# The Observable Part of a Network

Piet Van Mieghem and Huijuan Wang
Delft University of Technology, P.O. Box 5031, 2600 GA Delft, The Netherlands

***Abstract*— The union of all shortest path trees $G_{\cup spt}$ is the maximally observable part of a network when traffic follows shortest paths. Overlay networks such as peer to peer networks or virtual private networks can be regarded as a subgraph of $G_{\cup spt}$. We investigate properties of $G_{\cup spt}$ in different underlying topologies with regular i.i.d. link weights. In particular, we show that the overlay $G_{\cup spt}$ in an Erdös-Rényi random graph $G_p(N)$ is a connected $G_{p_c}(N)$ where $p_c \sim \frac{\log N}{N}$ is the critical link density, an observation with potential for ad-hoc networks.**

**Shortest paths and, thus also the overlay $G_{\cup spt}$, can be controlled by link weights. By tuning the power exponent $\alpha$ of polynomial link weights in different underlying graphs, the phase transitions in the structure of $G_{\cup spt}$ are shown by simulations to follow a same universal curve $F_T(\alpha) = \Pr[G_{\cup spt}$ is a tree$]$. The existence of a controllable phase transition in networks may allow network operators to steer and balance flows in their network. The structure of $G_{\cup spt}$ in terms of the extreme value index $\alpha$ is further examined together with its spectrum, the eigenvalues of the corresponding adjacency matrix of $G_{\cup spt}$.**

***Index Terms*— Overlay, observability, union of shortest paths**

## I. Introduction

In this paper, we study the subgraph formed by the union of all possible shortest paths in a graph $G(N, L)$ with $N$ nodes and $L$ links. The motivation to consider shortest paths, defined in Section II, is that, in most real world networks, transport is mainly carried along shortest paths. Even for the Internet, it is a reasonable assumption, since roughly 80% of the routes seem to correspond to shortest paths.

The shortest path tree ($SPT$) rooted at some node is the union of the shortest paths from that node to all the other nodes. The union of all shortest path trees $G_{\cup spt}$, which is also the union of the shortest paths between all possible pairs of nodes, can be regarded as the "transport overlay network" on top of the network topology or substrate. Also, $G_{\cup spt}$ can be regarded as the maximally observable part of a network. In the Internet, for example, the traffic is carried along the overlay $G_{\cup spt}$, a fraction of the links in the underlying network, which is just the maximal part of the Internet that we can actually observe by traceroute measurements. In the remainder of this introductory section, we discuss three potential applications that motivate a study of $G_{\cup spt}$: overlay networks, traffic engineering and Internet topology interference.

The importance of overlay networks is only believed to grow in the future. One example of an overlay network are peer to peer networks [5] with $n$ distributed systems sharing resources such as content, CPU cycles and storage, where $n$ is smaller than the number of nodes $N$ in the underlying network. The peer to peer overlay network can be regarded as a union of paths connecting these $n$ nodes. Another type of overlay network is a virtual private network (VPN), a private network that uses a public network (usually the Internet or the telephone network) to connect remote sites or users together. The physical networks traversed by both the peer to peer and the VPN overlay networks are a subgraph of $G_{\cup spt}$. The robustness in overlay networks, the persistence of epidemics [7] and the vulnerability to node failures and attacks [8] are depending on structural properties of $G_{\cup spt}$ that are studied in this paper.

The overlay $G_{\cup spt}$, not the substrate, determines the network's performance: any link removed in $G_{\cup spt}$ will definitely impact at least those flows of traffic that pass over that link. Here we show that, instead of changing the infrastructure of a network [6], the overlay network $G_{\cup spt}$ can be controlled by tuning the link weight structure [3], especially by changing the extreme value index $\alpha$, defined as

$$\alpha = \lim_{x \downarrow 0} \frac{\log F_w(x)}{\log x}$$

of a link weight distribution $F_w(x) = \Pr[w \leq x]$, for example, the polynomial distribution

$$F_w(x) = x^{\alpha} 1_{x \in [0,1]} + 1_{x \in (1,\infty)}, \quad \alpha > 0 \tag{1}$$

where the indicator function $1_y$ is 1 if the event $y$ is true else it is zero. If $\alpha \to \infty$, it follows from (1) that $w = 1$ almost surely for all links. The $\alpha \to \infty$ regime is entirely determined by the topology of the graph because the link weight structure does not differentiate between links. Here, the $\alpha \to \infty$ regime is not further considered. In the $\alpha \to 0$ regime, all flows are transported over the minimum possible fraction of links in the network and each shortest path tree coincides with the minimum spanning tree (MST). Any failure in a node or link disconnects the MST into two parts and may result in obstruction of transport in the network. The $\alpha \to 0$ regime may constitute a weak regime although it is highly efficient: only $N-1$ links are used which means that a minimum of links need to be controlled and/or secured. From a traffic engineering point of view, choosing $\alpha$ around 1 will lead to the use of more paths and, hence, a more balanced overall network load than in the $\alpha \to 0$ regime.

The final motivation applies to interfering the Internet topology. Recently, Lakhina *et al.* [18] have pointed to the effect of biases when trying to construct the Internet topology from a few source trees that span a huge number of destinations, basically because the links close to the source have a substantially higher probability to be detected than links close to the destination. The potential dramatic effect of biases was recently rigorously analyzed in a mathematical analysis first by Clauset and Moore [33], and later extended by Achlioptas *et al.* [32]. They showed that, irrespective of

*Email*: P.F.A.VanMieghem@tudelft.nl.
*Email*: H.Wang@ewi.tudelft.nl.

the degree distribution of the substrate, the inferred topology deduced from only a few source trees to many destinations is likely to possess a power law degree distribution. For example, a random network with a Poisson degree distribution and a regular graph (with constant degrees) all lead to an observed power law degree distribution. These analyses place doubt on the believed power law degree structure of the Internet. Biases can be circumvented if sources and destinations are regarded as equally important. By constructing the union of all paths between $m < N$ testboxes, an overlay network $G_{\cup_m spt}$ that is a subgraph of $G_{\cup spt}$ is obtained. An interesting, still open question is, how large needs $m$ to be such that $G_{\cup_m spt}$ sufficiently resembles properties of $G_{\cup spt}$ in which $m = N$. In this paper, we shed some light on how different $G_{\cup spt}$ can be compared to the underlying substrate and also show measurement results in Section VII of a partial overlay $G_{\cup_m spt}$.

The paper is outlined as follows. First, in Section II, we explain the notation. The general properties of the $G_{\cup spt}$ are analyzed in Section III. In Section IV, basic notions related to the link weight structure and three fundamentally different classes of underlying topologies are proposed. In these underlying topologies equipped with different link weight structures, Section V presents simulation results of (a) the number of links, (b) the degree distribution, and (c) the spectrum. Furthermore, in Section VI, by varying the extreme value index of the link weight distribution, the phase transition in $G_{\cup spt}$ is analyzed and compared among the three classes of topologies. Finally, our results are summarized in Section VII.

## II. TERMINOLOGY AND NOTATION

A network is represented by a graph $G(N, L)$, in short $G$, which consists of a set $\mathcal{N}$ of $N$ nodes and a set $\mathcal{L}$ of $L$ links. Following the notation in [11], a *path* from node $A$ to node $B$ with $k-1$ hops or links is the node list $P_{A\to B} = n_1 \to n_2 \to \cdots n_{k-1} \to n_k$ where $n_1 = A$ and $n_k = B$ and where $n_j \neq n_i \in \mathcal{N}$ for each index $i$ and $j$. We confine to connected graphs in which there is always a path between any pair of nodes in $G$.

To each link $l(i \to j) \in \mathcal{L}$, in short $i \to j$, from node $i \in \mathcal{N}$ to node $j \in \mathcal{N}$ in the network, we assign a link weight $w(i \to j)$, a non-negative real number, which quantifies a property of that link such as the delay incurred when traveling over that link, the distance, the capacity, etc. The set of all link weights is called the link weight structure of $G$. We consider only additive link weights such that the weight of a path $P$ is $w(P) = \sum_{(i\to j)\in P} w(i \to j)$, i.e. $w(P)$ equals the sum of the weights of the constituent links of $P$. The shortest path $P^*_{A\to B}$ from $A$ to $B$ is the path with minimal link weight, thus, $w(P^*_{A\to B}) \le w(P_{A\to B})$ for all $P_{A\to B}$.

A tree is a subgraph without cycles. Similarly as for a path, the weight of a tree $T$ is $w(T) = \sum_{(i\to j)\in T} w(i \to j)$. A spanning tree $T_G$ is a tree that contains or spans all $N$ nodes of the graph $G$. A minimum spanning tree (MST) $T^*_G$ is a minimum weight spanning tree in $G$ such that $w(T^*_G) \le w(T_G)$ for all $T_G$ in $G$. In general, there can be more than one shortest path and more than one minimum spanning tree. In particular, if $w(i \to j) = 1$ for all links $l(i \to j) \in \mathcal{L}$, the number of MSTs equals the complexity $\xi(G) = \frac{1}{N}\prod_{j=1}^{N-1}\mu_j$ of the graph which is the product of all positive eigenvalues $\mu_j$ of the Laplacian of $G$ divided by $N$ (see e.g. [34],[11, Appendix B]). The complexity $\xi(G)$ can be as high as $N^{N-2}$ (Caley's Theorem). Algorithms to compute a shortest path (such as Dijkstra's) and a MST (such as Prim's and Krushkal's) are nicely explained in [35] and applied to data communication networks in [36]. If the MST does not exist, the graph is disconnected.

Both the topology and the link weight structure of the graph are key determinants. A further discussion is deferred to Section IV.

## III. THE UNION OF THE SHORTEST PATH TREES: THEORY

Any set of links $l(i \to j)$, $l(j \to k)$ and $l(i \to k)$ between three nodes $i$, $j$ and $k$ in the union $G_{\cup spt}$ of the shortest path trees obeys the triangle inequality,

$$w(i \to k) \le w(i \to j) + w(j \to k)$$

otherwise the link $l(i \to k)$ is not the shortest path from $i$ to $k$ and, consequently, does not belong to $G_{\cup spt}$. Hence, if $l(i \to j)$, $l(j \to k)$, $l(i \to k) \in \mathcal{L}_{G_{Uspt}}$ where $\mathcal{L}_{G_{Uspt}}$ is the set of links of the graph $G_{\cup spt}$, then

$$\Pr\left[w(i \to j) + w(j \to k) \ge w(i \to k)\right] = 1$$

*Theorem 1:* A minimum spanning tree belongs to $G_{\cup spt}$.

**Proof:** The proof is by contradiction. Suppose that a MST does not belong to $G_{\cup spt}$. This means that there is at least one link $l(i \to j) \in MST$ which does not belong to the union of shortest path trees, $l(i \to j) \notin G_{\cup spt}$. Hence, the link $l(i \to j)$ is not the shortest path $P^*_{i\to j}$ from node $i$ to node $j$ implying that

$$w\left(P^*_{i\to j}\right) < w(i \to j)$$

In that case, we can lower the weight of the MST which equals

$$\begin{aligned} w_{MST} &= \sum_{(k\to l)\in MST} w(k \to l) \\ &= w(i \to j) + \sum_{k\to l\in MST; k\to l\neq i\to j} w(k \to l) \end{aligned}$$

by changing[1] $w(i \to j)$ for $w\left(P^*_{i\to j}\right)$. However, this is impossible since $w_{MST}$ is, by definition, the tree that minimizes the above sum. This proves the theorem. □

In case there are several MSTs, it is possible that there is more than one subgraph $G_{\cup spt}$. In the sequel, we confine ourselves to real link weights $w$ and undirected links $l(i \to j) = l(j \to i)$. The uncertainty about the underlying topology of complex networks leads us to consider both the underlying topologies and each of the link weights as random variables. If we assume in addition identically and independently distributed (i.i.d.) link weights then the probability to

[1] Actually, since $P^*_{i\to j}$ must consist of at least two links, only those that are necessary to obtain a tree are needed in the MST such that we can further lower $w_{MST}$.

have more than one shortest path or more than one MST is negligibly small.

Any link $l\,(i \to j)$ with link weight $w(i \to j)$ in the $G_{\cup spt}$ must be the shortest path $P^*_{i\to j}$ between $i$ and $j$ because a link in the $G_{\cup spt}$ must belong to a shortest path and a subsection of a shortest path is also a shortest path. Conversely, if a link $l\,(i \to j)$ is the shortest path $P^*_{i\to j}$ between $i$ and $j$, it must belong to the $G_{\cup spt}$, because the $G_{\cup spt}$ is the union of shortest paths between all possible source and destination nodes. Therefore, the event that a link $l\,(i \to j)$ is observed in the $G_{\cup spt}$ is equivalent to the event $\{P^*_{i\to j} = l\,(i \to j)\}$ that the link $l\,(i \to j)$ is the shortest path $P^*_{i\to j}$ between $i$ and $j$. Hence, $\Pr[P^*_{i\to j} = l\,(i \to j)]$ is also the probability that a link can be observed.

*Theorem 2:* In any graph with positive i.i.d. link weights $w$ specified by the probability density function $f_w\,(x)$, the probability that a link $l\,(i \to j)$ between node $i$ and $j$ is the shortest path $P^*_{i\to j}$ between $i$ and $j$ is

$$\Pr[P^*_{i\to j} = l\,(i \to j)] = \int_0^\infty \frac{f_w(x)\Pr[w(P^*_{i\to j}) > x]}{\Pr[w(i \to j) > x] + \frac{1-p_{ij}}{p_{ij}}}dx \tag{2}$$

where $p_{ij} = \Pr[l\,(i \to j) \text{ exists}]$.

**Proof:** See Section A. $\square$

Since $\Pr[P^*_{i\to j} = l\,(i \to j)] \le 1$ and $\int_0^\infty f_w(x)dx = 1$, we see in (2) that $\frac{\Pr[w(P^*_{i\to j})>x]}{\Pr[w(i\to j)>x]+\frac{1-p_{ij}}{p_{ij}}} \le 1$. A slightly tighter bound follows from the probability $P_c$ in (18) and the left hand side of (21), that is bounded by $p_{ij}$, such that

$$\Pr[w(P^*_{i\to j}) > x] \le p_{ij}\Pr[w(i \to j) > x] + 1 - p_{ij} \tag{3}$$

In words, the probability that the weight of the shortest path exceeds $x$ is always less than or equal to the probability that an arbitrary link weight exceeds $x$ (because of the assumption of i.i.d. link weights) multiplied by the probability of the existence of the link plus the probability $1 - p_{ij}$. The bound (3) is sharpest in case $p_{ij} = 1$, thus, in case the direct link $i \to j$ exists surely.

*Corollary 1:* In any graph with $N$ nodes and with positive i.i.d. link weights, we can write

$$\begin{aligned}\Pr[P^*_{i\to j} = l\,(i \to j)] &= \Pr\left[H_N = 1\right]\\ &= -\int_0^\infty f_{w(P^*_{i\to j})}\,(x)\log\left(1 - p_{ij}F_w\,(x)\right)dx\end{aligned} \tag{4}$$

where $H_N$ denotes the hopcount (or the number of links) of a shortest path and $F_w\,(x) = \Pr[w(i \to j) \le x]$ is the link weight distribution.

**Proof:** See Section B. $\square$

Corollary 1 establishes a relation between the probability that the hopcount of the shortest path equals 1 in terms of the distribution of the link weights and of the weight of the shortest path.

When multiplying all the link weights by a factor $\frac{1}{b}$ where $b > 0$, relation (2) remains unchanged. For, since $f_{\frac{w}{b}}\,(x) = bf_w\,(bx)$, we have

$$\begin{aligned}&\int_0^\infty f_{\frac{w}{b}}(x)\frac{\Pr\left[\frac{w(P^*_{i\to j})}{b} > x\right]}{\Pr\left[\frac{w(i\to j)}{b} > x\right] + A}dx\\ &= \int_0^\infty f_w\,(bx)\,\frac{\Pr\left[w(P^*_{i\to j}) > bx\right]}{\Pr\left[w(i \to j) > bx\right] + A}d\,(bx)\end{aligned}$$

and substitution of $u = bx$ leads to (2). This fact is, of course, natural because the shortest path does not change in structure and in the number of hops when all links are scaled or expressed in a different unit.

### A. *Example*

If link weights are exponentially distributed,

$$\Pr[w(i \to j) > x] = \exp(-\alpha x)$$

where $E\,[w] = \frac{1}{\alpha}$ and $\Pr[l\,(i \to j) \text{ exists}] = p_{ij}$ is the link density, then (4) gives with $W_N = w(P^*_{i\to j})$

$$\Pr\left[H_N = 1\right] = -\int_0^\infty f_{W_N}\,(x)\log\left(1 - p_{ij} + p_{ij}e^{-\alpha x}\right)dx \tag{5}$$

Applied to the complete graph $K_N$ where $p_{ij} = 1$ leads to

$$\Pr\left[H_N = 1\right] = \alpha\int_0^\infty x f_{W_N}\,(x)\,dx = \alpha E\left[W_N\right]$$

Invoking [11, Chapter 15]

$$E\left[W_N\right] = \frac{1}{\alpha(N-1)}\sum_{n=1}^{N-1}\frac{1}{n}$$

we find that

$$\Pr[P^*_{i\to j} = l\,(i \to j)] = \frac{E\left[W_N\right]}{E\left[w\right]} = \frac{1}{N-1}\sum_{n=1}^{N-1}\frac{1}{n} \tag{6}$$

Alternatively, the probability density function of hopcount of the shortest path in $K_N$ with exponential link weights is [11, Chapter 15]

$$\Pr[H_N = k] = \frac{N}{N-1}\frac{(-1)^{N-(k+1)}S_N^{(k+1)}}{N!}$$

where $S_N^{(k)}$ is the Stirling number of the first kind [14]. The second Stirling number of the first kind can be explicitly written as

$$S_N^{(2)} = (-1)^N(N-1)!\sum_{n=1}^{N-1}\frac{1}{n}$$

we obtain

$$\Pr[P^*_{i\to j} = l\,(i \to j)] = \Pr[H_N = 1] = \frac{1}{N-1}\sum_{n=1}^{N-1}\frac{1}{n}$$

which is, indeed, the same as (6).

### B. The number of observable links in a network

The number of observable links, or the number of links in $G_{\cup spt}$, denoted by $L_o$, in any network $G$ is, by definition,

$$L_o = \sum_{(i\to j)\in\mathcal{L}} 1_{\left\{l(i\to j)=P^*_{i\to j}\right\}} \tag{7}$$

By taking the expectation in (7), the average number of observable links is

$$E\left[L_o\right] = \sum_{(i\to j)\in\mathcal{L}} \Pr[P^*_{i\to j} = l\left(i\to j\right)]$$

and, if all links in $G$ have equal probability to exist ($p_{ij} = p$),

$$E\left[L_o\right] = L\Pr[P^*_{i\to j} = l(i\to j)] = L\Pr[H_N = 1] \tag{8}$$

Hence, the probability of link observability, $\Pr[P^*_{i\to j} = l(i\to j)]$, is equal to the average number of links in $G_{\cup spt}$ divided by the total number of links $L$ in the substrate network $G$.

Clearly, since $G_{\cup spt}$ is connecting all nodes and the number of links in a tree – which is the minimum number of links to connect all nodes – is $N-1$, we have

$$N-1 \leq L_o \leq L \tag{9}$$

The total number of links in a square two dimensional lattice with $N$ nodes is $L = 2(N-\sqrt{N})$. Applying the bounds (9) for $L_o$ to a 2D-lattice,

$$N-1 \leq L_o \leq 2N-2\sqrt{N} \tag{10}$$

which shows that $L_o = cN$ is linear to first order in $N$. An estimate of $c$ is given in Section V-A.

For exponential link weights, the average number of observable links in the complete graph $K_N$ equals

$$E\left[L_o\right] = \frac{N}{2}\sum_{n=1}^{N-1}\frac{1}{n} \simeq \frac{N}{2}\left(\gamma+\ln N\right) \tag{11}$$

which follows from (6), (8), and $L = \frac{N(N-1)}{2}$. For a polynomial link weight distribution (1), we have approximately [4] to highest order in $N$ and for $\alpha$ around 1 only, that

$$\Pr[H_N = k] \simeq \frac{1}{N}\frac{(\ln\alpha N)^{\alpha k}}{\Gamma(\alpha k+1)}$$

and, hence, for $\alpha$ around 1 and large $N$,

$$E\left[L_o\left(\alpha\right)\right] \simeq \frac{N}{2}\frac{(\ln\alpha N)^{\alpha}}{\Gamma(\alpha+1)}$$

If $\alpha\to\infty$, all link weights are the same and equal to 1, $L_o = L$, which shows that equality in (9) can occur. For that extreme case, (8) tells us that any link is a shortest path between its end nodes.

### C. The degree distribution and beyond

*Theorem 3:* The degree distribution of a node in the overlay $G_{\cup spt}$ is equal to its degree distribution when that node is the root of a $SPT$.

**Proof:** We construct $G_{\cup spt}$ in two steps. First, the $SPT$ rooted at a particular node, say node 1, is calculated. Second, the $SPT$ rooted at each other node is computed. Since $G_{\cup spt}$ is the union of the $SPT$s rooted at all the nodes, the union of shortest paths obtained in these two steps is just $G_{\cup spt}$. All the links in $G_{\cup spt}$ connected to node 1 are found in the first step. For, assume that the link $1\to j$ is only found in the second step. Since $1\to j\in G_{\cup spt}$, it must be the shortest path between 1 and $j$. Thus, it must belong to the $SPT$ rooted at 1 in the first step. Hence, the degree of a node in $G_{\cup spt}$ is equal to its degree when it is the root of a $SPT$. □

A direct consequence of this proof is

*Corollary 2:* A link $l\left(i\to j\right) \in G_{\cup spt}$ if and only if $l\left(i\to j\right)$ is a first hop link in the SPT rooted at node $i\in G_{\cup spt}$.

*Theorem 4:* The degree distribution in the overlay $G_{\cup spt}$ on top of the complete graph $K_N$ equipped with exponentially distributed link weights is

$$\Pr[D_{G_{\cup spt}} = k] = \frac{(-1)^{N-1-k}S^{(k)}_{N-1}}{(N-1)!} \tag{12}$$

**Proof:** In [11, Section 16.6.3], it is shown that the degree distribution of the root in the $SPT$ in the complete graph with exponential link weights is given by the right hand side of (12). Application of Theorem 3 proves this Theorem 4. □

Before extending Theorem 4, we review the following definition [11, Chapter 16]. A regular link weight distribution $F_w(x) = \Pr[w\leq x]$ has a Taylor series expansion around $x = 0$,

$$F_w(x) = f_w(0)x + O(x^2)$$

since $F_w(0) = 0$ and $F'_w(0) = f_w(0)$ exists. A regular link weight distribution is thus linear around zero. The uniform and exponential distributions belong to the regular link weight distribution with $f_w(0) = 1$.

*Corollary 3:* For large $N$, the degree distribution in the overlay $G_{\cup spt}$ on top of the Erdös-Rényi random graph $G_p(N)$ with link density $p$ above the disconnectivity threshold $p_c$ and equipped with i.i.d. regular link weights is

$$\Pr[D_{G_{\cup spt}} = k] \sim \frac{(-1)^{N-1-k}S^{(k)}_{N-1}}{(N-1)!} \tag{13}$$

**Proof:** In [11, Chapter 16], it is shown that the $SPT$ in the complete graph $K_N$ with exponential link weights is precisely a uniform recursive tree[2] $URT$ for any $N$. In [10], a $URT$ is shown to be asymptotically the $SPT$ in the Erdös-Rényi random graph $G_p(N)$ (see e.g. [15]) with *any* link density $p$ above the disconnectivity threshold $p_c \sim \frac{\log N}{N}$ and with exponential links weights. The exponential distribution is a regular distribution and the shortest path is mainly determined by small link weights in the substrate graph. For sufficiently large and dense, connected graphs, and in view of the i.i.d. assumption, there are enough small link weights very near zero. Hence, under these assumptions, any regular link weight distribution will lead asymptotically to the same SPT. In conclusion, with regular link weights and large $N$, we have, structurally, that $SPT_{K_N} \simeq SPT_{G_p(N)} \simeq URT$. Therefore, the degree distribution of $G_{\cup spt}$ in $K_N$ or $G_p(N)$ with regular

[2] A URT of size $N$ is a random tree rooted at some source node and where at each stage a new node is attached uniformly to one of the existing nodes until the total number of nodes is equal to $N$.

link weights, is asymptotically equal to the degree distribution of the root in the $URT$ [11], which is equal to (12). $\square$

*Conjecture 1:* For large $N$, the overlay $G_{\cup spt}$ on top of a connected Erdös-Rényi random graph $G_p(N)$ with link density $p \in (p_c, 1]$ and equipped with i.i.d. regular link weights is a **connected** Erdös-Rényi random graph $G_{p_c}(N)$ where $p_c$ the disconnectivity threshold.

**Arguments:** First, relation (6) states that each link in the underlying complete graph $K_N$ has a probability to appear in the overlay $G_{\cup spt}$ equal to $\frac{1}{N-1}\sum_{n=1}^{N-1}\frac{1}{n} \sim p_c$, for large $N$.

Second, for large $N$ and $p = \frac{\log N}{N}$, the binomial degree distribution of the Erdös-Rényi random graph $G_p(N)$ tends to a Poisson distribution with mean $\log N$. Hence, for large $N$,

$$\Pr\left[D_{G_{\frac{\log N}{N}}} = k\right] = \binom{N-1}{k} p^k (1-p)^{N-1-k}\Big|_{p=\frac{\log N}{N}} \sim \frac{(\log N)^k}{Nk!}$$

In addition, in [11, Section 16.3.1], it is shown that also (12) tends to a same Poisson distribution,

$$\Pr[D_{G_{\cup spt}} = k] = \frac{(-1)^{N-1-k} S_{N-1}^{(k)}}{(N-1)!} \sim \frac{(\log N)^k}{Nk!}$$

In summary, for large $N$, the URT is also asymptotically the shortest path tree in a connected Erdös-Rényi random graph $G_p(N)$ with link density $p \in (p_c, 1]$. Thus, since $G_{\cup spt}$ is surely connected[3], for large $N$, each link in the substrate $G_p(N)$ has a same probability of appearing in $G_{\cup spt}$ equal to $p \sim p_c$ and the degree distribution of $G_{\cup spt}$ equals that of $G_{p_c}(N)$.

In contrast to the Erdös-Rényi random graph $G_p(N)$ where all links *are* independent for any $N$, the links in $G_{\cup spt}$ on top of the complete graph $K_N$ are not independent because each of them is a shortest path link, a fact that correlates all these links. It still remains to prove that links in the overlay $G_{\cup spt}$ are asymptotically independent. Lemma's 1, 2, and 3 on the uncorrelation of links in $G_{\cup \text{URT}}$ are presented in Appendix C as partial arguments. If the asymptotic independence of links can be proved (which would turn the conjecture into a theorem), then, for large $N$, the three properties (a link density $p_c$, a Poissonean degree distribution and asymptotic independence of the links) together with the connectedness of $G_{\cup spt}$ will demonstrate that $G_{\cup spt}$ is a connected Erdös-Rényi random graph with $p = p_c$. $\square$

Simulations in Section VI further illustrate this Conjecture. We expect that Conjecture 1 may hold for a broader class than Erdös-Rényi random graphs: namely all substrate topologies that are *homogeneous* (i.e. the SPT rooted at any node has a same structure or any node perceives, views the network in a same way) and *dense* (i.e. "enough" link). Conjecture 1 explains why the role of the simple Erdös-Rényi random graph $G_p(N)$ is more important in overlay networks, such as e.g. peer-to-peer networks (see also Figures 11, 12), than in substrate topologies, where only a few complex networks belong to the class of Erdös-Rényi random graphs. Finally, the asymptotic results in this section motivate why a confinement to the complete graph (in later sections) is much less restrictive than it appears at first glance.

## IV. Basic notions and simulation scenarios

Three different classes of topologies are considered: the complete graphs $K_N$, lattices and power law graphs. An example of a two-dimensional lattice is shown in Figure 2. In D-dimensional lattices, all interior nodes have the same degree 2D, where D is the dimension. Here, we confine ourselves to hyper-cube D-lattices in which each edge is of equal size. Power law graphs are random graphs specified by a degree distribution $\Pr[d = i] = ci^{-\tau}$, where $c$ is a constant such that $\sum_{i=1}^{N-1} ci^{-\tau} = 1$. Beside the degree distribution, power law graphs are generated according to the Havel-Hakimi algorithm [12][17] and the Barabási-Albert model [1]. A Havel-Hakimi graph is constructed by successively connecting the node of highest degree to other nodes of highest degree, resorting remaining nodes by degree, and repeating the process. The resulting graph has a high degree-associativity. Hence, the Havel-Hakimi power law graph has a "dense core" and is more tree-like compared to $K_N$ and D-lattice. The main difference between the Havel-Hakimi algorithm and the Barabási-Albert model is that the first can also be applied to other degree distributions (with different exponent $\tau$, while $\tau_{BA} = 3$). Moreover, it shows already power law behavior for small $N$, while power law behavior in the Barabási-Albert model is only observed for large $N$. Furthermore, large bias occurs mainly in graphs with high average degree [32]. For a same number of a nodes $N$, the Havel-Hakimi graph for $\tau < 3$ has, on average more links (or higher $E[D]$) than the Barabási-Albert power law graph, which makes the former more attractive to study the effects of link weight tuning.

The link weights are chosen *independently* of the topology. Although in some biological networks, the link weight or strength of a link is coupled to the structure of the underlying topology, in many man-made large infrastructures such as the Internet and WWW, the link weight structure can be chosen independently. The latter allows us to control or steer transport in the network as shown in [3]. As assumed before, these undirected link weights are i.i.d. and additive which we deem a reasonable approximation in many large networks, with the exceptions of wireless networks[4]. As mentioned before, the $SPT$ is mainly sensitive to the smaller, non-negative link weights. The probability distribution of the link weights around zero will dominantly influence the properties of the resulting shortest path tree as well as that of $G_{\cup spt}$. The simplest distribution of the link weight $w$ with a distinct different behavior for small values is the polynomial distribution (1). The motivation to select a polynomial distribution is given earlier [2]. The exponent $\alpha$ is called the extreme value index of the probability distribution of $w$. When $\alpha \to \infty$, it follows from (1) that $w = 1$ almost surely for all links. When

[3] By definition, $\Pr[G_{p_c}$ is connected$] = \frac{1}{2}$ when $N \to \infty$. Hence, roughly half of the Erdös-Rényi random graphs $G_p(N)$ are connected if $p \sim p_c$ for large $N$.

[4] All nodes in the radio-range of some sending node (or base-station) are correlated by (a) the nature of electromagnetic waves and (b) wireless MAC protocols.

$\alpha \to 0$, all links will be close to 0, but, relatively, they differ significantly with each other. When $\alpha = 1$, the polynomial distribution (1) becomes a uniform distribution where link weights are regular as explained in Section III-C. By varying the exponent $\alpha$ over all non-negative real values, any extreme value index can be attained and a large class of corresponding $SPT$ and $G_{\cup spt}$ can be generated.

For each simulation, $10^4$ iterations are carried out. Within each iteration, the specified underlying topology is generated randomly and the polynomial link weights with parameter $\alpha$ are assigned independently to each link in the graph. The $G_{\cup spt}$ is found by calculating the shortest paths between all pairs of nodes with Dijkstra's algorithm [9] and with the high precision Dijkstra algorithm [3].

## V. Properties of $G_{\cup spt}$ with $\alpha = 1$

For $\alpha = 1$ in (1), we obtain the uniform distribution on $[0, 1]$. Three classes of topologies are considered: the complete graph $K_N$, for which exact results exist, the 2D-lattice and the power law graph. In both the simulation and the analysis, the average number of links $E[L_o]$ and the degree distribution of $G_{\cup spt}$ is examined.

### A. The average number of links

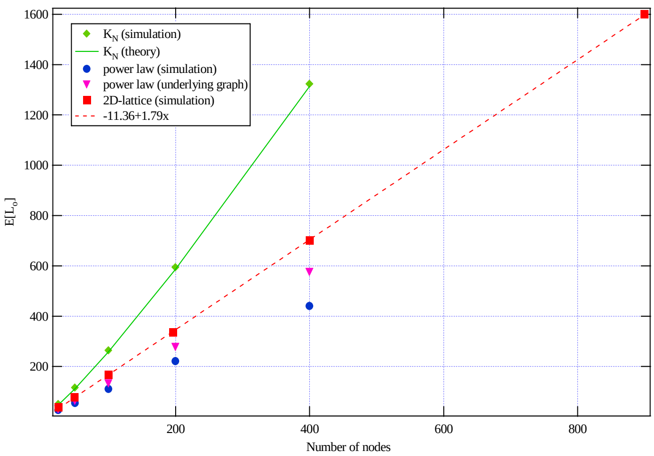


Fig. 1. Average number of links in the $G_{\cup spt}$.

Figure 1 shows that the simulation and the theory (6) of the average number of links $E[L_o]$ in $G_{\cup spt}$ of the complete graph $K_N$ nicely match.

Simulations in Figure 1 also show that, in 2D-lattices, the average number of the observed links via $G_{\cup spt}$ is linear with the number of nodes $N$. Hence, with (8)

$$\Pr[P^*_{i \to j} = l\,(i \to j)] \cdot L = cN \tag{14}$$

where fitting the simulations yields $c = 1.79$, while (10) implies that $1 \leq c \leq 2$.

We will further determine the constant $c$ in (14). As shown in Figure 2 (thick line), the link $l((x_3, y_2), (x_3, y_3))$ can be observed via $G_{\cup spt}$, if it is the shortest path between node $(x_3, y_2)$ and node $(x_3, y_3)$. The link $l$ can be dominated by a shorter path such as $P^1_{h=3} = P_{(x_3,y_2)\to(x_2,y_2)\to(x_2,y_3)\to(x_3,y_3)}$ or $P^2_{h=3} = P_{(x_3,y_2)\to(x_4,y_2)\to(x_4,y_3)\to(x_3,y_3)}$ which are the only three hops paths between $(x_3, y_2)$ and $(x_3, y_3)$. Furthermore, each link in these two paths can again be dominated by a three hops path, which is shown in bold line in Figure 2. Recursively, these links can also be dominated further. Each level of domination results in a longer hopcount of the shortest path between two adjacent nodes. If we define $w\,(P_{h=3})$ as the weight of a three hops path, then $w\,(P_{h=3})$ is the sum of three independent uniform random variables and its *pdf* [11] is

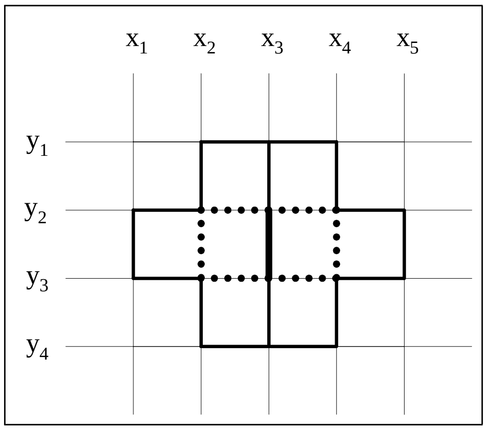


Fig. 2. Link Observability in a 2D-lattice.

$$f_{w(P_{h=3})}(x) = \sum_{j=0}^{3} \binom{3}{j} \cdot (-1)^j \cdot \frac{(x-j)^2}{2} 1_{(x-j)\geq 0}$$

and

$$\Pr[w\,(P_{h=3}) \leq x] = \int_0^x f_{w(P_{h=3})}(y)dy = \frac{x^3}{6}$$

Since the two 3 hops paths are independent, the probability that one of these two three hops paths is smaller than $x$ is

$$\Pr[\min_{1\leq k\leq 2} w\,(P^k_{h=3}) \leq x] = 1 - (1 - \Pr[w\,(P_{h=3}) \leq x])^2$$
$$= \frac{x^3}{3} - \frac{x^6}{36}$$

For any link, the probability that there exists a three hops path shorter than this direct link $w(l) \in [0, 1]$ is

$$\Pr[\min_{1\leq k\leq 2} w\,(P^k_{h=3}) \leq w(l)]$$
$$= \int_0^1 \Pr[\min_{1\leq k\leq 2} w\,(P^k_{h=3}) \leq x]dx = 0.08$$

If there exists a three hops path shorter than the direct link, the direct link is definitely not observed. However, the hopcount of the shortest path between these two adjacent nodes can be longer than 3, since links in the three hops path can be dominated on their turn. When both the three hops paths are longer than the direct link, the direct link is not necessary the shortest, because paths with hopcount larger than 3 can be even shorter, which is, however, very unlikely to happen for uniform i.i.d. link weights. Hence, the probability that a link can be observed can be approximated by the upper bound:

$$\lim_{N\to\infty}\Pr[P^*_{i\to j}{=}l(i\to j)] \leq 1{-}\Pr[\min_{1\leq k\leq 2} w(P^k_{h=3}){\leq}\, w(l)]{=}0.92 \quad (15)$$

where $N \to \infty$ means that each node has four neighbors and links at the border are not taken into account. Combining (15) and (14) results in

$$\lim_{N\to\infty} \Pr[P^*_{i\to j} = l(i \to j)] = \lim_{N\to\infty} \frac{c \cdot N}{L} = \lim_{N\to\infty} \frac{c}{2} \cdot \frac{\sqrt{N}}{\sqrt{N}-1} = \frac{c}{2}$$

and, in $c = 1.84$ which is close to the simulation result $c = 1.79$. In addition, (15) also shows that $L_o \simeq L$. In other words, the observable 2D-lattice is very near to the substrate, in contrast to $K_N$ as observed from (11).

In Figure 1, the underlying power law graph with $\tau = 2.4$ is shown to be sparse and $E[L_o]$ of the corresponding $G_{\cup spt}$ approaches $N - 1$. It is natural that the $G_{\cup spt}$ is close to a tree, because the sparse underlying power law graph is already tree-like as explained in Section IV.

### B. The degree distribution

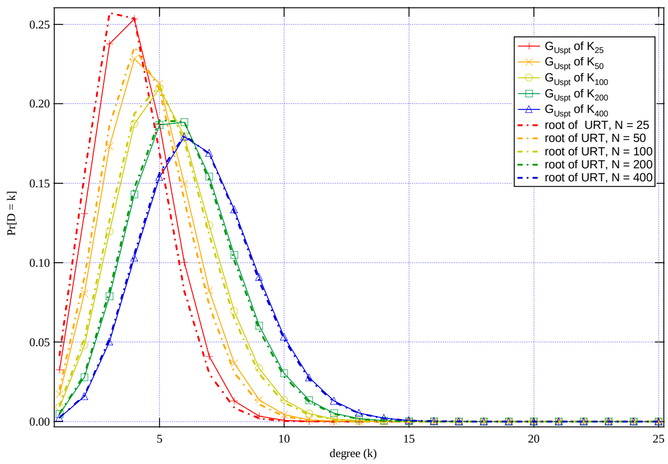


Fig. 3. Degree distribution of $G_{\cup spt}$ in $K_N$ and degree distribution of the root of the corresponding $URT$.

The simulation result of the degree distribution of $G_{\cup spt}$ in $K_N$ is shown in Figure 3 together with the result calculated by Theorem 4. As $N$ increases, the degree distribution of $G_{\cup spt}$ tends to that of the $URT$ root. These simulations support that $SPT_{K_N} \simeq URT$ with any regular link weights as explained in Section III-C.

The degree distribution of $G_{\cup spt}$ in 2D-lattices is shown in Figure 4. The good approximation (15) encourages us to further simplify the analysis for the degree distribution in a finite 2D-lattice. There are three kinds of nodes in a lattice: $n_2 = 4$ nodes with degree $d = 2$; $n_3 = 4 * (\sqrt{N} - 2)$ nodes with degree $d = 3$; $n_4 = N - 4 * (\sqrt{N} - 1)$ nodes with degree $d = 4$. Two kinds of links exist: (a) links at the border, that have a probability $p_2 \approx 1 - \int_0^1 \Pr[P_{h=3} \leq x]dx = 0.96$ to be observed via $G_{\cup spt}$ and (b) central links, that have a probability $p_1 \approx 0.92$ to be the shortest path between its end nodes. We assume that the observation of one link via $G_{\cup spt}$ will not influence the probability of other links being observed, since the hopcount of the shortest path in a 2D-lattice is usually large and a link with small link weight is unlikely to attract the other shortest paths to pass through it. Then the following can be obtained:

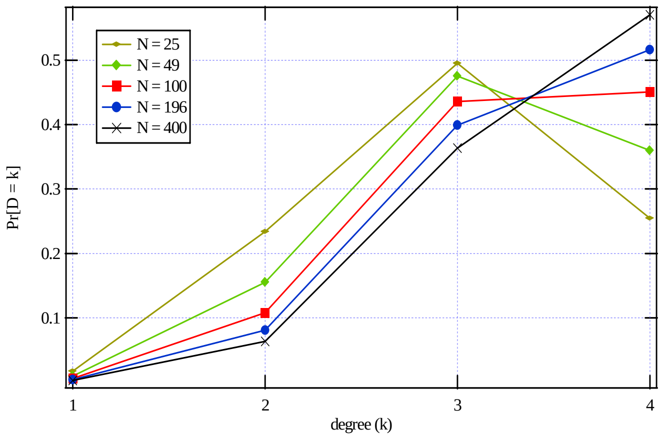


Fig. 4. Degree distribution of $G_{Uspt}$ in 2D-lattices.

$$\begin{aligned}
&\Pr[d=1] = 1 - \Pr[d=2] - \Pr[d=3] - \Pr[d=4] \\
&\Pr[d=2] = \tfrac{1}{N}\tbinom{4}{2} \cdot n_4 \cdot p_1^2(1-p_1)^2 + \\
&\tfrac{1}{N}\left(n_3 \cdot \left(p_2^2(1-p_1) + 2p_1p_2(1-p_2)\right) + n_2 \cdot p_2^2\right) \\
&\Pr[d=3] = \tfrac{1}{N}\left(n_3 \cdot p_1 p_2^2 + \tbinom{4}{1} \cdot n_4 \cdot p_1^3(1-p_1)\right) \\
&\Pr[d=4] = \tfrac{1}{N} \cdot n_4 \cdot p_1^4
\end{aligned} \quad (16)$$

We explain $\Pr[d = 3]$. A node with degree 3 in the $G_{\cup spt}$ is either a node with degree 3 in the 2D-lattice with all three links being observed, or a node with degree 4 in the 2D-lattice with three of its four links being observed. With the set (16), we calculate the degree distribution of $G_{\cup spt}$ in a 2D-lattice with $N = 400$ nodes and compare those values with the simulation result in the table below,

| | $\Pr[d = 1]$ | $\Pr[d = 2]$ | $\Pr[d = 3]$ | $\Pr[d = 4]$ |
|---|---|---|---|---|
| Simulation | 0.003 | 0.063 | 0.36 | 0.57 |
| theory | 0.004 | 0.062 | 0.354 | 0.580 |

which, again, shows a good agreement.

The degree distribution of $G_{\cup spt}$ in Havel-Hakimi power law graph substrates with $\tau = 2.4$ is shown in Figure 5. The degree distribution of a power law graph is $\Pr[d = i] = ci^{-\tau}$ (by definition) which is shown in bold line in the figure for $N = 100$. Nodes with degree 1 in the underlying graph must remain the same in the $G_{\cup spt}$ in order for $G_{\cup spt}$ to be connected. A node with higher degree in the underlying graph may have only one link in $G_{\cup spt}$, thus, degree 1 in $G_{\cup spt}$. Hence, as shown in Figure 5, compared to the degree distribution of the underlying topology, $\Pr[d = 1]$ in the $G_{\cup spt}$ increases while $\Pr[d = i]$ for $i > 1$ decreases. However, such difference is not substantial. The overlay network exhibits a degree distribution visually similar to the underlying graph, because as shown in Section V-A, the underlying topology is sparse and already

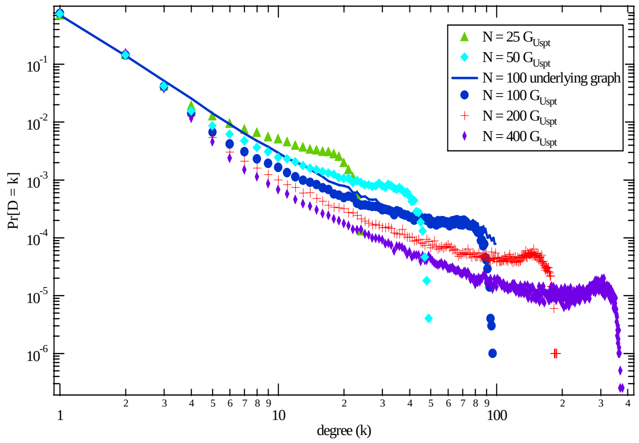


Fig. 5. Degree distribution of $G_{\cup spt}$ in Havel-Hakimi power law graphs with $\tau = 2.4$.

tree-like. A similar conclusion is reached for the Barabási-Albert preferential attachment model as illustrated in Figure 6, where we have chosen a larger number of $m = 4$ links that is attached at each time which is equivalently to a large link density. This demonstrates that the overlay on top of a power law graph is very close to the substrate in terms of degree distribution, even for $m = 4$. These observations are consistent with results in [18], where even a subgraph of $G_{\cup spt}$ has similar degree distribution as that of the underlying power law graph with $w = 1$ or $\alpha \to \infty$ [23].

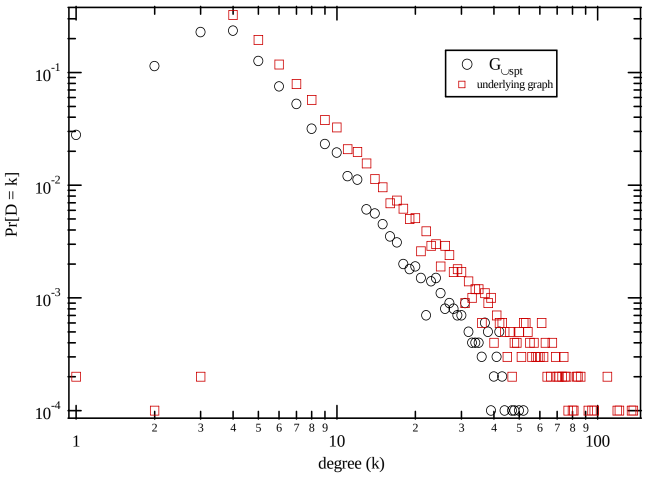


Fig. 6. Degree distribution in Barabási-Albert power law graphs where $N = 800$ and the initial number to start the preferential growth is $m_0 = m = 4$

## VI. Properties of $G_{\cup spt}$ with varying $\alpha$

Van Mieghem and Magdalena [3] have found that, by tuning the extreme value index $\alpha$ of a link weight distribution, a phase transition occurs around a critical extreme value index $\alpha_c$. The critical extreme value index $\alpha_c$ is defined as $F_T(\alpha_c) = \frac{1}{2}$ where $F_T(\alpha) = \Pr\left[G_{\cup spt(\alpha)} = MST\right]$. When $\alpha > \alpha_c$, the overlay $G_{\cup spt(\alpha)}$ contains more than $N-1$ links whereas for $\alpha < \alpha_c$, all transport traverses a critical backbone consisting of $N-1$ links, which is the minimum spanning tree $MST$, as follows from Theorem 1. Here, we extend the analysis of [3] in two ways: (a) we include, besides the complete graphs $K_N$ and 2D-lattices, also 3-lattices and Havel-Hakimi power law graphs; (b) by a spectral analysis, we further obtain insights in the structure of the overlay $G_{\cup spt(\alpha)}$.

### *A. Phase transition in the $G_{\cup spt(\alpha)}$ structure*

Instead of the number of links in $G_{\cup spt(\alpha)}$, when $\alpha$ is small, we study the probability that the overlay $G_{\cup spt(\alpha)}$ is a tree. As

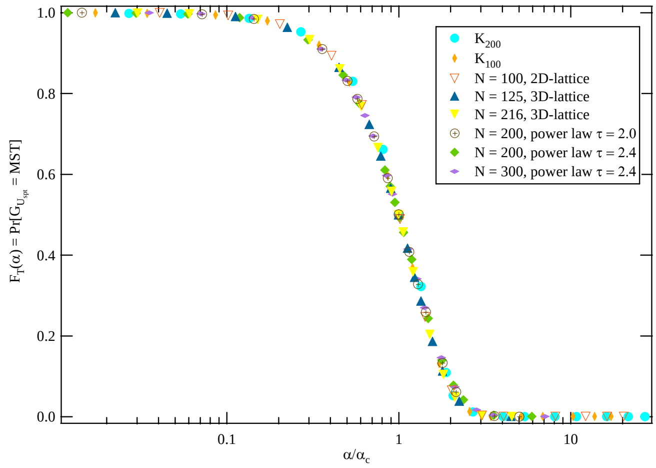


Fig. 7. The probability distribution $F_T(\alpha)$ as a function of the normalized $\alpha/\alpha_c$.

shown in Figure 7, normalized by $\alpha_c$, the same phase transition curve is observed for all these three topologies. As $\alpha$ increases, the transport is more likely to traverse over more links and the overlay $G_{\cup spt(\alpha)}$ will less likely become a tree. These additional simulations over those reported in [3] strengthen the belief that the curve $F_T(\alpha) \approx 2^{-\left(\frac{\alpha}{\alpha_c}\right)^2}$ is universal for all graphs that are not trees.

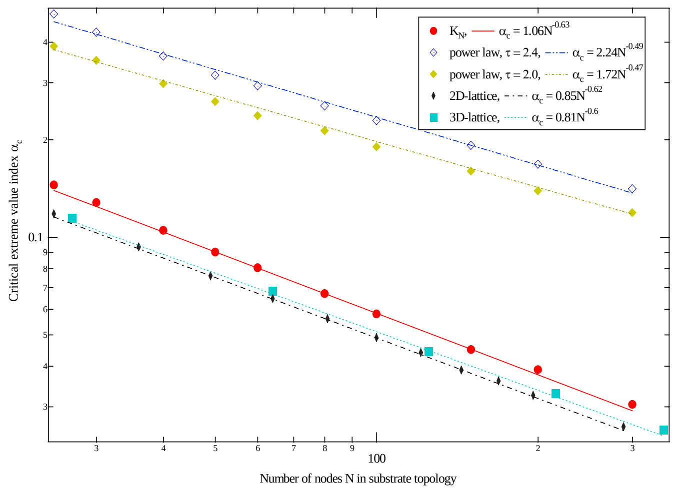


Fig. 8. The critical extreme value $\alpha_c$ as a function of $N$.

For each substrate topology, the critical extreme value $\alpha_c$ is shown in Figure 8 as a function of $N$ on a log-log scale.

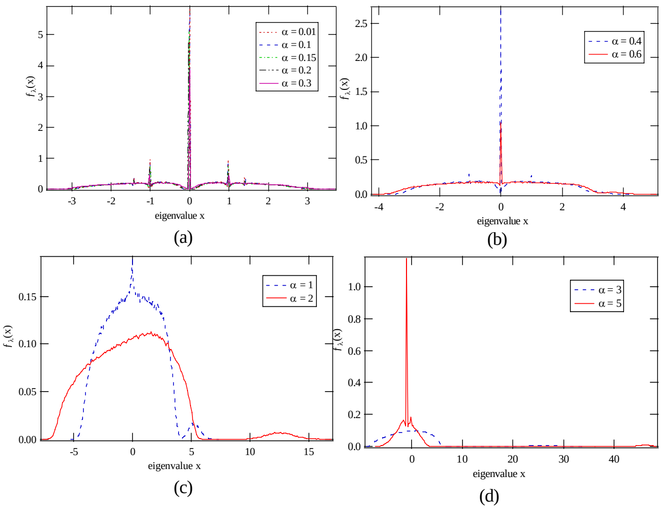


Fig. 9. The spectrum of the $G_{\cup spt(\alpha)}$ in the complete graph $K_{50}$: (a) $\alpha \in [0.01, 0.3]$; (b) $\alpha = 0.4, 0.6$; (c) $\alpha = 1, 2$;(d) $\alpha = 3, 5$.

Each curve is fitted with a line and indicates that $\alpha_c \simeq bN^{-\beta}$ where $b$ and $\beta$ depend on the underlying substrate. In spite of the fact that this phase transition is constructed, the exponent $\beta$ seems to lie in the interval $[\frac{1}{2}, \frac{2}{3}]$, which agrees surprisingly well with critical exponents observed in nature (see e.g. [31]). As explained in [3], it is computationally difficult to determine $\alpha_c$ for large $N$. This limits the extent to which a seemingly power law can be observed. It may also explain why, in Figure 8, the value of $\alpha_c$ at higher $N$ are likely less accurate such that the apparent deviation of a power law may be due to numerical inaccuracies. However, we do not have any theoretical argument why $\alpha_c$ should obey a strict power law.

The higher the $\alpha_c$-curve in Figure 8, the faster $G_{\cup spt(\alpha)}$ tends to a tree if $\alpha$ decreases. The power law graphs seem to possess the highest $\alpha_c$ because the power law graph is already tree-like. Although when $\alpha = 1$, the overlay $G_{\cup spt(\alpha)}$ of the complete graph $K_N$ contains more links than that of the 2D-lattice (compare (11) with (10)), Figure 8 illustrates that, since $\alpha_c\,(K_N) > \alpha_c\,(\text{2D-lattice})$, the overlay $G_{\cup spt(\alpha)}$ of the complete graph $K_N$ will tend faster to a tree if $\alpha$ decreases. We can only give a speculative explanation. When $\alpha$ is very small around the $\alpha_c$, link weights are small, but they differ significantly from each other. In fact, if $\alpha \rightarrow 0$, the ratio $\frac{\sqrt{\text{Var}[w]}}{\mathbb{E}[w]} \sim \frac{1}{\sqrt{\alpha}}$ diverges which means that, in this limit, the link weights possess extremely strong fluctuations. Since the number of possible trees in $K_N$ is much larger than that in the 2D-lattice, it is more probable to find a spanning tree only composed of these extremely small links and that tree is the MST.

### B. The spectrum of the adjacency matrix of $G_{\cup spt(\alpha)}$

The spectrum, the eigenvalues of the adjacency matrix, of $G_{\cup spt(\alpha)}$ in the complete graph $K_{50}$ are displayed in Figure 9. For $K_{50}$, the critical extreme value index is $\alpha_c = 0.09$. Figure 7 shows that the onset of the phase transition is somewhere between $3\alpha_c$ and $4\alpha_c$. Figure 9 (a) shows that, for $\alpha \in [0.01, 0.3]$, the corresponding $G_{\cup spt(\alpha)}$ have almost the same spectrum, except that, the peaks at $\pm 1.4$ diminish as $\alpha$ increases. The nodes with small degrees are most likely responsible [24] for the delta peak at[5] $\lambda = 0$. For example, the local configurations with two and more dead-end nodes produce eigenvalues $\lambda = 0$, where the dead-end node is a node with degree 1. The corresponding eigenvectors have non-zero components only at the dead-end nodes [25][26]. The spectrum of a tree is symmetric [27], because any tree is a bi-partite graph and any bi-partite graph is symmetric around $\lambda = 0$. In Figure 9 (b) when $\alpha = 0.4$, these two peaks at $\pm 1.4$ are smoothed out and the spectrum is not symmetric. It indicates that when $\alpha$ is smaller than the onset value of the phase transition (case (a) in Figure 9), the $G_{\cup spt(\alpha)}$ seem to possess similar topological, tree-like structure. Since very few trees can be uniquely specified by their spectrum [28], the spectrum is not well suited to reveal the specifics of the MST. The spectrum of the $G_{\cup spt(1)}$ in $K_{50}$ with regular link weights $\alpha = 1$ illustrated in Figure 9 (c) is close to the spectrum of a random graph according to the Wigner's Semicircle Law [16][11, Appendix B]. This correspondence is an additional illustration of Conjecture 1. When $\alpha$ is large, link weights are ineffective in that $\lim_{\alpha\rightarrow\infty} G_{\cup spt(\alpha)} = G_{\text{substrate}}$. The spectrum of $G_{\cup spt(\alpha)}$ in Figure 9 (d) is, indeed, close to the spectrum of substrate $K_N$, that has $N-1$ eigenvalues at $-1$ and 1 eigenvalue at $N-1$. While peaks in the spectrum reflect structure and regularity in the graph, a bulk almost symmetrical around zero, which ultimately tends to a semicircle, points to uncorrelated randomness. The latter is a characteristic property of an Erdös-Rényi random graph. Figure 9 thus shows, as a function of $\alpha$, transitions of $G_{\cup spt(\alpha)}$ between two graph types, a tree and the complete graph, with apparent maximum randomness for regular link weights ($\alpha = 1$).

The spectrum of the $G_{\cup spt(\alpha)}$ in a 2D-lattice with 49 nodes are displayed in Figure 10. When $\alpha$ is small, as shown in Figure 10 (a) and (b), the transition of the spectrum of $G_{\cup spt(\alpha)}$ is similar to that in the complete graph. As studied in Section V-A, on average, 92% of the links in the underlying 2D-lattice can be observed via the overlay $G_{\cup spt}$ when $\alpha = 1$. The spectrum of a 2D-lattice with $N$ nodes [27] comprises the eigenvalues

$$\lambda_{ij} = 2\cos\frac{2\pi}{\sqrt{N}}i + 2\cos\frac{2\pi}{\sqrt{N}}j \qquad i, j \in \{1, \ldots, \sqrt{N}\}$$

which correspond to the peaks in the spectrum of $G_{\cup spt(\alpha)}$ with $\alpha \geq 2$. Compared to the complete graph, in a 2D-lattice, the overlay $G_{\cup spt}$ approaches the underlying topology at a smaller $\alpha$-values, $\alpha \geq 2$ for the 2D-lattice while $\alpha \geq 5$ for $K_N$.

The spectrum of the $G_{\cup spt(\alpha)}$ in a Havel-Hakimi power law graph is almost the same as that of the underlying graph for any $\alpha$, because the underlying graph is already close to a tree. For the spectrum of a Barabási-Albert power law graph, we

[5] In general, each time when two rows in the adjacency matrix $A$ are the same, the rank of $A$ decreases with 1, which is equivalent to an increase in the multiplicity of the eigenvalue $\lambda = 0$, since $\det A = \prod_{j=1}^{N} \lambda_j$.

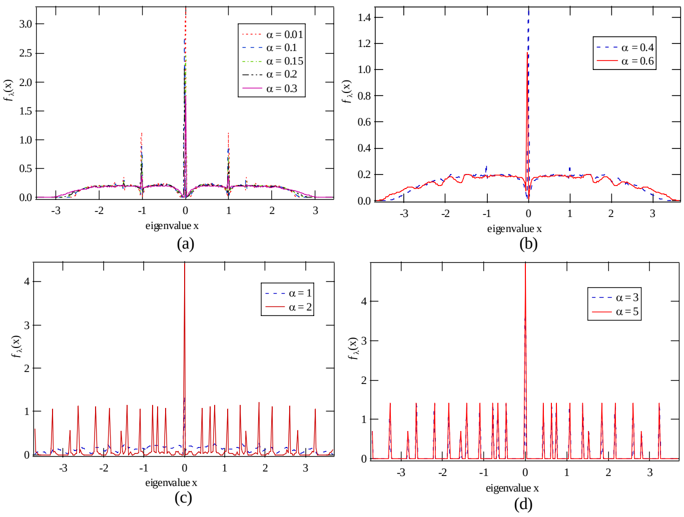


Fig. 10. The spectrum of the $G_{\cup spt}(\alpha)$ in a 2D-lattice with 49 nodes: (a) $\alpha \in [0.01, 0.3]$; (b) $\alpha = 0.4, 0.6$; (c) $\alpha = 1, 2$; (d) $\alpha = 3, 5$.

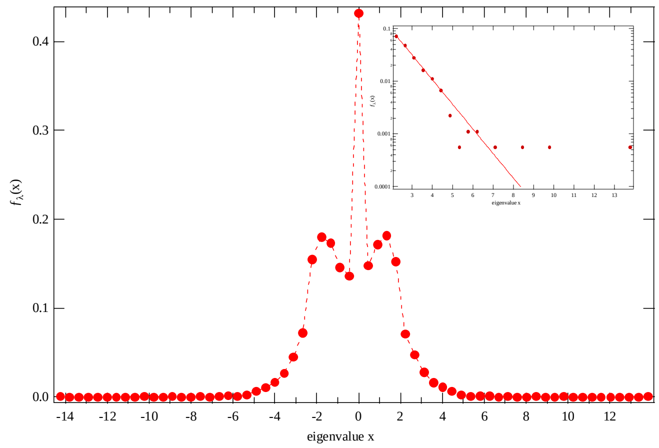


Fig. 11. The spectrum of $G_{\text{RIPE}}$ measured by RIPE. The insert graph is the tail part, fitted with $f_\lambda(x) \simeq 0.8e^{-x}$. The theory of [24] is not applicable to exponentially decaying tails in $f_\lambda(x)$.

refer to [29]. As mentioned before, since the spectrum for most trees is not unique [28], a spectral analysis of tree-like graphs is not the best way to deduce specific properties.

## VII. Conclusion

The union of all shortest path trees $G_{\cup spt}$ constitutes the observable part of a network provided traffic flows follow shortest paths. In this paper, we have studied two properties of the $G_{\cup spt}$, the average number of links $E[L_o]$ and the degree distribution, and simulated the spectrum of the adjacency matrix of $G_{\cup spt(\alpha)}$ as a function of the extreme value index of the link weight structure. Different underlying topologies and the link weight structure were treated independently. The minimum spanning tree belongs to the $G_{\cup spt}$. Any link $l\,(i \rightarrow j)$ with link weight $w(i \rightarrow j)$ in the $G_{\cup spt}$ must be the shortest path between $i$ and $j$ and if a link is the shortest path between its end nodes, it must belong to the $G_{\cup spt}$. Most of the theory is based on these two points. Apart from the Theorems and Corollaries presented, the Conjecture 1 and the seemingly universality of $F_T(\alpha) \approx 2^{-\left(\frac{\alpha}{\alpha_c}\right)^2}$ in the phase transition appearing in the structure of $G_{\cup spt}$ are considered important new findings. For example, Conjecture 1 which has assumed an i.i.d. link weight structure, claims the appearance of the random graph $G_p(N)$ in many application such as, for example, peer to peer networks [20] and ad-hoc networks [19]. The universality of $F_T(\alpha)$ in the phase transition points to the possibility to control the network structure or to steer or balance transport by tuning the link weight structure.

The overlay $G_{\cup spt}$ is, actually, the *maximally* measurable part of the substrate topology. For example, the RIPE traceroute measurement configuration [21] only measures the union $G_{\cup_m spt}$ of shortest paths between each pair of a small group of $m << N$ nodes, while the number of nodes in the underlying graph $N$ is much larger. Considerable attention has been devoted to the properties of graphs derived from Internet measurements. But how accurate does the measured subgraph reflect the underlying graph [18]?

The spectrum of the topology $G_{\text{RIPE}}$ measured by RIPE [22] is shown in Figure 11, and is very akin to that measured on Planet lab in Figure 12. More details are in the table below

| | RIPE | PlanetLab |
|---|---|---|
| date | 9-18-2005 | 11-10-2004 |
| $m$ | 70 | 79 |
| #spts in union | 67 | 76 |
| $N$ | 4058 | 4214 |
| $L$ | 6151 | 6994 |
| $\Pr[D=k]$ | $\sim e^{-0.39k}$ | $\sim e^{-0.44k}$ |

Each of the $m$ testboxes acts as a source and sends traffic to other testboxes. After removing error measurements in the trace-routes, the overlay $G_{\text{RIPE}}$ and $G_{\text{PlanetLab}}$ are constructed as the union of (only) #spts trees. Both $G_{\text{RIPE}}$ and $G_{\text{PlanetLab}}$ are subgraphs of $G_{\cup spt}$ on top of the underlying Internet topology.

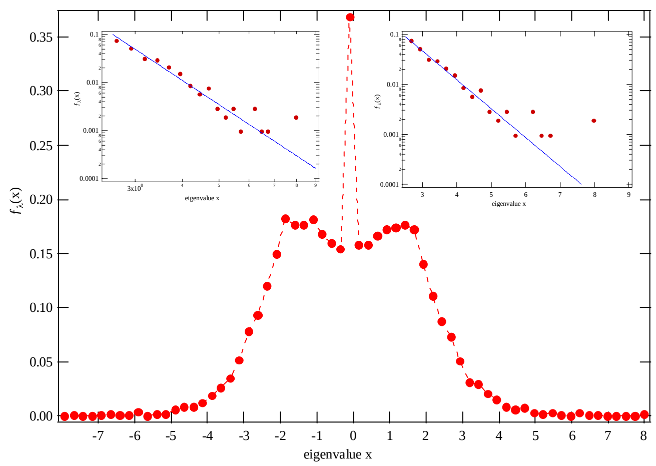


Fig. 12. The spectrum of the overlay measured on PlanetLab. The deep tail region can be fitted both by an exponential ($f_\lambda(x) \simeq 2.5e^{-1.3x}$) and a power law ($f_\lambda(x) \simeq 1.5\lambda^{-5.2}$). Applying the conversion $f_\lambda(x) \simeq 2xf_D(x^2)$ from [24] to the power law results in a degree pdf $f_D(x) \sim x^{-3.1}$. The power law exponent $\gamma = 3.1$ exceeds the commonly accepted $\gamma_{\text{Internet}} \in [2, 2, 2.5]$.

The spectra of $G_{\text{RIPE}}$ and $G_{\text{PlanetLab}}$ seem to give support for the Conjecture 1, since the partial overlay $G_{\cup_m spt}$ ($m << N$) seems approximately close to $G_{\cup spt}$ (see Figure 9 (c)) and reveals more features of the overlay than that of the underlying topology, which is overwhelmingly shown in the literature to belong to the class of power law or scale free graphs. The spectrum of scale-free networks exhibit a power law tail in the region of large eigenvalues [29][24]. The spectra of $G_{\text{RIPE}}$ and $G_{\text{PlanetLab}}$, however, may possess an exponential tail, most likely because the value of $m$ is still not large enough to cover $G_{\cup spt}$ sufficiently and to observe the power law behavior [30]. The relationship between a partial overlay $G_{\cup_m spt}$, the complete overlay $G_{\cup_N spt}$ and the underlying topology or substrate stands on the agenda for future work. In view of the potentially strong biases in current measurements of the Internet topology as mentioned in Section I, we deem it crucial to understand how accurately $G_{\cup spt}$ can be approximated by $G_{\cup_m spt}$. In particular, given a property $P$ of a graph $G$ and an accuracy $\epsilon$, what is the threshold $m_c(\epsilon)$ such that $|P(G_{\cup spt}) - P(G_{\cup_m spt})| < \epsilon$ for all $m > m_c(\epsilon)$.

As a final remark, Conjecture 1 implies that the observed network will have a low maximum degree and a Poisson degree distribution, contrary to the view, promoted by some papers [18][33][32], claiming that the bias may lead to an observed power law degree distribution irrespective of the degree distribution of the substrate. Most published work on sampling bias focuses on unweighted graphs and the bias originates purely from the sampling methods (such as the union of paths from a small set of sources to a relatively larger set of destinations). Here, we study the "bias" introduced by the link weight structure of the substrate. The overlay constructed as the union of shortest paths between all node pairs is exactly the same as the substrate if the substrate is unweighted ($\alpha \to \infty$ case). In weighted graphs, links with high weight rarely appear in the observed network.

## Acknowledgement

We would like to thank Xiaoming Zhou for providing the measurement results. This research was supported by the Netherlands Organization for Scientific Research (NWO) under project number 643.000.503 and partially supported by European Union CONTENT NoE (FP6-IST-038423) (www.ist-content.eu).

## Appendix

### A. Proof of Theorem 2

By the law of total probability, we can write

$$\Pr[P^*_{i\to j} = l(i \to j)] = \int_0^\infty f_{w(i\to j)}(x) \times \Pr[P^*_{i\to j} = l(i\to j) | w(i \to j) = x] dx \tag{17}$$

where $f_{w(i\to j)}(x)$ is the pdf of the weight of a link in the graph.

Let us assume that $P^*_{i\to j,h>1}$ is the shortest path among paths with more than 1 hop and let us denote the path weight by $w(P^*_{i\to j,h>1})$. Provided that the direct link between $i$ and $j$ exists, then the shortest path equals the direct link if its weight is smaller than any path with more than one hop and vice versa,

$$\begin{aligned} P^*_{i\to j} = P^*_{i\to j,h>1} \cdot 1_{\{w(i\to j) \geq w(P^*_{i\to j,h>1})\}} \\ + l(i\to j) \cdot 1_{\{w(i\to j) < w(P^*_{i\to j,h>1})\}} \end{aligned}$$

from which the conditional probability

$$\begin{aligned} \lim_{\Delta x \to 0} \Pr[P^*_{i\to j} = l(i\to j) | \{x \leq w(i\to j) \leq x + \Delta x\} \\ \cap \{l(i\to j) \text{ exists}\}] = \Pr[w(P^*_{i\to j,h>1}) > x] \end{aligned}$$

is immediate. Since the link weights are i.i.d. and also independent of a specific link, the conditional probability $P_c$ is

$$\begin{aligned} P_c &= \Pr[P^*_{i\to j} = l(i\to j) | \{x \leq w(i\to j) \leq x + \Delta x\} \\ &\quad \cap \{l(i\to j) \text{ exists}\}] \\ &= \frac{\Pr\left[P^*_{i\to j} = l(i\to j) | \{x \leq w(i\to j) \leq x + \Delta x\}\right]}{\Pr[l(i\to j) \text{ exists}]} \end{aligned} \tag{18}$$

where the last step follows from the fact that the event $\{P^*_{i\to j} = i \to j\}$ is contained in the event $\{l(i\to j) \text{ exists}\}$. Hence,

$$\begin{aligned} &\lim_{\Delta x\to 0} \Pr\left[P^*_{i\to j} = l(i\to j) | \{x \leq w(i\to j) \leq x + \Delta x\}\right] \\ &= \Pr[l(i\to j) \text{ exists}] \Pr[w(P^*_{i\to j,h>1}) > x] \end{aligned} \tag{19}$$

The sample space $\Omega$ consists of four mutually exclusive events:

$$\begin{aligned} &\{w(P^*_{i\to j,h>1}) \leq x, w(i\to j) \leq x \cap l(i\to j) \text{ exists}\} \\ &\{w(P^*_{i\to j,h>1}) > x, w(i\to j) \leq x \cap l(i\to j) \text{ exists}\} \\ &\{\{w(i\to j) > x \cap l(i\to j) \text{ exists}\} \cup \{l(i\to j) \text{ does not exist}\}, \\ &w(P^*_{i\to j,h>1}) \leq x\} \\ &\{\{w(i\to j) > x \cap l(i\to j) \text{ exists}\} \cup \{l(i\to j) \text{ does not exist}\}, \\ &w(P^*_{i\to j,h>1}) > x\} \end{aligned}$$

Therefore, the related probability measure is

$$\begin{aligned} &\Pr[w(P^*_{i\to j}) \leq x] \\ &= \Pr[w(P^*_{i\to j,h>1}) \leq x, \{\{w(i\to j) > x \cap l(i\to j) \text{ exists}\} \\ &\quad \cup \{l(i\to j) \text{ does not exist}\}] \\ &+ \Pr[w(P^*_{i\to j,h>1}) \leq x, w(i\to j) \leq x \cap l(i\to j) \text{ exists}] \\ &+ \Pr[w(P^*_{i\to j,h>1}) > x, w(i\to j) \leq x \cap l(i\to j) \text{ exists}] \\ &= \Pr[w(P^*_{i\to j,h>1}) \leq x, \{\{w(i\to j) > x \cap l(i\to j) \text{ exists}\} \\ &\quad \cup \{l(i\to j) \text{ does not exist}\}] \\ &+ \Pr[w(i\to j) \leq x] \Pr[l(i\to j) \text{ exists}] \end{aligned} \tag{20}$$

where in the last step, we have again used the law of total probability. Further, the event

$$\begin{aligned} &\{\{w(i\to j) > x \cap l(i\to j) \text{ exists}\} \\ &\quad \cup \{l(i\to j) \text{ does not exist}\}, w(P^*_{i\to j,h>1}) \leq x\} \\ &= \{w(P^*_{i\to j,h>1}) \leq x\} \cap \{\{w(i\to j) > x \cap l(i\to j) \text{ exists}\} \\ &\quad \cup \{l(i\to j) \text{ does not exist}\}\} \end{aligned}$$

and the first two events are independent because link weights are independently and identically distributed and $l(i\to j)$ is different from $P^*_{i\to j,h>1}$. Hence, (20) reduces to

$$\begin{aligned}&\Pr[w(P^*_{i\to j})\le x]\\ =&(\Pr[w(i\to j)>x]\cdot\Pr[l(i\to j)\text{ exists}]\\ &+\Pr[l(i\to j)\text{ does not exist}])\cdot\Pr[w(P^*_{i\to j,h>1})\le x]\\ &+\Pr[w(i\to j)\le x]\Pr[l(i\to j)\text{ exists}]\end{aligned}$$

from which

$$\begin{aligned}&\Pr[w(P^*_{i\to j,h>1})\le x]\\ &=\frac{\Pr[w(P^*_{i\to j})\le x]-\Pr[w(i\to j)\le x]\cdot p_{ij}}{\Pr[w(i\to j)>x]\cdot p_{ij}+1-p_{ij}}\end{aligned}$$

where $p_{ij}=\Pr[l(i\to j)\text{ exists}]$. Finally, the conditional probability (19) becomes

$$\begin{aligned}&\lim_{\Delta x\to 0}\Pr\left[P^*_{i\to j}=l(i\to j)|\{x\le w(i\to j)\le x+\Delta x\}\right]\\ &=\frac{p_{ij}\cdot\Pr[w(P^*_{i\to j})>x]}{\Pr[w(i\to j)>x]\cdot p_{ij}+1-p_{ij}}\end{aligned}\tag{21}$$

and substituted into (17) leads to (2). □

### B. Proof of Corollary 1

The event $\{P^*_{i\to j}=l(i\to j)\}$ is equivalent to the event $\{H_N=1\}$ because the direct link corresponds to a one hop shortest path. The second equality is demonstrated as follows. Since $\frac{d}{dx}\Pr[w(i\to j)>x]=-f_{w(i\to j)}(x)$, we can write (2) as

$$\begin{aligned}&\Pr[P^*_{i\to j}=l(i\to j)]\\ &=-\int_0^\infty\frac{\Pr[w(P^*_{i\to j})>x]}{\Pr[w(i\to j)>x]+A}d(\Pr[w(i\to j)>x]+A)\end{aligned}$$

where $A=\frac{1-p_{ij}}{p_{ij}}$. Partial integration yields,

$$\begin{aligned}&\Pr[P^*_{i\to j}=l(i\to j)]\\ &=-\Pr[w(P^*_{i\to j})>x]\log(\Pr[w(i\to j)>x]+A)\Big|_{x=0}^{x=\infty}\\ &\quad-\int_0^\infty f_{w(P^*_{i\to j})}(x)\log(\Pr[w(i\to j)>x]+A)dx\end{aligned}$$

The limit $x\to 0$ gives $\log(1+A)=-\log p_{ij}$ and it remains to show that the limit $x\to\infty$ vanishes. Since for any $x$ and any probability distribution $w$ holds that

$$\begin{aligned}-\log(\Pr[w>x]+A)&=\int_0^x\frac{f_w(u)\,du}{\Pr[w>u]+A}\\ &\le\frac{1}{\Pr[w>x]+A}\int_0^x f_w(u)\,du\\ &=\frac{\Pr[w\le x]}{\Pr[w>x]+A}\end{aligned}$$

we observe that

$$-\log(\Pr[w(i\to j)>x]+A)\le\frac{1+A}{\Pr[w(i\to j)>x]+A}-1$$

such that

$$\begin{aligned}&\Pr[w(P^*_{i\to j})>x]\log\Pr[w(i\to j)>x]\cdot(-1)\\ &\le\frac{(1+A)\Pr[w(P^*_{i\to j})>x]}{\Pr[w(i\to j)>x]+A}-\Pr[w(P^*_{i\to j})>x]\end{aligned}$$

Using (21), we have

$$\begin{aligned}&-\Pr[w(P^*_{i\to j})>x]\log\Pr[w(i\to j)>x]\\ &\le\lim_{\Delta x\to 0}\Pr\left[P^*_{i\to j}=l(i\to j)|\{x\le w(i\to j)\le x+\Delta x\}\right]\cdot\\ &\quad(1+A)-\Pr[w(P^*_{i\to j})>x]\end{aligned}$$

For $x\to\infty$, the both probabilities at right hand side tend to zero. Hence,

$$\begin{aligned}&\Pr[P^*_{i\to j}=l(i\to j)]\\ &=-\int_0^\infty f_{w(P^*_{i\to j})}(x)\log\left(-\Pr[w(i\to j)\le x]+\frac{1}{p_{ij}}\right)dx\\ &\quad-\log p_{ij}\end{aligned}$$

After writing

$$\begin{aligned}&\log\left(-\Pr[w(i\to j)\le x]+\frac{1}{p_{ij}}\right)\\ &=-\log p_{ij}+\log\left(1-p_{ij}F_w(x)\right)\end{aligned}$$

we arrive at (4). □

### C. Asymptotic uncorrelation of links in $G_{\cup URT}$

*Lemma 1:* All first hop links in a same URT are independent.

**Proof:** see [11, p. 371]. □

This independence in the URT is only true for the first hop nodes, and not for higher hop nodes since the latter depend on the specific structure of the URT.

*Lemma 2:* Two links in a same URT, of which only one is a first hop link, are asymptotically (for large $N$) uncorrelated.

**Proof:** Drmota and Hwang [37] have, for large $N$, computed the asymptotic correlation coefficient $\rho\left(X_N^{(k)},X_N^{(j)}\right)$ of the number of nodes $X_N^{(k)}$ at hopcount $k$ (called the $k$-th level set of the URT) from the root in a URT, based on the exact probability generating function $E\left[x^{X_N^{(k)}}y^{X_N^{(j)}}\right]$ derived by van der Hofstad *et al.* in [38]. For large $N$ and small hopcounts $k=o(\log N)$ and $k=o(j)$ where $j$ can range over all levels, $\rho\left(X_N^{(k)},X_N^{(j)}\right)$ tends to zero, which implies that the level set $k$ and $j$ are asymptotically uncorrelated[6]. Since there is a one-to-one correspondence between nodes and links in a tree because each node (apart from the root) has precisely one ancestor, the correlation between nodes transfers to a correlation between links. Hence, *all* higher hop links in the URT are asymptotically independent from the first hop links. This proves the Lemma. □

The proof actually demonstrates more than necessary: instead of uncorrelation between two links, it shows uncorrelation between all higher hop links.

[6]The correlation coefficient $\left|\rho\left(X_N^{(k)},X_N^{(j)}\right)\right|\to 1$ if $k=O(\log N)$ and $j=O(\log N)$, implying that levels around the average hopcount $E[H_N]\sim\log N$ (containing most of the nodes) are strongly correlated, and this is mainly a consequence of the growth rule of the URT [11, Sec. 16.2.2.] and of the "conservation of nodes over the levels", $\sum_{k=0}^{N-1}X_N^{(k)}=N$. Each URT of $G_{\cup spt}$ thus contains highly correlated level sets, but the individual links (not paths) in $G_{\cup spt}$ seem far less correlated as suggested by Conjecture 1.

*Lemma 3:* Two different arbitrary links in the overlay $G_{\cup spt}$ on top of the complete graph $K_N$ are asymptotically (for large $N$) pairwise uncorrelated.

**Proof:** We denote two arbitrary links of $G_{\cup spt}$ by $l_1 = l(n_1 \to x)$ and $l_2 = l(n_2 \to y)$ where node $x \neq y$. We distinguish between two cases, $n_1 = n_2$ and $n_1 \neq n_2$.

If $n_1 = n_2$, then both $l_1$ and $l_2$ are first hop links in the same URT and, by Lemma 1, independent.

If $n_1 \neq n_2$, the links $l_1$ and $l_2$ do not share a common node and there are two cases: (1) $l_2$ does not belong to the URT rooted at $n_1$ (and vice versa), in which case $l_1$ and $l_2$ are either independent or at most asymptotically uncorrelated (see proof Lemma 2) because both links $l_1$ and $l_2$ may appear in a URT rooted at another node $n_3$. (2) $l_2 \in URT_{n_1}$, the URT rooted at node $n_1$. Lemma 2 then shows that $l_2$ and $l_1$ are asymptotically independent. $\square$

Recall that pairwise uncorrelation is weaker than pairwise independence, which in turn does not necessarily imply independence.

## References


[1] R. Albert, A. Barabasi, "Emergence of scaling in random networks", Science 286, 509-512, 1999.

[2] P. Van Mieghem and S. van Langen, "Influence of the link weight structure on the shortest path", Physical Review E 71, 056113, 2005.

[3] P. Van Mieghem and S. M. Magdalena, "A Phase Transition in the Link Weight Structure of Networks", Physical Review E 72, 056138, 2005.

[4] P. Van Mieghem, G. Hooghiemstra and R. W. van der Hofstad, "A scaling law for the hopcount", Delft University of Technology, report2000125, 2000.

[5] S. Androutsellis-Theotokis and D. Spinellis, "A survey of peer-to-peer content distribution technologies", ACM Computing Surveys (CSUR), 36(4):335 – 371, December 2004.

[6] T. Watanabe, Y. Higashi, and A. Nakamura, "An approach to robust network construction from graph augmentation problems", In Proc. of 1990 IEEE Int'l Symp. on Circuits and Systems, pages 2861–2864, 1990.

[7] A. Ganesh, L. Massoulie and D. Towsley, "The effect of network topology on the spread of epidemics", Proc. IEEE Infocom, 2005.

[8] L. Zhao, K. Park, and Y.-C. Lai, "Attack vulnerability of scale-free networks due to cascading breakdown", Phys. Rev. E 70, 035101(R), 2004.

[9] E. W. Dijkstra, "A note on two problems in connection with graphs", Num. Math., 1:269–271., 1959.

[10] R. van der Hofstad, G. Hooghiemstra, and P. Van Mieghem, "First passage percolation on the random graph", Probability in the Engineering and Informational Sciences (PEIS), Vol. 15, pp. 225-237, 2001.

[11] P. Van Mieghem, *Performance Analysis of Communications Systems and Networks*, Cambridge University Press, 2006.

[12] G. Chartrand, Ortrud R. Oellermann, *Applied and Algorithmic Graph Theory*, Mcgraw-Hill College, 1992.

[13] M. Fiedler, "Algebraic connectivity of graphs", Czech. Math. J. 23, pp. 298-305, 1973.

[14] M. Abramowitz and J. A. Stegun, *Handbook of Mathematical Functions*, Dover Publications Inc., New York, 1999.

[15] B. Bollobas, *Random Graphs*, 2nd ed., Cambridge University Press, Cambridge, 2001.

[16] E. Wigner, "On the distribution of the roots of certain symmetric matrices", Ann. Math. 67, pp. 325-327, 1958.

[17] G. Chartrand and L. Lesniak, *Graphs and Digraphs*, Chapman and Hall/CRC, 1996.

[18] A. Lakhina, J. Byers, M. Crovella and P. Xie, "Sampling Biases in IP Topology Measurements", Proc. of IEEE INFOCOM 2003.

[19] R. Hekmat, and P. Van Mieghem, "Study of Connectivity in Wireless Ad-Hoc Networks with an Improved Radio Model", Proceedings of the Second Workshop on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks (WiOpt'04), University of Cambridge, UK, March 24-26, pp. 142-151, 2004.

[20] M. Castro, M. Costa, and A. Rowstron, "Should We Build Gnutella on a Structured Overlay", ASM SIGCOMM Computer Communications Review 34 (1), pp. 131-136, 2004.

[21] RIPE Test Traffic Measurements, http://www.ripe.net/ripencc/mem-services/ttm/.

[22] M. Janic, F. Kuipers, X. Zhou and P. Van Mieghem, "Implications for QoS provisioning based on traceroute measurements", Proceedings of 3nd International Workshop on Quality of Future Internet Services, QofIS2002, Zurich, Switzerland, October 16-18, pp. 3-14, 2002.

[23] W. Aiello, F. Chung, and L. Lu, "A Random Graph Model for Massive Graphs", in 32nd Annual Symposium in Theory of Computing, 2000.

[24] S. N. Dorogovtsev, A. V. Goltsev, J. F. F. Mendes and A. N. Samukhin, "Spectra of complex networks", Phys. Rev. E 68, 046109, 2003.

[25] L. He, X. Liu, and G. Strang, "Trees with Cantor Eigenvalue Distribution", Stud. Appl. Math. 110, 123, 2003.

[26] D. Vukadinovi, P. Huang, and T. Erlebach, "On the spectrum and structure of Internet topology graphs", Lect. Notes Comput. Sci. 2346, 83, 2002.

[27] D. M. Cvetković, M. Doob, and H. Sachs, *Spectra of Graphs*, (VEB Deutscher Verlag der Wissenschaften, Berlin, 1980).

[28] E. D. van Dam and W. H. Haemers, "Which graphs are determined by their spectrum?", Linear Algebra and its Applications 373, pp. 241-272, 2003.

[29] I. J. Farkas, I. Deréyi, A.-L. Barabási, and T. Vicsek, "Spectra of "real-world" graphs: Beyond the semicircle law", Phys. Rev. E 64, 026704, 2001.

[30] M. Janic and P. Van Mieghem, "On properties of multicast routing trees", International Journal of Communication Systems, vol. 19, No. 1, pp. 95-114, February 2006.

[31] J. J. Binney, N.J. Dowrick, A.J. Fisher and M.E.J. Newman, *The Theory of Critical Phenomena: An Introduction to the Renormalization Group*, Clarendon Press, Oxford, 1992.

[32] D. Achlioptas, A. Clauset, D. Kempe and C. Moore, "On the Bias of Traceroute Sampling: or, Power-law Degree Distributions in Regular Graphs", cond-mat/0504087, 2006.

[33] A. Clauset and C. Moore, "Accuracy and Scaling Phenomena in Internet Mapping", Phyiscal Review Letters, Vol. 94, p. 018701, 2005.

[34] N. Biggs, *Algebraic Graph Theory*, Cambridge University Press, Cambridge, U.K., second edition, 1996.

[35] T. H. Cormen and C. E. Leiserson and R. L. Rivest, *An Introduction to Algorithms*, MIT Press, Boston, 1991.

[36] P. Van Mieghem, *Data Communications Networking*, Techne Press, Amsterdam, 2006.

[37] M. Drmota and H-K. Hwang, "Profiles of Random Trees: Correlation and Width of Random Recursive Trees and Binary Search Trees", Advances in Applied Probability, Vol. 37, p. 321-341, 2005.

[38] R. van der Hofstad, G. Hooghiemstra and P. Van Mieghem, 2002, "On the covariance of the level sizes in random recursive trees, Random Structures and Algorithms, vol. 20, pp. 519-539.


PLACE PHOTO HERE


**P** iet Van Mieghem is professor at the Delft University of Technology with a chair in telecommunication networks and chairman of the section Network Architectures and Services (NAS). His main research interests lie in new Internet-like architectures for future, broadband and QoS-aware networks and in the modelling and performance analysis of network behavior and complex infrastructures. Professor Van Mieghem received a Master's and Ph. D. in Electrical Engineering from the K.U.Leuven (Belgium) in 1987 and 1991, respectively. Before joining Delft, he worked at the Interuniversity Micro Electronic Center (IMEC) from 1987 to 1991. During 1993 to 1998, he was a member of the Alcatel Corporate Research Center in Antwerp where he was engaged in performance analysis of ATM systems and in network architectural concepts of both ATM networks (PNNI) and the Internet.

He was a visiting scientist at MIT (department of Electrical Engineering, 1992-1993) and, in 2005, he was visiting professor at ULCA (department of Electrical Engineering). He was member of the editorial board of the journal Computer Networks from 2005-2006.

PLACE
PHOTO
HERE

**H** uijuan Wang received the M.Sc. degree in Electrical Engineering at the Delft University of Technology in August, 2005. She is currently pursuing the Ph.D. degree in the Network Architectures and Services group, TUDelft. Her Ph.D. work mainly focuses on performance analysis of complex networks and Next Generation Infrastructures.