John Hanacek

Hidden Distance Underlies Graph Topology

Some years ago I began to wonder: how does information propagate inside of networks? Does it have general dynamics that hold true across all networks? I wanted to know beyond the symptom level, I wanted to know what flowing “looks like” at the level of node-to-node communication. How does a given node “see" the inside of a network?

I was–and still am–troubled by the seeming inevitability of power laws in the distribution of fame, money and power. The microcosm of my thinking was Twitter. Twitter is a small world network where individuals can find short paths to many other people in the network. Preferential attachment–also known as the “rich get richer” effect–is highly present on Twitter causing clusters to form with certain accounts being highly connected (high follower counts) and a long tail of accounts with low followers. An account with more followers is more “visible” on the overall Twitter network; information transmitted from such an account has a much larger potential to reach more people. The concept of visibility is referred to colloquially, but I began to see it literally: as you would see a point on the top of a mountain more visibly than a point next to you on a flat plane. I wondered if clustering activity deformed an unobserved landscape that packets of information must travel. If so did this deformation influence how information flows in the topology? Does the deformation influence further connections, conveying advantage or disadvantages even?

My hypothesis was that there was actually a topography underlying the observable topology. I visualized this as a landscape with hills and valleys. I believed that there was a hidden “distance” and “elevation” between certain nodes that was related to how many connections a given node had. Like a river running through a landscape, might information take a path of least resistance when flowing through a network? I thought that information would be inherently more apt to flow toward highly connected central nodes, and also information would be more visible if it came from said nodes.

I found studies exploring the concept of hidden metric spaces mapped to geometry, but did not formally know social network analysis and so was stuck. I am still a beginning amateur, but with what I have learned in class so far I am much more equipped and have found that my intuitions were correct in some areas and demand more exploration in others.

Overall, my hope is that mapping metrics like clustering coefficient to geometry could allow new insights about the structure and evolution of edges in networks, and potentially shine light on information flow dynamics. At minimum, methods for displaying metrics like clustering and density of triangles in a component would allow people to use their natural spatial reasoning capabilities to analyze graphs in novel ways. By making network analysis metrics into visual representations embedded the graph views one could allow even non-experts to bring insight to the study of graphs.

The first paper I will discuss is titled “Curvature of co-links uncovers hidden thematic layers in the World Wide Web” by Jean-Pierre Eckmann and Elisha Moses from 2002. The paper explores cooperative linking between websites. “Beyond the information stored in pages of the World Wide Web, novel types of ‘meta-information’ are created when pages connect to each other.” (5825) The authors focus on the clustering coefficient of websites. “Reciprocal links (co-links) between pages signal a mutual recognition of authors and then [we] focus on triangles containing such links, because triangles indicate a transitive relation. The importance of triangles is quantified by the clustering coefficient.” (5825)

The authors deploy automatic web crawling agents, known as robots to discover triads (triangles in their language) and measure the clustering coefficient. The paper employed a robot to generate the graph by crawling over 300,000 URLs and clustering was implemented by determining whether a link is local or remote. “A local link is basically one inside a given site and any other link is remote.” (5827) The authors set out to define a geometric curvature tied to the clustering coefficient of triads in order to reveal thematic elements in the web: high curvature characterizes a common topic.

Using the law of cosines, a general definition of mathematical curvature is applied to walks in the sample graph of the WWW. “Triangles capture transitivity, which we measure by the associated notion of curvature.” (5825) “The average number of triangles c_n at a node n is related to the average distance between any two of its nearest neighbors (say n’ and n”). This distance is measured by counting links in the shortest path from one node to another, namely n’ and n”.” (5826) Thus the authors implement the metric of minimal number of hops between nodes to arrive at definition of curvature, which allows them to visualize the WWW geometrically. “We shall say that it looks like a hyperboloid when it is tree-like, with exponential separation of branches that ‘fan out’. In contrast, a highly connected region looks

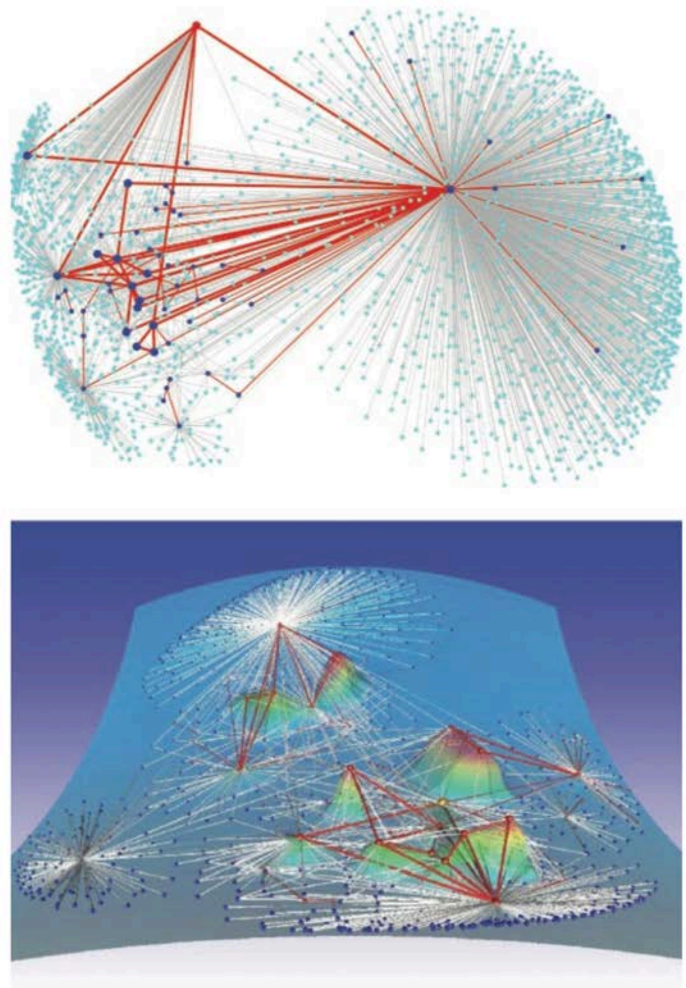

Fig. 3. (*Top*) The nodes (2,125) and links (2,755) of a site specializing in tango (specifically the music of Piazzolla). The co-links are shown in red, congruence triangles are enhanced, and one-way links are black. The anchor is red. The sparsity and high connectivity of the congruence graph are striking. (*Bottom*) A view of the curvature for the 947 nodes and the 1,307 links of a notorious Swiss revisionist (anti-Semitic) site. Height above the blue surface is proportional to curvature (of congruence triangles). The one-way graph is clearly less connected, its links are more likely to be dangling ends, and it is generally dominated by the appearance of hubs.

more like a sphere that is 'closed' on itself." (5826) Nodes that are mutually recognized by peers express high curvature thus, "curvature captures contextual vicinity rather than notoriety of a set of pages."

The curvature of the graph was heavily tied to and influenced by triangle co-links containing at least one reciprocal side. "This curvature is not produced by triangles with three co-links alone, because of their small relative number, but rather from triangles that have a combination of co- and one-way links." (5827) The density of triangles, thus average curvature, of a random graph tends toward zero with the growth in the size of the graph. Curvature as the authors define it is a local phenomenon. "Highly connected nodes […] such as hubs and authorities [tend] to have, on average, a low curvature. This finding agrees with our intuition that such nodes contribute perhaps to global connectivity but not to the local interest groups that curvature identifies." (5828) This locality "develops a different direction than global notions like scaling in the WWW and the 'small-world' effect." (5829) The authors contend, "By using curvature, we show that practically all the connective information depends on co-links. Thus, our method reveals a new geometric view of the locus of information in networks, which is not 'put there' but 'happens' as a collective effect." (5825) Further they believe that codifying the clustering coefficient as geometry has promise for freeing us from the "subjectivity of contextual concepts, such as meaning, content, and the like." (5829)

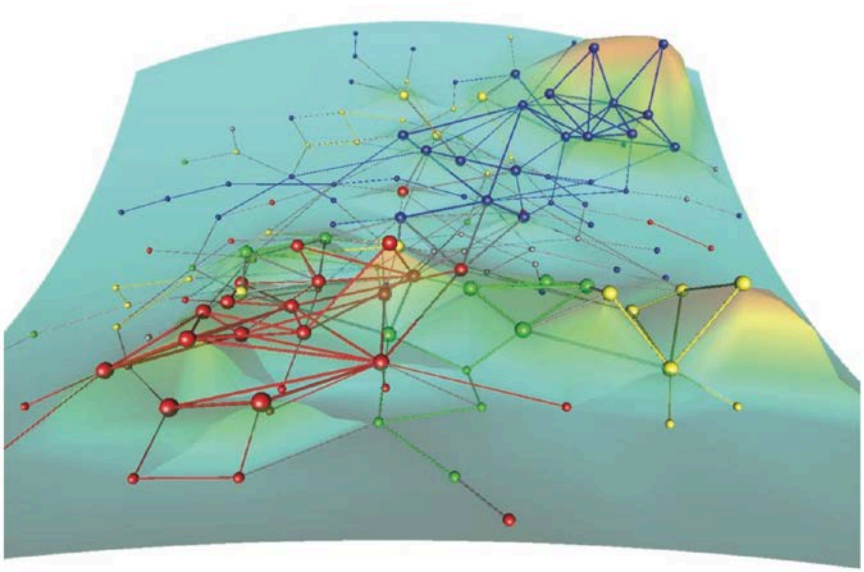

Fig. 5. The worm brain. The height is proportional to curvature. The red nodes are amphid cells, the yellow nodes are other sensory neurons of the head, and blue nodes are motor neurons of the nerve ring. Only co-links are shown, and triangles are enhanced.

What I found interesting about the paper was the perspective of the robot script and what that revealed about efficiency of traveling inside networks: “a random walk by the robot in the graph will tend to get trapped somewhat longer in highly curved regions.” (5827) This implies that their notion of curvature is not just a useful metric for visualization, but may actually be a real feature of the structure of the local topology. As the authors conclude: “Our study further shows that robots perceive the Web differently than the people who actually write and use the pages. Our robots go back and forth between various sites, gaining a more coherent view of their relations. We have shown that the geometric properties of the space in which they roam and the landscape that they reconstruct reveal new connective meta-information, hidden from the common user.” (5829) Indeed this is what the second paper I looked at explores in a broader theoretical scope.

“Navigability of Complex Networks” by Marián Boguñá, Dmitri Krioukov and K. C. Claffy, explores transport inside of networks with a focus on routing, which is the concept of signaling information propagation paths through a complex network maze. Observed real world efficiency of the routing process across network examples led the authors to ask: “how is this efficiency achieved?”(74) They expand the problem: “When each element of the system has a full view of the global network topology, finding short routes to target destinations is a well-understood computational process. However, in many networks observed in nature, including those in society and biology (signaling pathways, neural networks and so on), nodes efficiently find intended communication targets even though they do not possess any global view of the system.”(74) Their work was inspired by Stanley Milgram’s small world experiments, in which participants routed information through a complex network without global information about the network.

The authors “assume the existence of a hidden metric space, an underlying geometric frame that contains all nodes of the network, shapes its topology and guides routing decisions.” (75) The hidden metric is based on node similarity, defined by shared intrinsic characteristics of given nodes, this manifests as a hidden distance that guides signaling (routing) efforts. Complimentary to the work of Eckmann and Moses, here the authors employ the measure of the number of triangles in a network topology, known as clustering; this defines the shared intrinsic characteristics. The metric space allows an easy mapping of clustering: the more powerful the influence of the network’s underlying metric space on observable topology, the more strongly it is clustered.

In the hidden metric space, distances between nodes abstract to their similarity as measured by clustering. The hidden distances guide routing decisions by creating a kind of landscape. “These distances influence both the observable topology and routing function: (1) the smaller the distance between two nodes in the hidden space, that is, the more similar the two nodes, the more likely they are connected in the observable topology; (2) nodes also use hidden distances to select, as the next hop, the neighbour closest to the destination in the hidden space.” (75) Using this understanding to route information on a node-to-node basis is called “greedy routing,” it is simply the act of

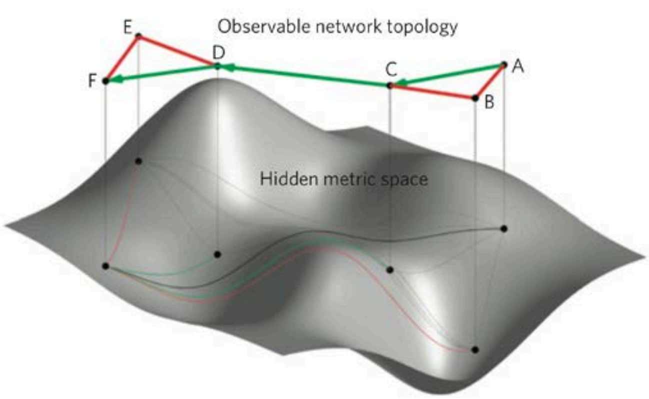

Figure 1 | How hidden metric spaces influence the structure and function of complex networks. The smaller the distance between two nodes in the hidden metric space, the more likely they are connected in the observable network topology. If node A is close to node B, and B is close to C, then A and C are necessarily close because of the triangle inequality in the metric space. Therefore, triangle ABC exists in the network topology with high probability, which explains the strong clustering observed in real complex networks. The hidden space also guides the greedy-routing process: if node A wants to reach node F, it checks the hidden distances between F and its two neighbours B and C. Distance CF (green dashed line) is smaller than BF (red dashed line); therefore, A forwards information to C. Node C then carries out similar calculations and selects its neighbour D as the next hop on the path to F. Node D is directly connected to F. The result is path A → C → D → F shown by green edges in the observable topology.

minimizing the distance to the destination at each hop. The authors determine that strong clustering improves navigability of information forwarded in the greedy routing paradigm. This is due to an observed "zoom-out/zoom-in" mechanism that occurs. This is discussed in the context of air travel networks, where the goal is to travel according to the greedy routing strategy: using geography as the underlying metric space, at each airport one chooses the next-hop airport that is geographically closest to the destination. "The turning point between the [zoom-in/zoom-out] phases appears naturally: once we are in a hub near the destination, the probability that it is connected to a bigger hub closer to the destination sharply decreases, but at this point we do not need hubs anyway, and greedy routing directs us to smaller airports at shorter distances next to the destination." (77) This is because "preference for connections between nodes nearby in the hidden space means that low-degree nodes are less likely to have connectivity to distant low-degree nodes; only high-degree nodes can have long-range connection that greedy routing will effectively select." (78) Greedy routing in this way only works if there are sufficient hubs and sufficient clustering conditions in the network.

This paper is an important conceptual leap because it allows us to think more clearly about how information is routed and new connections are formed using spatial reasoning. The authors find that "the connection cost increases with hidden distance, thus discouraging long-range links." (78) Particularly interesting is that "in making connections, rich (well-connected, high- degree) nodes care less about distances (connection costs) than poor nodes." (75) This speaks to power dynamics in social settings, with well-connected people able to traverse vast social distance at minimal cost, while non-connected individuals may get stuck in local minima.

The emphasis in Doguña et. al is on navigation, specifically on enabling scalable internet packet routing schemes. They discuss how current routing database methods are not scaling as routers communicate constantly in an attempt to retain a global view of the Internet. “Discovery of the Internet’s hidden metric space would remove this bottleneck, eliminating the need for the inherently unscalable communication of topology changes. Instead routers would be able to just forward packets greedily to the destination based on hidden distances.” (79) This work has implications beyond the Internet and indeed applies to any scale-free complex network; it is a new frontier of network science.

I now am now confident that my intuition was correct: there is a hidden metric space underlying network topologies. However I am entirely unsure just what that means for my ideas of topography and information flow. These papers set the groundwork for a better understanding of how complex scale-free small world networks function, such as societies, the Internet, and brains. Doguña et. al focus on routing and information transport, while Eckmann and Moses focus on edges between web pages; this work ought to be synthesized into a more robust exploration. I think the next step is to explore mapping the hidden metric space over time. Does the specific geometry of the hidden metric space of given clusters allow us to predict specifics of its future connection evolution? Are hidden metric spaces purely emergent from observed topology, or do they “push back” and set clusters along path dependent conditions. That is, are nodes more likely to connect with clusters that are deforming the local metric space environment? How closely can we tie geometry to topology? I am struck by the potential of applying a gravity model to hidden metric spaces and nodes embedded in them. Like cosmic objects deform the “hidden metric” called space-time to create gravity wells, do strongly clustered sets of nodes manifest a kind of gravity drawing new connections in? What of information?

After reading the two papers above, I also uncovered a study that employed a gravity model to create a recommendation algorithm within clusters and reported positively on the efficacy of the gravity analogy in clustering. The authors concluded that “Numerical analyses [..] proved the gravity mechanism can characterize the structure of real networks better than two baseline stochastic methods, [Erdős–Rényi] ER and [preferential attachment] PA models.” (6) Hardly a slam-dunk backing me up, but a solid lead: clearly there are more phenomena inside of network topologies than was previously considered, and gravity may even be lurking inside. I think the study of hidden metric spaces also demands a study of the dynamics of interactions in the metric space. As it matures, this approach could be imported into the sociological sciences and economics allowing them to ask empirical questions about network flow dynamics in society. We could ask better questions about how to design social systems that minimize “superstar economics” and inequal power-law distributions of wealth.

Putting aside huge scales and left-field ideas about network gravitation since they are very speculative, I am at least convinced that utilizing geometry to explore landscapes underlying given topologies can yield new understanding of how information flows inside of networks. The distance of a walk it seems has a deeper meaning to it, and this can be visualized and explored geometrically. The hidden metric space approach can inform people as they seek to understand the signaling propagation in the brain and other biological networks as well as allow people to build much more flexibly scalable complex communications networks. I hope it will bring us closer to understanding how information transmission functions in a more abstract way that is generally implementable across more kinds of complex networks, as well as giving us insight into a potentially deeper nature of preferential attachment. The study of hidden geometries underlying network topologies is just beginning and I think it could be some of the most important contemporary work in network science.

Works Cited

Boguñá, Marián, Dmitri Krioukov, and K. C. Claffy. "Navigability of Complex Networks." *Nature Physics* 5.1 (2008): 74-80. Web. <http://complex.ffn.ub.es/ckfinder/userfiles/files/nphys1130.pdf>.

Eckmann, J.-P., and Elisha Moses. "Curvature of Co-links Uncovers Hidden Thematic Layers in the World Wide Web." *Proceedings of the National Academy of Sciences* 99.9 (2002): 5825-829. Web. <http://www.weizmann.ac.il/complex/EMoses/pdf/CurvatureOfCo-links.pdf>.

Liu J-H, Zhang Z-K, Chen L, Liu C, Yang C, et al. (2014) Gravity Effects on Information Filtering and Network Evolving. PLoS ONE 9(3): e91070. doi:10.1371/journal.pone.0091070