Hidden Metric Spaces and Navigability of Complex Networks

Dmitri Krioukov
CAIDA/UCSD
F. Papadopoulos, M. Boguñá, A. Vahdat, kc claffy
Science or engineering?

- Network science vs. network engineering
- Computer science vs. computer engineering
- Study existing networks vs. designing new ones
- We cannot really design truly large-scale systems (e.g., Internet)
  - We can design their building blocks (e.g., IP)
  - But we cannot fully control their large-scale behavior
  - At their large scale, complex networks exhibit some emergent properties, which we can only observe: we cannot yet fully understand them, much less predict, much less control
- Let us study existing large-scale networks and try to use what we learn in designing new ones
  - Discover “nature-designed” efficient mechanisms that we can reuse (or respect) in our future designs
Internet

- **Microscopic view ("designed constraints")**
  - IP/TCP, routing protocols
  - Routers
  - Per-ISP router-level topologies

- **Macroscopic view ("non-designed emergent properties")**
  - Global AS-level topology is a cumulative result of local, decentralized, and rather complex interactions between AS pairs
  - Surprisingly, in 1999, it was found to look completely differently than engineers and designers had thought
    - It is not a grid, tree, or classical random graph
    - It shares all the main features of topologies of other complex networks
      - scale-free (power-law) node degree distributions \( P(k) \sim k^{-\gamma}, \gamma \in [2,3] \)
      - strong clustering (large numbers of 3-cycles)
Problem

“Designed parts” have to deal with “emergent properties”

- For example, BGP has to route through the existing AS topology, which was not a part of BGP design
Routing practice

- **Global (DFZ) routing tables**
  - 300,000 prefix entries (and growing)
  - 30,000 ASs (and growing)

- **Routing overhead/convergence**
  - BGP updates
    - 2 per second on average
    - 7000 per second peak rate
  - Convergence after a single event can take up to tens of minutes

- **Problems with design?**
  - Yes and no
Routing theory

- There can be no routing algorithm with the number of messages per topology change scaling better than linearly with the network size in the worst case.
- Small-world networks are this worst case.

Is there any workaround?

If topology updates/convergence is so expensive, then may be we can route without them, i.e., without global knowledge of the network topology?

What about other existing networks?

CCR, v.37, n.3, 2007
Navigability of complex networks

- In many (if not all) existing complex networks, nodes communicate without any global knowledge of network topologies; examples:
  - Social networks
  - Neural networks
  - Cell regulatory networks

- How is this possible???
Hidden metric space explanation

- All nodes exist in a metric space
- Distances in this space abstract node similarities
  - More similar nodes are closer in the space
- Network consists of links that exist with probability that decreases with the hidden distance
  - More similar/close nodes are more likely to be connected
Mathematical perspective: Graphs embedded in manifolds

- All nodes exist in “two places at once”:
  - graph
  - hidden metric space, e.g., a Riemannian manifold

- There are two metric distances between each pair of nodes: observable and hidden:
  - hop length of the shortest path in the graph
  - distance in the hidden space
Greedy routing (Kleinberg)

- To reach a destination, each node forwards information to the one of its neighbors that is closest to the destination in the hidden space.
Observable network topology

Hidden metric space
Result #1: Hidden metric space do exist

Their existence appears as the only reasonable explanation of one peculiar property of the topology of real complex networks – self-similarity of clustering

Result #2: Complex network topologies are navigable

- Specific values of degree distribution and clustering observed in real complex networks correspond to the highest efficiency of greedy routing.
- Which implicitly suggests that complex networks do evolve to become navigable.
- Because if they did not, they would not be able to function.

Nature Physics, v.5, p.74-80, 2009
Result #3:
Successful greedy paths are shortest

- Regardless the structure of the hidden space, complex network topologies are such, that all successful greedy paths are asymptotically shortest
- But: how many greedy paths are successful does depend on the hidden space geometry

Phys Rev Lett, v.102, 058701, 2009
Result #4: In hyperbolic geometry, all paths are successful

- Hyperbolic geometry is the geometry of trees; the volume of balls grows exponentially with their radii.
- Greedy routing in complex networks, including the real AS Internet, embedded in hyperbolic spaces, is always successful and always follows shortest paths.
- Even if some links are removed, emulating topology dynamics, greedy routing finds remaining paths if they exist, without recomputation of node coordinates.
- The reason is the exceptional congruency between complex network topology and hyperbolic geometry.
Result #5: Emergence of topology from geometry

- The two main properties of complex network topology are direct consequences of the two main properties of hyperbolic geometry:
  - Scale-free degree distributions are a consequence of the exponential expansion of space in hyperbolic geometry
  - Strong clustering is a consequence of the fact that hyperbolic spaces are metric spaces
Shortest paths in scale-free graphs and hyperbolic spaces
In summary

- Complex network topologies are congruent with hidden hyperbolic geometries
  - Greedy paths follow shortest paths that approximately follow shortest hidden paths, i.e., geodesics in the hyperbolic space
    - Both topology and geometry are tree-like
- This congruency is robust w.r.t. topology dynamics
  - There are many link/node-disjoint shortest paths between the same source and destination that satisfy the above property
    - Strong clustering (many by-passes) boosts up the path diversity
  - If some of shortest paths are damaged by link failures, many others remain available, and greedy routing still finds them
Conclusion

- To efficiently route without topology knowledge, the topology should be both hierarchical (tree-like) and have high path diversity (not like a tree)
- Complex networks do borrow the best out of these two seemingly mutually-exclusive worlds
- Hidden hyperbolic geometry naturally explains how this balance is achieved
Applications

- Greedy routing mechanism in these settings may offer virtually infinitely scalable information dissemination (routing) strategies for future communication networks
  - Zero communication costs (no routing updates!)
  - Constant routing table sizes (coordinates in the space)
  - No stretch (all paths are shortest, stretch=$1$)

- Interdisciplinary applications
  - systems biology: brain and regulatory networks, cancer research, phylogenetic trees, protein folding, etc.
  - data mining and recommender systems
  - cognitive science