Statement of Research

Van K. Nguyen

My research interests lie in three broad areas of Networks: (i) Modeling real-world complex networks (ii) Random structures and algorithms, and (iii) Designs of large-scale dynamic networks. These areas act as the three vertices of a triangle with mutual inspiration between them. For example, while creating modeling tools for (i), I have been developing my best technical skills and expertise in (ii), more specifically in small-world random structures. Interestingly, many new theoretical results obtained here inspire fruitful approach and useful techniques in solving some practical problems in (iii). In a reverse direction, applications also suggest new questions and extensions to the models. Below I discuss my motivation and contribution in each area. I start with my work in Small-world random structures, a technical core of my research. Building on that, I discuss my new model(s) for real-world random networks, then new constructions and designs for routing in large-scale computer networks. Finally I present my future research plan.

1. Small-world Random Structures

Background. Four decades have passed since Milgram (1967 [8]) confirmed a (probably, centuries old) folklore that we are in a small-world where two strangers can be linked by a short chain of acquaintances. The recent extensive multi-disciplinary research initiated by Watts and Strogatz's seminal work [14] shows that small-world properties are common in many large-scale real-world networks such as social networks, biological networks and the Internet networks. Yet, this striking phenomenon was little understood (and contemplated) till recently when Kleinberg [5] produces a nice model for the other striking aspect: such a short chain can be found using limited local information only [1].

Starting with Watts and Strogatz, much work has been done recently on models where random links are added to a simple graph which is rich in local contacts. While these local contacts are likely the source for high clustering, the random links (uniform in early models) create long-range contacts which shrink the graph diameter. Kleinberg uses a special *non-uniform* distribution of random links, which favor closer nodes over more distant ones. His model initiates an active research branch on *decentralized search* where routing using local information only is possible. This decentralized search is desired in many Internet-related scenarios such as in peer-to-peer networking.

My contribution. We envision a new direction of research on this type of random structures, which are formed by adding a non-uniform distribution of random links to a simple local-contact graph [2]. We study general rules and characteristics for making small-world (and related) properties. Particularly, we consider the abstract properties of the random link distributions which can introduce short paths (typically, with length poly-logarithmic in the size of the graph) between the sites of the local-contact graph. Our work significantly extends previous study of small-world random structures and contributes to the classical study of random graphs. Non-uniformity (of random links) has been the main technical challenge that we had to confront. We developed a number of analysis techniques that can be generally useful.

A framework to construct small-worlds graphs. In our abstraction of small-world graphs, we introduce a general criterion for random link distributions, namely *expansion property.* We propose an analysis technique where if the expansion property is satisfied and there are sufficient local links in the local contact graph, then the augmented graph has expected diameter $O(\log n)$. We also develop a general framework to construct small-world graphs, featuring a hierarchical family of classes of random structures where *short paths are available and can be found using decentralized routing strategies in the more refined classes* [9, 13]. Our abstract class(es) can be refined to re-construct a wide range of existing

---

[1] In Migram's experiment the acquaintance chain is formed (in a search for a particular person in a given destination) by having people forward a message to someone they know just *based on a first-name basis.*

[2] Later we also aim to use this new approach to contribute network design applications by introducing constructions which simultaneously optimize many practical factors.

small-world structures including Watts and Strogatz's [14], Kleinberg's, and peer-to-peer Distributed Hash Table (DHT) settings using (or imitating) random links, e.g. Viceroy network [6]. Despite their different topologies, we successfully capture their common characteristic of neighborhood expansion, and show that they all have expected logarithmic diameter.

A thorough analysis of Kleinberg's small-world model [7, 9]. If the local contact graph is a D-dimensional grid and random arcs are added from u to v with probability $\propto d^{-r}(u,v)$, we show that *with high probability* the model still has poly-log diameter when $0 < r < 2D$, but has polynomial expected diameter when $r > 2D$. This shows *a phase-transition between small-world and "large-world"* states. We also show that the (expected) $s - t$ route length, obtained by using Kleinberg's greedy algorithm, is $\theta(\log^2 n)$, and *can be improved to* $O(\log^{1+\frac{1}{D}} n)$, using *our new routing algorithm*, which uses more local information (size $O(\log n)$ compared to $O(1)$ in Kleinberg's routing).

2. Models for real-world networks: localities plus (distance-bias) random links

There has been extensive research on characterizing and modeling real-world random networks, where the small-world phenomenon and the power-law degree distribution have been widely recognized as common properties. We focus more on the Internet topology and related networks and observe that geographical factors are not only important in modeling these, but also crucial in designing new virtual networks or backbone infrastructures. Falousous et al. [2] suggest a form of *bounded growth* in neighborhood expansion: the expansion around a node can have size (number of nodes) approximated by a fixed power of the radius distance if this radius distance is significantly smaller than the network diameter. *Bounded growth* has also been found in peer-to-peer networks and wireless networks. Moreover, a recent study (e.g. [15]) suggests a distance bias in the distribution of long-range links: these favor closer nodes over more distant ones). Although many models have been developed to study small-world properties and power-law degrees (and even all together) none looks at geographical factors at the same time.

My contribution. In [11], we propose *a new general model for small-world properties which also considers geographical factors, yet can feature power-law degrees.* We also provide a set of technical tools for modeling and analyzing such random networks. Our new model is a significant extension of our abstract framework of section 1. Particularly, our new general model features arbitrary expected degrees (so, can be refined to create power-law degrees), a new class of base graphs - the growth bounded graphs - and a family of *distance-bias distributions*: we define a semi-metric function $d(u,v)$ and generate random links between any two nodes u and v with probability $Pr[u \to v] \propto d^{-r}(u,v)$. Various (known and new) specific results can be shown using our model. *One of our main results is to analyze* a structure, called a *distance-bias structure*, where a growth-bounded base graph is augmented with a distance-bias distribution of random links. For these *distance-bias* structures, we show how the small-world effect may occur or change (i.e. in the graph diameter) *depending on the coordination between the distance-bias parameter (r) and the (bounded) range of the 'growth degree'* in neighborhood expansion in the base graphs. This helps to see why the Internet graph is considered as a small-world with low diameter, but is locally growth bounded.

3. Using small-worlds to design efficient routing networks

Our earlier work provides a foundation for *network design*: a clever choice of distribution (or fixed arrangement) of long links results in network properties which are desired in important network design problems. *We provide a general framework for designing routing networks where the geographical distance (or an appropriate distance measure* [3]*) has an important role*: long links tend to appear less often for farther distances. We design a new network (or add additional long links to an existing one) using a distance-bias distribution of long links. We also give a new routing strategy using a distributed routing database of limited size designed to best exploit this distance-bias tendency.

Especially, we consider *constructions which optimize several measures simultaneously*: low graph weight (total edge weight), small routing diameter, bounded degrees and low congestion [4] [12]. There

---

[3] Many Internet measures e.g. the transmission delay in the Internet, also forms a metric.

[4] In a computer network scenario, low graph weight means cheap cost for connecting cable (or bandwidth backbone

are complex trade-offs between these factors which makes it challenging to balance all factors. Previous work in network design usually focuses on a smaller set of factors, so may ignore important issues. A full approach can be useful for different classical areas of network designs, such as building a network from scratch, or building a virtual private network over an existing infrastructure.

A novel hierarchical routing strategy based on a new network partition. We propose a new network partitioning scheme (specific for the grid setting in [12] and more general for growth-bounded base graphs in [11]) that we use as a building block for network designs and routing. Intuitively, in our partitioning scheme, an $n \times n$ grid is recursively partitioned such that a (parent) block of size Δ is partitioned into child blocks of size Δ^{μ}, for a fixed constant exponent $\mu \in (0, 1)$. This hierarchy has only a poly-log number of levels and yet, is far less steep than a b-ary tree (for constant $b \geq 2$) where nodes near the root can be overloaded and highly congested. We propose a *hierarchical routing strategy* which is very efficient for a large class of our *distance-bias* structures: a short $s - t$ route within a parent region can be formed by combining a few sub-routes (just two in the grid setting) within a few sub-regions. Based on this, our routing scheme [11] uses a *small distributed routing database and finds routes of poly-log expected length.*

Hybrid wireless ad-hoc networks. We propose network constructions for the new paradigm of building hybrid ad-hoc networks by adding a wired infrastructure to an unstructured (ad-hoc) wireless network [4]. Our basic model [12] is to 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 to buy long links, we choose links so the routing diameter is a poly-log, 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 weight.

Routing with dynamic capabilities with an example in optical networks. In [11], we consider routing with Quallity of Service provisioning in a network where the links are associated with (bandwidth) capabilities which may change arbitrarily. 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). Dolev et al. [1] propose the first dynamic routing scheme for high-speed networks and rank dynamic schemes by an *adaptivity* measure defined as the maximum number of sites to be updated upon a single topology change. We propose a new approach with the assumption that links tend to favor closer distances and therefore obtain a significantly better adaptivity. Ideally, we can achieve adaptivity $O(1)$ which improves the result in [1] (for a general topology) of $O(3^k n^{1/k})$ for an internal parameter k.

4. Future research plan

I envision 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. My work so far has resulted in a number of significant results, but there are many additional issues to explore.

Modeling issues. I initially start with path-length features (small diameter) and geographical factors (distance-bias random links and growth-bounded base graphs) and later, focus more on degrees. In the next big step, I intend to develop models which *combine distance-bias distributions of random links with power-law degrees* (besides other existing features such as growth-bounded localities). If successful, such an approach would be the first to analyze the role of geographical factors behind the well-known small-world properties and power-law degrees. Such a new model would also be a powerful tool in many important problems, including developing generators of the Internet topology or other related networks. With respect to power-law degrees, our routing algorithms do not exploit high-degree nodes (unlike most recent work in power-law research) but are primarily based on localities (a connected component in a neighborhood) of rich enough long (random) links. This suggests that we can develop strategies to avoid

---

renting), small routing diameter means limited hops in a path, and bounded degree means a bounded number of physical links connected to a node. Although the weight of a link in our abstract model can be naturally seen as (or proportional to) the Euclidean distance between nodes, this can also be realized by other specific measures, e.g. the Internet transmission delay relying too much on high-degree nodes, which then improves network load-balancing (or reduces congestion) and network robustness and resilience (to accidental collapses of high-degree nodes). Although our adoption of bounded-growth in localities and distance-bias links is justified, we need to consider experimental tasks to validate our assumptions and tune our parameters. If the distributions we use are too simple to closely model the long links in practice, we need to extend our models to more complicated settings.

Design issues. We have only provided concrete designs for simple grid or grid-like settings ([12]), where we assume a uniform distribution of nodes and uniform node-to-node communication demands. I plan and have already begun to work on concrete designs for a more general setting, namely Euclidean metrics [10], where nodes can be distributed non-uniformly and there can be big "holes on the surface" (like lakes and seas in practice) – areas without nodes. I also have already begun to work with more general settings, where node-to-node communication demands are not uniform. An approach using multicommodity network flows is already initiated [12]. Distance bias in real-world networks may also be too complicated to model closely by a well-defined distribution. I intend to refine our general routing scheme for a general topology (without distance-bias assumption) and evaluate its efficiency by experiments with practical data for real-world networks.

Other applications. Our design framework can be useful as a natural platform to combine several mutual restrictive factors in our routing schemes. Thus, our study on routing networks can also be extended for some related topics such as Euclidean spanners, compact routing. I also intend to consider the issue of finding vertex-disjoint paths between a given source-destination pair in our network models. This helps to design secure protocols in some private communication scenarios, where users and secure communication links (e.g. using shared secret keys) can form small-worlds [3].

References

[1] S. Dolev, E. Kranakis, D. Krizanc, and D. Peleg, "Bubbles: Adaptive routing scheme for high-speed dynamic networks," *SIAM J. Comput.*, vol. 29, no. 3, pp. 804–833, 1999.

[2] C. Faloutsos, P. Faloutsos, and M. Faloutsos, "On power-law relationships of the internet topology," in *ACM SIGCOMM*, 1999.

[3] M. Franklin and C. Martel, "Small-world networks and secure distributed protocols," Tech. Rep., UC-Davis, 2002.

[4] A. Helmy, "Small worlds in wireless networks," *IEEE Commun. Lett.*, vol. 7, no. 10, pp. 490–492, Oct. 2003.

[5] J. Kleinberg, "The small-world phenomenon: An algorithmic perspective," in *Proc. of ACM Symp. on Theory of Computing (STOC)*, 2000, pp. 163–170.

[6] D. Malkhi, M. Naor, and D. Ratajczak, "Viceroy: A scalable and dynamic emulation of the butterfly," in *Proc. of ACM Symp. on Princ. of Dist. Comp. (PODC)*, 2002, pp. 183–192.

[7] 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.

[8] S. Milgram, "The small world problem," *Psychology Today*, vol. 22, pp. 61–67, 1967.

[9] V. Nguyen and C. Martel, "Analyzing and characterizing small-world graphs," in *Proc. of ACM Symp. on Discrete Algorithms (SODA)*, 2005, pp. 311–320.

[10] ——, "Efficient dynamic routing schemes in euclidean metrics," Tech. Rep., In prepartion.

[11] ——, "Localities plus (distance-bias) random links: Small-world models and applications," Submitted to STOC'06.

[12] ——, "Designing low cost networks with short routes and low congestion," To apprear in INFOCOM'06.

[13] ——, "Non-uniform random links in small-world graphs: Models, analysis and applications in network designs," To be submitted to SIAM J. on Computing.

[14] D. Watts and S. Strogatz, "Collective dynamics of small-world networks," *Nature*, vol. 393, pp. 440–32, 1998.

[15] 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.