Numerical results for matching agents that are located on the nodes of Barabási–Albert scale-free networks with $n = 200$. (a) The variation of $\bar{P}$ (red), $\bar{w}^{\text{ue}}$ (green) and $\bar{w}^{\text{so}}$ (blue) with the number of nodes m each new node connects with upon joining the network, for fixed $N = 10$. (b) The variation of $\bar{P}$ with m for $N = 10$ (red), 20 (purple) and 30 (brown).

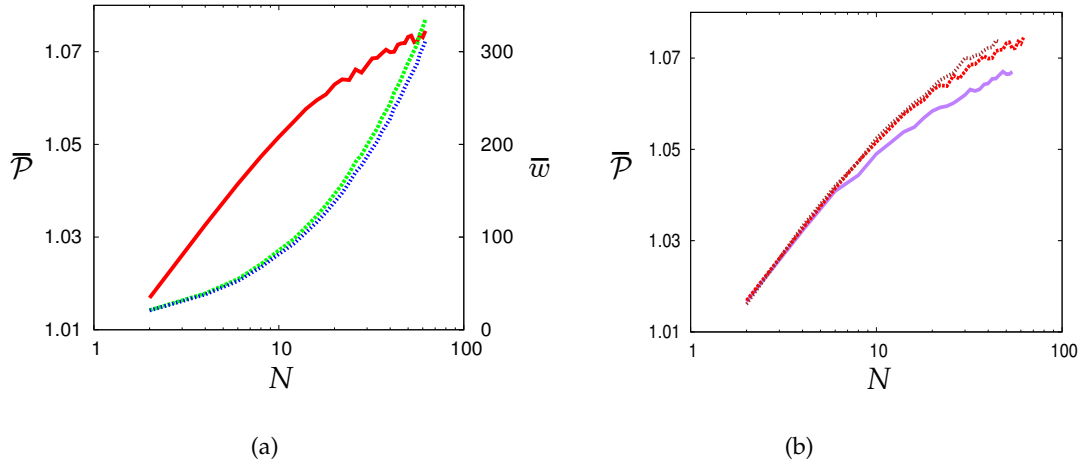


FIG. 5.4.10: Numerical results for matching agents that are located on the nodes of Barabási–Albert scale-free networks with $n = 200$. (a) The variation of $\bar{P}$ (red), $\bar{w}^{\text{ue}}$ (green) and $\bar{w}^{\text{so}}$ (blue) with the number of pairs of agents N , for fixed $m = 3$. (b) The variation of $\bar{P}$ with N for $m = 2$ (purple), 3 (red) and 4 (brown). Inspection finds that the function $f(N) = a + b \log N + c\sqrt{N}$, for some constants a, b and c , provides a good fit for $\bar{P}$.

Figure 5.4.10a displays the variation of $\bar{\mathcal{P}}$ (red), $\bar{w}^{\text{ue}}$ (green) and $\bar{w}^{\text{so}}$ (blue) with the number of pairs of agents N , for fixed $m = 3$ and $n = 200$. Figure 5.4.10b displays the variation of $\bar{\mathcal{P}}$ with N for $m = 2$ (purple), 3 (red) and 4 (brown). Similarly to the previous cases, inspection finds that the function $f(N) = a + b \log N + c\sqrt{N}$, for constants a, b and c (see Table 5.1), provides a good fit for $\bar{\mathcal{P}}$.

Figure 5.4.11a displays the variation of $\bar{\mathcal{P}}$ (red), $\bar{w}^{\text{ue}}$ (green) and $\bar{w}^{\text{so}}$ (blue) with the number of nodes n in the network, for fixed $N = 10$ and $m = 3$. It can be seen that $\bar{\mathcal{P}}$ monotonically decreases for all n , while the matching weights monotonically increase with n . Figure 5.4.11b shows that the variation of $\bar{\mathcal{P}}$ with n displays similar behaviour for $N = 5$ (purple), 10 (red) and 15 (brown), with fixed $m = 3$.

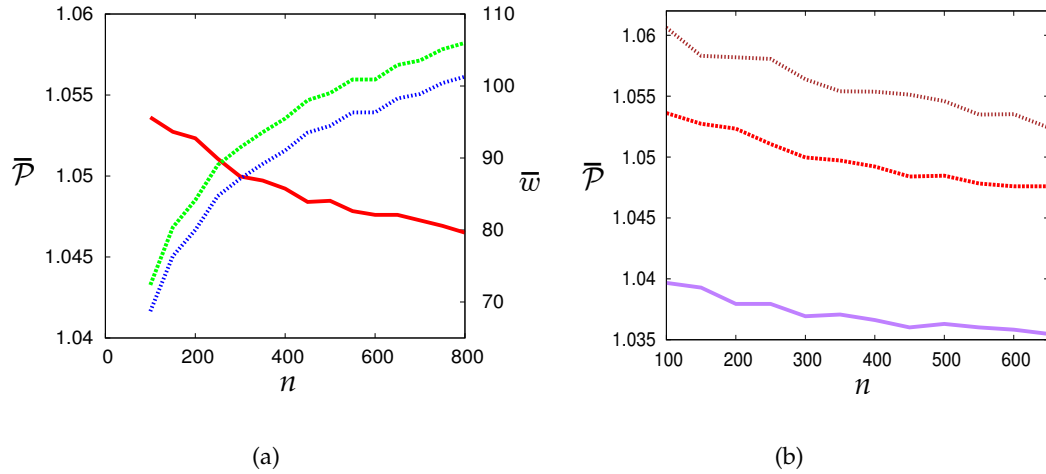


FIG. 5.4.11: Numerical results for matching agents that are located on the nodes of Barabási–Albert scale-free networks. (a) The variation of $\bar{\mathcal{P}}$ (red), $\bar{w}^{\text{ue}}$ (green) and $\bar{w}^{\text{so}}$ (blue) with the number of nodes n in the network, for fixed $N = 10$ and $m = 3$. (b) The variation of $\bar{\mathcal{P}}$ with n for $N = 5$ (purple), 10 (red) and 15 (brown), for fixed $m = 3$.

5.4.5 Summary

This section has considered the matchings of agents on networks with particular topologies, including a diamond lattice structure, Erdős–Rényi (ER) random networks, Watts–Strogatz (WS) small-world networks, and Barabási–Albert (BA) scale-free networks.

			a	b	c
ER		$p_e = 0.05$	1.002	0.033	-0.006
		$p_e = 0.11$	1.001	0.030	-0.006
		$p_e = 0.6$	1.000	0.057	-0.011
WS	$d = 4$	$p_w = 0.05$	0.986	0.050	-0.006
		$p_w = 0.45$	0.995	0.028	-0.004
		$p_w = 0.8$	1.005	0.036	-0.009
	$p_w = 0.05$	$d = 6$	0.987	0.049	-0.006
		$d = 10$	0.984	0.049	-0.007
BA		$m = 2$	1.009	0.025	-0.006
		$m = 3$	1.007	0.027	-0.006
		$m = 4$	1.005	0.029	-0.006

Table 5.1: The function $f(N) = a + b \log N + c\sqrt{N}$ provided a good fit for the variation of $\bar{\mathcal{P}}$ with the number of pairs of agents N . This holds for Erdős–Rényi (ER) random networks, Watts–Strogatz (WS) small-world networks and Barabási–Albert (BA) scale-free networks. The above displays the specific values of a, b and c for all cases considered in this section.

For the diamond lattice network, the variation of the price of anarchy displayed very similar behaviour to the empty domain case considered in the previous section. By contrast, the price of anarchy displayed novel behaviours for ER, WS and BA networks. For all of these network structures, the variation of $\bar{P}$ with the number of pairs of agents N was well fitted by the function $f(N) = a + b \log N + c\sqrt{N}$. Moreover, it can be seen from Table 5.1 that the values of a , b and c vary little across the different network structures. When the network was large enough to be connected, the price of anarchy decreased with network size. The variation of $\bar{P}$ with the connection parameter p_e for ER networks and with the rewiring parameter p_w for WS networks displayed non-trivial behaviour. For BA networks, the price of anarchy monotonically increased with the density of edges.

The next section considers a system in which agents arrive at different times, and analyses the inefficiency of matching agents at their entry times.

5.5 The price of impatience

5.5.1 Introduction

A notion of temporality will now be included in the model such that agents enter the system at different times. This extension is particularly relevant to the context of ride-sharing, where requests from riders occur at different times and the availability of drivers varies over time. In this case, dispatching the nearest available driver at the moment a request is made may not be the most efficient approach. One concern is whether a sequence of matching assignments — even if each matching is the SO at a fixed point in time — could be made more efficient by delaying dispatches until further requests have been made. If so, then what are the optimum times to implement the matchings? These issues will be addressed in this section.

When agents enter the system at different times, a sequence of SO matchings is considered. Since we are only interested in sequences of SO matchings for the remainder of this section, the reader should take any reference to a “matching” to mean a SO matching. The waiting time between an agent entering the system and being matched is considered to be a cost to the agent. Hence the weight of a sequence of matchings depends not only on the distance between paired agents, but also the times between agents entering the system and their being matched. Consider the example of two pairs of agents with regular spacing x , as shown in Figure 5.5.1a. Agents $a_1 \in A$

and $b_1 \in B$ enter at time $t = 0$, and agents $a_2 \in A$ and $b_2 \in B$ enter at time $t = \hat{t}$. When agents are matched at their entry times, the resultant sequence of matchings $M_0 = (\{(a_1, b_1)\}, \{(a_2, b_2)\})$ has weight $4x$. When both pairs of agents are matched at time $\hat{t}$, the matching $M_{\hat{t}} = \{(a_1, b_2), (a_2, b_1)\}$ has weight $2x$. Suppose the weight of the “wait time” of agents a_1 and b_1 under $M_{\hat{t}}$ is directly proportional to time, then $w(M_{\hat{t}}) = 2x + k\hat{t}$ for some dimensional constant $k > 0$. Clearly, if $\hat{t} > 2x/k$ the sequence of matchings M_0 is better, and if $\hat{t} < 2x/k$ the single matching $M_{\hat{t}}$ at time $\hat{t}$ is better.

5.5.2 Methodology

The problem of finding the sequence of matchings with minimum weight can be non-trivial. Consider N pairs of agents (each pair a rider and a driver) whose entry times are distributed according to a Poisson process, and where each agent’s spatial location is distributed uniformly at random in a square, planar domain. The assumption that each rider request occurs simultaneously with a driver becoming available is clearly a simplification of ride-sharing in the real world. However, the aim of Uber’s “surge pricing” policy [49] — in which driver service fares increase/decrease with increasing/decreasing rider demand — is to achieve parity between rider demand and driver supply, so our assumption that the system always contains the same number of riders and drivers is quite reasonable. We also assume that each rider request is stochastically independent to all other rider requests, hence the Poisson process (widely used in queuing theory) is appropriate for the distribution of entry times.

Denote the two sets of N agents by $A := \{a_1, a_2, \dots, a_N\}$ and $B := \{b_1, b_2, \dots, b_N\}$ where each pair (a_i, b_i) has entry time $t_i < t_{i+1} \forall i = 1, 2, \dots, N-1$. Suppose only the wait times of the agents in A (i.e. the riders) carry weight, so that the weight of a matching M implemented at time t is given by

$$w(M, t) = \sum_{m \in M} (d(m) + k(t - t_{i_m})), \quad (5.5.1)$$

where t_{i_m} is the entry time of $a_{i_m} \in A \cap m$. Since the average distance between points on a plane decreases with the density of points, it follows that the weight of distances between matched agents (given by the first term in the sum of (5.5.1)) decreases with the number of a - b pairs that are matched in one go. On the other hand, the weight of the waiting times (the second term in the sum of (5.5.1)) increases with the number of a - b pairs before a matching is implemented. The optimum sequence of matchings is

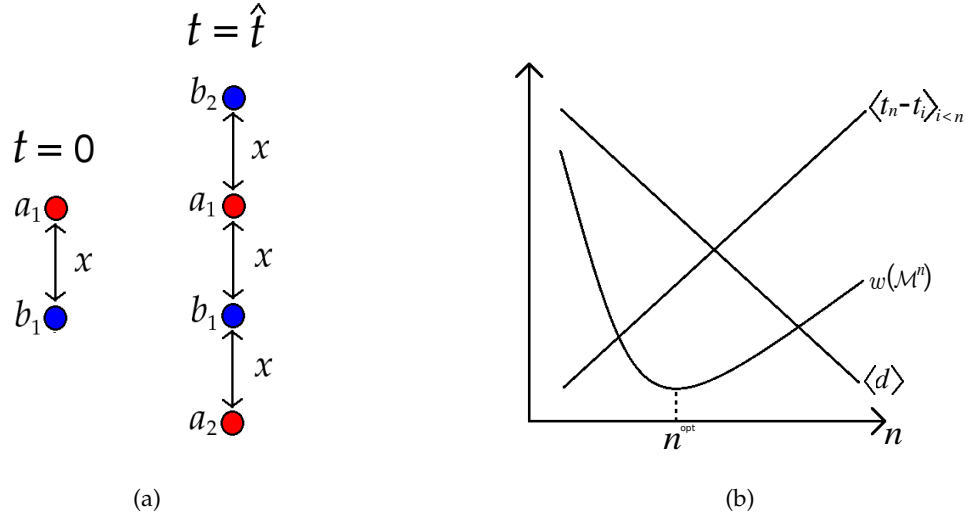


FIG. 5.5.1: (a) A simple example demonstrating the significance of time in implementing matchings. When pairs are matched at their entry times, the resultant sequence of matchings $(\{(a_1, b_1)\}, \{(a_2, b_2)\})$ has weight $4x$. When both pairs are matched at time $\hat{t}$, the optimum matching $\{(a_1, b_2), (a_2, b_1)\}$ has weight $2x + k\hat{t}$, for some $k > 0$. (b) A schematic diagram of the average distance between matched agents $\langle d \rangle$, the average wait time for agents $\langle t_n - t_i \rangle_{i < n}$, and the weight of the sequence of matchings $w(\mathcal{M}^n)$, as functions of n .

determined by the balance between the above two factors (see Figure 5.5.1b).

For any i, j such that $1 \leq i \leq j \leq N$, define $A_i^j := \{a_i, a_{i+1}, \dots, a_j\} \subseteq A$, $B_i^j := \{b_i, b_{i+1}, \dots, b_j\} \subseteq B$, and M_i^j as the matching of the sets A_i^j and B_i^j . Consider the sequence of matchings $\mathcal{M}^n$ such that each matching is implemented after every n pairs of a - b agents have entered (whereupon the n matched pairs are removed from the system), given by:

$$\mathcal{M}^n = (M_1^n, M_{n+1}^{2n}, M_{2n+1}^{3n}, \dots).$$

Pseudocode for such a sequence of matchings is given in Algorithm 3.

Algorithm 3 Pseudocode for computing the sequence of matchings after every n pairs of agents have entered the system.

```

for ( $i = 1; i < N + 1; i++$ ) {
    add agents  $a_i$  and  $b_i$  into the system
    if ( $n$  divides  $i$ ) {
        match all agents and remove them from the system
    }
}
match all remaining agents
END
    
```

The weight of $\mathcal{M}^n$ is

$$w(\mathcal{M}^n) = \sum_{i=1}^{\lfloor \frac{N}{n} \rfloor} w(M_{(i-1)n+1}^{i \cdot n}, t_{i \cdot n}) + w(M_\alpha^N, t_N), \quad (5.5.2)$$

where $\alpha = N + 1 - n(\frac{N}{n} - \lfloor \frac{N}{n} \rfloor)$, so that the second term of (5.5.2) accounts for the case in which n does not divide N and all remaining a - b pairs are matched at t_N . We are interested in finding $n = n^{\text{opt}}$ that minimises (5.5.2) (see Figure 5.5.1b). This will depend on the rate λ at which pairs of agents enter the system, which will be assumed to follow a Poisson point process. So the probability that x pairs of agents enter the system over any time interval of length t is given by $P(X = x) = (\lambda t)^x e^{-\lambda t} / x!$, with expectation λt . Since the average wait time of a rider for an Uber vehicle is approximately 3 minutes [50], the side length S of the square domain is chosen so that the average distance between pairs of agents matched immediately on entry is also approximately

3 (given by $S = 6$). The dimensional constant for the waiting times of agents is fixed at $k = 1 \text{ min}^{-1}$.

5.5.3 Results

Figure 5.5.2 displays results for simulations taken over $N = 10^6$ a - b pairs for each value of λ . Figure 5.5.2a shows that n^{opt} increases with the rate λ , which is to be expected. When λ is small, the average time between entries of a - b pairs is large so the optimum sequence of matchings is to implement at each entry time, *i.e.* $n^{\text{opt}} = 1$. As λ increases, the average time between fresh entries becomes smaller and the cost of waiting times is outweighed by the benefit of matching a greater number of a - b pairs immediately. Analogous to the price of anarchy, we introduce the *Price of Impatience* (POI) to quantify the inefficiency of matching a - b pairs at their entry times, given by:

$$\text{POI} = \frac{w(\mathcal{M}^1)}{w(\mathcal{M}^{n^{\text{opt}}})} \geq 1$$

Figure 5.5.2a shows that the POI increases with the rate λ , as it becomes increasingly more efficient to allow more a - b pairs to enter the system before implementing a matching. Figure 5.5.2b shows the average time between the matchings for the optimum sequence of matchings $\mathcal{M}^{n^{\text{opt}}}$, given by

$$\bar{t} = \frac{\sum_{i=1}^{\lfloor \frac{N}{n^{\text{opt}}} \rfloor} t_{i \cdot n^{\text{opt}}} - t_{(i-1) \cdot n^{\text{opt}}}}{\lfloor \frac{N}{n^{\text{opt}}} \rfloor},$$

where $t_0 = 0$. Recall that the expected number of pairs of agents entering the system within an interval of length t is λt . Since $\bar{t}$ is the average time for n^{opt} pairs of agents to appear in the system, it is expected that $\bar{t} \approx n^{\text{opt}} / \lambda$, which is confirmed in figure 5.5.2b. Note, however, that there is a significant variance of the time t between the matchings of $\mathcal{M}^{n^{\text{opt}}}$, particularly for smaller values of λ . Figure 5.5.3 shows the probability density functions of t for $\lambda = 1, 2, 10$, and 50 . Since these are the waiting times until n^{opt} Poisson events have occurred, they follow precisely a Gamma (Erlang) distribution with probability function $P(t) = \lambda(\lambda t)^{n^{\text{opt}}-1} e^{-\lambda t} / (n^{\text{opt}} - 1)!$.

5.6 Summary

This chapter has considered inefficiency in bipartite graph matchings. When agents are paired by mutual agreement to their nearest available neighbour, inefficiency can occur if the resultant matching does not minimise the sum of the distances between all

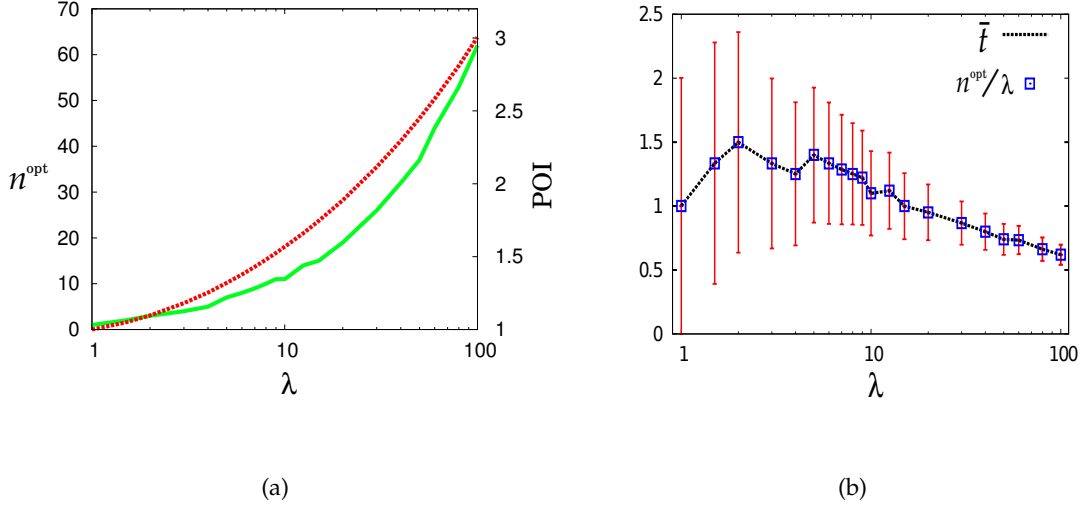


FIG. 5.5.2: (a) A semi-log plot of the POI (red) and n^{opt} (green) against the rate λ .
 (b) A semi-log plot of the average time $\bar{t}$ between the matchings for the optimum sequence of matchings $\mathcal{M}^{n^{\text{opt}}}$ against λ .

paired agents. Continuing the theme of previous chapters, the variation of the price of anarchy was analysed with respect to certain parameters of the system. The first case considered agents located within an empty, rectangular domain where the distance between agents was given by the Euclidean metric, which gave the following results:

1. $\bar{\mathcal{P}}$ is invariant with the area of the domain.
2. $\bar{\mathcal{P}}$ increases logarithmically with the number of pairs of agents being matched.
3. The variation of $\bar{\mathcal{P}}$ with the domain aspect ratio is maximised at approximately 7, *i.e.* when the length of the rectangle is 7 times the width (or vice versa).
4. The variation of $\bar{\mathcal{P}}$ with the ratio between the numbers of the two types of agents is maximised at 1, *i.e.* when the two sets of agents have equal size.

This analysis was then repeated where agents were located on the nodes of a network, and the distances between pairs of agents were given by the lengths of the geodesics between the nodes that they occupy. When the network had a diamond lattice structure, the variation of $\bar{\mathcal{P}}$ displayed the same behaviour as for the empty domain case.

Other network structures were then considered, including Erdős–Rényi (ER) random networks, Watts–Strogatz (WS) small-world networks, and Barabási–Albert (BA) scale-free networks. In each of these cases, the variation of the price of anarchy with the

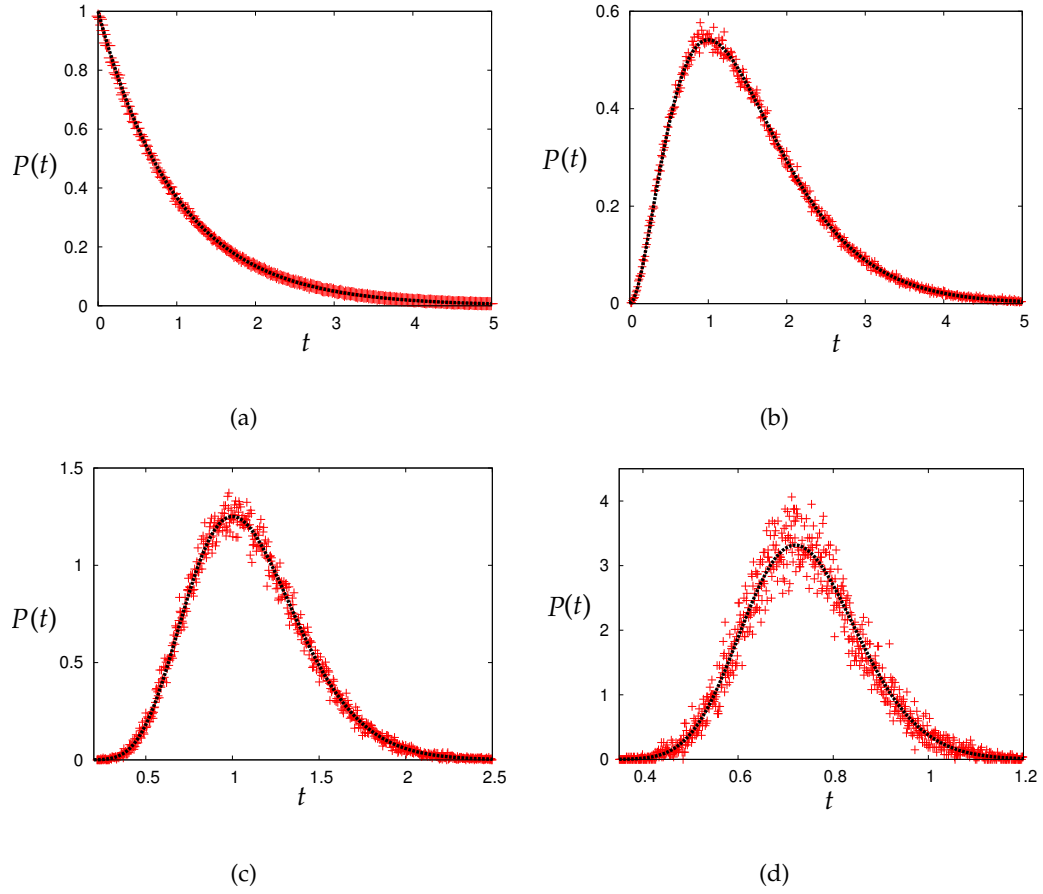


FIG. 5.5.3: Probability density functions (red crosses) for the time t between the matchings of $\mathcal{M}^{n^{\text{opt}}}$, for (a) $\lambda = 1$ ($n^{\text{opt}} = 1$); (b) $\lambda = 2$ ($n^{\text{opt}} = 3$); (c) $\lambda = 10$ ($n^{\text{opt}} = 11$); and (d) $\lambda = 50$ ($n^{\text{opt}} = 37$). The black lines indicate the corresponding Gamma distributions.

number of pairs of agents N was well fitted by the function $f(N) = a + b \log N + c\sqrt{N}$, with the values of a , b and c varying little across the different network structures. The price of anarchy and matching weights were also analysed as functions of parameters relating to the network. When the network was large enough such that all agents were matched, it was shown that the price of anarchy decreased with the network size. For WS networks, a parallel was drawn between the initial fall in the price of anarchy and in the global clustering coefficient with the rewiring parameter p_w . This may also explain the increasing behaviour of $\bar{P}$ with the density of edges for BA networks, and for ER networks in the limit $p_e \rightarrow 1$. However, the non-monotonic behaviour of $\bar{P}$ as a function of the connection parameter p_e (for ER networks) and the rewiring probability p_w (for WS networks) could not be fully explained.

An extension to the model then considered agents entering the system at different times via a random Poisson process. This extension is applicable to the context of ride-sharing, where requests from riders occur at different times and the availability of drivers varies over time. Sequences of SO matchings were considered such that agents were matched after a given number of pairs had entered the system. The weight of a sequence of matchings contained two factors:

1. The distance between matched agents.
2. The time between agents entering the system and being matched.

On the one extreme, matching agents at their entry times would have a high weight from factor 1) and no weight from factor 2). On the other extreme, matching agents after many pairs had entered the system would have a relatively small weight from factor 1) and a high weight from factor 2). The optimum sequence of matchings would find the right balance between the two factors. The inefficiency of matching agents at their entry times was considered relative to the optimum sequence of matchings, and quantified by the “Price of Impatience” (POI). The POI was analysed as a function of the rate λ at which agents entered the system. Numerical results showed that the POI quickly increases with λ , suggesting that there is significant merit in delaying matchings until multiple pairs of agents have entered the system. It was also shown that the optimum number of pairs between matchings increases with λ , and that the average optimum time between matchings decreases with λ . For ride-sharing companies, our analysis implies that service efficiency could be greatly improved by delaying driver dispatches until further rider requests have been made, and that this is particularly true when the rate of incoming rider requests is high.

This chapter concludes the research of this thesis. The next chapter will summarise the main findings and discuss possible opportunities for further research.

Conclusions and Further Work

This thesis has considered inefficiency within the context of agents acting on networks. A measure known as the “Price of Anarchy” (PoA) was used to quantify the inefficiency that occurs when agents seek to optimise their individual pay-offs while neglecting the performance of the society as a whole. The purpose of this thesis was to investigate how the PoA varies with respect to certain system parameters, and to identify the causes of high inefficiency. To this end, the PoA was analysed within the context of a traffic network model and a bipartite matching “ride-sharing” model. The following sections will summarise the main findings from the research into the two models, and will discuss possible opportunities for further research.

6.1 Traffic network model

6.1.1 Summary

The first model considered in this thesis consisted of a traffic network with a diamond lattice structure, originally studied in [4]. The network comprised two types of edges — those with a variable cost function equal to the amount of traffic the edge carries, and those with a fixed cost of 1 — and a traffic load T was routed across the width of the lattice. Chapter 2 recreated the main result of [4] and discovered that the variation of the PoA with the potential congestion parameter p exhibited a small local maximum at small p . This was a new observation that could not be explained by the conclusions in [4]. Several extensions to the model were then considered, including the effect of the lattice aspect ratio and the total traffic volume on the PoA. When the height of the lattice was significantly greater than the width, the variation of the PoA with p was shown to exhibit “ripples”, with the number of ripples being equal to the path length across the lattice. The variation of the PoA with the total traffic volume T was also shown

to exhibit fluctuations. The numerical results therefore displayed novel behaviours that could not be explained by the conclusions in [4]. Moreover, the non-monotonic behaviour of the variation of the PoA has significant implications for traffic network management, since it shows that an adjustment to the value of either parameter p or T could increase or decrease inefficiency.

Chapter 3 provided an investigation into the cause of the non-trivial behaviours displayed by the PoA in Chapter 2. A simplified model was introduced that retained the important features of the original model, but also allowed for analysis of the traffic flows. Paths between sources and sinks could be easily categorised into “classes” depending on the number of variable-cost edges they contain. Traffic-thresholds for paths of a given class to carry traffic were derived, and it was shown that these thresholds — known as “route transition points” — occurred at different values for the User Equilibrium (UE) and System Optimum (SO) traffic flows. It was described how this disparity caused the fluctuations in the PoA, and numerical simulations confirmed that the PoA was maximised at route transition points for the UE. This result was then applied to the original lattice network model, and the maxima of the PoA were successfully located.

Chapter 4 considered a method for reducing the PoA by applying “caps” to the most congested edges in the network under the UE flow. The optimum number of caps for each network realisation was dependent on the cost of implementing a cap. The reduction to the UE cost and the PoA was analysed with respect to the initial UE cost and the initial PoA, respectively, as well as the optimum number of caps. It was shown that cap implementation was more effective at reducing the UE cost when the UE flow was less efficient. Moreover, the optimum number of caps was the minimum number required to cap all k^* -paths. Finally, the average optimum number of caps and the average reduction in the UE cost were considered as functions of the cost per cap, showing a cut-off point where cap implementation ceases to be effective. The implication for road traffic management is that inefficiency could be greatly reduced by regulating the traffic on a few of the most congested roads.

6.1.2 Further research

Since road traffic networks are the main application for the traffic network model, the model could be modified so that the network structure further resembles the structures of real road networks. One property inherent to road networks is *planarity*, meaning that the network lies on the plane such that its edges meet only at nodes and nowhere

else. Clearly, this property holds for the diamond lattice. Research into the structure of real road networks has found the following common properties:

1. Nodes with degree ≥ 5 are very rare and the proportion of nodes with degree 4 is disproportionately large.
2. Angles between edges are generally around 90° or 180° , with the former being more abundant.
3. The distribution of cell areas $P(A_c)$ follows a power law of the form $P(A_c) = A_c^{-1.9}$.

The lattice network satisfies properties 1 and 2 to a degree. However, it completely fails in satisfying property 3, since all the cell areas in the lattice are the same size. One way of improving the lattice model could be to remove some segments of adjacent nodes and their edges from the lattice. For example, the removal of a single node and its 4 edges would create a cell with an area 4 times that of the standard lattice cell; the removal of 2 adjacent nodes would create a cell with an area 6 times that of the standard cell, *etc.* This removal process could be adjusted so that the power law cell-area distribution of property 3 is satisfied. A further advantage of this process is that the resultant network would satisfy property 1 to a greater extent, since nodes with degree < 4 would be formed yet the network would retain a high proportion of nodes with degree 4. It would also satisfy property 2 to a greater extent, since angles of 180° between edges would be created yet most angles would still be 90° .

Another way in which the model could be improved is by adopting a more sophisticated traffic assignment structure. While there is a lack of empirical data on the structure of traffic routing in real road networks, an assignment in which traffic is routed from one side of the network to the other — and moves only in one direction — seems to be very simple. Other routing structures that could be considered are the following:

1. Routing traffic in the vertical direction — between the top nodes and the bottom nodes — as well as horizontally.
2. Restricting the traffic so that a certain amount emanates from each source node, or that a certain amount terminates at each destination node.
3. Modifying the source and destination nodes by choosing them at random from any of the nodes within the network.

Finally, the method of reducing the PoA by applying caps to network edges could be considered on real road networks. The analysis of [6] found Braess’s paradox to occur on several of the roads of a London road network, meaning that the removal of these roads would improve the efficiency of the flow. Intuitively, it would be expected that applying “caps” on these roads would improve efficiency further. It would be interesting to analyse whether the PoA could be significantly reduced by applying caps to only a few of the roads. Such research could have useful implications for road traffic management, particularly with continued advances in technology.

6.2 Bipartite matching model

6.2.1 Summary

The second model considered in this thesis consisted of the pairwise matching of agents in an empty domain and on networks. While this model has a number of applications (see Section 1.7), the main motivation of the model was the context of ride-sharing, where riders are matched with drivers. The variation of the PoA was again analysed with respect to various system and network parameters. When the agents were located within an empty domain, numerical results yielded the following observations:

1. The PoA is invariant with the area of the domain.
2. The PoA increases logarithmically with the number of pairs of agents being matched.
3. The variation of the PoA with the domain aspect ratio is maximised at approximately 7, *i.e.* when the length of the rectangle is 7 times the width (or vice versa).
4. The variation of the PoA with the ratio between the numbers of the two types of agents is maximised at 1, *i.e.* when the two sets of agents have equal size.

The PoA was then considered when agents were located on the nodes of networks of particular topologies. Network structures considered included the diamond lattice (as in the traffic model), Erdős–Rényi (ER) random networks, Watts–Strogatz (WS) small-world networks, and Barabási–Albert (BA) scale-free networks. The diamond lattice case displayed very similar behaviour to the empty domain case. By contrast, novel behaviours were observed for ER, WS and BA networks. For all of these network structures, the variation of the PoA with the number of pairs of agents N was well fitted by the function $f(N) = a + b \log N + c\sqrt{N}$, with the values of a , b and c varying little across the different network structures. It was shown that, when the network

was large enough to be connected, the PoA decreased with network size. The variation of the PoA with the connection parameter p_e for ER networks was computed, and displayed several prominent features:

1. a local maximum at $p_e \approx \log(n)/n$, where n is the number of nodes comprising the network.
2. a local minimum at $p_e \approx 0.11$.
3. an increase with p_e to a plateau as $p_e \rightarrow 1$.

The variation of the PoA with the rewiring parameter p_w for WS networks also displayed several prominent features:

1. a significant drop for $0 < p_w < 0.45$.
2. a minimum at $p_w \approx 0.45$.
3. a local maximum at $p_w \approx 0.8$.

A parallel was observed between the initial fall in the PoA with varying p_w and the behaviour of the clustering coefficient (originally observed in [24]). It was also observed that the PoA monotonically increased with the density of edges for BA networks.

Finally, an extension to the model considered pairs of agents entering the system at different times. In this case, matching agents at their entry times did not necessarily represent the optimum sequence of matchings. The implication for ride-sharing companies is that service efficiency could be greatly improved by delaying driver dispatches until further rider requests have been made, particularly when the rate of incoming rider requests is high. It was shown that both the inefficiency of matching agents at their entry times — termed the “Price of Impatience” (PoI) — and the optimum number of pairs of agents to match, increase with the rate at which agents enter the system.

6.2.2 Further research

Chapter 5 presented an extensive survey of numerical results for the bipartite matching model, and explanations were provided for some of the behaviours displayed by the PoA. However, the causes for some of the observations were left unexplained. Further research could attempt to find reasons for the following:

1. The PoA increases logarithmically with the number of pairs of agents N for the empty domain and lattice network cases.

2. The PoA as a function of the aspect ratio of the domain is maximised at ≈ 7 for the empty domain and lattice network cases.
3. The PoA is well fitted by the function $f(N) = a + b \log N + c\sqrt{N}$ in each of the random network cases considered.
4. The non-trivial behaviour of the PoA as a function of the connection parameter p_e for ER networks.
5. The non-trivial behaviour of the PoA as a function of the rewiring parameter p_w for WS networks.

An extension to the model would be to consider many-to-one matchings, in which the agents of one of the sets are each matched to multiple agents from the other set. Such an extension would be pertinent in the context of ride-sharing for the analysis of multiple riders sharing a driver. Another opportunity for further research is to analyse the PoI on the same random network structures used in the analysis of the PoA. Finally, the PoA and the PoI could be analysed on real road networks, analogous to the research of [6] for the traffic-flow model.

Appendix

Chapter 2 considered the price of anarchy within a diamond lattice traffic network model. The price of anarchy $\mathcal{P}$ was computed as a function of the congestion potential parameter p and the traffic volume T , and displayed non-trivial behaviours. Chapter 3 provided an explanation for the variation of $\mathcal{P}$ that was consistent with the observation that $\mathcal{P}$ was maximised at route transition points under the UE. Since route transition points determine which classes of paths carry traffic, it was proposed that the variation of $\mathcal{P}$ is dependent on the classes of paths being utilised under the UE and SO. For a more detailed understanding of the variation of $\mathcal{P}$, several simple proofs that support the conjectures made in Chapter 3 are provided below. The total costs under the UE and the SO (and hence $\mathcal{P}$) are also solved for the simplified traffic model introduced in Chapter 3.

A.1 Variation of $\mathcal{P}$ with p

It was shown in Chapter 2 that when the number of source/destination nodes is much larger than the path length, *i.e.* $2I > J$, the variation of $\bar{\mathcal{P}}$ with respect to p exhibits ripples. Local maxima occur at the UE-RTP's for $(k^* - 1)$ -paths given by

$$N_{k^*}^{\text{ue}} = Tk^*,$$

such that $\mathcal{P}$ increases for $N_{k^*} < Tk^*$ and decreases for $Tk^* < N_{k^*} < 2Tk^*$ (and is at unity for $N_{k^*} \geq 2Tk^*$). One consequence of having $2I > J$ is that the number of paths is large relative to the path length. This means that for any $k^* = i < J$, as p increases, the average number of i -paths becomes relatively large before there emerges an $(i + 1)$ -path (and the value of k^* increments by 1). In other words, network realisations that exhibit a small N_{k^*} will also have a large value of N_{k^*-1} . Given this, an abstract case of the simplified model will now be considered where $N_{k^*-1} \rightarrow \infty$. This limit will prove useful for determining the sign of the derivative of $\mathcal{P}$. For convenience, define $C^{\text{ue}} := C(\tau^{\text{ue}})$ and $C^{\text{so}} := C(\tau^{\text{so}})$.

A.1.1 $\mathcal{P}(N_{k^*})$ increasing

It was shown in Chapter 3 that when $N_{k^*} < Tk^*$, all traffic is routed to $(k^* - 1)$ -paths and k^* -paths under both the UE and the SO, which gives

$$\tau_{k^*-1}N_{k^*-1} + \tau_{k^*}N_{k^*} = T.$$

Solving for τ_{k^*-1} :

$$\tau_{k^*-1} = \frac{T - \tau_{k^*}N_{k^*}}{N_{k^*-1}}. \quad (\text{A.1.1})$$

Consider the UE. Since all traffic-carrying paths have equal cost c , and $\tau_{k^*-1} > 0$ and $\tau_{k^*} > 0$, it follows that

$$c_{k^*-1} = c_{k^*} = c. \quad (\text{A.1.2})$$

The cost of traversing a k -path is given by $c_k = k\tau_k + (J - k)$, in view of which, solving (A.1.2) for τ_{k^*-1} gives

$$\tau_{k^*-1} = \frac{k^*\tau_{k^*} - 1}{k^* - 1}. \quad (\text{A.1.3})$$

From (A.1.1) and (A.1.3):

$$\frac{T - \tau_{k^*}N_{k^*}}{N_{k^*-1}} = \frac{k^*\tau_{k^*} - 1}{k^* - 1}.$$

Solving for τ_{k^*} :

$$\tau_{k^*} = \frac{T(k^* - 1) + N_{k^*-1}}{k^*N_{k^*-1} + (k^* - 1)N_{k^*}}.$$

In the limit $N_{k^*-1} \rightarrow \infty$:

$$\tau_{k^*} \xrightarrow{N_{k^*-1} \rightarrow \infty} \frac{1}{k^*}.$$

So the total cost under the UE is

$$C^{\text{ue}} = Tc = Tc_{k^*} = T(k^*\tau_{k^*} + (J - k^*)) = T(1 + J - k^*).$$

APPENDIX

Now consider the SO. Since $\tau_{k^*-1} > 0$ and $\tau_{k^*} > 0$, it follows from (3.3.6) that

$$C_{k^*-1}'(\tau_{k^*-1}) = C_{k^*}'(\tau_{k^*}), \quad (\text{A.1.4})$$

where $C_k(\tau_k) = \tau_k c_k(\tau_k)$ is the total cost of the traffic on each k -path, giving $C_k'(\tau_k) = 2k\tau_k + (J - k)$. Solving (A.1.4) for τ_{k^*-1} :

$$\tau_{k^*-1} = \frac{2k^*\tau_{k^*} - 1}{2(k^* - 1)}. \quad (\text{A.1.5})$$

From (A.1.1) and (A.1.5):

$$\frac{T - \tau_{k^*}N_{k^*}}{N_{k^*-1}} = \frac{2k^*\tau_{k^*} - 1}{2(k^* - 1)}.$$

Solving for τ_{k^*} :

$$\tau_{k^*} = \frac{2T(k^* - 1) + N_{k^*-1}}{2k^*N_{k^*-1} + 2(k^* - 1)N_{k^*}}.$$

In the limit $N_{k^*-1} \rightarrow \infty$:

$$\tau_{k^*} \xrightarrow{N_{k^*-1} \rightarrow \infty} \frac{1}{2k^*}.$$

From (A.1.1):

$$\tau_{k^*-1}N_{k^*-1} = T - \tau_{k^*}N_{k^*} = T - \frac{N_{k^*}}{2k^*}.$$

So the total cost under the SO is

$$\begin{aligned}
 C^{\text{so}} &= \tau_{k^*-1} N_{k^*-1} c_{k^*-1} + \tau_{k^*} N_{k^*} c_{k^*} \\
 &= \left(T - \frac{N_{k^*}}{2k^*} \right) \left((k^* - 1) \tau_{k^*-1} + (J - (k^* - 1)) \right) + \frac{N_{k^*}}{2k^*} \left(k^* \tau_{k^*} + (J - k^*) \right) \\
 &= \left(T - \frac{N_{k^*}}{2k^*} \right) (J - (k^* - 1)) + \frac{N_{k^*}}{2k^*} \left(\frac{1}{2} + J - k^* \right) \\
 &= T(J - (k^* - 1)) - \frac{N_{k^*}}{4k^*}. \tag{A.1.6}
 \end{aligned}$$

The price of anarchy is therefore:

$$\mathcal{P} = \frac{C^{\text{ue}}}{C^{\text{so}}} = \frac{T(J - (k^* - 1))}{T(J - (k^* - 1)) - \frac{N_{k^*}}{4k^*}},$$

which is increasing with N_{k^*} .

A.1.2 $\mathcal{P}(N_{k^*})$ decreasing

Now consider $T \leq N_{k^*}/k^* < 2T$, **first for the UE**. In this case, it was shown in Chapter 3 that all traffic is routed to the k^* -paths, so $\tau_{k^*} N_{k^*} = T$ and

$$C^{\text{ue}} = Tc = Tc_{k^*} = T(k^* \tau_{k^*} + (J - k^*)) = T \left(\frac{Tk^*}{N_{k^*}} + (J - k^*) \right).$$

Under the SO, traffic utilises both $(k^* - 1)$ -paths and k^* -paths and the total cost is therefore given by (A.1.6). By the quotient rule, the sign of $\mathcal{P}$ is determined by

$$\begin{aligned}
 C^{\text{so}} \cdot \frac{dC^{\text{ue}}}{dN_{k^*}} - C^{\text{ue}} \cdot \frac{dC^{\text{so}}}{dN_{k^*}} &= -\frac{T^2 k^*}{N_{k^*}^2} \left[T(J - (k^* - 1)) - \frac{N_{k^*}}{4k^*} \right] + \frac{1}{4k^*} \left[T \left(\frac{Tk^*}{N_{k^*}} + (J - k^*) \right) \right] \\
 &= T(J - k^*) \left(\frac{1}{4k^*} - \frac{T^2 k^*}{N_{k^*}^2} \right) + \frac{T^2}{N_{k^*}} \left(\frac{1}{2} - \frac{Tk^*}{N_{k^*}} \right). \tag{A.1.7}
 \end{aligned}$$

APPENDIX

It follows from $N_{k^*} < 2Tk^*$ that

$$\frac{T^2k^*}{N_{k^*}^2} > \frac{T^2k^*}{(2Tk^*)^2} = \frac{1}{4k^*},$$

and

$$\frac{Tk^*}{N_{k^*}} > \frac{Tk^*}{2Tk^*} = \frac{1}{2}.$$

Moreover, since $J \geq k^*$, it follows that the first term of (A.1.7) is ≤ 0 and the second term of (A.1.7) is < 0 . Hence (A.1.7) is negative and $\mathcal{P}$ is decreasing with N_{k^*} .

While the example that has been considered in this section is too abstract to be proof that $\mathcal{P}$ is maximised at $N_{k^*} = Tk^*$ in either the lattice model or the simplified model, it does provide support for this conjecture. Before considering the variation of $\mathcal{P}$ with respect to T , it will be useful to first solve for the UE and the SO total costs, which will be done in the next section.

A.2 Solving for $\mathcal{P}$ for the simplified model

The abstract case of $N_{k^*-1} \rightarrow \infty$ will now be discarded and the standard simplified model considered. Full solutions to the UE and the SO total costs (and hence $\mathcal{P}$) will be derived.

First consider the UE. Recalling that $\hat{k}^{\text{ue}}$ is the number of variable edges within traffic-carrying paths with the fewest variable edges (under the UE), it follows that

$$T = \sum_{j=\hat{k}^{\text{ue}}}^{k^*} N_j \tau_j. \quad (\text{A.2.1})$$

Denoting by c the cost of traversing paths that carry traffic,

$$c_j = c \quad \text{for } j = \hat{k}^{\text{ue}}, \hat{k}^{\text{ue}} + 1, \dots, k^*.$$

Since $c_j = j\tau_j + (J - j)$, solving for τ_j gives:

$$\tau_j = \frac{c + j - J}{j} \quad \text{for } j = \hat{k}^{\text{ue}}, \hat{k}^{\text{ue}} + 1, \dots, k^*.$$

Substituting this into (A.2.1):

APPENDIX

$$T = \sum_{j=\hat{k}^{\text{ue}}}^{k^*} \frac{c+j-J}{j} N_j,$$

which on solving for c gives

$$c = \frac{T + \sum_{j=\hat{k}^{\text{ue}}}^{k^*} N_j \left(\frac{J}{j} - 1\right)}{\sum_{j=\hat{k}^{\text{ue}}}^{k^*} \frac{N_j}{j}}.$$

So the total cost is given by

$$C^{\text{ue}} = Tc = \frac{T^2 + T \sum_{j=\hat{k}^{\text{ue}}}^{k^*} N_j \left(\frac{J}{j} - 1\right)}{\sum_{j=\hat{k}^{\text{ue}}}^{k^*} \frac{N_j}{j}}. \quad (\text{A.2.2})$$

Now consider the SO, where

$$T = \sum_{j=\hat{k}^{\text{so}}}^{k^*} N_j \tau_j. \quad (\text{A.2.3})$$

By (3.3.6), it follows that for all k such that $\tau_k > 0$, $C'_k(\tau_k)$ have equal value. Denoting this value by $\hat{c}$,

$$C'_j(\tau_j) = \hat{c} \quad \text{for } j = \hat{k}^{\text{so}}, \hat{k}^{\text{so}} + 1, \dots, k^*.$$

Solving for τ_j :

$$\tau_j = \frac{\hat{c} + j - J}{2j} \quad \text{for } j = \hat{k}^{\text{so}}, \hat{k}^{\text{so}} + 1, \dots, k^* \quad (\text{A.2.4})$$

Substituting this into (A.2.3):

$$T = \sum_{j=\hat{k}^{\text{so}}}^{k^*} \frac{\hat{c} + j - J}{2j} N_j,$$

which on solving for $\hat{c}$ gives

$$\hat{c} = \frac{2T + \sum_{j=\hat{k}^{\text{so}}}^{k^*} N_j \left(\frac{J}{j} - 1\right)}{\sum_{j=\hat{k}^{\text{so}}}^{k^*} \frac{N_j}{j}}. \quad (\text{A.2.5})$$

APPENDIX

The total cost is given by

$$\begin{aligned}
 C^{\text{so}} &= \sum_{j=\hat{k}^{\text{so}}}^{k^*} C_j(\tau_j) N_j \\
 &= \sum_{j=\hat{k}^{\text{so}}}^{k^*} c_j \tau_j N_j \\
 &= \sum_{j=\hat{k}^{\text{so}}}^{k^*} (j\tau_j + (J-j)) \tau_j N_j
 \end{aligned}$$

Substituting (A.2.4) into the above and expanding:

$$C^{\text{so}} = \hat{c}^2 \sum_{j=\hat{k}^{\text{so}}}^{k^*} \frac{N_j}{4j} - \sum_{j=\hat{k}^{\text{so}}}^{k^*} \frac{(J-j)^2}{4j} N_j$$

Substituting (A.2.5) into the above and expanding:

$$C^{\text{so}} = \frac{4T^2 + 4T \sum_{j=\hat{k}^{\text{so}}}^{k^*} N_j \left(\frac{J}{j} - 1 \right) + \left(\sum_{j=\hat{k}^{\text{so}}}^{k^*} N_j \right)^2 - \left(\sum_{j=\hat{k}^{\text{so}}}^{k^*} j N_j \right) \left(\sum_{j=\hat{k}^{\text{so}}}^{k^*} \frac{N_j}{j} \right)}{4 \left(\sum_{j=\hat{k}^{\text{so}}}^{k^*} \frac{N_j}{j} \right)}.$$

A.3 Variation with T

In Chapter 3, it was conjectured that $\mathcal{P}$ is increasing with T when $\hat{k}^{\text{so}} < \hat{k}^{\text{ue}}$, and is either constant or decreasing with T when $\hat{k}^{\text{so}} = \hat{k}^{\text{ue}}$. Consider first the case $\hat{k}^{\text{so}} = \hat{k}^{\text{ue}} = \tilde{k}$.

A.3.1 $\mathcal{P}(T)$ constant or decreasing

By (A.2.2),

$$\frac{dC^{\text{ue}}}{dT} = \frac{2T + \sum_{j=\tilde{k}}^{k^*} N_j \left(\frac{J}{j} - 1 \right)}{\sum_{j=\tilde{k}}^{k^*} \frac{N_j}{j}} = \frac{2T - \sum_{j=\tilde{k}}^{k^*} N_j \left(1 - \frac{J}{j} \right)}{\sum_{j=\tilde{k}}^{k^*} \frac{N_j}{j}}. \quad (\text{A.3.1})$$

By (A.2.5) and (A.3.1) it follows that $\hat{c} = \frac{d}{dT} C^{\text{ue}}$. The derivative of C^{so} with T is given

APPENDIX

by:

$$\begin{aligned}
 \frac{dC^{\text{so}}}{dT} &= \frac{d}{dT} \left(\sum_{j=\tilde{k}}^{k^*} C_j(\tau_j) N_j \right) \\
 &= \sum_{j=\tilde{k}}^{k^*} N_j \frac{d}{dT} C_j(\tau_j) \\
 &= \sum_{j=\tilde{k}}^{k^*} N_j C'_j(\tau_j) \frac{d\tau_j}{dT} \\
 &= \hat{c} \sum_{j=\tilde{k}}^{k^*} N_j \frac{d\tau_j}{dT}.
 \end{aligned} \tag{A.3.2}$$

From (A.2.4) and (A.2.5),

$$\frac{d\tau_j}{dT} = \frac{1}{2j} \frac{d\hat{c}}{dT} = \frac{1}{j \sum_{k=\tilde{k}}^{k^*} \frac{N_k}{k}} \quad \text{for } j = \tilde{k}, \tilde{k} + 1, \dots, k^*.$$

Substituting the above into (A.3.2):

$$\frac{dC^{\text{so}}}{dT} = \frac{\hat{c} \sum_{j=\tilde{k}}^{k^*} \frac{N_j}{j}}{\sum_{k=\tilde{k}}^{k^*} \frac{N_k}{k}} = \hat{c} = \frac{dC^{\text{ue}}}{dT}.$$

Since C^{ue} and C^{so} increase with T at the same rate, it follows that $\mathcal{P}(T)$ is constant (and equal to one) when $C^{\text{ue}} = C^{\text{so}}$ and decreasing when $C^{\text{ue}} > C^{\text{so}}$.

References

- [1] G. Owen, *Game Theory*, 3rd ed. (Academic Press, 1995).
- [2] S. Torre, J. Kalda, R. Kitt, and J. Engelbrecht, arXiv:1602.04352 [physics.soc-ph] (2016).
- [3] C. Papadimitriou, Proceedings of the 33rd Annual ACM Symposium on the Theory of Computing (2001), pp. 749-753.
- [4] B. Skinner, Phys. Rev. E **91**, 052126 (2015).
- [5] Y. Sheffi, *Urban Transportation Networks* (Prentice-Hall, 1985).
- [6] H. Youn, M. T. Gastner, and H. Jeong, Physical Review Letters **101**, 128701 (2008).
- [7] T. Roughgarden and E. Tardos, JACM **49**, 236 (2002).
- [8] T. Roughgarden, JCSS **67**, 341 (2003).
- [9] V. A. Knight and P. R. Harper, European Journal of Operational Research **230**, 122 (2013).
- [10] B. Skinner, Journal of Quantitative Analysis in Sports **6**, 3 (2010).
- [11] B. Skinner and B. Carlin, Significance **10**, 25 (2013).
- [12] A. Nagurney, Transportation Research Part E **42**, 293-316 (2006).
- [13] A. Roth and M. Sotomayor, *Two-sided Matching: A study in game-theoretic modeling and analysis* (Cambridge University Press, 1989).
- [14] B. Maggs and R. Sitaraman, ACM SIGCOMM Computer Communication Review **45**, 3 (2015).
- [15] D. Gale and L. S. Shapley, American Mathematical Monthly **69**, 9-15 (1962).
- [16] R. Albert and A. Barabasi, Reviews of Modern Physics **74**, 47-97 (2002).

REFERENCES

- [17] A. Fronczak, P. Fronczak and J. Holyst, Phys. Rev. E **70**, 056110 (2004).
- [18] J. Nieminen, Scandinavian Journal of Psychology **15**, 332-336 (1974).
- [19] G. Sabidussi, Psychometrika **31**, 581-603 (1966).
- [20] M. O. Jackson, *Social and Economic Networks* (Princeton University Press, 2008).
- [21] L. C. Freeman, Sociometry **40**, 35-41 (1977).
- [22] P. Erdos and A. Renyi, Publicationes Mathematicae (Debrecen) **6**, 290-297 (1959).
- [23] S. Milgram, Psychology Today **1**, 61 (1967).
- [24] D. J. Watts and S. H. Strogatz, Nature **393**, 440-442 (1998).
- [25] A. Barabasi and R. Albert, Science **286**, 509-512 (1999).
- [26] A. D. Broido and A. Clauset, arXiv:1801.03400v1 [physics.soc-ph] (2018).
- [27] <https://www.barabasilab.com/post/love-is-all-you-need> (date accessed 11/05/18).
- [28] M. O. Jackson and B. W. Rogers, American Economic Review **97**, 890-915 (2007).
- [29] W. O. Kermack and A. G. McKendrick, Proceedings of the Royal Society A **115**, 700-721 (1927).
- [30] J. G. Wardrop, Proceedings of the Institute of Civil Engineers **2**, 325-378 (1952).
- [31] J. J. Wu, Z. Y. Gao, H. J. Sun, and H. J. Huang, Europhysics letters **74**, 560 (2006).
- [32] X. M. Zhao and Z. Y. Gao, Chinese Phys. Lett. **24**, 283 (2007).
- [33] J. J. Wu, Z. Y. Gao, and H. J. Sun, Physica A **387**, 1025 (2008).
- [34] J. J. Wu, X. Guo, H. Sun, and B. Wang, Mathematical Problems in Engineering **2014**, 490483 (2014).
- [35] A. L. Peressini, F. E. Sullivan, and J. J. Uhl, *The Mathematics of Nonlinear Programming*. (Springer-Verlag, 1988).
- [36] M. Beckman, C. B. McGuire, and C. B. Winsten, *Studies in the Economics of Transportation* (Yale University Press, 1956).
- [37] D. Braess, A. Nagurney, and T. Wakolbinger, Transportation Science **39**, 446-450 (2005).

REFERENCES

- [38] N. E. Stier-Moses, J. R. Correa and A. S. Schulz, *Games and Economic Behaviour* **64**, 457 (2008).
- [39] Bureau of public roads, *Traffic assignment manual for application with a large, high speed computer*, U.S. Dept. of Commerce, Bureau of Public Roads, Office of Planning, Urban Planning Division (1964).
- [40] S. J. O'Hare, R. D. Connors, and D. P. Watling, *Transportation Research Part B* **84**, 55 (2016).
- [41] A. E. Roth and M. A. O. Sotomayor, *Two-sided matching* (Cambridge University Press, 1990).
- [42] Michel X. Geomans, *Combinatorial Optimization: Lecture notes on bipartite graph matching*. (Massachusetts Institute of Technology, 2009).
- [43] <https://uberpeople.net/threads/another-uber-management-meeting.153809/> (date accessed 29/04/18).
- [44] E. M. Reingold and R. E. Tarjan, *Society for Industrial and Applied Mathematics* **10**, 4 (1981).
- [45] M. Poloczek and M. Szegedy, *FOCS '12 Proceedings of the 2012 IEEE 53rd Annual Symposium on Foundations of Computer Science* (2012), pp. 708-717.
- [46] S. Miyazaki, *Information Processing Letters* **114**, 714-717 (2014).
- [47] B. Skinner, *arXiv:1404.2935 [cond-mat.dis-nn]* (2014).
- [48] A. Rose, R. O'Dea, and K. I. Hopcraft, *Phys. Rev. E* **94**, 032315 (2016).
- [49] <https://www.uber.com/en-GB/drive/partner-app/how-surge-works/> (date accessed 13/01/18).
- [50] <http://www.newsweek.com/exclusive-heres-how-long-it-takes-get-uber-across-us-cities-289133> (date accessed 13/01/18).

List of Symbols and Abbreviations

$\mathcal{P}$	price of anarchy
T	total traffic volume
p	potential congestion parameter
p_c	lattice percolation threshold
I	number of nodes down the vertical side of the lattice
J	edge path length spanning the width of the lattice
$2I/J$	lattice aspect ratio
τ	arbitrary traffic flow of the network
τ^{ue}	The User Equilibrium (UE) traffic flow
τ^{so}	The System Optimum (SO) traffic flow
$\mathbb{E}$	set of edges of the network
$\mathcal{Q}$	set of paths between source nodes and desination nodes
k -path	a path that contains k variable-cost edges
N_k	number of k -paths
k^*	the number of variable-cost edges within paths with the greatest number of variable-cost edges
τ_e	traffic routed to edge $e \in \mathbb{E}$
τ_q	traffic routed to path $q \in \mathcal{Q}$
τ_k	traffic routed to each k -path
$c_e(\tau_e)$	cost of traversing edge $e \in \mathbb{E}$
$c_q(\tau)$	cost of traversing path $q \in \mathcal{Q}$
$c_k(\tau_k)$	cost of traversing a k -path
c	cost of traversing any traffic-carrying path
$C_k(\tau_k)$	total cost of the traffic on each k -path
$C(\tau)$	total cost of the traffic flow
$\mathcal{T}_i^{\text{ue}}$	route transition point for i -paths under the UE
$\mathcal{T}_i^{\text{so}}$	route transition point for i -paths under the SO

$\mathcal{N}$	optimum number of caps for reducing $\mathcal{P}$
N_{cap}	upper bound on the number of caps
δC	reduction in the UE traffic flow cost after i^* caps have been applied
μ	cost of applying each cap
$\delta \mathcal{P}$	reduction in the price of anarchy after i^* caps have been applied
$w(M)$	weight of matching M
w^{ue}	weight of the UE matching
w^{so}	weight of the SO matching
A, B	the two sets of agents (riders and drivers) to be matched
N	number of pairs of agents
S	side length of square domain containing the agents
S_1, S_2	side lengths of rectangular domain containing the agents
V	aspect ratio of domain
R	ratio of riders to drivers
n	number of nodes comprising the network
p_e	network connectivity parameter for Erdős–Rényi random graphs
p_w	rewiring parameter for Watts–Strogatz small-world networks
d	initial degree of each node for Watts–Strogatz small-world networks
$Cl(p_w)$	clustering coefficient for Watts–Strogatz small-world networks
m	initial degree of a newly added node in the generation of Barabási–Albert scale-free networks
λ	rate at which agents enter the time-dependent system
$\mathcal{M}^n$	sequence of SO matchings implemented after every n agents have entered the system
n^{opt}	optimum number of agents between matchings
t	times between matchings for the optimum sequence of matchings
POI	the price of impatience