

# Price of anarchy on heterogeneous traffic-flow networks

A. ROSE, R. O'DEA, K. I. HOPCRAFT

*School of Mathematical Sciences, University of Nottingham, Nottingham, NG7 2RD, UK*

September 23, 2016

The efficiency of routing traffic through a network, comprising nodes connected by links whose cost of traversal is either fixed or varies in proportion to volume of usage, can be measured by the 'price of anarchy'. This is the ratio of the cost incurred by agents who act to minimise their individual expenditure to the optimal cost borne by the entire system. As the total traffic load and the network variability—parameterised by the proportion of variable-cost links in the network—changes, the behaviours that the system presents can be understood with the introduction of a network of simpler structure. This is constructed from classes of non-overlapping paths connecting source to destination nodes that are characterised by the number of variable-cost edges they contain. It is shown that localised peaks in the price of anarchy occur at critical traffic volumes at which it becomes beneficial to exploit ostensibly more expensive paths as the network becomes more congested. Simulation results verifying these findings are presented for the variation of the price of anarchy with the network's size, aspect-ratio, variability and traffic load.

<https://journals.aps.org/pre/abstract/10.1103/PhysRevE.94.032315>

# 1 Introduction

The principles of least action or minimum energy are fundamental to the determination of equilibria in the physical sciences. However, in systems comprising competitive agents who can interact with their environment and among themselves, the attainable equilibrium configurations can be influenced substantially through the local strategies that agents adopt and be different from those accessible through globally-imposed strictures. Hence there can be a difference between an equilibrium that optimises the advantages of the *society* or ensemble of agents compared with an equilibrium that optimises the advantages gained for the *individual* agents. This difference can be gauged by determining the cost (measured in time or resources) for establishing these different classes of equilibria. The ratio of these costs has been termed the ‘Price of Anarchy’ (denoted by  $\mathcal{P}$ ) [1] whose deviation from unity measures the percentage of resource wasted through the agents’ selfish behaviour. This concept is applicable to and important for systems in which agents compete for a limited resource, and for mediating or ameliorating the effects of self-interest-driven inefficiency. Illustrative examples of such systems include: the use of public services such as health centres [2], the implementation of strategies in sports [3,4], transportation [5–15] and computer networks [16–21], and contagion dynamics [22]. The interactions of competitive agents in such systems can be understood in terms of dynamics on networks, and the techniques of this field can be brought to bear in their analysis. The desire to minimise  $\mathcal{P}$  motivates investigating how its value is affected by the agents’ actions and influenced by the environment that connects them to each other. Such knowledge would provide the capability to achieve efficiency and robustness to individuals’ actions through manipulating the network structure or, if this is not feasible, to enforce the social optimum through a central authority influencing the quantity it carries.

The archetypal study of this problem was presented in [23] in the context of road traffic flow. A simple network model was proposed in which ‘roads’ and ‘traffic’ provided a proxy for any congestible system of links or edges bearing a flux of some quantity. The network consists of two nodes connected by two parallel edges (roads), and a total traffic  $T$  is routed from one node (the source) to the other (the destination). The two edges are different in function: the cost incurred by a network user for traversing the *fixed-cost* edge is constant, while the cost of traversing the *variable-cost* edge is equal to the amount of traffic it carries. Given  $T = t_f + t_v$ , where  $t_f$  and  $t_v$  are the amounts of traffic routed to the fixed- and variable-cost edges respectively, then  $c_f(t_f) = c_f$  and  $c_v(t_v) = \alpha t_v$  are the respective costs of traversing the fixed- and variable-cost edges. The total cost incurred by all users is then given by  $C(t_f, T) = t_f c_f + (T - t_f) c_v(T - t_f)$ . The traffic flow with minimum total cost is called the *System Optimum* (SO). This is achieved when  $t_f = T - c_f/2\alpha$  and the associated total cost is  $C^{\text{so}} = \alpha(T^2 - t_f^2)$ . However, when network users choose their routes to minimise their individual costs, analogous to the concept of a Nash equilibrium [24] in game theory, the resultant traffic flow is called the *User Equilibrium* (UE). At the SO, the fixed-cost edge incurs a cost  $c_f$  whereas the variable-cost edge incurs a cost  $c_f/2$ . There is evidently an incentive for users to switch from the fixed to the variable-cost edge. At the UE, if  $T \leq c_f/\alpha$  then all the traffic takes the variable-cost edge giving a total cost  $C^{\text{ue}} = \alpha T^2$ . Consequently  $\mathcal{P} \equiv C^{\text{ue}}/C^{\text{so}} \geq 1$  in general, with  $\mathcal{P} = 1$  corresponding

to the absence of any inefficiency driven through selfish routing. In [23] the network was analysed for  $T = \alpha = c_f = 1$  which gave  $\mathcal{P} = 4/3$ , this being an upper bound on the value of  $\mathcal{P}$  achievable for any network with linear cost functions [17]. Effort has been expended on deriving bounds on  $\mathcal{P}$  for various nonlinear cost functions [25–28], although these issues will not be the concern here.

More recent investigations have considered the behaviours that  $\mathcal{P}$  can exhibit for large and complex networks. In [5] the efficiency of real transportation networks was considered, using road networks for parts of Boston-Cambridge, New York City and London as examples. In each case, the value of  $\mathcal{P}$  was calculated as a function of the traffic load  $T$  which was routed between a selected pair of origin (source) and destination (sink) locations. The routes between the two locations were a series of roads, all of which incurred a variable cost for carrying traffic. The value of  $\mathcal{P}$  exhibited generically similar features irrespective of the road structure or form of  $c_v(t_v)$  that was appropriate to that city, increasing from unity through a series of local maxima or ‘ripples’ to a principal peak value  $\sim 1.3$  before falling monotonically to unity at high values of  $T$ . The number of ripples exhibited was particular to the city. This general trend was repeated for abstract networks having different topologies by virtue of the degree distribution [29]. The ripple structures smoothed out to leave a single peak in  $\mathcal{P}$  when averages over different sources and sinks were considered as an ensemble. Similar behaviour was seen in a study of the public health service in Wales [2], which examined the inefficiency of allowing patients to choose which health centre they visit rather than having it designated to them by central authority. The value of  $\mathcal{P}$  arising in a network representing public and private healthcare provision—with the former corresponding to variable-cost edges and the latter to fixed-cost edges—was investigated as a function of the demand for treatment, this being analogous to  $T$ . Here too  $\mathcal{P}$  exhibited ripples before declining monotonically beyond a principal peak value  $\mathcal{P} \sim 1.2$ . In recent work [14], the dependence of  $\mathcal{P}$  on  $T$  for a two-node network with parallel links of varying costs was investigated, and analytical locations of the peaks in  $\mathcal{P}$  were given.

The investigation in [30] sought to gain a more fundamental understanding of the influence of a congestible network structure on its efficiency. In place of networks aimed at modelling real-world systems a regular lattice was considered and the effect on  $\mathcal{P}$  to changing the proportion of variable-cost edges whilst keeping  $T$  fixed was examined. This enabled the influence of the network microscopic-structure on its macroscopic manifestations to be probed. The lattice structure considered is depicted in Figure 1, the key features of which being that traffic emanates from  $I$  left-most nodes, terminates at  $I$  right-most nodes, between which paths of  $L$  edges are traversed. The traffic is one-way, always travelling from source to destination, and the edges joining the nodes have probability  $p$  and  $(1 - p)$  of belonging to the variable- or fixed-cost type respectively. In [30] the particular case for which  $L = 2I$  was considered and the value of  $\mathcal{P}$  (whose average over many network realisations is  $\bar{\mathcal{P}}$ ) as a function of  $p$  was shown to display similar behaviour to that occurring in [5], rising from  $\bar{\mathcal{P}} = 1$  to a peak value  $\sim 1.05$  at  $p \sim 0.6$  before falling rapidly to unity for larger values of  $p$ , but the ripple structures were absent. The location of the peak in  $\bar{\mathcal{P}}$  was shown to occur at the value of  $p$  which corresponds to the directed-percolation threshold of the lattice, *i.e.* at the emergence of paths crossing the entire network compris-

ing entirely variable-cost edges. Consequently a connection was drawn between this and the more general insights gleaned from the extensive study of idealized lattice models. The conclusions of [30] apply for a network of sufficiently large size that edge effects could be ignored and a suitably defined thermodynamic limit thereby inferred.

The qualitative similarity of the behaviours exhibited by  $\mathcal{P}$  to variation in the flux upon, and inhomogeneity of, the network prompts enquiring whether the effect of some underlying principle is being manifested, and if so whether this can be employed to explain the quantitative differences. This paper seeks to provide a more complete study of the dependence of the efficiency of random heterogeneous networks on the network’s structure and the volume of traffic it bears. This will be achieved by considering the idealized lattice structure of [30] to explore the effects that network microstructure and macroscale geometry has on  $\mathcal{P}(p, T)$ . Evidently the value of  $\mathcal{P}$  when viewed as a function of two variables possesses a richer structure than has been revealed hitherto. This work will unveil that structure both by computational and analytical means.

We highlight that the ripple structures observed in [2,5] for  $\mathcal{P}(p = \text{const.}, T)$  also occur for  $\bar{\mathcal{P}}(p, T = \text{const.})$  in smaller networks with aspect ratio  $R = 2l/L$  different from unity. We analyse the dependence of  $\bar{\mathcal{P}}$  on  $p, T$ , and network size and shape, highlighting regions in  $(p, T)$ -space where  $\mathcal{P}$  attains both local and global maxima, thereby providing diagnostics to predict and manipulate those conditions causing inefficiency through the action of dynamical processes on a heterogeneous network. We formalise and extend the analysis of [14] to consider a lattice network and thereby characterise the ripples in  $\mathcal{P}(p, T)$ , a key ingredient of which being to recognise that the lattice network under consideration can be distilled to a simpler network that permits analysis.

The contents are organized as follows. Section 2 describes the lattice structure and the procedures through which the optimum and Nash equilibrium costs, and thereby  $\mathcal{P}$ , are determined. Section 3 explores the effects of network size and aspect ratio in combination with variation in the homogeneity of the links and overall traffic volume respectively. The full landscape of  $\bar{\mathcal{P}}(p, T)$  is also considered, and approximately invariant structures that exist within that terrain are highlighted. In Section 4, an explanation for the occurrence of the ripple structure and location is provided. The final section discusses the implications and generalizations of the work. Technical details are assigned to Supplementary Material.

## 2 A heterogeneous network model

The network structure studied here is that considered by [30] and described in Section 1. Periodic boundary conditions are applied in the vertical direction and, for a particular realization, users have complete information about the network’s structure and all other users’ actions.

For the remainder of this article we shall refer to “variable-cost” and “fixed-cost” edges simply as “variable” and “fixed” edges, respectively. If  $j \in P$  denotes an edge of a particular path  $P$ , comprising both variable and fixed edges, the cost for an individual user to

traverse that path will be

$$c_P = \sum_{j \in P} \left( \chi_j c_v(\tau_j) + (1 - \chi_j) c_f \right). \quad (1)$$

where  $0 \leq \tau_j \leq T$  is the traffic carried on edge  $j$  and  $\chi_i$  is an indicator function:

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

We consider a network in which a proportion  $p$  of randomly-chosen edges have variable cost function  $c_v(\tau) = \tau$ , with the remainder having constant cost  $c_f = 1$ . In this way, the parameter  $p$  denotes the level of *variability* of the network. The cost of traversing a path  $P$  is then

$$c_P = \sum_{j \in P} \left( \chi_j \tau_j + (1 - \chi_j) \right). \quad (2)$$

The total cost incurred by *all* traffic on path  $P$  is

$$C_P = \sum_{j \in P} \left( \chi_j \tau_j^2 + (1 - \chi_j) \tau_j \right). \quad (3)$$

If  $E$  denotes the set of all edges of the network, then the total cost of a traffic flow on the network is given by

$$C = \sum_{j \in E} \left( \chi_j \tau_j^2 + (1 - \chi_j) \tau_j \right).$$

To reiterate, at UE all network users choose the cheapest path available to them while the SO traffic flow minimises  $C$ . Denoting these total network costs  $C^{\text{ue}}$  and  $C^{\text{so}}$ , respectively, the price of anarchy is  $\mathcal{P} = C^{\text{ue}} / C^{\text{so}}$ .

An iterative algorithm (after [30]) is employed to compute the UE and SO traffic flows—full details are given in Appendix A, but the essentials are as follows. The network is initialised with a uniform distribution of traffic over all paths of the lattice. Then, an iterative procedure is implemented whereby two paths  $A$  and  $B$  are chosen at random and some traffic  $\delta\tau$  is transferred from each edge of  $B$  to each edge of  $A$ . The calculation of  $\delta\tau$  is dependent on whether the UE or SO is being computed: for UE, the transfer of  $\delta\tau$  results in  $c_A = c_B$ ; for SO, the transfer minimises  $(C_A + C_B)$ . Due to the structure of the lattice network under consideration, the number of paths between source nodes and destination nodes is very large even for ostensibly small networks. The computational cost of computing the traffic flows is therefore relatively high – details of which can be found in Appendix A.

### 3 Results

This section presents results demonstrating the influence that the network variability  $p$ , traffic  $T$ , size  $I \times L$ , and aspect ratio  $R = 2I/L$ , have upon the price of anarchy  $\mathcal{P}$ . Ensemble

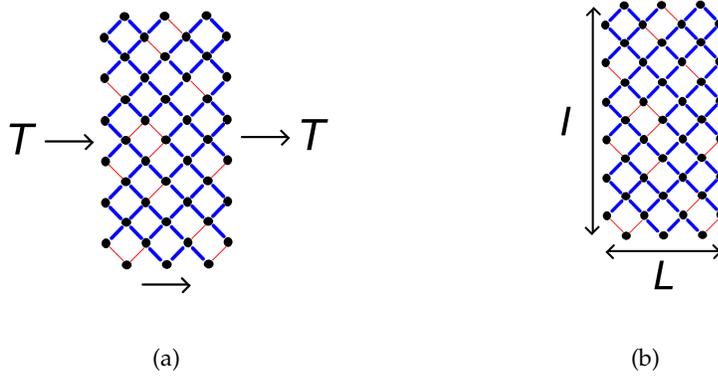


FIG. 1: (Colour online.) The lattice network structure, with a volume of traffic  $T$  routed from  $I$  source nodes to  $I$  destination nodes via paths of length  $L$  that comprise randomly interspersed fixed edges (blue/thick) and variable edges (red/thin).

averages for  $\bar{\mathcal{P}}$  are formed from 2000 realisations. Figure 2 shows ensemble average results  $\bar{\mathcal{P}}(p, T = 1)$ , for lattices with aspect ratios  $R > 1$ . In 2(a)  $\bar{\mathcal{P}}$  is presented for  $I = 80$  and  $L = 5$  and 6: the value of  $\bar{\mathcal{P}}$  increases with  $p$  to a maximum of  $\bar{\mathcal{P}} \sim 1.14$  through a series of ripples of increasing size before rapidly falling to unity, at which the UE and SO traffic flows coincide. The number of ripples is always equal to the path length, with one ripple which is not easily visible occurring at  $p \approx 0$ . Figure 2(b) indicates the variation of  $\mathcal{P}$  within the network realisations associated with the results in Figure 2(a) for  $L = 6$ , by inclusion of an envelope of one standard deviation, thereby highlighting that while the details in  $\mathcal{P}$  vary, the essential features described above are generic. The inset shows the inconspicuous ripple at  $p \approx 0$ .

The results presented in Figure 2 correspond to a relatively small network, and with large aspect ratio; however, such a restrictive choice is not required, the ripples being preserved for networks of increasing size and for smaller  $R$ . In each case, the number of ripples is increased due to the longer path lengths involved.

The ripples in  $\bar{\mathcal{P}}(p, T = \text{constant})$  are a novel feature. The variation of  $\bar{\mathcal{P}}$  with network variability is more subtle than the correspondence between maximum inefficiency and the lattice percolation threshold, highlighted in [30] in the thermodynamic limit of large network size and for  $R = 1$ . Moreover, [30] studied variations of network variability carrying unit traffic ( $T = 1$ ), which we will show is less generic than one might initially suppose.

Figure 3 shows the price of anarchy as a function of  $T$  for fixed  $p$ . Figure 3(a) shows ensemble results  $\bar{\mathcal{P}}$  corresponding to a lattice with  $I = 80$  and  $L = 6$ , for two different values of  $p$ . Here,  $\bar{\mathcal{P}}$  attains a principal peak before saturating to unity. However, for  $p = 0.2$  there is additional structure that recalls the ripples seen in Figure 2. Figure 3(b) shows  $\mathcal{P}(p = 0.2, T)$  for a single realisation of the network. While peaks are evident in  $\bar{\mathcal{P}}(p = \text{constant}, T)$  the sharply-defined structures shown in Figure 3(b) have been smoothed due to variability occurring in each realisation, with only the dominant final peak remaining consistently.

Figure 4 examines the full landscape of  $\bar{\mathcal{P}}(p, T)$  for different choices of network size and

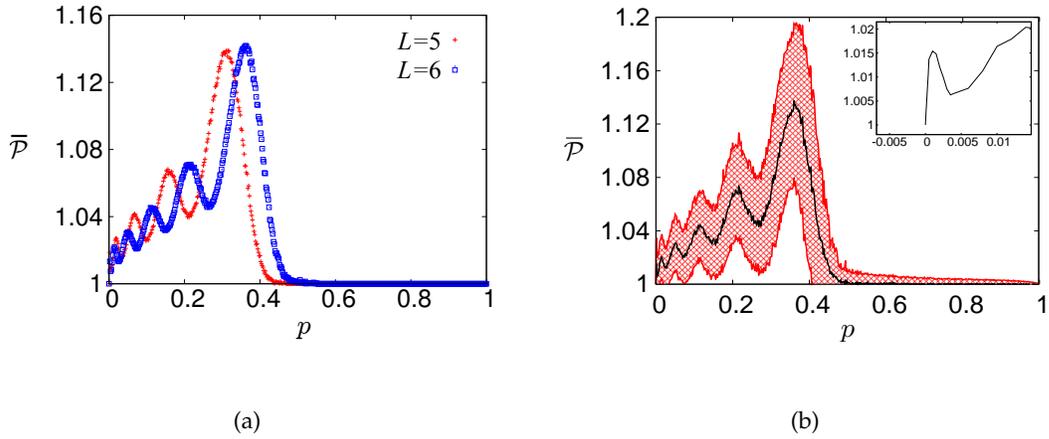


FIG. 2: The variation of  $\bar{\mathcal{P}}$  with respect to  $p$  (for  $T = 1$ ) for a lattice with dimensions (a)  $I = 80, L = 5, 6$ ; (b)  $I = 80, L = 6$ . In (b) the shaded area indicates an envelope of one standard deviation, while the inset displays the first peak which occurs at very small  $p$ .

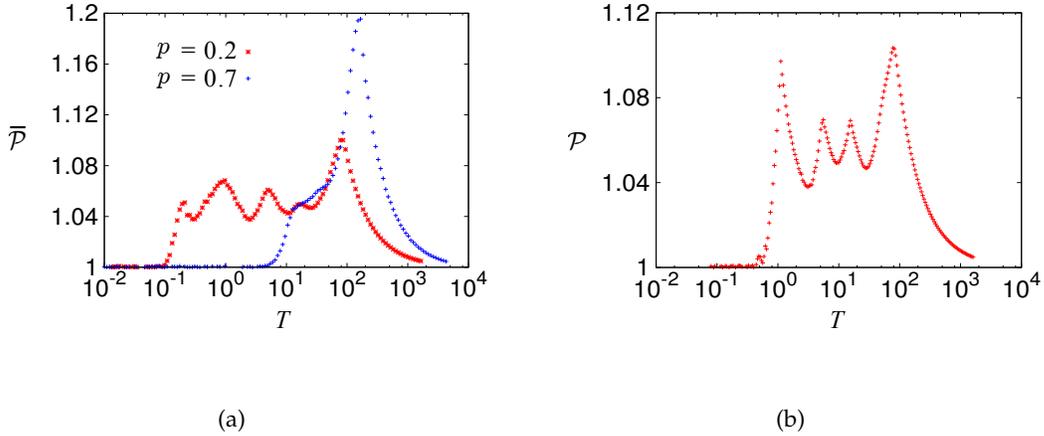


FIG. 3: The variation of  $\mathcal{P}$  with  $T$ . (a)  $\bar{\mathcal{P}}(p = \text{constant}, T)$  for  $I = 80, L = 6$  and different network variability:  $p = 0.2, 0.7$ . (b)  $\mathcal{P}(p = 0.2, T)$  for a single realisation of the network with  $I = 80, L = 6$ .

aspect ratio, with  $0 \leq p \leq 1$  and  $T$  ranging over 4 decades. The top row shows results for square networks, while the bottom row indicates those for  $R > 1$ ; network size increases 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  $R = 1$  it occurs at  $T \sim 10$ , whereas for  $R > 1$  its location moves to  $T \sim 100$ ). The second more complex feature is more strongly dependent on network characteristics, in particular the aspect ratio. For a square network it has a localised and globular morphology concentrated around

$p = 0.6$ ,  $T = 0.1$ . For  $R > 1$  we observe striations through the  $(p, T)$ -landscape. The results of Figures 4(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 aspect ratio  $R < 1$ , 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.

Figures 2 and 3, which were for networks of moderate size, reveal a rich structure in  $\mathcal{P}$  that is directly linked to the network path length and aspect ratio. Figure 4 highlights that these features are replicated in larger networks. Therefore the finite-size effect is not restricted to artificially small networks but applies more widely to larger structures encountered in real networks of finite size.

## 4 Network analysis

### 4.1 Network simplification: a collection of non-overlapping paths

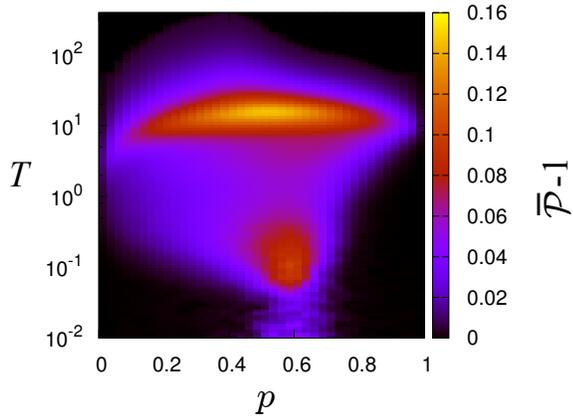
This section provides a fundamental understanding of the phenomenology shown in Section 3. This is achieved by distilling the lattice network to a simpler and more essential entity that comprises a set of  $I$  non-overlapping paths of length  $L$ , each of which connect a source node to a destination node via  $L$  edges. The other assumptions about the network and the rules for the traffic flow are unchanged. A key consequence of this simplification is that the traffic borne by each edge in a given path is identical. Moreover, under both UE and SO all paths with the same number of variable edges (greater than 0) carry equal traffic. This is evident for the UE case: if two paths with the same number of variable edges carried unequal traffic, one path would cost more to traverse than the other and the traffic flow would not satisfy the Nash equilibrium (see §2). It must also be true for the SO case because the cost of variable edges is an increasing function of traffic carried, hence the total cost of the flow can only be minimised if the traffic is uniformly distributed within ‘classes’ of paths which have the same number of variable edges.

In this simplified representation, the UE and SO on a given network realisation can be characterised by the amount of traffic allocated to paths (or ‘classes’ of paths) depending on the number of variable edges they comprise. Denote by  $\tau_k$  the quantity of traffic routed to each path which comprises  $k$  variable edges and  $(L - k)$  fixed edges (referred to as ‘ $k$ -paths’), and denote by  $N_k$  the number of such paths in the network. For a given realisation, denote by  $k^* \leq L$  the number of variable edges within the path(s) with the greatest number of variable edges. The user cost incurred by traversing a  $k$ -path is, from equations (2) and (3),

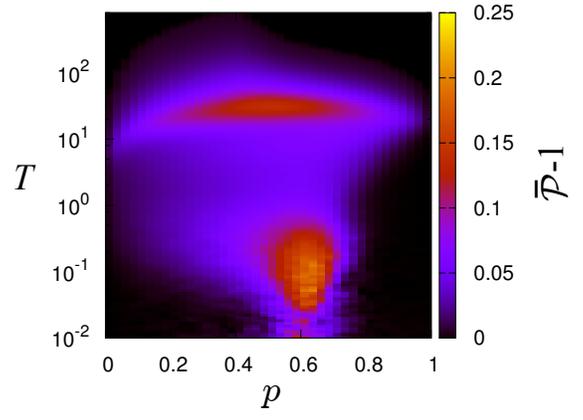
$$c_k(\tau_k) = k\tau_k + (L - k), \quad (4)$$

and the total cost incurred by traffic using that path is

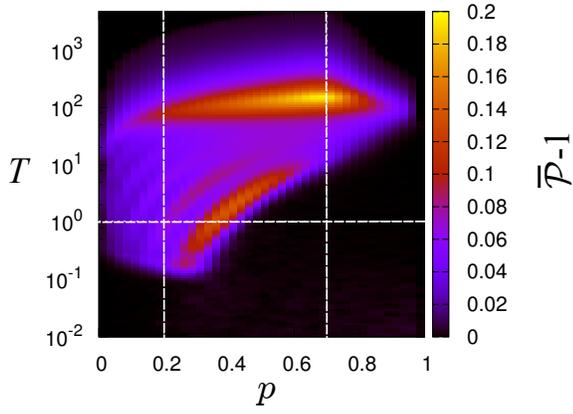
$$C_k(\tau_k) = \tau_k c_k(\tau_k) = k\tau_k^2 + (L - k)\tau_k. \quad (5)$$



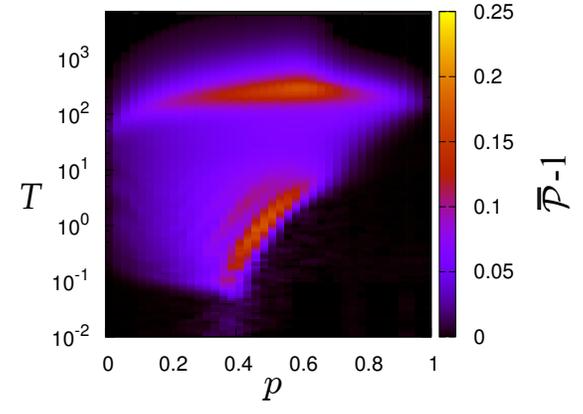
(a)



(b)



(c)



(d)

FIG. 4: The  $\bar{P}$  landscape for different choices of aspect ratio and network size.  $(\bar{P} - 1)$  is presented here to enhance contrast. Top row,  $R = 1$ ; bottom row,  $R > 1$ . (a)  $I = 10, L = 20$ ; (b)  $I = 20, L = 40$ ; (c)  $I = 80, L = 6$ ; (d)  $I = 160, L = 12$ . Dashed white lines in (c) highlight parameter choices corresponding to results presented in Figures 2 and 3.

Since all paths in a class carry equal traffic, the total cost of a UE or SO flow is given by

$$C = \sum_{k=0}^{k^*} C_k(\tau_k) N_k. \quad (6)$$

The network variability  $p$  dictates the set of  $N_k, k = 0 \dots k^*$ , with  $N_k$  in general increasing with  $p$ . Note that  $T$  and  $p$  have opposite effects on the total cost  $C$ . The total cost increases with  $T$  because the total cost is an aggregate cost of all the traffic, hence an increase in the traffic load results in a greater total cost. Moreover, increasing the traffic load results in greater congestion on the variable edges and therefore greater user-travel costs. By contrast, the total cost decreases with respect to  $p$  because variable edges are cheaper to traverse than fixed edges (assuming that the traffic demand in the system is not too large and that the traffic load on variable edges is  $\leq 1$ ), hence an increase in the number of variable edges results in a decrease to the total aggregate cost. Both UE and SO flows depend on which  $k$ -paths carry traffic, which is determined by the relationship between  $T$  and  $N_k$ . From this relationship we infer thresholds in the traffic load at which unused  $k$ -paths become utilised. These thresholds, themselves depending on  $N_k$ , differ for the UE and SO flows and so affect  $\mathcal{P}$ . The following sections establish these thresholds and discuss their consequences. A rigorous derivation of the thresholds and associated bounds is given in Appendix B.

## 4.2 $\mathcal{P}(p = \text{constant}, T)$

We examine here the landscape of  $\mathcal{P}(p, T)$  when  $T$  varies with a fixed value of  $p$  by analysing how an increase in the total traffic from  $T = 0$  influences its routing at each equilibrium.

Consider first the UE, at which the network users choose the path with minimum cost. This means that all traffic-carrying paths incur equal cost irrespective of the number of variable edges they contain. This cost is less than or equal to the cost of any unused path. As  $T$  increases from zero, the paths comprising entirely fixed edges are the most expensive to traverse, with cost  $L$ , and the *least* expensive paths to use are the  $k^*$ -paths, with minimum cost  $L - k^*$ . Consequently, the  $k^*$ -paths will be utilised first, with all others remaining unused. This will continue until the traffic increases to such an extent that any additional routing along the  $k^*$ -paths would become more costly than routing onto paths containing fewer variable edges but carrying no traffic. There is evidently a critical value of the traffic routed along a  $j$ -path at which the cost is identical to that of an unused  $i$ -path ( $i < j$ ). Denoting this critical value to be  $\tau_{ji}^{\text{ue}}$  and appealing to (4) this occurs when  $j\tau_{ji}^{\text{ue}} + L - j = L - i$ , that is:

$$\tau_{ji}^{\text{ue}} = \frac{j-i}{j}. \quad (7)$$

When traffic carried by the  $j$ -paths attains this critical value, any further increase in  $T$  will result in some traffic being routed to  $i$ -paths. Since there are  $N_j$  such  $j$ -paths and if  $i$ -paths are unused then all  $k$ -paths with  $k < i$  incur an even greater cost and are also unused, the critical traffic load under UE at which traffic begins to use  $i$ -paths is given by

$$\mathcal{T}_i^{\text{ue}} = \sum_{j=i+1}^{k^*} \tau_{ji}^{\text{ue}} N_j. \quad (8)$$

This enumerates how much traffic is carried on  $k$ -paths, for  $k = i + 1, \dots, k^*$ , at the point when it becomes beneficial for network users to employ  $i$ -paths which carry no traffic. Provided that the total network traffic is such that  $\mathcal{T}_{i+1} < T \leq \mathcal{T}_i$ , all the  $k$ -paths in the range  $i + 1 \leq k \leq k^*$  carry traffic whereas those higher-cost paths with fewer variable edges are unused. Note that  $\mathcal{T}_i^{\text{ue}} \leq \mathcal{T}_{i-1}^{\text{ue}}$ , since paths with more variable edges are always used preferentially.

Now consider the SO traffic flow, which minimises the *total* cost  $C$  given by (19) and which differs in general from the total cost at the UE—in particular,  $C^{\text{ue}}$  can exceed  $C^{\text{so}}$  (so that  $\mathcal{P} > 1$ ) due to paths with a high number of variable edges having too much traffic routed to them. The SO traffic flow must undergo a similar series of transitions with varying  $T$ . For sufficiently small  $T$  all traffic is routed to the  $k^*$ -paths; further increases in  $T$  then lead to some traffic being routed sequentially, beginning with  $(k^* - 1)$ -paths, to previously unused paths with fewer variable edges. Thresholds at which these transitions occur can therefore be constructed analogously to those in the case of UE, by pairwise comparison between the cost of traffic-carrying paths and the unused paths with the most variable edges. To determine these, we provide the following heuristic argument for the SO equilibrium condition and attendant critical traffic loads, a formal justification for which is given in Appendix B.

Given a SO flow such that  $\tau_j > 0$ , consider rerouting an amount of traffic  $N_j \delta\tau > 0$  from the set of  $j$ -paths to the set of  $i$ -paths. Because the flow before the rerouting was SO, the total cost of the flow cannot be decreased by the rerouting, so

$$N_i C_i(\tau_i) + N_j C_j(\tau_j) \leq N_i C_i \left( \tau_i + \frac{N_j}{N_i} \delta\tau \right) + N_j C_j(\tau_j - \delta\tau) \quad (9)$$

where the factor  $N_j/N_i$  reflects the possible disparity between the number of paths in each class. Since this is true for any  $\delta\tau > 0$ , this provides  $C_j'(\tau_j) \leq C_i'(\tau_i)$ . Equality occurs if the  $i$ - and  $j$ -paths both carry traffic. However if  $\tau_i = 0$ , the critical value at which it becomes cost effective to route additional traffic to both  $i$ - and  $j$ -paths, denoted by  $\tau_{ji}^{\text{so}}$ , occurs when  $C_j'(\tau_j) = C_i'(\tau_i = 0)$ , and from (5) this is

$$\tau_{ji}^{\text{so}} = \frac{j-i}{2j}. \quad (10)$$

Hence the critical thresholds for the SO are given by

$$\mathcal{T}_i^{\text{so}} = \sum_{j=i+1}^{k^*} \tau_{ji}^{\text{so}} N_j. \quad (11)$$

Note that  $\mathcal{T}_i^{\text{so}} = \mathcal{T}_i^{\text{ue}}/2$ , meaning that the traffic load for which traffic begins to use  $i$ -paths under SO is precisely half that for UE. Conditions similar to thresholds (8) and (11) were obtained in [14] for a two-node network with parallel edges. These thresholds therefore extend and generalise the treatment of [14] to a random network comprising an ensemble of heterogeneous non-overlapping paths.

	SO traffic flow	UE traffic flow	$\mathcal{P}$
$\frac{1}{2}\mathcal{T}_i^{\text{ue}} < T \leq \mathcal{T}_i^{\text{ue}}$	$\tau_i > 0.$	$\tau_i = 0.$	Increasing
	$C^{\text{ue}}$ increases with $T$ at a higher rate than $C^{\text{so}}$ .		
$T = \mathcal{T}_i^{\text{ue}}$	$\tau_i > 0.$	$\tau_i = 0.$ $c_{i+1}(\tau_{i+1}) = c_i(\tau_i = 0).$	Maximised
$\mathcal{T}_i^{\text{ue}} < T \leq \frac{1}{2}\mathcal{T}_{i-1}^{\text{ue}}$	$\begin{cases} \tau_i > 0; \\ \tau_{i-1} = 0. \end{cases}$	$\begin{cases} \tau_i > 0; \\ \tau_{i-1} = 0. \end{cases}$	Decreasing
	$C^{\text{ue}}$ increases with $T$ at the same rate as $C^{\text{so}}$ .		

Table 1: Description of the variation of  $\mathcal{P}$  with  $T$ .

These critical thresholds can be employed to explain the behaviour of  $\mathcal{P}$  observed in Figure 3 as follows. At the SO, for  $\mathcal{T}_i^{\text{ue}}/2 < T \leq \mathcal{T}_i^{\text{ue}}$ , only paths containing  $i$  or more variable edges carry traffic, whereas at UE only paths containing  $(i + 1)$  or more variable edges carry traffic. Given that  $i$ -paths are utilised at SO but not UE,  $C^{\text{ue}}$  increases with  $T$  at a higher rate than does  $C^{\text{so}}$ , and so  $\mathcal{P}$  increases.  $\mathcal{P}$  continues to increase until  $T$  reaches the critical threshold  $\mathcal{T}_i^{\text{ue}}$  at which point traffic is routed to  $i$ -paths at UE,  $C^{\text{ue}}$  increases with  $T$  at the same rate as  $C^{\text{so}}$  (see Appendix C) and hence  $\mathcal{P}$  decreases. The cycle begins to repeat itself when  $T$  exceeds the threshold  $\mathcal{T}_{i-1}^{\text{ue}}/2$ , at which point  $(i - 1)$ -paths carry traffic at SO but not UE, and  $\mathcal{P}$  increases with  $T$ . In this way a series of  $k^*$  local maxima are created in  $\mathcal{P}(p = \text{constant}, T)$ . This behaviour is caricatured in Figure 5 and summarised in Table 1, highlighting in particular the key influence on  $\mathcal{P}$  of the difference between the thresholds  $\mathcal{T}_i^{\text{ue}}$  and  $\mathcal{T}_i^{\text{so}}$  at which the flows in the network exhibit significant transition.

The above indicates that the variation in  $\mathcal{P}$  for fixed  $p$  can be understood entirely in terms of  $C^{\text{ue}}$ . Recall that for the UE all traffic-carrying paths incur the same cost. Denoting this cost by  $c$ , the total network cost arising from an imposed traffic load  $T$  is  $C^{\text{ue}} = Tc$ . At  $T = \mathcal{T}_j^{\text{ue}}$ ,  $k$ -paths with  $k > j$  carry traffic, the remainder being unused. Moreover the cost of traffic-carrying paths is equal to that of the unused  $j$ -paths so that  $c = L - j$ . Consequently  $\mathcal{P}$  will be maximised when

$$\frac{C^{\text{ue}}}{T} = L - j, \quad (12)$$

*i.e.*, when  $C^{\text{ue}}/T$  is an integer, with unit increments corresponding to sequential allocation to paths with increasingly many fixed edges.

Recall that Figure 3a displays  $\bar{\mathcal{P}}(p = \text{constant}, T)$ , which is an average over multiple network realisations for a given  $p$ . The peaks of  $\bar{\mathcal{P}}(p = 0.2, T)$  are less acute than those exhibited by  $\mathcal{P}$  in Figure 3b for a corresponding single realisation. This is because the corresponding thresholds  $\mathcal{T}_i^{\text{ue}}$  occur at different values for different network realisations, while remaining sufficiently grouped to lead to  $\bar{\mathcal{P}}(p = 0.2, T)$  displaying smoothed peaks.

For  $\bar{\mathcal{P}}(p = 0.7, T)$ , the network consists mostly of variable edges, most of which belong to at least one  $k^*$ -path. Therefore the network essentially comprises only  $k^*$ -paths and

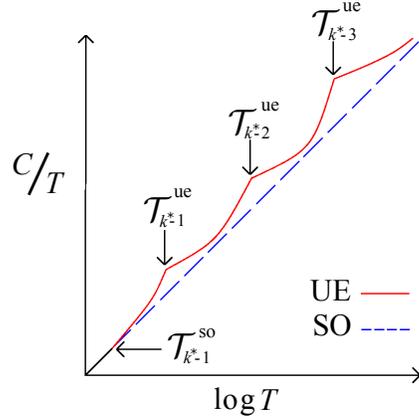


FIG. 5: (Colour online.) A schematic diagram showing the variation of the total costs (per unit traffic) at UE (red/solid line) and SO (blue/dashed line) with the traffic load  $T$ . The cusps of the UE curve correspond to the critical thresholds  $\mathcal{T}_i^{\text{ue}}$  where  $\mathcal{P}$  exhibits local maxima.

0-paths. Hence there is only one  $\mathcal{T}_i^{\text{ue}}$  which takes approximately the same value for all realisations, leading to a single, well-defined peak.

### 4.3 $\mathcal{P}(p, T = \text{constant})$

Equations (8) and (11) highlight the intimate relationship between  $T$  and  $p$  in determining the cost of traffic flows in the network, and hence the variations in  $\mathcal{P}$ . The behaviour of  $\mathcal{P}(p, T)$  for a fixed value of  $T$  may therefore be understood by exploiting the insight obtained from the previous section, with the focus now on  $k$  and  $N_k$  in place of  $T$ .

The primary critical threshold of  $T$  for the UE for a given realisation of the network is given by equations (7) and (8) as

$$\mathcal{T}_{k^*-1}^{\text{ue}} = \frac{N_{k^*}}{k^*}.$$

For a given  $k^*$ , the value of  $N_{k^*}$  increases with  $p$  as more  $k^*$ -paths are formed. Hence the critical threshold for the UE at fixed  $T$  occurs at the value of  $p$  for which  $N_{k^*}$  and  $k^*$  satisfy

$$N_{k^*}^{\text{ue}} = Tk^*. \quad (13)$$

Similarly the critical threshold at SO is given by

$$N_{k^*}^{\text{so}} = 2Tk^*. \quad (14)$$

Since traffic  $T$  and network variability  $p$  have opposite effects on the total cost of a traffic flow we can infer the behaviour of  $\mathcal{P}$  as  $N_{k^*}$  varies around these critical thresholds. Taking

	SO traffic flow	UE traffic flow	$\mathcal{P}$
$N_{k^*} < Tk^*$	$\begin{cases} \tau_i > 0, & \text{for some } i < k^*; \\ \tau_{k^*} > 0. \end{cases}$ $C^{\text{ue}}$ decreases with $p$ at a lower rate than $C^{\text{so}}$ .		Increasing
$N_{k^*} = Tk^*$	$\begin{cases} \tau_i > 0, & \text{for some } i < k^*; \\ \tau_{k^*} > 0. \end{cases}$	$\begin{cases} \tau_i = 0, & \text{for all } i < k^*; \\ \tau_{k^*} > 0. \end{cases}$	Maximised
$Tk^* \leq N_{k^*} < 2Tk^*$	$\begin{cases} \tau_i > 0, & \text{for some } i < k^*; \\ \tau_{k^*} > 0. \end{cases}$		Decreasing
	$C^{\text{ue}}$ decreases with $p$ at a higher rate than $C^{\text{so}}$ .		
$N_{k^*} \geq 2Tk^*$	$\begin{cases} \tau_i = 0, & \text{for all } i < k^*; \\ \tau_{k^*} > 0. \end{cases}$		Equals unity

Table 2: Description of the variation of  $\mathcal{P}$  with  $p$ .

$i = k^* - 1$  in Table 1, it can be seen that  $\mathcal{P}$  is increasing with  $T$  for  $N_{k^*}/2k^* < T \leq N_{k^*}/k^*$ , and therefore  $\mathcal{P}$  is decreasing with  $p$  for  $Tk^* \leq N_{k^*} < 2Tk^*$ . Similarly, since  $\mathcal{P}$  is decreasing with  $T$  for  $N_{k^*}/k^* < T$ , it follows that  $\mathcal{P}$  is increasing with  $p$  for  $N_{k^*} < Tk^*$ . For  $N_{k^*} \geq 2Tk^*$ , traffic is routed only to  $k^*$ -paths at both the UE and SO, and  $\mathcal{P} = 1$ . However, Figure 2 shows that  $N_{k^*}$  reaches  $N_{k^*}^{\text{so}}$  only for  $k^* = L$ . For  $k^* < L$ , a path with more variable edges than all other paths is formed before  $N_{k^*}$  can reach  $N_{k^*}^{\text{so}}$ , at which point  $k^*$  increments by one,  $N_{k^*} = 1$  and the cycle repeats itself. This process, summarised in Table 2, produces a series of  $L$  local maxima in  $\mathcal{P}$  before falling to unity.

The above considerations apply to an individual network realisation. When considering an ensemble of statistically similar networks of the same size, aspect ratio and traffic load, maxima in  $\mathcal{P}$  occur according to (13) for each realisation. Variation in each network implies that maxima in  $\bar{\mathcal{P}}$  occur at values of  $p$  for which  $\langle |N_{k^*} - Tk^*|/k^* \rangle$  has a local minimum.

#### 4.4 Heterogeneous network

The results of the preceding sections provide exact locations of the peaks in  $\mathcal{P}$  under variation of  $T$  and  $p$  in a simple network comprising a set of non-overlapping paths. This has provided insight into the cause of the maximisation of  $\mathcal{P}$  – namely, at thresholds in which a given class of paths carries traffic under UE – which holds for more general networks. This section investigates the applicability of the results of Sections 4.2 and 4.3 to the lattice network of Section 3. While the lattice network does not have a complex structure, the paths across the network overlap one another thereby making the traffic flows analytically intractable.

Translation of the results of Sections 4.2 and 4.3 to networks with overlapping traffic flows is not trivial. With regards to  $\bar{\mathcal{P}}(p, T = \text{constant})$ , we know from the case of non-overlapping paths that  $\mathcal{P}$  is maximised by condition (13). However, this property fails to hold for the regular-lattice network if  $N_{k^*}$  is calculated simply by counting the number of  $k^*$ -paths. This is because paths across the lattice can overlap one another by containing one or more of the same edges, in which case the overall variability of these two paths is between that of a single  $k^*$ -path and two non-overlapping  $k^*$ -paths, *i.e.*  $1 < N_{k^*} < 2$ . Therefore peak locations given by (13) hold for more general networks only if we assign a more sophisticated definition to  $N_{k^*}$ , namely that it is the *effective variability* of the  $k^*$ -paths, rather than the number of such paths. To obtain the value of  $N_{k^*}$  under this definition, for a given realisation of the network, we implement a computational procedure given in Appendix D. Nevertheless, in the case for  $\mathcal{P}(p = \text{constant}, T)$ , the insight of (12) enables the location of the peak to be understood entirely in terms of the UE and traffic  $T$ .

Figures 6 and 7 show the correspondence between the peak locations given by equations (12) and (13) and those observed in numerical experiments both in simple networks of non-overlapping paths and in lattice networks. Figure 6 shows that in each case the local minima in  $\langle |N_{k^*} - k^*| / k^* \rangle$  align with local maxima in  $\bar{\mathcal{P}}(p, T = 1)$ . Figure 7 shows that the peaks in  $\mathcal{P}$  do indeed coincide with the locations given by equation (12).

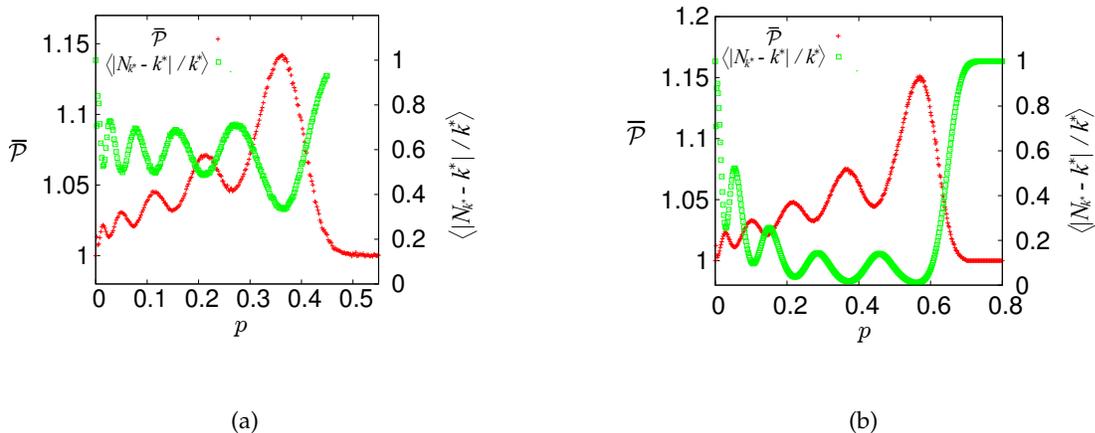


FIG. 6: The correspondence between the peaks in  $\bar{\mathcal{P}}(p, T = 1)$  and (13) in (a) a lattice network with  $I = 80$  and  $L = 6$ ; (b) a network of non-overlapping paths with  $I = 160$  and  $L = 6$ .

Figure 8 displays the peak locations of  $\bar{\mathcal{P}}(p, T = \text{constant})$  and those given by equation (13) for lattice networks varying in size and with the same aspect ratio  $R = 80/3$ . This is considered for  $T = 0.2$ ,  $T = 1$  and  $T = 5$ . The locations given by equation (13) correspond to the actual peaks of  $\bar{\mathcal{P}}(p, T = \text{constant})$  with good accuracy regardless of network size or value of  $T$ . It can also be observed that peak locations increase both with network size and with  $T$ , and tend to the lattice percolation threshold only as  $T$  becomes small.

The insight obtained from the above analysis may now be employed to explain the quasi-invariant structure of  $\bar{\mathcal{P}}(p, T)$  observed in Figure 4. This corresponds to the critical threshold  $\mathcal{T}_0^{\text{ue}}$ , the largest traffic load for which traffic is routed to 0-paths for the SO but

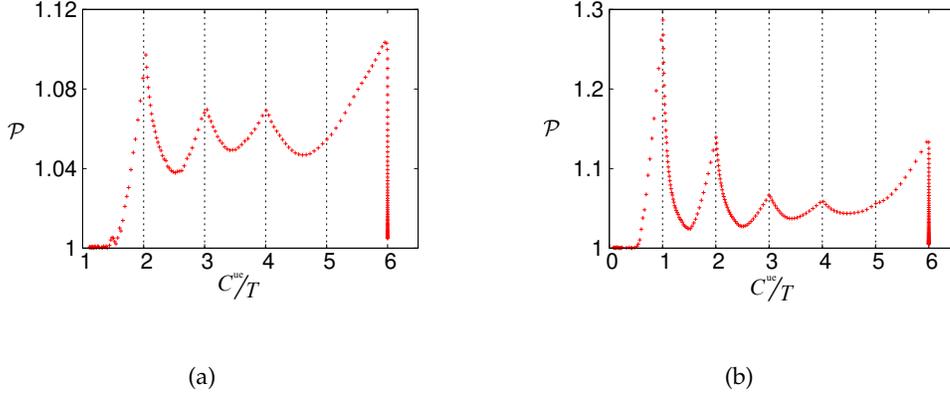


FIG. 7: The correspondence of peaks in  $\mathcal{P}(p = \text{constant}, T)$  and (12) in (a) a lattice network with  $I = 80, L = 6, p = 0.2$ ; (b) a network of non-overlapping paths with  $I = 160, L = 6, p = 0.5$ .

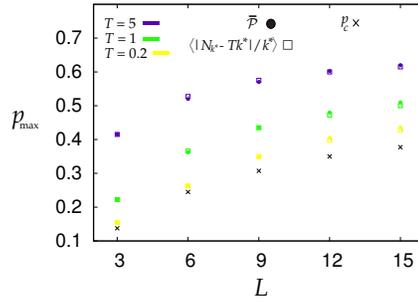


FIG. 8: (Colour online.) The peak locations  $p_{\max}$  of  $\bar{\mathcal{P}}(p, T = \text{constant})$  (circles) in a lattice network and those given by equation (13) (squares) for  $T = 0.2$  (yellow/light gray),  $T = 1$  (green/medium gray) and  $T = 5$  (purple/dark gray) over a range of network sizes. The lattices all have the same aspect ratio  $R = 80/3$ . The percolation threshold  $p_c$  (black crosses) is estimated as the value of  $p$  in which precisely half of the network realisations contain an  $L$ -path.

not for the UE. With  $i = 0$  in equations (7) and (8) this is

$$\mathcal{T}_0^{\text{ue}} = \sum_{j=1}^{k^*} N_j, \quad (15)$$

which is the total number of paths that contain at least one variable edge. The quasi-invariant structure of  $\overline{\mathcal{P}}(p, T = \mathcal{T}_0^{\text{ue}})$  therefore shows that the total number of paths that contain at least one variable edge is approximately constant over a wide range of  $p$ . From (15) we infer that changes in  $p$  within this region alter the distribution of  $N_j$  without altering the total sum  $\sum_{j=1}^{k^*} N_j$ .

## 5 Summary

This paper has presented an analysis of how the efficiency for routing traffic on a network is influenced by the degree of network heterogeneity and the traffic load. The lattice network comprises both fixed- and variable-cost edges with nodes that are arranged in a columnar array such that their connecting edges form tilted square cells. Employing the ‘price of anarchy’  $\mathcal{P}$  as a metric, the system efficiency was analysed as a function of traffic  $T$  and network variability characterised by the proportion of variable edges  $p$ , considering both individual networks and ensemble averages over a large number of realisations.

The numerical results serve to highlight a clear finite size effect in the structure of  $\mathcal{P}$ , that is induced by the network structure and not observed in studies focussing on the thermodynamic limit of large networks. Specifically, the peaked structures observed in [2, 5, 14] for  $\mathcal{P}(p = \text{const.}, T)$  also occur for  $\mathcal{P}(p, T = \text{const.})$  in networks with aspect ratio different from unity. The emergence of such structures persists in finite networks of all sizes, including large networks which are of relevance to the analysis of real-world phenomena.

The full landscape of  $\mathcal{P}$  in  $(p, T)$ -space reveals a rich structure. This includes a peak in  $\mathcal{P}$  that occurs for large traffic volumes and is quasi-invariant with network variability  $p$ . Here, the network inefficiency cannot be manipulated by changes to the network variability but only by changing the traffic volume. A complex feature at lower traffic volumes is more strongly dependent on network characteristics, taking either ‘globular’ or ‘striated’ form depending on the lattice aspect ratio. This is associated with the ripples in  $\mathcal{P}(p, T = \text{const.})$  that we have highlighted.

In addition to these numerical experiments, we have provided an explanation for the development of the peaks in  $(p, T)$ -space. By considering traffic flows that take place within a simplified network of non-overlapping paths, we obtain exact formulæ (8) and (13) for the peak locations in  $(p, T)$ -space. Under suitable re-definition of the terms, we showed that these results hold for a network with complex traffic flows: a simple computation provides, with good accuracy, equivalent peak locations that correspond to the random lattice network.

This work therefore provides an improved understanding of the routing efficiency in random networks. In particular, our semi-analytical characterisation of the dependence of peaks in  $\mathcal{P}$  on network variability and traffic load provides a powerful tool for the analysis of general network structures.

There are a number of important extensions to this work that warrant investigation, and will form the focus for future work. These include investigations of the robustness of our analyses to alternative network structures, including perturbations to the ordered lattice considered here. Also the methodology allows more general random network topologies to be considered, and the effects of nonlinear cost functions.

## Acknowledgements

A.R. acknowledges the support of the UK Engineering and Physical Sciences Research Council (EPSRC) under Grant RS2244.

## References

1. C. Papadimitriou, *Proceedings of the 33rd Annual ACM Symposium on the Theory of Computing* (ACM, New York, 2001), pp. 749–753.
2. V. A. Knight and P. R. Harper, *European Journal of Operational Research* **230**, 122 (2013).
3. B. Skinner, *Journal of Quantitative Analysis in Sports* **6**, 3 (2010).
4. B. Skinner and B. Carlin, *Significance* **10**, 25 (2013).
5. H. Youn, M. T. Gastner, and H. Jeong, *Physical Review Letters* **101**, 128701 (2008).
6. X. M. Zhao and Z. Y. Gao, *Chinese Phys. Lett.* **24**, 283 (2007).
7. J. J. Wu, X. Guo, H. Sun, and B. Wang, *Mathematical Problems in Engineering* **2014**, 490483 (2014).
8. J. J. Wu, Z. Y. Gao, H. J. Sun, and H. J. Huang, *Europhysics letters* **74**, 560 (2006).
9. J. J. Wu, Z. Y. Gao, and H. J. Sun, *Physica A* **387**, 1025 (2008).
10. J. Cohen and F. Kelly, *J. Appl. Prob.* **27**, 730 (1990).
11. H. J. Sun, H. Zhang, and J. J. Wu, *Nonlinear Dyn.* **69**, 2097 (2012).
12. A. S. Schulz and N. S. Moses, *Proceedings of the 14th annual ACM-SIAM symposium on Discrete algorithms* (ACM, New York, 2003), pp. 86-87.
13. Z. H. Zhu, J. F. Zheng, Z. Y. Gao, and H. M. Du, *Physica A* **400**, 200 (2014).
14. S. J. O’Hare, PhD thesis, University of Leeds (2015).
15. S. J. O’Hare, R. D. Connors, and D. P. Watling, *Transportation Research Part B* **84**, 55 (2016).
16. E. Koutsoupias and C. Papadimitriou, *Comp. Sci. Rev.* **3**, 65 (2009).
17. T. Roughgarden and E. Tardos, *JACM* **49**, 236 (2002).
18. S. Sreenivasan, R. Cohen, E. Lopez, Z. Toroczkai, and H. E. Stanley, *Phys. Rev. E* **75**, 036105 (2007).
19. G. Yan, T. Zhou, B. Hu, Z. Q. Fu, and B. H. Wang, *Phys. Rev. E* **73**, 046108 (2006).

20. B. Danila, Y. Yu, J. A. Marsh, and K. E. Bassler, *Phys. Rev. E* **74**, 046106 (2006).
21. L. Zhao, Y. C. Lai, K. Park, and N. Ye, *Phys. Rev. E* **71**, 026125 (2005).
22. C. Nicolaidis, L. Cueto-Felgueroso and R. Juanes, *J R Soc Interface* **10** (2013).
23. A. C. Pigou, *The economics of welfare* (Macmillan, 1920).
24. G. Owen, *Game Theory, 3rd ed.* (Academic Press, 1995).
25. N. E. Stier-Moses, J. R. Correa and A. S. Schulz, *Games and Economic Behaviour* **64**, 457 (2008).
26. K. M. Sim and C. K. Chau, *Operations Research Letters* **31**, 327 (2003).
27. M. Haviv and T. Roughgarden, *Operations Research Letters* **35**, 421 (2007).
28. T. Roughgarden, *Journal of Computer and System Sciences* **67**, 341 (2003).
29. P. Erdős and A. Rényi, *Publicationes Mathematicae (Debrecen)* **6**, 290 (1959);  
D. J. Watts and S. H. Strogatz, *Nature* **393**, 440 (1998);  
A.-L. Barabási and R. Albert, *Science* **286**, 509 (1999).
30. B. Skinner, *Phys. Rev. E* **91**, 052126 (2015).
31. J. G. Wardrop, *Proceedings of the Institution of Civil Engineers* (Thomas Telford, London, 1952), Vol. 1, pp. 325-362.
32. A. L. Peressini, F. E. Sullivan, and J. J. Uhl, *The Mathematics of Nonlinear Programming* (Springer-Verlag, 1988).

## Appendix A: Computing the traffic flows

We used an algorithm from an earlier version of [30] to obtain the UE and SO traffic flows. Denote by  $\mathbb{E}$  the set of all edges of the network, and  $\mathbb{P}$  the set of all paths. For an edge  $i \in \mathbb{E}$ , we 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  $P \in \mathbb{P}$  is

$$c_P = \sum_{i \in P} \left( \chi_i \tau_i + (1 - \chi_i) \right),$$

where  $\tau_i$  is the traffic on edge  $i$ . The total cost of all traffic on path  $P$  is given by:

$$C_P = \sum_{i \in P} \left( \chi_i \tau_i^2 + (1 - \chi_i) \tau_i \right).$$

The algorithm for computing the traffic flows is as follows. Firstly, the traffic is uniformly distributed over all the paths. Then, an iterative procedure is 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 UE or SO is being computed. Recall that under UE all paths which carry traffic incur the same cost to

traverse. Therefore, to compute 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 SO the total cost of the traffic flow is minimised, so in this case  $\delta\tau$  is calculated such that  $(C_A + C_B)$  is minimised. (See below for detailed derivations of  $\delta\tau$ .) 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, the total cost of the flow is calculated by:

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

We note that this simple numerical method is not guaranteed to converge to the exact UE or SO flow. In order to minimise the effect of errors, we iterated over 50,000 pairs of paths per network realisation and took the average of  $\mathcal{P}$  over 2000 realisations for each choice of parameters  $(p, T)$ . These choices are based on a convergence criteria given in [30].

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

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

Solving for  $\delta\tau$ :

$$\delta\tau = \frac{\sum_{i \in B} \left( \chi_i \tau_i + (1 - \chi_i) \right) - \sum_{i \in A} \left( \chi_i \tau_i + (1 - \chi_i) \right)}{\sum_{i \in A} \chi_i + \sum_{i \in B} \chi_i}. \quad (16)$$

Note that  $\delta\tau$  may take a negative value here, which corresponds to 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 (16). Finally, to avoid oscillatory behaviour in the numerical computation, we scale  $\delta\tau^{\text{ue}}$  with a damping parameter  $\lambda = 0.5$ .

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

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

so

$$\frac{d}{d(\delta\tau)} (C_A + C_B) = 0. \quad (17)$$

Solving (17) for  $\delta\tau$ :

$$\delta\tau = \frac{\sum_{i \in B} (2\chi_i \tau_i + (1 - \chi_i)) - \sum_{i \in A} (2\chi_i \tau_i + (1 - \chi_i))}{2 \left( \sum_{i \in A} \chi_i + \sum_{i \in B} \chi_i \right)}. \quad (18)$$

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 (18), and we scale  $\delta\tau^{\text{so}}$  with a damping parameter  $\lambda = 0.5$ .

The results of this paper were produced using computer programs written in C++. With regards to the cost of computation: the results of Figure 2 and Figure 3a were each obtained from 20 hours of a Core i3 Central Processing Unit; Figure 3b pertains to a single realisation and was obtained from an hour of computation; the landscapes displayed in Figure 4 were each obtained from approximately half a week of computation.

## Appendix B: Conditions for $k$ -paths to carry traffic

The UE traffic flow adheres to Wardrop's first principle of traffic flow for which no single user can decrease their personal cost by changing to a different path [31]. We thereby define a UE traffic flow as follows.

### Definition 1

A traffic flow is UE if and only if for all  $i \leq k^*$ ,

$$\tau_i > 0 \implies c_i \leq c_j \text{ for all } j \leq k^*.$$

### Remark 1

Under UE,

$$\tau_i, \tau_j > 0 \implies c_i = c_j.$$

The SO traffic flow adheres to Wardrop's second principle of traffic flow in which the total cost of all users is minimised [31]. Given that the cost functions of all the edges in the network under consideration here are convex, finding an SO flow represents a problem of optimising the following convex program:

$$\text{minimise } \sum_{k=0}^{k^*} C_k(\tau_k) N_k \quad (19)$$

subject to:

$$\sum_{k=0}^{k^*} \tau_k N_k = T, \quad \tau_k \geq 0 \quad \text{for } k = 0, 1, \dots, k^*.$$

Consider a traffic flow in which some path  $P_1$  carries an amount of traffic  $\tau_{P_1}$ . A necessary condition for this traffic flow to be *locally* optimal is that for any path  $P_2$ , a transfer of any amount of traffic  $\epsilon$ , with  $0 < \epsilon \leq \tau_{P_1}$ , from  $P_1$  to  $P_2$  does not result in a decrease to the total cost of the traffic flow. In other words, for any transfer of traffic from  $P_1$  to  $P_2$ , the decrease in the total cost of traffic on  $P_1$  is not greater than the increase in the total cost of traffic on  $P_2$ . Hence the flow must satisfy  $C_{P_1}'(\tau_{P_1}) \leq C_{P_2}'(\tau_{P_2})$ , where  $C_P(\tau_P)$  denotes the total cost of traffic on path  $P$ . A necessary and sufficient condition for the traffic flow to be locally optimal is that for all  $P_1$  such that  $\tau_{P_1} > 0$ ,  $C_{P_1}'(\tau_{P_1}) \leq C_{P_2}'(\tau_{P_2})$  for all  $P_2$ . Since the local and global minima of a convex function on a convex set coincide [32], and since the objective function of (19) is convex, this condition is also necessary and sufficient for the flow to be *globally* optimal, i.e. SO. This can be generalised to classes of paths: a traffic flow is SO if and only if for any  $i$  such that  $\tau_i > 0$ ,  $C_i'(\tau_i) \leq C_j'(\tau_j)$  for all  $j$ , where  $C_k(\tau_k)$  is the total cost on a  $k$ -path. We thereby define a SO traffic flow as follows.

**Definition 2**

A traffic flow is SO if and only if for all  $i \leq k^*$ ,

$$\tau_i > 0 \implies C_i'(\tau_i) \leq C_j'(t_j) \quad \text{for all } j \leq k^*.$$

**Remark 2**

Under SO,

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

**Remark 3**

Under both UE and SO,

$$\tau_i = 0 \implies \tau_j = 0 \quad \text{for all } j \leq i.$$

Here we prove the following bounds associated with the critical thresholds (8) and (11) given in the main text:

For any  $i \leq k^*$  with  $N_i > 0$ ,

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

$$T \leq \sum_{j=i+1}^{k^*} \frac{j-i}{j} N_j. \tag{20}$$

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

$$T \leq \frac{1}{2} \sum_{j=i+1}^{k^*} \frac{j-i}{j} N_j. \quad (21)$$

**Proof**

(i) We first prove the “if” part. If  $\tau_k = 0$  for all  $k \leq i$ , then  $\tau_i = 0$  and  $c_i = L - i$ . For any  $j$  with  $i < j \leq L$ , if  $\tau_j = 0$  then  $c_j = L - j < L - i$ ; if  $\tau_j > 0$  then by Definition 1 it follows that  $c_j \leq c_i = L - i$ . Hence

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

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.

We now prove the “only if” part, by contraposition. Suppose there exists a  $k \leq i$  such that  $\tau_k > 0$ . Then by the contrapositive of Remark 3 it follows that  $\tau_j > 0$  for all  $j \geq k$ . In particular,  $\tau_i > 0$  and

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

Since  $\tau_j > 0$  for all  $j > i$ , by Remark 1 it follows that

$$\begin{aligned} c_j &= c_i \quad \text{for } j = i + 1, i + 2, \dots, k^*. \\ \implies \tau_j &\geq \frac{j-i}{j} \quad \text{for } j = i + 1, i + 2, \dots, k^*. \end{aligned}$$

So

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

as required.

(ii) Again, we first prove the “if” part. If  $\tau_k = 0$  for all  $k \leq i$ , then  $\tau_i = 0$  and  $C_i'(\tau_i = 0) = L - i$ . For any  $j$  with  $i < j \leq L$ , if  $\tau_j = 0$  then  $C_j'(\tau_j = 0) = L - j < L - i$ ; if  $\tau_j > 0$  then by Definition 2 it follows that  $C_j'(\tau_j) \leq C_i'(\tau_i = 0) = L - i$ . Hence

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

So

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

as required.

We now prove the “only if” part, by contraposition. Suppose there exists a  $k \leq i$  such that  $\tau_k > 0$ . Then by the contrapositive of Remark 3 it follows that  $\tau_j > 0$  for all  $j \geq k$ . In particular,  $\tau_i > 0$  and

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

Since  $\tau_j > 0$  for all  $j > i$ , by Remark 2 it follows that

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

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

So

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

as required.

## Appendix C: Properties of the total cost under UE and SO

For  $\mathcal{T}_i^{\text{ue}} < T \leq \frac{1}{2}\mathcal{T}_{i-1}^{\text{ue}}$ , under both UE and SO we have that  $\tau_i > 0$  and  $\tau_{i-1} = 0$ , hence

$$T = \sum_{j=i}^{k^*} \tau_j N_j.$$

Consider the UE case. Denote by  $c$  the cost of traversing paths which carry traffic, then

$$c_j = c \text{ for } j = i, i + 1, \dots, k^*$$

$$\implies \tau_j = \frac{c + j - L}{j} \text{ for } j = i, i + 1, \dots, k^*$$

$$\implies T = \sum_{j=i}^{k^*} \frac{c + j - L}{j} N_j,$$

which on solving for  $c$  gives

$$c = \frac{T - \sum_{j=i}^{k^*} N_j \left(1 - \frac{L}{j}\right)}{\sum_{j=i}^{k^*} \frac{N_j}{j}}.$$

Since  $C^{\text{ue}} = Tc$ , it follows that:

$$\frac{d}{dT}C^{\text{ue}} = \frac{2T - \sum_{j=i}^{k^*} N_j(1 - \frac{L}{j})}{\sum_{j=i}^{k^*} \frac{N_j}{j}}.$$

Now consider the SO case. Recall that for all  $k$  such that  $\tau_k > 0$ ,  $C_k'(\tau_k)$  have equal value, and denote this value by  $c_2$ , then

$$\begin{aligned} C_j'(\tau_j) &= c_2 \quad \text{for } j = i, i+1, \dots, k^* \\ \implies \tau_j &= \frac{c_2 + j - L}{2j} \quad \text{for } j = i, i+1, \dots, k^* \\ \implies T &= \sum_{j=i}^{k^*} \frac{c_2 + j - L}{2j} N_j, \end{aligned} \tag{22}$$

and solving for  $c_2$  gives

$$c_2 = \frac{2T - \sum_{j=i}^{k^*} N_j(1 - \frac{L}{j})}{\sum_{j=i}^{k^*} \frac{N_j}{j}} = \frac{d}{dT}C^{\text{ue}}. \tag{23}$$

Moreover,

$$C^{\text{so}}(T) = \sum_{k=i}^{k^*} C_k(\tau_k) N_k.$$

Hence

$$\begin{aligned} \frac{d}{dT}C^{\text{so}}(T) &= \sum_{k=i}^{k^*} N_k \frac{d}{dT}C_k(\tau_k) \\ &= \sum_{k=i}^{k^*} N_k C_k'(\tau_k) \frac{d\tau_k}{dT} \\ &= c_2 \sum_{k=i}^{k^*} N_k \frac{d\tau_k}{dT}. \end{aligned}$$

From equations (22) and (23),

$$\frac{d\tau_k}{dT} = \frac{1}{k \sum_{j=i}^{k^*} \frac{N_j}{j}}.$$

Hence

$$\frac{d}{dT}C^{\text{so}}(T) = \frac{c_2 \sum_{k=i}^{k^*} \frac{N_k}{k}}{\sum_{j=i}^{k^*} \frac{N_j}{j}} = c_2 = \frac{d}{dT}C^{\text{ue}}(T).$$

So for any  $T$  such that  $\mathcal{T}_i^{\text{ue}} < T \leq \frac{1}{2}\mathcal{T}_{i-1}^{\text{ue}}$  and  $C^{\text{ue}} > C^{\text{so}}$ , the price of anarchy  $\mathcal{P}$  is decreasing with  $T$ .

## Appendix D: Finding the value of $N_{k^*}$ for the diamond lattice

In the context of the diamond lattice, the definition of  $N_{k^*}$  is not simply the number of  $k^*$ -paths but the *effective variability* of the  $k^*$ -paths. Attaining the value of  $N_{k^*}$  for a given realisation is a non-trivial task because  $k^*$ -paths can overlap one another, in which case variable-cost edges are utilised by multiple  $k^*$ -paths. The greater the extent to which variable-cost edges are shared between  $k^*$ -paths, the greater the variability of the  $k^*$ -paths is diminished. Hence the task of calculating  $N_{k^*}$  cannot be achieved simply by counting the number of  $k^*$ -paths; the extent to which variability is diminished by paths overlapping must be accounted for.

Consider (20) which is applicable to the network of non-overlapping paths. Taking  $i = (k^* - 1)$  in (20) provides:

$$\text{Under UE, } \tau_k = 0 \quad \forall k \leq (k^* - 1) \iff T \leq \frac{N_{k^*}}{k^*}. \quad (24)$$

The above can be made applicable to the lattice network by making the following adjustment. Denote by  $\mathbb{P}_k$  the set of  $k$ -paths and denote by  $T_k$  the total traffic on all  $k$ -paths, given by:

$$T_k := \sum_{P \in \mathbb{P}_k} \tau^P,$$

where  $\tau^P$  is the traffic routed to path  $P$ . Then (24) can be restated as:

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

In particular:

$$T_{k^*-1} = 0 \iff T \leq \frac{N_{k^*}}{k^*}. \quad (25)$$

Denote by  $c^P$  the cost of traversing path  $P$  and let  $\hat{P} \in \mathbb{P}_{k^*-1}$  be a  $(k^* - 1)$ -path which does not overlap with any  $k^*$ -paths, then

$$c^{\hat{P}} \begin{cases} = L - (k^* - 1), & \text{if } T_{k^*-1} = 0; \\ > L - (k^* - 1), & \text{if } T_{k^*-1} > 0. \end{cases} \quad (26)$$

Recall that under 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, then, that:

$$c^{\hat{p}} \begin{cases} \geq c, & \text{if } T_{k^*-1} = 0; \\ = c, & \text{if } T_{k^*-1} > 0. \end{cases} \quad (27)$$

From (26) and (27):

$$c \begin{cases} \leq L - (k^* - 1), & \text{if } T_{k^*-1} = 0; \\ > L - (k^* - 1), & \text{if } T_{k^*-1} > 0. \end{cases} \quad (28)$$

From (25) and (28):

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

The total cost under UE is given by  $C^{\text{ue}} = Tc$ , so

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

This result allows for the use of a numerical method for computing  $N_{k^*}$ , which works as follows. First, the UE traffic flow is computed with  $T = 1/k^*$  so that (since  $N_{k^*} \geq 1$ )  $C^{\text{ue}}/T \leq L - (k^* - 1)$ . Then  $T$  is increased in incremental steps until  $C^{\text{ue}}/T = L - (k^* - 1)$  at which point it is known that  $T = N_{k^*}/k^*$  and  $N_{k^*}$  can be computed by  $N_{k^*} = Tk^*$ . Denote by  $\mathbb{E}$  the set of all the edges of the lattice and by  $\tau^i$  the traffic routed to edge  $i \in \mathbb{E}$ , then the numerical method for obtaining  $N_{k^*}$  is summarised as follows:

1. Compute the UE traffic flow with  $T = 1/k^*$ . If  $C^{\text{ue}}/T = L - (k^* - 1)$  then  $N_{k^*} = 1$ ; otherwise go to step 2.
2. For all  $i \in \mathbb{E}$ ,  $\tau^i \mapsto \tau^i + \tau^i \epsilon / T$ , for small  $\epsilon > 0$ .  $T \mapsto T + \epsilon$ .
3. If  $C^{\text{ue}}/T < L - (k^* - 1)$  go to step 2; otherwise  $N_{k^*} = Tk^*$ .

Note from Step 2 that only those edges which carry traffic before an incremental increase to  $T$  will carry traffic after the increase. This is true because the algorithm terminates when  $T$  reaches the critical threshold at which  $(k^* - 1)$ -paths begin to carry traffic, so only edges which belong to  $k^*$ -paths carry traffic throughout the procedure.