

Research Statement

Van K. Nguyen

My research interests lie in theoretical computer science (algorithms, graph theory, random structures), distributed computing and networking. In the theoretical side, the core of my research is on random structures, which are suggested by the recent studies in real-world complex networks. In the applied side, I work with Internet and related problems in peer-to-peer, wireless, and optical networks, where I design new network constructions with efficient routing algorithms. Below, I briefly present a brief background on the study of real-world complex networks, which has motivated the main theme of my research. Then I briefly discuss my main results and present my research plan.

Background and Motivation. A large inter-disciplinary community has emerged recently to study real-world complex networks. In 1967, Milgram confirmed a folklore that we are in a small-world where two strangers are linked by a short chain of acquaintances. In 1998, Watts and Strogatz [9] produced a graph model for small-world networks, which features this property of small diameter (i.e. paths between any two nodes are short) and high clustering coefficient (two nodes with a shared neighbor are likely connected by a link). Recent study finds these small-world properties in many large-scale real-world networks such as in social networks, biological networks and the Internet. In 2000, Kleinberg [3] extended Watts and Strogatz’s model to help explain another striking aspect: that such a short path can be found using a simple greedy strategy with limited local information only (this strategy plays the role of a search based on a first-name basis used in Milgram’s experiment). In Kleinberg’s basic model, a grid graph is augmented by random links, where a link is more likely to go to closer nodes. Specifically, each node u has one random link which goes to a node v with probability inversely proportional to $d^2(u, v)$, where $d(u, v)$ is the lattice distance between u and v . Besides, the power-law feature of vertex degrees has also been widely recognized as another common property. For example, the distribution of vertex degrees in the Internet graphs has a power-law shape [1].

Inspired by these studies, especially by Kleinberg’s work on small-world networks, I began (a) modeling real-world complex networks. This led me to develop a new approach for studying (b) small-world random structures. I believed that my small-world study with new network constructions can contribute applications, and later discovered these in (c) designing large-scale dynamic computer networks. Also, my modeling work using geographical factors suggests why the Internet has a small diameter. Below I briefly discuss my results in (b), then (a) and (c).

Random structures with small-world properties. I study generalizations of Watts and Strogatz’s and Kleinberg’s models. I am interested in this type of general graph construction: a base graph, e.g. a grid, is augmented by a generally non-uniform distribution of random links. I assume a metric distance d between the nodes. In [5] and [7] we (Professor Martel and I) introduce general conditions on the base graphs and random link distributions which produce constructions with small diameter, e.g. $O(\log n)$ with n as the vertex size. We are particularly interested in, what we call, *local-contact base graphs* and *distance-bias distributions* (of random links). In our local-contact graphs, each node is in a ‘dense enough’ neighborhood so that these base graphs (and the augmented new graphs) are likely to have high clustering. In our distance-bias distributions, random links ‘favor’ closer distances over more distant ones, which we characterize by using this family of distributions: a random link from a node u goes to a node v with probability $\propto d^{-r}(u, v)$, for parameter r as a given arbitrary constant. Our framework can be used to construct new small-world structures and to prove that many different existing networks have small diameters.

We provide a thorough analysis of Kleinberg’s small-world model, its generalizations and many other settings [4, 5, 6], as well as a number of analysis techniques that can be generally useful (in settings with non-uniform random links). In the random structure where a D -dimensional grid is augmented with a distance-bias distribution ($Pr[“u linked to v”] \propto d^{-r}(u, v)$), we show that *with high probability* the structure still has poly-log diameter when $0 < r < 2D$, but has polynomial diameter when $r > 2D$. This shows a *phase-transition between small-world and “large-world”* states. Our new routing algorithm also

improves the expected route length in Kleinberg’s greedy algorithm ($O(\log^{1+\frac{1}{D}} n)$ compared to $\theta(\log^2 n)$ in a D -dimensional setting), by using more local information (size $O(\log n)$ compared to $O(1)$).

Modeling real-world random networks. I focus on the Internet topology and related networks where geographical factors are important modeling issues. The geographical factors I consider include the *distance-bias tendency* (links appear more for closer distances) and the property of *bounded growth* in neighborhood expansion. For example, Faloutsos et al. [1] observed in the Internet topology that a ‘ball’ of neighbors expanding around a node can have size approximated by R^α where the radius distance R is small enough. In my model, I control this bounded growth by two parameters (lower and upper bounds of α). Note that most existing models for real-world random networks feature small-world properties and power-law degrees (or even both together) but none looks at geographical factors at the same time.

In [7], we propose and analyze *a new general model for small-world and power-law properties which also considers geographical factors*. Various results can be drawn from our model. One interesting result is for a refined model where a growth-bounded base graph is augmented with a distance-bias distribution of random links. We show if the small-world effect may occur and how the diameter changes *depending on the coordination between the distance-bias parameter (r) and the two bounded-growth parameters*.

This helps explain *why the Internet graph has low diameter but is locally growth bounded*. We can think of the Internet graph as a set of localities, i.e. local neighborhoods with significant clustering (e.g. LANs), augmented with a collection of distance-bias random long links ($r \approx 1$ as suggested by [10]).

Designing large-scale routing networks. Based on my above results, I observed that my random graph constructions provide a natural platform for designing efficient routing networks. Using my distance-bias structures (growth-bounded base graphs plus distance-bias random links) I developed two main building blocks: (i) an efficient routing strategy and (ii) a mechanism to optimize network performance by changing construction parameters. Based on these, I introduce two applications in (iii) hybrid wireless networks and (iv) optical networks.

For (i), in [8] we give an efficient *hierarchical routing* based on our *new network partitioning* scheme for the grid setting, and in [7] we generalize this to growth-bounded base graphs. Our routing scheme in [7] uses a small distributed routing database with memory size $O(1)$ per node and finds routes of poly-log expected length.

For (ii), we optimize several measures simultaneously by cleverly choosing distribution of long links. For example, we obtained low cost (total edge weight), short routes, bounded degrees and low congestion in [8]. Prior designs always focused on a more limited set of criteria. Our approach can be useful for building a network from scratch or in building a virtual private network over an existing infrastructure. Thus we consider two main scenarios: 1) adding new long links to an existing network, and 2) building a new network, where we need to specify both the base graph (local links) and the long links. Geographical factors (distance bias and bounded growth) play a crucial role in our design approach.

For (iii), we propose network constructions for building hybrid ad-hoc networks by adding a wired infrastructure to an (ad-hoc) wireless network [8]. In our basic model we consider *adding longs links* to a grid-like network with uniform traffic demand between any two nodes, where the cost of a link is proportional to its weight (length). For a given budget for long links, we choose links so the route length is *poly-logarithmic*, while the congestion ratio is as small as possible: by adding $O(1)$ long links to each node we can maintain a *near optimal trade-off* between congestion ratio and the total cost of additional links. To produce our best trade-off, we select fixed long links using an arrangement which is similar to a distance-bias distribution of random links.

For (iv), we consider routing with Quality of Service provisioning in a dynamic network where the links are associated with (bandwidth) capabilities which may change arbitrarily [7]. Our working example here is in optical wavelength division multiplexing (WDM) networks where the capability of fiber-optic cable is enormous while communication is mainly restricted by limited resources at the hosts (switching devices and other expensive hardware). Using our distance-bias assumption, we propose a new approach which can obtain an ideal *adaptivity* which measures the maximum number of sites to be updated upon a single capability change.

Future Plan. For modeling issues, I intend to use experimental tasks to validate my assumptions on long link distributions, tune parameters, and develop extended settings with more complicated dis-

tributions. Similarly, as with bounded-growth modeling, I will extend the models by introducing a *local* distance-bias degree (r), varying between different neighborhoods, and control it by two bound parameters.

For future designs, I have already begun work on concrete designs for a more general setting, where nodes can be distributed non-uniformly and node-to-node communication demands are not uniform. An approach using multi-commodity network flows is already initiated in [8]. I also intend to refine my general routing scheme for a general topology (without distance-bias assumption) and evaluate its efficiency by experiments with practical data for real-world networks.

My design framework is a useful platform to combine several conflicting factors in a routing network. Thus, my work on routing networks can also be extended to related topics such as Euclidean spanners and compact routing. I also intend to consider security issues such as those in [2] where one need to find multiple vertex-disjoint paths between a given source-destination pair.

For a long-term direction, I foresee two significant lines. It is difficult to produce a mathematical model which can simulate many common features at the same time. A major next step would be to develop models which *combine distance-bias distributions of random links with power-law degrees*, besides other existing features such as growth-bounded localities. Such a new model would be a powerful platform to study many important problems. For example, this would create a new approach to study generators for the Internet topology and other related networks - a fundamental issue in many Internet related areas.

I can foresee a potential algorithmic perspective upon new discoveries of important structural properties in the Internet and related networks. Such new knowledge would introduce new ways to evaluate the performance of basic algorithms, e.g. search and routing, and to develop new efficient algorithms. However, unlike a similar current direction in the power-law research, my approach aims at strategies which avoid relying too much on high-degree nodes. I believe this is a crucial issue in algorithm design for dynamic networks. Good solutions in this direction would improve network robustness and resilience, e.g., to accidental collapses of high-degree nodes, and would also improve network load-balancing and reduce congestion.

I also intend to migrate into related active areas around the theme of the Internet as a marketplace and its dynamic processes, where algorithmic game theory or stochastic models for the Web processes are key technical tools.

Summary. I am interested in a new branch of algorithmic research on random structures using non-uniform random links, to model real-world random networks and to design large-scale computer networks. This line of research blossomed from a desire to understand the striking aspects of the small-world effect. My work significantly extends previous study of small-world random structures and contributes to the classical study of random graphs, in both theoretical and application aspects.

References

- [1] C. Faloutsos, P. Faloutsos, and M. Faloutsos, “On power-law relationships of the internet topology,” in *ACM SIGCOMM*, 1999.
- [2] M. Franklin and C. Martel, “Small-world networks and secure distributed protocols,” Tech. Rep., UC-Davis, 2002.
- [3] J. Kleinberg, “The small-world phenomenon: An algorithmic perspective,” in *Proc. of ACM Symp. on Theory of Computing (STOC)*, 2000, pp. 163–170.
- [4] C. Martel and V. Nguyen, “Analyzing kleinberg’s (and other) smallworld models,” in *Proc. of ACM Symp. on Princ. of Dist. Comp. (PODC)*, 2004, pp. 179–188.
- [5] V. Nguyen and C. Martel, “Analyzing and characterizing small-world graphs,” in *Proc. of ACM Symp. on Discrete Algorithms (SODA)*, 2005, pp. 311–320.
- [6] —, “Non-uniform random links in small-world graphs: Models, analysis and applications in network designs,” In internal review, to be submitted to SIAM J. on Computing.
- [7] —, “Localities plus (distance-bias) random links: Small-world models and applications,” Submitted to STOC’06.
- [8] —, “Designing low cost networks with short routes and low congestion,” To appear in INFOCOM’06.
- [9] D. Watts and S. Strogatz, “Collective dynamics of small-world networks,” *Nature*, vol. 393, pp. 440–32, 1998.
- [10] S. Yook, H. Jeong, and A. Barabasi, “Modeling the internets large-scale topology,” in *Proceedings of the National Academy of Sciences*, vol. 99, 2002, pp. 13 382–13 386.