

## Topology of the Fittest Transportation Network

Jayanth R. Banavar,<sup>1</sup> Francesca Colaiori,<sup>2</sup> Alessandro Flammini,<sup>3</sup> Amos Maritan,<sup>4</sup> and Andrea Rinaldo<sup>5</sup>

<sup>1</sup>*Department of Physics and Center for Materials Physics, 104 Davey Laboratory, The Pennsylvania State University, University Park, Pennsylvania 16802*

<sup>2</sup>*Department of Theoretical Physics, 1 Keble Road, University of Oxford, OX1-3NP, Oxford, United Kingdom*

<sup>3</sup>*Cavendish Laboratory, Madingley Road, University of Cambridge, CB3-0HE, Cambridge, United Kingdom*

<sup>4</sup>*International School for Advanced Studies (S.I.S.S.A.), Via Beirut 2-4, 34014 Trieste, Italy*  
*and INFN and the Abdus Salam International Center for Theoretical Physics, 34014 Trieste, Italy*

<sup>5</sup>*Dipartimento IMAGE, Universita' di Padova, via Loredan 20, I-35131 Padova, Italy*

(Received 15 October 1999)

The presence or absence of loops in the emergent transportation networks, that are characterized by a minimum overall cost, is shown to depend on the convexity of the cost function for the local transportation of material. Our results are directly applicable to a variety of situations across disciplines.

PACS numbers: 45.70.Vn, 05.60.Gg, 05.65.+b, 87.23.Kg

Transportation networks are all around us and play a vital role in virtually all aspects of everyday life [1,2]. Examples of such networks include circuits [3] that transport electric charge, mammalian circulatory systems that convey nutrients to the body through blood circulation [4], and channel networks associated with river drainage basins [5]. In networks of resistances, typically, the flow pattern is one in which a nonzero current flows through each resistor (and loops are present). The flow patterns in the circulatory system and the channel networks, in contrast, normally correspond to spanning trees, which are loopless. Here, we prove generally that the shape (convexity or concavity) of the cost function for local transportation of material impacts directly on the topology (the presence or absence of loops) of the emergent networks that minimize the total cost of transportation. Our results suggest an explanation for the observed topologies in the previously mentioned situations, the directed Abelian sandpile model [6] of self-organized criticality [7], a model of force fluctuations in granular bead packs [8], and are also related to save-wire Steiner trees [9,10] and minimum cost network flows [11].

Consider a transportation network connecting a set of locations (sites) to an outlet site denoted by  $O$ . Each of the sites is connected to one or more of its nearby locations (or neighbors), which results in a network that spans the system. Such a network may be a well-connected one with loops or merely a spanning tree. At the local level, let the cost function for transporting a quantity of material  $i_b$  along a link  $b$  be denoted by  $C_b(i_b)$  with  $C_b(0) = 0$ . In the simplest scenario, that is characteristic of many of the physical examples detailed below,  $C_b$  is an increasing algebraic function of  $i_b$ ,

$$C_b = k_b |i_b|^\gamma, \quad (1)$$

where  $k_b$  is a possibly link-dependent positive amplitude. (In the electrical circuit case,  $C_b$  is the power dissipated

when a current  $i_b$  flows through a resistor  $k_b$  with the exponent  $\gamma$  equal to 2.) More generally, our results for the topology of the network hold even when  $C_b$  is a function that is concave upwards or convex corresponding to situations with  $\gamma > 1$  and  $\gamma < 1$ , respectively. Note that in the concave case the incremental cost for transporting an additional amount of material increases as the amount of material increases, whereas the reverse is true for the convex case. For the linear cost function with  $\gamma = 1$ , the incremental cost is constant. Our focus is on studying the emergent topological properties of the networks that minimize the total cost of transportation. We prove that when  $\gamma > 1$  it is most efficient to use all possible transportation pathways and loops emerge, whereas when  $\gamma < 1$ , it is more economical to send all the material from a given site to one neighboring site rather than to multiple neighboring sites, which leads, globally, to a network with a tree topology.

Let a nonzero "current" (this could represent a water current, an electrical current, or the number of people per unit time that need to be transported to  $O$  from the location  $X$ )  $I_X$  be injected into the network at each site  $X$ . The injected current flows through the network and is extracted at  $O$ . The network spans the system because of the requirement that a route ought to exist from any site  $X$  to the outlet  $O$ . Let the scalar quantity  $|i_b|$  represent the magnitude of the current on the  $b$ th link. The flow into the outlet has a rate that exactly equals the sum of all the injected currents (the  $I_X$ 's). At a junction of the links, a conservation law for the net flow holds—the outflow must exactly balance the inflow plus the amount injected at the site. These conservation laws at all junctions do not uniquely determine the flow on each link for an arbitrary network. The number of degrees of freedom in the choice of the flow pattern is controlled by the number of independent loops.

The total transportation cost in the network per unit time is given by

$$E = \sum_b k_b |i_b|^\gamma. \quad (2)$$

Here, we study the current flow pattern for arbitrary non-negative values of  $\gamma$  with the requirements that the conservation law holds and  $E$  be minimized. (When  $\gamma$  is negative, the incremental cost for sending additional material along a link becomes negative with the result that the energy is minimized by having a proliferation of loops with virtually infinite material being transported in a circle. If one excludes loops in the network, the preferred topology is that of a spiral network because the aggregation of the material is maximized.) We find that, for  $\gamma > 1$ , there is a unique flow pattern that is chosen with nonzero currents along all the links [3,12]. For  $0 < \gamma < 1$ , there are multiple solutions, each of which has the topology of a spanning tree (loops are excluded). The  $\gamma = 1$  case is rich. For the case, when  $\gamma$  approaches 1 from above, all directed flow patterns are solutions, whereas for the complementary case of  $\gamma$  approaching 1 from below, all directed spanning trees form degenerate solutions [13].

*Proof.*—We focus first on the cost function of (2) and give a detailed proof of the results in this case. We will analyze how the results extend to more general cost functions later.

The distance,  $L_X$ , between the outlet and any given site  $X$  is defined to be the minimum number of sites encountered among all the routes along the network from  $O$  to  $X$ . As a consequence, two neighboring sites  $X$  and  $Y$  satisfy the constraint that  $|L_X - L_Y| = 0$  or 1. We now define a convention for the orientation of each of the links. The link between sites  $X$  and  $Y$  is defined to have an orientation that is directed toward the outlet. For cases in which  $L_X = L_Y$ , the orientation is chosen to be one of the two possibilities. A directed flow pattern is, then, defined to be one in which  $i_b$  is zero on all links that are not directed towards the outlet.

Given an arbitrary graph, the number,  $l$ , of independent loops is given by

$$l = \#(\text{bonds}) - \#(\text{sites}) + \#(\text{connected components}). \quad (3)$$

Because our graphs are spanning structures, the number of connected components equals 1.

We will first show that each spanning tree is a local minimum of (2) for  $0 < \gamma < 1$  and that there are no other minima. The proof is straightforward if one chooses a spanning tree and adopts the independent variables to be the currents flowing in the bonds absent in that tree. The number of bonds absent in a tree spanning a graph is equal to the number of loops in that graph and given by Eq. (3). Consider a general structure with  $l$  independent loops. In order to be able to differentiate the energy (2) with respect to the current safely, we introduce the regularized form

$$E_\varepsilon = \sum_b k_b (i_b^2 + \varepsilon^2)^{\gamma/2}. \quad (4)$$

We will show that this energy has a local minimum in each of the treelike spanning configurations (a tree is henceforth denoted by the symbol  $T$ ) with currents  $i_b \sim \varepsilon^{2-\gamma}$  in all  $b \notin T$ . If we overlap  $T$  to the fully connected structure, one may assign loop currents  $x_1, x_2, \dots, x_l$  to the bonds  $b \notin T$ . All other currents are determined by the conservation law at each site in terms of  $\vec{x}$ . Thus  $E_\varepsilon = E_\varepsilon(\vec{x})$ . We will prove that  $\partial E_\varepsilon / \partial x_i = 0$  at  $x_i \sim \varepsilon^{2-\gamma}$  and that the Hessian is positive definite. Indeed

$$\frac{\partial E_\varepsilon}{\partial x_i} = \gamma \sum_b k_b \frac{i_b}{(i_b^2 + \varepsilon^2)^{1-\gamma/2}} \frac{\partial i_b}{\partial x_i}, \quad (5)$$

where  $i_b$  is linear in  $x_i$  and  $\frac{\partial i_b}{\partial x_i}$  takes on one of the three values 0, +1, or -1. In the sum (5) the only link not belonging to the tree  $T$  that gives a nonzero contribution is the one for which  $i_b = x_i$ . The only other terms that could contribute are the links of the tree through which nonzero currents flow. To deduce one class of minima (we will later on show that this is the only class), we postulate that each of the  $x_i$  is infinitesimal and demonstrate this is a consistent solution. Setting  $\frac{\partial E_\varepsilon}{\partial x_i}$  equal to 0 leads to

$$x_i = -\varepsilon^{2-\gamma} \sum_b' \frac{k_b}{k_i} |i_b|^{\gamma-1} \text{sgn}(i_b) \frac{\partial i_b}{\partial x_i} \simeq \varepsilon^{2-\gamma}, \quad \varepsilon \sim 0, \quad (6)$$

and the sum  $\sum_b'$  is over all the links of the tree  $T$ . The second derivative

$$H_{ij} = \frac{\partial^2 E_\varepsilon}{\partial x_i \partial x_j} = \gamma \sum_b' k_b \frac{(\gamma-1)i_b^2 + \varepsilon^2}{(i_b^2 + \varepsilon^2)^{2-\gamma/2}} \frac{\partial i_b}{\partial x_i} \frac{\partial i_b}{\partial x_j}, \quad i \neq j. \quad (7)$$

Thus, each of the  $\frac{\partial^2 E_\varepsilon}{\partial x_i \partial x_j}$  is nonzero in the  $\varepsilon \rightarrow 0$  limit. When  $i = j$

$$H_{ii} = \frac{\partial^2 E_\varepsilon}{\partial x_i \partial x_i} = \gamma k_i \varepsilon^{\gamma-2} + \text{less singular terms}, \quad (8)$$

and thus  $H$  has the form

$$H = \gamma \varepsilon^{\gamma-2} K(I + \varepsilon^{2-\gamma} T), \quad (9)$$

where  $I$  is the identity matrix,  $K_{ij} = k_i \delta_{ij}$ ,  $T_{ii} = 0$ , and  $T_{ij} = \frac{1}{\gamma k_i} H_{ij}$ ,  $i \neq j$ . This implies that the eigenvalues of  $H$  for sufficiently low  $\varepsilon$  are all positive, confirming that (6) is indeed a minimum.

Furthermore, one can see that closing one or more loops with nonzero [ $\mathcal{O}(\varepsilon^0)$ ] currents (taking one or more of the  $x_i$  to be nonzero) causes the appearance of at least one negative eigenvalue in the Hessian matrix for each added bond. If  $x_i$  is the current in this added bond, then the first derivative is (on setting  $\varepsilon = 0$ )

$$\frac{\partial E_\varepsilon}{\partial x_i} = \gamma k_i |x_i|^{\gamma-1} \text{sgn}(x_i) + \gamma \sum_b^l k_b |i_b|^{\gamma-1} \frac{\partial i_b}{\partial x_i} \text{sgn}(i_b) \quad (10)$$

and the second derivative is

$$\frac{\partial^2 E_\varepsilon}{\partial x_i^2} = \gamma(\gamma - 1) \left[ k_i |x_i|^{\gamma-2} + \sum_b^l k_b |i_b|^{\gamma-2} \frac{\partial i_b^2}{\partial x_i} \right] < 0. \quad (11)$$

$H_{ii} < 0$  implies that  $\sum_{kj} H_{kj} z_k z_j < 0$  if  $z_i \neq 0$  and  $z_k = 0, k \neq i$ . This shows that in this case the quadratic form  $H_{ij}$  is not positive definite and no local minima can be found when there are loops with nonzero currents. This completes the proof for the case  $0 < \gamma < 1$ .

The situation is quite different in the case  $\gamma > 1$ : first of all one can easily see that loopless structures cannot be the minima any more. In fact, setting  $\frac{\partial E_\varepsilon}{\partial x_i}$  equal to 0 [where  $\frac{\partial E_\varepsilon}{\partial x_i}$  is given by Eq. (5)], and assuming the set of independent currents to be of order  $\varepsilon^\alpha$ , leads to an inconsistency for any  $\alpha > 0$ . Therefore, there must be loops. Moreover, it is possible to show that there exists just one minimum. The uniqueness of the minimum for  $\gamma > 1$  can be proved using a convexity argument: fix a spanning tree  $T$  and, as before, take the currents  $i_b$  on the bonds  $b \notin T$  to be the independent variables. Let  $j_b, b \in T$  be the other currents, determined by continuity. Now, let  $\{i_b^{(1)}\}$  and  $\{i_b^{(2)}\}$  be two sets of values for the independent currents  $\{i_b\}$ . It is easy to see that  $E$  is a convex function, namely, that for any  $\alpha \in (0, 1)$ ,

$$\alpha E(\{i_b^{(1)}\}) + (1 - \alpha) E(\{i_b^{(2)}\}) > E(\{\alpha i_b^{(1)} + (1 - \alpha) i_b^{(2)}\}). \quad (12)$$

Noting that

$$E(\{i_b\}) = \sum_{b \notin T} k_b |i_b|^\gamma + \sum_{b \in T} k_b |j_b|^\gamma, \quad (13)$$

the convexity of  $E$  follows from the convexity of the function  $|x|^\gamma$  (the inequality holds term by term) and the fact that, because the  $\{j_b\}$ 's are linear in the  $\{i_b\}$ 's,

$$j_b(\{\alpha i_b^{(1)} + (1 - \alpha) i_b^{(2)}\}) = \alpha j_b(\{i_b^{(1)}\}) + (1 - \alpha) j_b(\{i_b^{(2)}\}). \quad (14)$$

The strict convexity property (12) implies the uniqueness of the minimum. In fact, if two minima were present, then, along any path connecting the two, there must be at least one point in which the inequality (12) fails. This completes the proof.

Our analysis applies to much more general forms of the cost function. Let us write in general  $E$  as a sum of the contributions of the single bonds:  $E = \sum_b F(i_b)$ , where we assume that the contribution  $F(i_b)$  of bond  $b$  to the cost function is an increasing function of the current in that bond, and that  $F(0) = 0$ .

Let us consider first the  $0 < \gamma < 1$  case. In order to retain the result that spanning trees are minima of the cost function it is sufficient to require that  $F(x)$  behaves as  $|x|^\gamma$

with  $0 < \gamma < 1$  when  $x \rightarrow 0$ . For a given spanning tree  $T$ , let  $\{x_b\}$  be the currents on the bonds  $b \notin T$ . Because the derivatives of  $F(x_b)$  for  $b \notin T$  become singular when  $\{x_b\} = \{0\}$ , the contribution of the currents in the other bonds turns out to be irrelevant to the behavior of  $E$  in a sufficiently small neighborhood of  $\{x_b\} = \{0\}$ , where  $E$  can be approximated by  $E = \sum_b |i_b|^\gamma$ , and thus has a minimum. The result that the spanning trees are the only minima does not hold for any  $F(x)$ : we need to further assume that  $F(x)$  is a concave function (in order to have  $E$  piecewise concave).

The proof outlined above for  $\gamma > 1$  generalizes trivially to any convex function  $F(x)$ , because the convexity was the only property of  $|x|^\gamma$  used in the proof.

*Applications.*— We now turn to physical applications of these results and point out that several, quite distinct, situations can be described as optimization problems with a cost function of the form of Eq. (2) and with network topologies in accord with our predictions. The  $\gamma = 2$  case corresponds to electrical networks. There is a unique optimal current flow pattern, which contains loops. (Models with general values of  $\gamma$  have been considered previously in the context of fractal nonlinear conductor networks [12,14].) In the limit of  $\gamma$  approaching 0 from above, one obtains the problem of Steiner trees (for nonuniform values of  $k_b$ ) [9,10], random spanning trees (for a constant  $k_b$ ) [15,16], or equivalently [17] a zero-state Potts model [18] in statistical mechanics which can be used to determine the effective resistance between two node points of a network of linear resistors [17].

A wide spectrum of channel networks in the drainage basins of rivers may be characterized by a value of  $\gamma$  equal to  $1/2$  [5]. This follows from the observation that the potential energy lost as water tumbles down from one location to a neighboring one is proportional to the amount of water and the height difference between the locations. The latter is known [5] on average, to scale with a negative power of the height gradient with an exponent close to  $1/2$  in most geologic settings (the slope-discharge law). In this case, as predicted by our theorem, one merely has spanning trees as the network of choice for the transportation of water. This is in complete accord with the vast experimental data on the drainage basins of rivers (see, for example, [5], and references therein).

The  $\gamma = 1$  case is special in that it separates the two classes of behaviors (networks with loops and trees). One can prove [19] that, in this case, all directed networks are optimal and degenerate. This situation describes network flow problems [11] and the circulatory networks in mammals (and plant vascular systems) based on the assumption that the total amount of blood in circulation is as small as possible within the constraint of providing nourishment to all parts of the body and provides an explanation of quarter power allometric scaling in living organisms [19]. It is noteworthy that the appearance of loops has been associated with vascular network formation in tumor tissue [20].

The  $\gamma = 1$  optimization problem reduces to a network model that has been used to describe force fluctuations in granular bead packs [8], which is based on the assumption that all directed load networks are equally likely. The heterogeneous packing leads to a given grain being supported unequally by other grains below it leading to a directed network (with loops) of force fields. In this situation forces that significantly exceeded the mean value are found but are exponentially rare [8]. As  $\gamma$  approaches 1 from below, one obtains a host of equivalent situations. These include the Scheidegger model for river networks [21], the random agglomeration model [22,23], the Voter model [24], models of animal aggregation consistent with group size distributions of tuna fish, sardinellas, and American buffaloes [25], and the directed Abelian sandpile model of self-organized criticality [6,7]. In this case, all directed spanning trees occur with equal probability. This has been solved exactly for uniform  $k_b$  in arbitrary dimensionality [22,23]. A new universality class is obtained for the random  $k_b$  case [13]. In both situations, one obtains nontrivial algebraic distributions for physically relevant quantities [13,22,23].

It is important to underscore that while our theorem is rather general, it does not apply to a range of other interesting physical phenomena that involve networks. These include branching patterns such as those found in diffusion-limited aggregation, whose origin is a dynamic mechanism of the screening of the growing tips, optimization problems in which the local cost function convexity is not simply defined, and cases which need to be generalized to include the location and number of nodes themselves as optimization variables.

This work was supported by EPSRC Grant No. GR/K97783, INFN sez. di Trieste, MURST 40% Trasporto di sedimenti ed evoluzione morfologica di corsi d'acqua, estuari e lagune alle diverse scale temporali, NASA, NATO, and The Donors of the Petroleum Research Fund administered by the American Chemical Society.

---

[1] P.S. Stevens, *Patterns in Nature* (Little, Brown, Boston, 1974).

[2] P. Ball, *The Self-Made Tapestry: Pattern Formation in Nature* (Oxford University Press, Oxford, 1998).

- [3] P.G. Doyle and J.L. Snell, *Random Walk and Electric Networks* (American Mathematical Society, Providence, 1989).
- [4] T.A. McMahon and J.T. Bonner, *On Size and Life* (Scientific American Library, New York, 1983).
- [5] I. Rodriguez-Iturbe and A. Rinaldo, *Fractal River Basins: Chance and Self-Organization* (Cambridge University Press, New York, 1997).
- [6] D. Dhar and R. Ramaswamy, *Phys. Rev. Lett.* **63**, 1659–1662 (1989).
- [7] P. Bak, *How Nature Works: The Science of Self-organized Criticality* (Copernicus, New York, 1996).
- [8] C.H. Liu, S.R. Nagel, D.A. Schecter, S.N. Coppersmith, S. Majumdar, O. Narayan, and T.A. Witten, *Science* **269**, 513–515 (1995).
- [9] M.W. Bern and R.L. Graham, *Sci. Am.* **260**, No. 1, 66–71 (1989).
- [10] A.L. Barabasi, *Phys. Rev. Lett.* **76**, 3750–3753 (1996).
- [11] R.K. Ahuja, T.L. Magnanti, and J.B. Orlin, *Network Flows* (Prentice Hall, New Jersey, 1993).
- [12] J.P. Straley and S.W. Kenkel, *Phys. Rev. B* **29**, 6299–6305 (1984).
- [13] A. Maritan, F. Colaiori, A. Flammini, M. Cieplak, and J.R. Banavar, *Science* **272**, 984–986 (1996).
- [14] R. Blumenfeld, Y. Meir, A.B. Harris, and A. Aharony, *J. Phys. A* **19**, L791–L796 (1986).
- [15] R.L. Shreve, *J. Geol.* **74**, 17 (1966).
- [16] S.S. Manna, D. Dhar, and S.N. Majumdar, *Phys. Rev. A* **46**, R4471–R4474 (1992).
- [17] C.M. Fortuin and W. Kasteleyn, *Physica (Utrecht)* **57**, 536–564 (1972).
- [18] F.Y. Wu, *Rev. Mod. Phys.* **54**, 235–268 (1982).
- [19] J.R. Banavar, A. Maritan, and A. Rinaldo, *Nature (London)* **398**, 130–132 (1999).
- [20] Y. Gazit, D.A. Berk, M. Leunig, L.T. Baxter, and R.K. Jain, *Phys. Rev. Lett.* **75**, 2428–2431 (1995).
- [21] A.E. Scheidegger, *Bull. Assoc. Sci. Hydrol.* **12**, 15 (1967).
- [22] H. Takayasu, M. Takayasu, A. Provata, and G. Huber, *J. Stat. Phys.* **65**, 725–745 (1991); S.N. Majumdar and C. Sire, *Phys. Rev. Lett.* **71**, 3729 (1993).
- [23] M.R. Swift, F. Colaiori, A. Flammini, A. Maritan, A. Giacometti, and J.R. Banavar, *Phys. Rev. Lett.* **79**, 3278–3281 (1997).
- [24] J.M. Liggett, *Interacting Particle Systems* (Springer-Verlag, New York, 1985).
- [25] E. Bonabeau, L. Dagorn, and P. Freon, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 4472–4477 (1999).