Supplementary Information for Navigability of complex networks

Marián Boguñá

Departament de Física Fonamental, Universitat de Barcelona, Martí i Franquès 1, 08028 Barcelona, Spain

Dmitri Krioukov and kc claffy

Cooperative Association for Internet Data Analysis (CAIDA), University of California, San Diego (UCSD), 9500 Gilman Drive, La Jolla, CA 92093, USA

I. THE MODEL VS. REAL NETWORKS: THE AUTONOMOUS SYSTEM LEVEL MAP OF THE INTERNET AND THE US AIRPORT NETWORK

The model we use in this work is not meant to reproduce any particular system but to generate a set of general properties, like heterogeneous degree distributions, high clustering, and a metric structure lying underneath. Yet, despite its simplistic assumptions, the model generates graphs that are surprisingly close to some real networks of interest, in particular the Internet at the Autonomous System level (AS) [1, 2] and the network of airline connections among airports within the United States during 2006 (USAN) [3]. In the case of the Internet, we use two different data sets, the Internet as viewed by the Border Gateway Protocol (BGP) [1] and the DIMES project [2]. The BGP (DIMES) network has a size of $N \sim 17446$ (N = 19499) ASs, average degree $\langle k \rangle = 4.7 \; (\langle k \rangle = 5), \text{ and average clustering } C = 0.41$ (C = 0.6). The US Airport Network is composed of US airports connected by regular flights (with more than 1000 passengers per year) during the year 2006. This results in a network of N = 599 airports, average degree $\langle k \rangle \sim 10.8$ and average clustering coefficient C = 0.72.

Figures 1 and 2 show a comparison of the basic topological properties of these networks with graphs generated with the model. In the case of the AS map, we use a truncated power law distribution $\rho(\kappa) \sim \kappa^{-\gamma}, \kappa < \kappa_c$ with exponent $\gamma = 2.1$ and κ_c such that the maximum degree of the network is $k_c = 2400$. For the USAN, we use $\gamma = 1.6$ and a maximum degree $k_c = 180$, as observed in the real network. As it can be appreciated in both figures, the matching of the model with the empirical data is surprisingly good except for very low degree vertices. This is particularly interesting since we are not enforcing any mechanism to reproduce higher order statistics like the average nearest neighbours degree $k_{nn}(k)$ or the degree-dependent clustering coefficient $\bar{c}(k)$. This can be understood as a consequence of the high heterogeneity of the degree distribution that introduces structural constraints in the network [4, 5].

The airport network differs in several ways from our modelled networks: the distribution of airports in the geographic space is far from uniform; the airport degree distribution does not perfectly follow a power law; and it exhibits a sharp high-degree cut-off. However, the structure of greedy paths is surprisingly similar to that in our modelled networks in Fig. 6. The success ratio $p_s \approx 0.64$ and average length of successful paths $\tau \approx 2.1$ are also similar to those in our modelled networks of the corresponding size, clustering, and degree distribution exponent. These similarities indicate that the network navigability characteristics depend on clustering and heterogeneity of the airport degree distribution, and less so on how perfectly it follows a power law.

II. HIERARCHICAL ORGANIZATION OF MODELED NETWORKS

The routing process in our framework resembles guided searching for a specific object in a complex collection of objects. Perhaps the simplest and most general way to make a complex collection of heterogenous objects searchable is to classify them in a hierarchical fashion. By "hierarchical," we mean that the whole collection is split into categories (i.e., sets), sub-categories, sub-subcategories, and so on. Relationships between categories form (almost) a tree, whose leaves are individual objects in the collection [6–9]. Finding an object reduces to the simpler task of navigating this tree.

k-core decomposition [12, 13] is possibly the most suitable generic tool to expose hierarchy within our modeled networks. The k-core of a network is its maximal subgraph such that all the nodes in the subgraph have kor more connections to other nodes in the subgraph. A node's coreness is the maximum k such that the k-core contains the node but the k+1-core does not. The k-core structure of a network is a form of hierarchy since a k+1core is a subset of a k-core. One can estimate the quality of this hierarchy using properties of the k-core spectrum, i.e., the distribution of k-core sizes. If the maximum node coreness is large and if there is a rich collection of comparably-sized k-cores with a wide spectrum of k's, then this hierarchy is deep and well-developed, making it potentially more navigable. It is poor, non-navigable otherwise.

In Fig. 3 we feed real and modeled networks to the Large Network visualization tool (LaNet-vi) [10] which utilizes node coreness to visualize the network. Fig. 3 shows that networks with stronger clustering and smaller exponents of degree distribution possess stronger k-core hierarchies. These hierarchies are directly related to how networks are constructed in our model, since nodes with



FIG. 1: Degree distribution P(k), average nearest neighbours' degree $\bar{k}_{nn}(k)$, and degree-dependent clustering coefficient $\bar{c}(k)$ generated by our model with $\gamma = 2.1$ and $\alpha = 2$ compared to the same metrics for the real Internet map as seen by BGP data and the DIMES project.

higher κ and, consequently, higher degrees have generally higher coreness, as we can partially see in Fig. 3.

III. THE ONE-HOP PROPAGATOR OF GREEDY ROUTING

To derive the greedy-routing propagator in this appendix, we adopt a slightly more general formalism than in the main text. Specifically, we assume that nodes live in a generic metric space \mathcal{H} and, at the same time, have intrinsic attributes unrelated to \mathcal{H} . Contrary to normed spaces or Riemannian manifolds, generic metric spaces do not admit any coordinates, but we still use the coordinate-based notations here to simplify the exposition below, and denote by \mathbf{x} nodes' coordinates in \mathcal{H} and by ω all their other, non-geometric attributes, such as their expected degree κ . In other words, hidden variables \mathbf{x} and ω in this general formalism represent some collections of nodes' geometric and non-geometric hidden



FIG. 2: Degree distribution P(k), average nearest neighbours' degree $\bar{k}_{nn}(k)$, and degree-dependent clustering coefficient $\bar{c}(k)$ generated by our model with $\gamma = 1.6$, $\alpha = 5$ and a cut-off at $k_c = 180$ compared to the same metrics for the real US airport network.

attributes, not just a pair of scalar quantities. Therefore, integrations over **x** and ω in what follows stand merely to denote an appropriate form of summation in each concrete case.

As in the main text, we assume that \mathbf{x} and $\boldsymbol{\omega}$ are independent random variables so that the probability density to find a node with hidden variables $(\mathbf{x}, \boldsymbol{\omega})$ is

$$\rho(\mathbf{x},\omega) = \delta(\mathbf{x})\rho(\omega)/N,\tag{1}$$

where $\rho(\omega)$ is the probability density of the ω variables and $\delta(\mathbf{x})$ is the concentration of nodes in \mathcal{H} . The total number of nodes is

$$N = \int_{\mathcal{H}} \delta(\mathbf{x}) d\mathbf{x},\tag{2}$$

and the connection probability between two nodes is an integrable decreasing function of the hidden distance between them,

$$r(\mathbf{x},\omega;\mathbf{x}',\omega') = r[d(\mathbf{x},\mathbf{x}')/d_c(\omega,\omega')], \qquad (3)$$



FIG. 3: *k*-core decompositions of real and modeled networks. The first two rows show LaNet-vi [10] network visualizations. All nodes are color-coded based on their coreness (right legends) and size-coded based on their degrees (left legends). Higher-coreness nodes are closer to circle centers. The third row shows the *k*-core spectrum, i.e., the distribution S(k) of sizes of node sets with coreness *k*. The first column depicts two real networks: the AS-level Internet as seen by the Border Gateway Protocol (BGP) in [1] and the Pretty Good Privacy (PGP) social network from [11]. The rest of the columns show modeled networks for different values of power-law exponent γ in cases with weak ($\alpha = 1.1$) and strong ($\alpha = 5.0$) clustering. The network size *N* for all real and modeled cases is approximately 10⁴. Similarity between real networks and modeled networks with low γ and high α is remarkable.

where $d_c(\omega, \omega')$ a characteristic distance scale that depends on ω and ω' .

We define the one-step propagator of greedy routing as the probability $G(\mathbf{x}', \omega' | \mathbf{x}, \omega; \mathbf{x}_t)$ that the next hop after a node with hidden variables (\mathbf{x}, ω) is a node with hidden variables (\mathbf{x}', ω') , given that the final destination is located at \mathbf{x}_t .

To further simplify the notations below, we label the set of variables (\mathbf{x}, ω) as a generic hidden variable h and undo this notation change at the end of the calculations according to the following rules:

$$\begin{array}{cccc}
(\mathbf{x},\omega) &\longrightarrow h \\
\rho(\mathbf{x},\omega) &\longrightarrow \rho(h) \\
d\mathbf{x}d\omega &\longrightarrow dh \\
r(\mathbf{x},\omega;\mathbf{x}',\omega') &\longrightarrow r(h,h').
\end{array}$$
(4)

We begin the propagator derivation assuming that a particular network instance has a configuration given by $\{h, h_t, h_1, \dots, h_{N-2}\} \equiv \{h, h_t; \{h_j\}\}$ with $j = 1, \dots, N-2$, where h and h_t denote the hidden variables of the current hop and the destination, respectively. In this particular network configuration, the probability that the current node's next hop is a particular node i with hidden variable h_i is the probability that the current node is connected to i but disconnected to all nodes that are closer to the destination than i,

$$\operatorname{Prob}(i|h, h_t; \{h_j\}) = r(h, h_i) \prod_{j \neq i}^{N-2} [1 - r(h, h_j)]^{\Theta[d(h_i, h_t) - d(h_j, h_t)]},$$
(5)



FIG. 4: Probability $P_{up}(\omega/d^{1/2}, d)$.

where $\Theta(\cdot)$ is the Heaviside step function. Taking the average over all possible configurations $\{h_1, \dots, h_{i-1}, h_{i+1}, \dots, h_{N-2}\}$ excluding node *i*, we obtain

$$\operatorname{Prob}(i|h, h_t; h_i) = r(h, h_i) \left(1 - \frac{1}{N-3} \bar{k}(h|h_i, h_t) \right)^{N-3},$$
(6)

where

$$\bar{k}(h|h_i, h_t) = (N-3) \int_{d(h_i, h_t) < d(h', h_t)} \rho(h') r(h, h') dh'$$
(7)

is the average number of connections between the current node and nodes closer to the destination than node i, excluding i and t.

The probability that the next hop has hidden variable h', regardless of its label, i.e., index i, is

$$Prob(h'|h, h_t) = \sum_{i=1}^{N-2} \rho(h') Prob(i|h, h_t; h').$$
(8)

In the case of sparse networks, $\bar{k}(h|h', h_t)$ is a finite quantity. Taking the limit of large N, the above expression simplifies to

$$Prob(h'|h, h_t) = N\rho(h')r(h, h')e^{-\bar{k}(h|h', h_t)}.$$
 (9)

Yet, this equation is not a properly normalized probability density function for the variable h' since node h can have degree zero with some probability. If we consider only nodes with degrees greater than zero, then the normalization factor is given by $1 - e^{-\bar{k}(h)}$. Therefore, the properly normalized propagator is finally

$$G(h'|h,h_t) = \frac{N\rho(h')r(h,h')e^{-\bar{k}(h|h',h_t)}}{1 - e^{-\bar{k}(h)}}.$$
 (10)

We now undo the notation change and express this propagator in terms of our mixed coordinates:

$$G(\mathbf{x}', \omega' | \mathbf{x}, \omega; \mathbf{x}_t) = \frac{\delta(\mathbf{x}')\rho(\omega')}{1 - e^{-\bar{k}(\mathbf{x}, \omega)}} r \left[\frac{d(\mathbf{x}, \mathbf{x}')}{d_c(\omega, \omega')} \right] e^{-\bar{k}(\mathbf{x}, \omega | \mathbf{x}', \mathbf{x}_t)},\tag{11}$$

with

$$\bar{k}(\mathbf{x},\omega|\mathbf{x}',\mathbf{x}_t) = \int_{d(\mathbf{x}',\mathbf{x}_t)>d(\mathbf{y},\mathbf{x}_t)} d\mathbf{y} \int d\omega' \delta(\mathbf{y}) \rho(\omega') r\left[\frac{d(\mathbf{x},\mathbf{y})}{d_c(\omega,\omega')}\right].$$
(12)

In the particular case of the \mathbb{S}^1 model, we can express this propagator in terms of relative hidden distances instead of absolute coordinates. Namely, $G(d', \omega'|d, \omega)$ is the probability that an ω -labeled node, e.g., a node with expected degree $\kappa = \omega$, at hidden distance *d* from the destination has as the next hop an ω' -labeled node at hidden distance *d'* from the destination. After tedious calculations, the resulting expression reads:

$$G(d',\omega'|d,\omega) = \begin{cases} \frac{(\gamma-1)}{\omega'^{\gamma}} \left[\frac{1}{(1+\frac{d-d'}{\mu\omega\omega'})^{\alpha}} + \frac{1}{(1+\frac{d+d'}{\mu\omega\omega'})^{\alpha}} \right] \exp\left\{ \frac{(1-\gamma)\mu\omega}{\alpha-1} \left[\mathcal{B}(\frac{d-d'}{\mu\omega},\gamma-2,2-\alpha) - \mathcal{B}(\frac{d+d'}{\mu\omega},\gamma-2,2-\alpha) \right] \right\} ; d' \leq d \\ \frac{(\gamma-1)}{\omega'^{\gamma}} \left[\frac{1}{(1+\frac{d'-d}{\mu\omega\omega'})^{\alpha}} + \frac{1}{(1+\frac{d+d'}{\mu\omega\omega'})^{\alpha}} \right] \exp\left\{ \frac{(1-\gamma)\mu\omega}{\alpha-1} \left[\frac{2}{\gamma-2} - \mathcal{B}(\frac{d'-d}{\mu\omega},\gamma-2,2-\alpha) - \mathcal{B}(\frac{d+d'}{\mu\omega},\gamma-2,2-\alpha) \right] \right\} ; d' > d \end{cases}$$

$$(13)$$

where we have defined function

$$\mathcal{B}(z,a,b) \equiv z^{-a} \int_0^z t^{a-1} (1+t)^{b-1} dt, \qquad (14)$$

which is somewhat similar to the incomplete beta function $B(z, a, b) = \int_0^z t^{a-1} (1-t)^{b-1} dt.$

One of the informative quantities elucidating the structure of greedy-routing paths is the probability $P_{up}(\omega, d)$ that the next hop after an ω -labeled node at distance dfrom the destination has a higher value of ω . The greedyrouting propagator defines this probability as

$$P_{up}(\omega, d) = \int_{\omega' \ge \omega} d\omega' \int_{d' < d} dd' G(d', \omega' | d, \omega), \quad (15)$$

- P. Mahadevan, D. Krioukov, M. Fomenkov, B. Huffaker, X. Dimitropoulos, kc claffy, and A. Vahdat, Comput Commun Rev 36, 17 (2006).
- [2] Y. Shavitt and E. Shir, Comput Commun Rev **35** (2005).
- [3] Data available at http://www.transtats.bts.gov/.
- [4] J. Park and M. E. J. Newman, Phys. Rev. E 68, 026112 (2003).
- [5] M. Boguñá, R. Pastor-Satorras, and A. Vespignani, European Physical Journal B 38, 205 (2004).
- [6] D. J. Watts, P. S. Dodds, and M. E. J. Newman, Science 296, 1302 (2002).
- [7] M. Girvan and M. E. J. Newman, Proc. Nat. Acad. Sci. USA 99, 7821 (2002).
- [8] A. Clauset, C. Moore, and M. E. J. Newman, Nature 453, 98 (2008).

and we show $P_{up}(\omega/d^{1/2}, d)$ in Fig. 4. We see that the proper scaling of $\omega_c \sim d^{1/2}$, where ω_c is the critical value of ω above which $P_{up}(\omega, d)$ quickly drops to zero, is present only when clustering is strong. Furthermore, $P_{up}(\omega, d)$ is an increasing function of ω for small ω 's only when the degree distribution exponent γ is close to 2. A combination of these two effects guarantees that the layout of greedy routes properly adapts to increasing distances or graph sizes, thus making networks with strong clustering and γ 's greater than but close to 2 navigable.

- [9] D. Krioukov, F. Papadopoulos, M. Boguñá, and A. Vahdat, Efficient navigation in scale-free networks embedded in hyperbolic metric spaces (2008), arXiv:0805.1266.
- [10] J. I. Alvarez-Hamelin, L. Dall'Asta, A. Barrat, and A. Vespignani, in *Advances in Neural Information Processing Systems 18*, edited by Y. Weiss, B. Schölkopf, and J. Platt (MIT Press, Cambridge, MA, 2006), pp. 41–50.
- [11] M. Boguñá, R. Pastor-Satorras, A. Díaz-Guilera, and A. Arenas, Phys Rev E 70, 056122 (2004).
- [12] B. Bollobás, Modern Graph Theory (Springer-Verlag, New York, 1998).
- [13] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes, Phys Rev Lett 96, 040601 (2006).