EUROPHYSICS LETTERS 15 October 2005

Europhys. Lett., **72** (2), pp. 308–314 (2005)
DOI: 10.1209/epl/i2005-10232-x

Minimum spanning trees of weighted scale-free networks

P. J. Macdonald, E. Almaas and A.-L. Barabási(*)
Center for Network Research and Department of Physics, University of Notre Dame Notre Dame, IN 46556, USA

received 18 July 2005; accepted 26 August 2005
published online 14 September 2005

PACS. 89.75.Fb – Structures and organization in complex systems.
PACS. 89.75.Hc – Networks and genealogical trees.
PACS. 89.20.Hh – World Wide Web, Internet.

Abstract. – A complete characterization of real networks requires us to understand the consequences of the uneven interaction strengths between a system's components. Here we use minimum spanning trees (MSTs) to explore the effect of correlations between link weights and network topology on scale-free networks. Solely by changing the nature of the correlations between weights and network topology, the structure of the MSTs can change from scale-free to exponential. Additionally, for some choices of weight correlations, the efficiency of the MSTs increases with increasing network size, a result with potential implications for the design and scalability of communication networks.

Introduction. – The study of many complex systems has benefited from representing them as networks [1]. There is now extensive empirical evidence indicating that the degree (or connectivity) distribution of the nodes in many networks follows a power law, strongly influencing properties from network robustness [2] to disease spreading [3]. However, to fully characterize these systems, we need to acknowledge the fact that the links can differ in their strength and importance [4–7]. This is demonstrated, *e.g.*, in social networks where the relationship between two long-time friends presumably differs from that between two casual business associates [8], and in ecological systems where the strength of a particular species pair-interaction is crucial for the population dynamics [9, 10].

The fact that links in complex networks are weighted rather than binary (present or absent) must be taken into account when considering dynamical network processes like multicast or broadcast, which have important applications in modern computer networks [11]. For instance, when trying to broadcast a message like a new routing table, one must factor in that different paths and links are characterized by varying degrees of time delay, bandwith, and transmission costs. Consequently, one would attempt to reach all nodes preferably by using as few connections as possible, while at the same time keeping the total weight (*e.g.*, delay) of the traversed links minimal. Similarly, the effective spreading of a computer or biological virus to all nodes in a network would also opt for the low-weight paths. These broadcasting problems all boil down to finding and characterizing the minimum spanning tree (MST) of a

(*) E-mail: alb@nd.edu

network [12,13], where the MST is defined as the connected, loopless subgraph consisting of $(N-1)$ links reaching all N nodes while minimizing the sum of the link weights.

Additionally, MSTs have received attention as examples of optimal path spanning trees [14, 15], and recently spanning trees have been invoked to explain properties of traceroute measurements of the Internet [16]. At a more fundamental level, MSTs are closely connected to invasion bond percolation [17], which has been studied extensively on regular lattices with both random and correlated link weights [18], turning IBP into a key model of non-equilibrium statistical mechanics [19].

In this paper, we present the first results on how the structure and efficiency of MSTs change as a function of the *correlations* between the link weights and the network topology. We start by examining the correlations between the weights and the network structure in several real systems. We use the resulting insights to generate weighted scale-free networks, the MSTs of which are either scale-free or exponential depending on the nature of the link-weight correlations. In contrast, when correlations between weights and topology are removed, we find that the MST degree distribution follows a power law with a degree exponent close to that of the original network, independent of the weight distribution. Finally, we find that the exponential MSTs are increasingly more efficient with increasing network size N, while the efficiency of the scale-free MSTs quickly saturates at a finite value.

Topology correlated weights. – To uncover the relationship between network topology and link weights, in fig. 1 we display the dependence of the weights on the node degrees for the *E. coli* metabolic network, where the link weights represent the optimal metabolic fluxes [20]; the US Airport Network (USAN) where the weights reflect the total number of passengers travelling between two airports between 1992 and 2002; and the link betweenness-centrality (BC), representing the number of shortest paths along a link for the Barabási-Albert (BA) scale-free

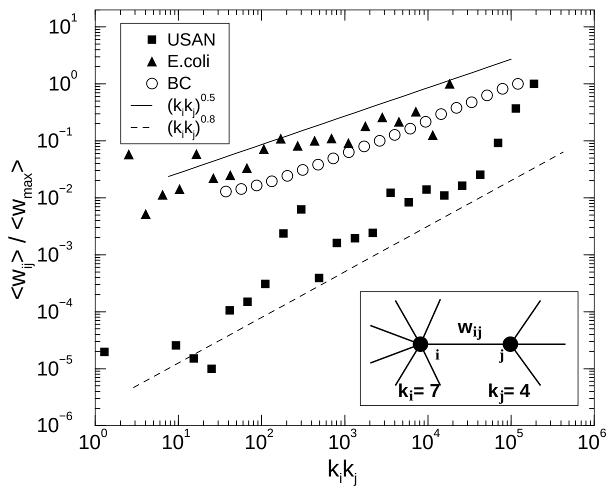

Fig. 1 – The average weight of a link between nodes i and j shown as function of the link end-point degree product k_ik_j: i) the USAN with number of passengers as link weights; ii) *E. coli* metabolic network with optimized flux as link weights [20]; iii) Barabási-Albert model with betweenness-centrality (BC) as link weights. The solid ($w \sim (k_ik_j)^{0.5}$) and the dashed lines ($w \sim (k_ik_j)^{0.8}$) serve as guides to the eye. Inset: The weights are determined by the end-point degrees k_i and k_j: i) $w_{ij} = k_ik_j$, ii) $w_{ij} = \max(k_i, k_j)$, iii) $w_{ij} = \min(k_i, k_j)$ or the inverse thereof iv)-vi).

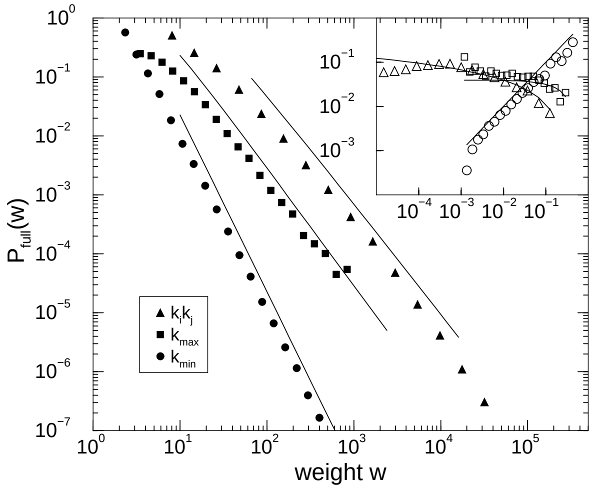

Fig. 2 – Distribution of link weights on $N = 10^5$ node scale-free networks. Link-weight choice i) (triangles), ii) (squares) and iii) (circles) are all heavy tailed. The analytical predictions (eqs. (1)-(3)) are indicated as solid lines. Note that the solid curves have been shifted vertically without changing the character of the scaling law. Inset: The inverse weight distributions iv)-vi) (triangles, squares and circles, respectively) and the analytical predictions shown as continuous lines.

model [21]. For each of these systems the weight distribution follows a power law (not shown) and, as fig. 1 shows, the average link weight scales with the degrees of the nodes on the two ends of a link as $\langle w_{ij} \rangle \sim (k_i k_j)^\theta$, similar to the scaling found for the World Airport Network [6].

These empirical observations allow us to assign weights to the links of a network for which we only have the network topology. To systematically study the role of the weight correlations on the structure of the MSTs, we use several weight assignments. i) First, we choose $w_{ij} = k_i k_j$ (see inset of fig. 1). Note that the MST generated by this weight assignment is identical to the MST obtained for weights $w'_{ij} = (w_{ij})^\theta$ with any $\theta > 0$, as it is the rank of the weights and not their absolute value that determines the MST [22]. We have also studied the two extreme cases of topology-correlated weights, distributed according to ii) $w_{ij} \sim k_{max}$ and iii) $w_{ij} \sim k_{min}$, where $k_{min} = \min(k_i, k_j)$ and $k_{max} = \max(k_i, k_j)$ and with $k_{min}^2 \leq k_i k_j \leq k_{max}^2$. For an important class of weighted network models choice iii) follows naturally from the local growth rules [23]. Finally, we investigate the structure of the MSTs after transforming cases i)-iii) as $w'_{ij} = 1/w_{ij}$, resulting in the link-weight choices iv) $w_{ij} \sim 1/k_i k_j$, v) $w_{ij} \sim 1/k_{max}$ and vi) $w_{ij} \sim 1/k_{min}$.

Weight distributions. – First, we grow scale-free networks according to the BA model [21], the resulting networks having a degree distribution $P(k) \sim k^{-\gamma}$ with $\gamma = 3$. We then assign a weight to each link according to i)-vi). Assuming that the degrees at the two ends of a link are uncorrelated, we can determine the weight distribution analytically using order statistics [24], finding

$$P_{k_ik_j}(w) = (\gamma-2)^2 \; m^{2(\gamma-2)} \; w^{-\gamma+1} \; \ln(w/m^2), \tag{1}$$

$$P_{k_{max}}(w) = 2(\gamma-2) m^{\gamma-2} w^{-\gamma+1} \left[1 - \left(\frac{w}{m}\right)^{-\gamma+2}\right], \tag{2}$$

$$P_{k_{min}}(w) = 2(\gamma-2) m^{2(\gamma-2)} w^{-2\gamma+3}, \tag{3}$$

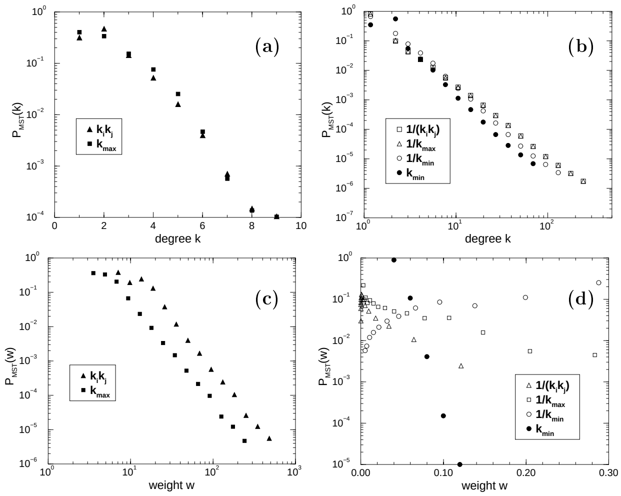

Fig. 3 – Degree and weight distribution of $N = 10^4$ node MSTs. (a) The degree distribution for weights proportional to either $k_i k_j$ (i) or k_{max} (ii) are dominated by an exponential cut-off, while (b) it is heavy-tailed for weights proportional to k_{min} (iii) and inversely proportional to either $k_i k_j$ (iv), k_{max} (v) and k_{min} (vi). (c) The distribution of link weights on the MSTs is a power law for i), ii) and vi) ($w \times 10^3$), while (d) it is dominated by an exponential cut-off for iii) ($w \times 0.02$), iv) and v). For each weight choice we averaged over 10^4 different MSTs.

where m is the number of links connecting a new node to the existing nodes in the BA model. Corresponding expressions for the inverse degree correlations are obtained after the variable change $w' = 1/w$. In fig. 2 we compare the numerically determined weight distributions with the scaling predicted by our analytical expressions. The numerical curves display a w-dependency close to that of eqs. (1)-(3), unaffected by the degree-degree correlations in the BA model [25, 26]. Note that power law weight distributions like those in fig. 2 have been observed for a wide range of network based dynamical processes [27].

Minimum spanning trees. – The MSTs were generated using Prim's algorithm [28]: starting from a randomly selected node, at each time step we add the link (and hence a node) with the smallest weight among the links connected to the already accepted nodes. Whenever n links with the same minimal weight are encountered, we break the degeneracy by randomly selecting one among them with probability $1/n$.

The numerical results indicate that the topology of the resulting MSTs falls into two distinct classes [29,30]. Correlated weight choices i) and ii) give rise to MSTs with exponential degree distributions (fig. 3a), while choices iii)-vi) result in MSTs with power law distributions (fig. 3b). We can understand the exponential nature of weight choice i) and ii) MSTs by the following argument: Since the MST tends to avoid links with large weights, it effectively shuns the hubs for the cases $w_{ij} = k_i k_j$ and $w_{ij} = k_{max}$, utilizing instead, whenever possible, links connecting low degree nodes (fig. 3a). Consequently, all the hubs are marginalized and the MST degree distribution must have a narrow range. This is supported by figs. 4a and b, where we show examples of MSTs for weight choices i) and ii), respectively. The sizes of the nodes in the figure reflect their degree in the original network. It is evident that the majority of the hubs are located on the branches ($k = 1$ degree nodes) of the MST (fig. 4). Additionally, this reliance on the small nodes and tendency to avoid the hubs forces the MSTs

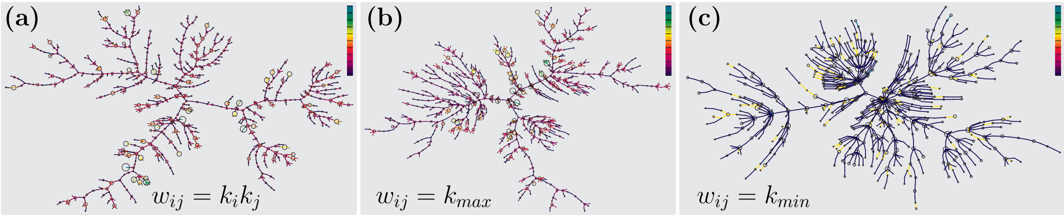

Fig. 4 – (Color online) Minimum spanning trees of a $N = 10^3$ node scale-free network for weight choices (a) k_ik_j, (b) k_{max} and (c) k_{min}. The size of a node represents its degree in the full network, and the color of a link represents its weight from low (black) to high (green). Note that the MST degree distribution is exponential for (a) and (b) and a power law for (c).

generated through method i) and ii) to be very similar. Indeed, on comparison we find that any two MSTs generated from the same original network, but with weights assigned using either method i) or ii), have 87% of the links in common. This explains the similar visual appearance of the two MSTs (figs. 4a and b).

The second class of MSTs is well represented by weight choices iii)-vi), resulting in MSTs with power law degree distributions. The similarity between weight schemes iv) and v) is emphasized by the fact that their MST degree distributions have the same power law exponent, $\gamma = 2.4$ [31] (fig. 3b). The links with the lowest weights are now connected to the hubs of the original network, and the MST grows utilizing these hubs extensively. Hence, the hubs of the full network experience only a slight reduction in their degree and are found at the center of the resulting MSTs (fig. 4c), while the intermediate-degree nodes sustain large losses of neighbors and are found at the branches of the network with only a few neighbors (fig. 4c).

The weight distribution on the MSTs also displays two distinct behaviors: it is either power law (fig. 3c) or exponential (fig. 3d). It is interesting to note that MSTs with exponential *degree* distribution (fig. 3a) have power law *weight* distributions with best-fit exponents of

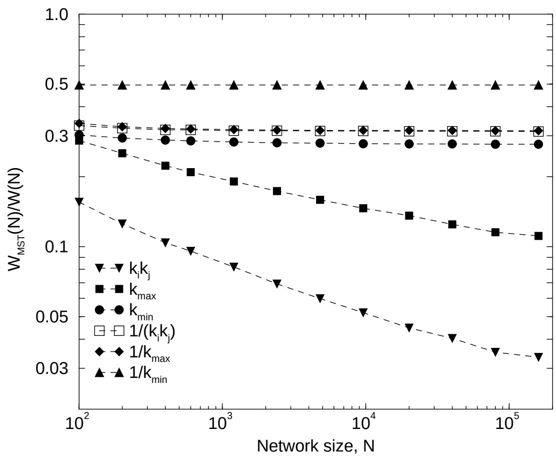

Fig. 5 – The efficiency of the MSTs as a function of network size. The average relative weight on the MSTs for choice i) and ii) displays a power law decrease, while weight choices iii)-vi) all saturate at a finite value.

$\sigma = 3.1$ (case i)) and $\sigma = 3.0$ (case ii)) (fig. 3c). Similarly, for weight choices iii)-v) the MST degree distribution is power law while the weight distribution is exponential or stretched exponential (fig. 3d). However, for weight choice vi) $w_{ij} = 1/k_{min}$ both the degree and the weight distribution of the MST are power law. Since there is a high degree of degeneracy for this MST, it is no surprise that the resulting degree distribution is very similar to the case of random weight assignment [29].

Finally, we characterize the efficiency of the MSTs by calculating their average total weight relative to that of the full network $W_{MST}(N)/W(N)$, where $W(N) = \sum w_{ij}$, for each of the weighting methods (see fig. 5). We find that the efficiency of an MST depends strongly on the type of weight-topology correlation present: While the relative total weight for choices iii)-vi) saturates at a finite value, we find a power law decrease for choices i) and ii). Hence for the empirically motivated case i), the MST becomes increasingly more efficient the larger the network is.

Discussion. – As networks play an increasing role in the exploration of complex systems, there is an imminent need to understand the interplay between network dynamics and topology. While focusing on the MSTs of scale-free networks, our results emphasize the significance of correlations between link-weights and local network structure. In the presence of correlations two classes of MSTs exist for scale-free networks, having either a power law or an exponential degree distribution, while the removal of correlations renders the MSTs scale-free. Additionally, the efficiency of MSTs, which is important for many routing and broadcast applications, varies dramatically with the nature of the correlations. Note that our findings are quite general, *e.g.* the MST of the US airport network, which unlike the BA model has strong degree correlations, is (power law) exponential when the link-weight is taken to be (inversely) proportional to the number of passengers. This is also true for the metabolic network of *E. coli*.

Our findings could serve as a natural starting point towards the systematic exploration of weighted networks. For example, while we have assumed that the weights are static, incorporating their time dependence may reveal novel dynamical rules. Second, we model the weights as solely dependent on the topology, potentially overlooking correlations among the weights themselves. Uncovering the role of such correlations remains a challenge for future research.

* * *

We thank J. KERTÉSZ, S. HAVLIN, P. L. KRAPIVSKY, G. SZABÓ, and T. VICSEK for discussions. We also thank M. A. DE MENEZES for sharing the US airport data. This work has been supported by grants from DoE and NSF (EA and ALB) and the REU program at Notre Dame (PJM).

REFERENCES

[1] STROGATZ S. H., *Nature*, **410** (2001) 268; ALBERT R. and BARABÁSI A.-L., *Rev. Mod. Phys.*, **74** (2002) 47; NEWMAN M. E. J., *SIAM Rev.*, **45** (2003) 167; NEWMAN M. E. J., BARABÁSI A.-L. and WATTS D. J., *The Structure and Growth of Networks* (Princeton University Press, Princeton) 2005.
[2] ALBERT R., JEONG H. and BARABÁSI A.-L., *Nature*, **406** (2000) 378; COHEN R., EREZ K., BEN-AVRAHAM D. and HAVLIN S., *Phys. Rev. Lett.*, **85** (2000) 4626; CALLAWAY D. S., NEWMAN M. E. J., STROGATZ S. H. and WATTS D. J., *Phys. Rev. Lett.*, **85** (2000) 5468.
[3] PASTOR-SATORRAS R. and VESPIGNANI A., *Phys. Rev. Lett.*, **86** (2001) 3200.
[4] YOOK S. H., JEONG H., BARABÁSI A.-L. and TU Y., *Phys. Rev. Lett.*, **86** (2001) 5835.
[5] GOH K.-I., KAHNG B. and KIM D., *Phys. Rev. Lett.*, **87** (2001) 278701.

[6] Barrat A., Barthélemy M., Pastor-Satorras R. and Vespignani A., *Proc. Natl. Acad. Sci. U.S.A.*, **101** (2004) 3747.
[7] Toroczkai Z. and Bassler K. E., *Nature*, **428** (2004) 716.
[8] Granovetter M., *Am. J. Soc.*, **78** (1973) 1360.
[9] Berlow E. L., *Nature*, **398** (1999) 330.
[10] Kilpatrick A. M. and Ives A. R., *Nature*, **422** (2003) 65.
[11] Oliveira C. A. S. and Pardalos P. M., *Comput. Oper. Res.*, **32** (2005) 1953.
[12] Abdel-Wahab H., Stoica I., Sultan F. and Wilson K., *Inf. Sci.*, **101** (1997) 47.
[13] Wan P. J., Calinescu G., Li X. Y., Frieder O., *Wireless Netw.*, **8** (2002) 607.
[14] Braunstein L. A. *et al.*, *Phys. Rev. Lett.*, **91** (2003) 168701.
[15] Sreenivasan S. *et al.*, *Phys. Rev. E*, **70** (2004) 046133.
[16] Clauset A. and Moore C., *Phys. Rev. Lett.*, **94** (2005) 018701.
[17] Barabási A.-L., *Phys. Rev. Lett.*, **76** (1996) 3750.
[18] Schmittbuhl J., Vilotte J.-P. and Roux S., *J. Phys. A*, **26** (1993) 6115.
[19] Vicsek T.,*Growth Phenomena* (World Scientific, Singapore) 1992; Stauffer D. and Aharony A., *Introduction to Percolation Theory* (CRC Press) 1994.
[20] Almaas E. *et al.*, *Nature*, **427** (2004) 839.
[21] Barabási A.-L. and Albert R., *Science*, **286** (1999) 509.
[22] Dobrin R. and Duxbury P. M., *Phys. Rev. Lett.*, **86** (2001) 5076.
[23] Barrat A. and Barthelemy M. and Vespignani A., *Phys. Rev. E*, **70** (2004) 066149.
[24] Hogg R. V. and Craig A. T., *Introduction to Mathematical Statistics*, 5th edition (Macmillan, New York) 1995.
[25] Krapivsky P. L. and Redner S., *Phys. Rev. E*, **63** (2001) 066123.
[26] Taking into account the degree-degree correlations of our model [25], we find the same $w \gg 1$ limiting behavior as that of eqs. (1)-(3).
[27] Barabási A.-L., de Menezes M. A., Balensiefer S. and Brockman J., *Eur. Phys. J. B*, **38** (2004) 143.
[28] Prim R. C., *Bell Syst. Tech. J.*, **36** (1957) 1389.
[29] Szabó G., Alava M. and Kertész J., *Physica A*, **330** (2003) 31.
[30] The acceptance function $P_{MST}(w)/P(w)$ for the case of $w_{ij} = k_{max}$ and $w_{ij} = 1/k_{max}$ was obtained independently by Szabó, Alava, and Kertész. Kertész J., private communication.
[31] Kim D.-H., Noh J. D. and Jeong H., *Phys. Rev. E*, **70** (2004) 046126.