What are the best routes for us to use for driving home tonight in rush hour traffic?

Bogdan Danila,† Yudong Sun, and Kevin E. Bassler

1Department of Physics, University of Houston, Houston TX 77204-5005
2Texas Center for Superconductivity, University of Houston, Houston TX 77204-5002

We show that the capacity of a complex network that models a city street grid to support congested traffic can be optimized by using routes that collectively minimize the maximum ratio of betweenness to capacity in any link. Networks with a heterogeneous distribution of link capacities and with a heterogeneous transport load are considered. We find that overall traffic congestion and average travel times can be significantly reduced by a judicious use of slower, smaller capacity links.

PACS numbers: 89.75.Hc, 89.75.Da, 05.60.-k

Choosing the best route for driving home from work is a task that many of us face everyday. Knowing the best choice is also important to urban planners who design city transportation networks. To answer this, and other important questions, a vast amount of research has been devoted in recent years to the analysis and optimization of transport on complex networks [1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]. If we would collectively choose to use the best routes, then the traffic congestion and our commute times could be reduced. When traveling in a city, the best route to use is presumably the one that takes the least time to go from origin to destination. If there is little or no traffic, this route is simply the so-called shortest-path route. On the complex network that describes the city's streets, the shortest-path route is defined as the path for which the sum of the weights of each of the links along the path is minimal. The weight of a link can be defined as the time it takes to travel from one end of the link to the other when traffic is light. Generally, these weights are inversely related to the transport capacities of the links (streets that can handle a lot of traffic are traveled faster). On the other hand, when traffic is heavy some links may become congested, and then the quickest route to the destination may be a longer one involving smaller capacity links. It is not obvious, however, exactly at what level of congestion the use of alternative routes becomes advantageous, or how much the improvements in network transport capacity and average travel time amount to. In this paper we present methods to help answer these important questions and demonstrate their effectiveness within the framework of a simple model that captures the most important characteristics of urban street networks.

Specifically, we use the complex network model introduced by Barthélémy and Flammini in [18]. This model has been shown to produce planar geometric networks with an exponential distribution of the node distances from the origin and other characteristics similar to the actual street networks of various cities. We use this model to demonstrate the effectiveness of our methods, rather than a particular realistic example, in order to make a statistical analysis by taking ensemble averages over many example city networks. Unlike previous studies [2, 11, 12, 13, 14, 15], which assumed that traffic is limited by node congestion, here we assume that the capacity of the network is limited by the amount of traffic each link can support. This important and realistic difference requires a nontrivial variation in our methods. As we have shown previously [12, 14, 15], when transport is limited by node congestion, the transport capacity of the network is maximized by using a set of routes that minimize the maximum betweenness of the nodes. We also introduced an algorithm for finding such a set of routes, and demonstrated that, for a number of commonly studied network topologies, it finds routes for which the capacity, at least, scales optimally with system size. In this case, the transport capacity of the network is maximized by minimizing the maximum betweenness to capacity ratio of links. Here we also consider heterogeneous link capacities and uneven traffic demands between the various pairs of origin and destination nodes. In particular, we study the case of a rush hour traffic burst emanating from a central location. To obtain our results, we use a variant of our previous routing optimization algorithm. Furthermore, we prove a formula that allows the quick computation of the average of the sum along the path of any link-related quantities and use this formula to compute average travel times.

In our network transport dynamics, particles are assumed to travel along the network according to static routing rules. To demonstrate the importance of our results, we consider two different types of routing. One is the “natural” shortest path (SP) routing computed with link weights in inverse proportion to their respective capacities. The other one is the “optimal” routing (OR) for congested traffic resulting from the application of a routing optimization algorithm which finds the set of routes that maximizes the transport capacity of the network. Note that the OR is the collective optimum that occurs if everyone uses the best routes for collective results. As such, it describes an important limit of collective behavior. This contrasts with the goal of the multitude of traffic studies that seek to optimize results individually through a learning process within the framework of evo-
lutionary game theory [4].

Here we assume that each outgoing link of a node has a separate “first in, first out” (FIFO) particle queue. Transport on the network proceeds in discrete time steps and is driven by inserting new particles at the nodes. The average number of new particles inserted per time step at node \( i \) with destination at node \( j \) is \( r_{ij} \) and we denote \( r_i = \sum_j r_{ij} \). Each new particle is inserted at the end of the appropriate queue at its node of origin, namely the queue corresponding to the first link it has to traverse on its way to destination. The transport capacity of a link \((i, j)\) is \( C_{ij} \), defined as the average number of particles transported per time step assuming an infinite number of particles in its queue. We use a stochastic sequential updating of the positions of particles that erases the correlations between particle arrival times and ensures that both the arrival and the delivery of particles are Poisson processes [14]. The load of the network is defined as the average \( \langle r \rangle \) over all nodes on the network of \( r_i \). The network transport capacity is the critical value \( \langle r \rangle_c \) of the load [2] above which the network becomes jammed.

Routing on a network is characterized by the set of probabilities \( P_{ij}^{(t)} \) for a particle with destination \( t \) currently at node \( i \) to be forwarded to node \( j \). The betweenness \( b_{ij}^{(s,t)} \) of a node \( i \) with respect to a source node \( s \) and a destination node \( t \) is defined as the sum of the probabilities of all paths between \( s \) and \( t \) that pass through \( i \). Node betweenness can be computed in terms of the probabilities for complete networks \[ B_{ij} = \sum_{s,t=1}^{N} b_{ij}^{(s,t)} \rho_{st}. \] (5)

The average number of particles crossing link \((i, j)\) in the course of a time step is given by

\[ w_{ij} = \sum_{s,t=1}^{N} b_{ij}^{(s,t)} \rho_{st}. \] (2)

To account for an uneven traffic demand pattern between the nodes while still using a single parameter to characterize the load of the network it is convenient to write \( r_{ij} = C \langle r \rangle \rho_{ij} \), where \( \langle r \rangle \) is the average number of particles per node generated in the course of a time step and \( \rho_{ij} \) are nonnegative demand weights (\( \rho_{ii} = 0 \)). If the weights are normalized such that \( \sum_{i,j=1}^{N} \rho_{ij} = N(N-1) \) we find

\[ r_{ij} = \frac{\langle r \rangle}{N-1} \rho_{ij} \] (3)

and, using Eq. 2 the flow of particles through link \((i, j)\) becomes

\[ w_{ij} = \frac{\langle r \rangle}{N-1} B_{ij}. \] (4)

Avoiding congestion means \( w_{ij} \leq C_{ij} \) for every link \((i, j)\). Consequently, maximum transport capacity is achieved when the highest betweenness-to-capacity ratio on the network \((B/C)_{\text{max}}\) is minimized and is given by \( \langle r \rangle_c = (N-1)/(B/C)_{\text{max}} \). The minimization of \((B/C)_{\text{max}}\) can be achieved by various methods including, but not limited to, simulated annealing and extremal optimization.

The average travel time can be computed from link betweennesses assuming the arrival and the delivery of particles at every node are Poisson processes. We will now prove a general formula for calculating averages over the entire set of routes considered on a network and then apply this formula to calculate the travel time. Let \( Q_{ij} \) be any quantity associated with the links of a network. To calculate the average over all particle routes of the sum of \( Q_{ij} \) along the route it is convenient to write the betweenness in terms of the probabilities for complete routes. Let \( \pi_n(s,t) \) be the ordered set of all nodes along the \( n \)-th distinct route between \( s \) and \( t \) (including \( s \) but excluding \( t \)) and \( p_n(s,t) \) be the probability for a particle

\[ \sum \text{weights are normalized such that} \]

\[ \text{particles per node generated in the course of a time step} \]

\[ \text{write} \]

\[ \text{characterize the load of the network it is convenient to} \]

\[ \text{between the nodes while still using a single parameter to} \]

\[ \text{probabilities} P \]

\[ \text{going from} \]

\[ \text{The average number of particles crossing link} \]

\[ \text{becomes} \]

\[ \text{with the total weighted betweenness of link} \]

\[ \text{Avoiding congestion means} \]

\[ \text{Consequently, maximum transport capacity is} \]

\[ \text{minimum and is given by} \]

\[ \text{The average travel time can be computed from link} \]

\[ \text{be any quantity associated with the links of a network.} \]

\[ \text{To calculate the average over all particle routes of the sum} \]

\[ \text{Let} \]

\[ \text{be the ordered set of all nodes along the} \]

\[ \text{and} \]

\[ \text{be the probability for a particle} \]

\[ \text{FIG. 1: (Color online) Average travel time (in time steps)} \]

\[ \text{vs. network load (in particles per time step per node) for a} \]

\[ \text{typical network realization with} N = 200 \text{ nodes. Results are} \]

\[ \text{shown for shortest-path routing, SP (blue dashed lines), and} \]

\[ \text{for optimal routing, OR (red solid lines), at three different} \]

\[ \text{high capacity link densities, (a)} P_h = 0, \text{ (b) } P_h = 0.3, \text{ and (c) } P_h = 0.7. \]

\[ \text{with} \]

\[ \text{and} \]

\[ \text{by} \]

\[ \text{Avoiding congestion means} \]

\[ \text{Consequently, maximum transport capacity is} \]

\[ \text{minimum and is given by} \]

\[ \text{The minimization of} \]

\[ \text{can be achieved by various methods including, but not limited to, simulated annealing and extremal} \]

\[ \text{The average travel time can be computed from link} \]

\[ \text{be any quantity associated with the links of a network. To calculate the average over all particle routes of the sum} \]

\[ \text{Let} \]

\[ \text{be the ordered set of all nodes along the} \]

\[ \text{and} \]

\[ \text{be the probability for a particle} \]

\[ \text{FIG. 1: (Color online) Average travel time (in time steps)} \]

\[ \text{vs. network load (in particles per time step per node) for a} \]

\[ \text{typical network realization with} N = 200 \text{ nodes. Results are} \]

\[ \text{shown for shortest-path routing, SP (blue dashed lines), and} \]

\[ \text{for optimal routing, OR (red solid lines), at three different} \]

\[ \text{high capacity link densities, (a)} P_h = 0, \text{ (b) } P_h = 0.3, \text{ and (c) } P_h = 0.7. \]
The inner sum on the right-hand side of Eq. 8 is the sum over all routes between $s$ and $t$.

By substituting Eq. 6 into Eq. 7, we find the number of nodes at $P_h$ where

\[ \pi_n \text{ defined by } \]

where $B_{ij}$ are defined in Eq. 5 and we used the fact that the sum of all $\rho_{st}$ equals $N(N - 1)$. The derivation above parallels the one presented in [14], but the difference in the way one accounts for the links along a route in Eqs. 6 and 8 as opposed to accounting for nodes is non-trivial.

The average time, measured in steps, required for a particle to traverse link $(i, j)$ in the absence of additional delays is

\[ T_{ij} = \frac{1 + \langle q_{ij} \rangle}{C_{ij}} \]  \hspace{1cm} (10)

where $\langle q_{ij} \rangle$ is the average length of the queue associated with this link and $C_{ij}$ is the link capacity. Since both the arrival and the delivery of particles at every queue are well approximated by Poisson processes, the average queue length is given by

\[ \langle q_{ij} \rangle = \frac{w_{ij}}{C_{ij} - w_{ij}}. \]  \hspace{1cm} (11)

By using Eqs. 10 and 11 in 9 we obtain the average travel time as a function of load

\[ T_{avg} = \frac{1}{N} \sum_{i,j=1}^{N} \frac{B_{ij}}{(N - 1)C_{ij} - (r) B_{ij}}. \]  \hspace{1cm} (12)

Additional time delays associated with traveling along the links may also be included in the calculation of $T_{avg}$ by using Eq. 9.

To find the set of routes that maximizes transport capacity we use a variant of extremal optimization [26,27]. Specifically, we modified an earlier algorithm that has been shown to converge to near-optimal routing in the case of both geometric and small-world networks if traffic is limited by the processing capacities of the nodes $[12,14,15]$. The original algorithm aims at minimizing the maximum betweenness of any node on the network. The variant employed here minimizes the maximum betweenness-to-capacity ratio of any link on the network. It is, therefore, capable of maximizing the transport capacity of networks with traffic limited by arbitrary link capacities. The algorithm works as follows:

1. Assign an initial weight equal to one to every link and compute the shortest paths between all pairs of nodes and the betweenness of every link.

2. Find the link that has the highest betweenness-to-capacity ratio $(B/C)_{max}$ and increase its weight by one.

3. Recompute the shortest paths and the betweennesses. Go back to step 2.

The computational complexity of the algorithm is $O(N^3 \log N)$ (it is $O(N^2 \log N)$ for one iteration and requires $O(N)$ iterations to converge), which is the same as that of the variant that applies to transport limited by
node capacity \( l \). The algorithm, therefore, provides an approximate solution in only polynomial time to what is known to be an NP-hard problem \( \{13\} \). As we will show, the approximate solutions we find, at least, scale optimally with network size \( N \). Furthermore, because the resulting optimal routing is static, it only needs to be calculated once. Thus, the algorithm is useful even for very large networks.

For simplicity, we consider networks with a binary distribution of link capacities. Then, “streets” are low capacity links and “highways” are high capacity links. Each network realization consists of a network topology and a set of capacities, generated by choosing any link to be a highway with probability \( P_h \). Networks are generated using the algorithm introduced by Barthélémy and Flammini \( \{18\} \), which is based on the idea of links (highways or streets) growing gradually towards population centers. All links grow at the same constant rate towards the centroid of their adjacent population centers, and new population centers are generated at constant average rate. When a link reaches the centroid of its adjacent population centers it splits into separate links directed towards each center. At the end, the nodes of the network will be the population centers and the points where links have been split. This model has been shown to generate networks with characteristics similar to those of real-world street networks, particularly if the distances of the population centers with respect to the city center are exponentially distributed. These similar characteristics include the distribution of the node degrees, as well as size and shape distributions of the areas delimited by streets.

Using this model rather than a particular real-world example allows the calculation of ensemble averages over thousands of model city networks. An ensemble of network realizations is characterized by the average number of nodes and shape distributions of the areas delimited by streets. Using this model rather than a particular real-world example allows the calculation of ensemble averages over thousands of model city networks. An ensemble of network realizations is denoted by angular brackets. Averaging over an ensemble of link capacity configurations results in an ensemble average of network realizations.

In Fig. 2 results are presented for the ensemble averages of \( (B/C)_{\text{avg}} \) and \( (B/C)_{\text{max}} \) for networks with \( \langle N \rangle \approx 250 \). All error bars represent \( 2\sigma \) estimates. The subscripts \( \text{avg} \) and \( \text{max} \) denote the average and respectively maximum over all links of a network with a given link capacity configuration. Averaging over an ensemble of network realizations is denoted by angular brackets. This is an average over an ensemble of network topologies and over an ensemble of link capacity configurations for each network topology. An ensemble of network realizations is characterized by the average number of nodes \( \langle N \rangle \) and by the probability \( P_h \) for a link to be a highway. Figure 2 shows a log-log plot of \( \langle (B/C)_{\text{avg}} \rangle \) and \( \langle (B/C)_{\text{max}} \rangle \) for both SP and OR at \( P_h = 0.3 \). Since the lines are nearly straight, the ensemble averages of these quantities scale with average network size as a power law. Note that any decrease in \( \langle (B/C)_{\text{max}} \rangle \) can only be obtained at the expense of an increase in \( \langle (B/C)_{\text{avg}} \rangle \) since avoiding congestion along the shortest path means
taking longer routes that contribute to the betweenness of more links. With OR, the slopes of \( \langle (B/C)_{\text{avg}} \rangle \) and \( \langle (B/C)_{\text{max}} \rangle \) are essentially the same. Thus, the capacity of the routes we find, at least, scales optimally with system size. This behavior is similar to that of the node betweenness when transport is limited by node processing capacity. However, the finite size effects are stronger in the current case and the error bars under estimate the true error, since they are calculated assuming that the values of \( (B/C)_{\text{avg}} \) and \( (B/C)_{\text{max}} \) for the various network realizations are normally distributed while our simulations show they are not. Figure (2b) shows the exponents of the power law scaling of the four quantities as functions of \( P_h \). Note that the exponents for \( \langle (B/C)_{\text{avg}} \rangle \) and \( \langle (B/C)_{\text{max}} \rangle \) are essentially equal over the whole range of \( P_h \) which argues in favor of the optimality of routing. Note also that the exponent for \( \langle (B/C)_{\text{OR}} \rangle \) exhibits a dip around \( P_h = 0.9 \). This is an interesting feature, indicating that the SP routing works unusually well when there is a small but nonzero concentration of low capacity links.

Figure (3) shows \( < B/C > \) versus \( P_h \) for \( (N) \approx 250 \). The error bars in this figure also represent \( 2\sigma \) estimates. Note that \( \langle (B/C)_{\text{avg}} \rangle \) varies monotonically from \( P_h = 0 \) and \( P_h = 1 \) corresponding to all streets and all highways, respectively, while \( \langle (B/C)_{\text{max}} \rangle \) exhibits a midrange maximum in the case of SP routing that corresponds to a dip of the network transport capacity. This type of behavior is due to the fact that SP routing forgoes the use of low capacity links as long as they are not strictly needed to achieve connectivity (just as we tend to do in real life), which increases congestion on highways. On the other hand, by optimally using all links, it is possible to avoid this phenomenon.

A comparison between the SP and OR average travel times for one network with uneven traffic demand is presented in Fig. 4. Specifically, we look at a “rush hour traffic burst” with particles originating from the innermost 10% of the nodes that we use to model a “downtown”. Their destinations are chosen uniformly from among all nodes with the only constraint being that the node of origin must be different from the destination. This is the same network that was used for Fig. 1 and again the results are representative for networks with an average of about 200 nodes. Note the lower maximum capacity of the network, which is due to the uneven distribution of traffic. Nevertheless, a judicious use of the low capacity links again results in a significant increase of the transport capacity. The factor by which transport capacity is increased may be even higher than in the case of uniform traffic demand. This result is particularly important since traffic congestion usually develops is situations of uneven traffic demand.

In summary, we have studied how to optimize collective transport on planar networks that model a city street grid. Traffic on these networks is limited by link congestion. The network model accounts for many topological features seen in real life urban street networks and allows for heterogeneous link capacities and uneven traffic demands. The question posed in the title is answered in statistical terms, as we show that a judicious use of all links, including low capacity ones, can significantly alleviate traffic congestion. The results we present show that the average factor by which transport capacity can be increased varies with network size as a power law, and that the routes we found are ones that allow the capacity to, at least, scale optimally with system size. Moreover, the average travel times on congested networks can be significantly decreased. We determine the network load at which using the low capacity links becomes advantageous from the point of view of the average travel time. Most importantly, we show that significant improvements in traffic routing can be achieved not only in the case of uniform traffic demand, but also in situations of extremely uneven traffic. Our results are important not only for quantifying the routing improvements that can be achieved on existing networks, but also for the design of future street networks. For example, numerical experimentation with optimal traffic routing on a model of a real city street network can pinpoint the links for which an increase in capacity will result in the highest network.
transport capacity.

This work was supported by the NSF through grant No. DMR-0427538 and by the Texas Advanced Research Program through grant No. 95921.

* dbogdan@mail.uh.edu
† bassler@uh.edu

[1] A.-L. Barabási and R. Albert, Science 286, 509 (1999).
[2] R. Guimerà, A. Díaz-Guilera, F. Vega-Redondo, A. Cabrales, and A. Arenas, Phys. Rev. Lett. 89, 248701 (2002).
[3] M. E. J. Newman, SIAM Review 45, 167 (2003).
[4] D. Helbing and K. Nagel, Contemp. Phys. 45 (5), 405 (2004).
[5] Z. Toroczkai and K. E. Bassler, Nature 428, 716 (2004).
[6] M. Catanzaro, M. Boguñá, and R. Pastor-Satorras, Phys. Rev. E 71, 056104 (2005).
[7] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, Phys. Rev. E 70, 056105 (2004).
[8] P. Echenique, J. Gómez-Gardeñes, and Y. Moreno, Europhys. Lett. 71, 325 (2005).
[9] G. Korniss, M. B. Hastings, K. E. Bassler, M. J. Berryman, B. Kozma, D. Abbott, Phys. Lett. A 350 (5-6), 324 (2006).
[10] B. Danila, Y. Yu, S. Earl, J. A. Marsh, Z. Toroczkai, and K. E. Bassler, Phys. Rev. E 74, 046114 (2006).
[11] G. Yan, T. Zhou, B. Hu, Z.-Q. Fu, and B.-H. Wang, Phys. Rev. E 73, 046108 (2006).
[12] B. Danila, Y. Yu, J. A. Marsh, and K. E. Bassler, Phys. Rev. E 74, 046106 (2006).
[13] S. Sreenivasan, R. Cohen, E. López, Z. Toroczkai, and H. E. Stanley, Phys. Rev. E 75, 036105 (2007).
[14] B. Danila, Y. Yu, J. A. Marsh, and K. E. Bassler, Chaos 17, 026102 (2007).
[15] Y. Yu, B. Danila, J. A. Marsh, and K. E. Bassler, EPL 79, 48004 (2007).
[16] M. Woolf, Z. Huang, and R. J. Mondragon, New J. Phys. 9, 174 (2007).
[17] Z. Toroczkai, B. Kozma, K. E. Bassler, N. W. Hengartner, and G. Korniss, J. Phys. A 41, 155103 (2008).
[18] M. Barthélémy and A. Flammini, Phys. Rev. Lett. 100, 138702 (2008).
[19] J. Scholz, W. Krause, and M. Greiner, Physica A 387, 2987 (2008).
[20] L. Danon, A. Arenas, and A. Díaz-Guilera, Phys. Rev. E 77, 036103 (2008).
[21] A. L. Pastore y Piontti, C. E. La Rocca, Z. Toroczkai, L. A. Braunstein, P. A. Macri, and E. Lopez, New J. Phys. 10, 093007 (2008).
[22] A. Barrat, M. Barthélémy, and A. Vespignani, Dynamical Processes on Complex Networks, (Cambridge University Press, 2008).
[23] M. E. J. Newman, Phys. Rev. E 64, 016132 (2001).
[24] M. Girvan and M. E. J. Newman, PNAS 99 (12), 7821 (2002).
[25] A. O. Allen, Probability, Statistics and Queuing Theory with Computer Science Application, 2nd ed. (Academic Press, New York, 1990).
[26] P. Bak and K. Sneppen, Phys. Rev. Lett. 71 (24), 4083 (1993).
[27] S. Boettcher and A. G. Percus, Phys. Rev. Lett. 86 (23), 5211 (2001).