Harish Chintakunta

hkchinta@illinois.edu
919-576-0844
www.harishchintakunta.com

Research Statement

My research interests include the areas of graph theory and network analysis, topological data analysis (TDA), signal processing, and distributed algorithms.

TDA is a relatively new and upcoming field where the objects of interest are topological invariants in contrast to geometric invariants classically used in machine learning, image processing and computer vision. Example topological invariants would be number of holes on a surface, number of voids enclosed in a point cloud, etc, which are invariant not only to the coordinate system chosen, but also to continuous deformations. From topological perspective, the surface of a bunny would be equivalent to the surface of a sphere! My research experience detailed below exemplifies the applications where such level of generality is not only useful, but most appropriate.

Topological invariants come in the form of algebraic objects, such as groups and vector spaces, the formal study of which is categorized as algebraic topology. It relies heavily on combinatorial objects, simplicial complexes for instance, which may be interpreted as generalization of graphs: graphs specify selected pair-wise combinations of the given vertices, whereas simplicial complexes generalize this to higher order combinations. As an engineer myself, the challenging part was to overcome the complex jargon associated with such rich (and often, quite abstract) mathematical subject and distill the useful concepts to linear algebra. In my writings, I focus on simplifying the subject in special cases and demonstrating its applications, thereby extending the reach of TDA.

Past Research

Sensor Networks:
There are certain problems in sensor networks which are inherently topological in nature. Detecting coverage holes is one such example, where coverage holes refer to regions which are not monitored by any sensors. Note that existence of coverage holes is a topological event, in the sense that it is invariant to continuous deformation of the coverage space. Algebraic topology offers meaningful formalization of topological properties such as coverage area of a sensor network in a coordinate free setting. As I show in [1], the problem of detecting a coverage hole may be expressed in terms of properties of elements in the null space of a certain matrix, which can be distributively tested using cleverly designed graph theoretic methods and distributed algorithms. The algorithm developed here is much faster than other existing techniques, and with the help of topological persistence, is the only one that is robust to errors in proximity detection. My collaboration [5] shows topological techniques such as zig-zag persistence can be used to track coverage holes and boundaries in mobile networks too.

Power consumption is important in the context of sensor networks, and one way to help solve this problem is to have a small number of sensors switched on at any given time, but without compromising the coverage. Once again, this problem has a good formulation in topological terms: Is there a subcomplex of the given network which has the same topology?, and can this subcomplex be computed efficiently? If a subcomplex has the same topology as the original network, this means selecting only the subcomplex does not create extra coverage holes. Minimizing the number of sensors involved in other signal processing and communication tasks will also help in conserving power. We show in [6] that TDA techniques can be employed to compute such a subcomplex efficiently and distributively.

Unlike the most general case of no geometric information available as discussed above, distributive techniques exists when either the proximity distances between the nodes or their precise coordinates are known. Two of the interesting techniques arise from 1) restricted Delaunay triangulations, which are planar graphs

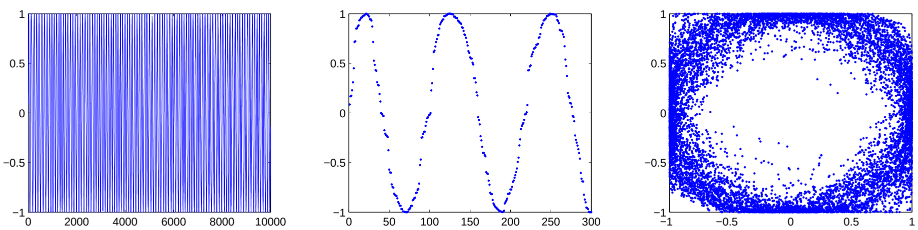

Figure 1: Figure shows the conversion of a time series to a point cloud. A non-uniformly sampled periodic signal $s(t)$ (left) is converted into a point cloud using the delay embedding $(s(t), s(t-\tau))$. The figure in the middle shows a "zoomed-in" version of the time signal. The problem of detecting periodicity is converted into detecting a hole or loopy structure in the point cloud. The latter type of analysis can be effectively performed by TDA techniques, and is robust to non-uniform sampling and missing data.

originally developed for efficient routing, and 2) alpha shapes, which were originally developed in the computer vision community to define shapes of arbitrary point cloud. Using topological ideas, I show in [2] that these two objects are closely related, thereby providing deeper insight into the success of these methods.

Biomedical Applications:
Technology of self-sustainable and wearable sensors will play a major role in the future of medical care. One application of this technology is to detect wheezing in the patients breathing, and particularly a type of wheeze called strider, which is usually a symptom of a life threatening event. Wheezes can be distinguished from regular breathing by their quasi periodicity. Using delay embedding, the problem of detecting this quasi-periodicity in a time signal can be converted to that of detecting loopy structure in a point cloud, which in turn is performed using topological techniques. Figure 1 such a conversion of a time signal into a point cloud. Our research [3] shows topological methods have better complexity and accuracy compared to other existing techniques based on time signal processing, and are robust to missing data.

Social Networks:
My recent work in social networks sheds light on the role of topology in community detection and extracting the large scale structure of the network. Detecting community structure helps in inferring important demographics and communal organization in the society under study, which would otherwise be very expensive to obtain. The large scale structure of the network complements this information by describing how the individual communities are connected to each other. The concept "large scale structure" has rather straightforward topological implications. This was our motivation to explore the role of topology in social networks, and analogous to my work in sensor networks, we wanted to see if there a core subnetwork which has the same topology as the entire network.

This led to interesting notions of a core and periphery structure of a network. We found strong evidence [4] to support a meaningful core-periphery structure in the sense that the periphery contains most of the communities and the distances between vertices in the core are independent of vertices in the periphery. We also provide a fast and distributed algorithm to compute the core and periphery decomposition. Furthermore, related research from a very different line of investigation shows a similar core-periphery structure in social networks.

Current Research

My current research is focused on 2 aspects of topological data analysis, 1) detecting transient events in dynamical systems, and 2) topology in social networks.

Transients in dynamical systems:
In the first direction, the goal is to transform the observation signal of the dynamical system into a high dimensional point cloud using delay embedding and detect transient events via changes in topology of the

resulting point cloud. Transient events and phase changes in the states of dynamical system often indicate a significant event, and timely detection of such events have many applications. The effectiveness of TDA methods in wheeze detection in my recent research strengthens my motivation to pursue this direction. The approach is to select a window in the time signal and embed this window in a high dimensional space. Repeating this by sliding the window produces a point cloud. The phase change and transients in the system are then modeled in terms of topological properties of the resulting point cloud. The techniques developed may potentially be applied to any dynamical system where transients are informative.

Biomedical applications: We are currently conducting experiments on electroencephalogram (EEG) and intra-cranial electroencephalogram (iEEG) recordings of cortical activity. EEG signals are obtained by attaching electrodes at suitable locations on the scalp, and iEEG signals are those obtained by attaching electrodes directly on the surface of the brain under the scalp. EEG signals are used in studying brain wave patterns and diagnosing several medical conditions. The current research goal is to explore correlation between events in cortical activity and topological properties of the point cloud obtained from embedding EEG signals in high dimensional space.

Cyber-Physical Systems: Cyber-physical systems usually refer to distributed control systems comprising two interconnected networks: 1) a physical network which consists of the operating equipment, and 2) a cyber network which consists of control units with communication and processing capabilities. The power grid is one such example. Real-time monitoring and early detection of events in power-grids, and consequently failures, is a rapidly growing field of research. There are several factors which lead to the current interest in distributed control of power grids, including 1) the enormous cost of a large scale failures, 2) the potential of significantly reducing the impact by early detection and taking the right control actions, and 3) recent increase in deployment of phasor measurement units (PMUs) in the power grids.

My current research focuses on detecting large scale oscillatory events in the network. The motivation for applying TDA techniques to detect such oscillatory events stems directly from its success in detecting similar events in other dynamical system, such as in wheeze detection in breathing signals [3].

Social Networks:

The current research in social networks is to 1) develop efficient community detection algorithms, and 2) characterize properties of the core which could accurately model information flow in real world networks, with applications including epidemiology.

Despite huge body of literature in community detection algorithms, some of which perform very well in practice, these algorithms can be computationally intensive specially due to large size of the networks. This necessitates the design of efficient algorithms. An effective strategy for detecting relatively small communities in large networks is to partition the network into smaller sub-networks and search for communities within these smaller sub networks. The core periphery-decomposition resulting from my previous research [4] offers a solution to this problem. Since the periphery contains most of the communities in the network, we are currently designing algorithms which search for communities within the connected components of the periphery. If successful, the resulting algorithms will provide very efficient algorithms for community detection since the core-periphery decomposition itself can be computed distributively and with low computational complexity. Furthermore, closer study of the structure of communities which are inside the core will provide more insight into different kinds of communities.

Our preliminary results also provided evidence to support the fact that the core of the network has larger expansion compared to the entire network. Intuitively, larger expansion for a network implies that it is difficult to divide the network. That is, it takes the removal of a large number of edges in the network to divide it two large (disconnected) subnetworks. Several lines of evidence, including statistical analysis and studying diffusion properties, show that larger edge expansion implies larger rate of information flow in the networks. In the problem of disease spread in a social network where the edges are determined by contact, higher expansion of the network would imply quicker spread of the disease and makes early detection imperative to control it. We are currently investigating the correlation between expansion properties of the core and rate of information flow in many real-world social networks. If the results show that the core plays an important part in the spread of information (disease for example), vertices in the core will be suitable places to take action in order to control the spread of this information in the network.

Future Research

Since TDA is still in its infancy, lots of avenues are still to be explored. The major areas I want to work in the near future are 1) social networks, 2) biomedical applications, and 3) Cyber-physical systems.

Social Networks:
In social networks, identifying generative models that will give rise to a network with specified topological properties, can contribute greatly to our understanding of how social networks are formed. This research will explore how well the existing generative models will create the core-periphery structure observed in real world networks, and design new generative models if necessary. There are interesting questions concerning the topology of individual communities. Existing research shows that graphs representing real world communities share some graph theoretic properties, such as low conductance and high triangle participation ratio for example. It is then natural to ask if there are any graphs which also share these properties but not topological properties observed in real networks. An affirmative answer in this case will mean that it is necessary to consider topological properties in order to characterize real world communities.

Network evolution modeling is another promising direction of TDA in social networks, and likely also in other types of networks. Network evolution models are important in order the predict and control the future state of the network, with applications including designing policies to control and fighting spread of infectious diseases. Existing research has shown that real networks have a subgraph, which is also called a "core", which is stable over time, and a "periphery" which changes over time. Therefore, the core in this sense serves, besides other purposes, to estimate the properties of the network which remain invariant over time. Once gain, it is natural to ask if the core-periphery structure determined according to network evolution matches with that obtained using topological considerations.

Biomedical applications:
Use of TDA in detecting transients in dynamical systems and in network analysis, will play a major role in biomedical applications. Biological systems, gene regulatory networks for example, are usually are very complex, with many variables influencing one another. Graphs are natural representation and analysis tools in these contexts as they can effectively capture the pair wise relationships amongst the variables. Once represented as a graph, the practitioners have available to them a massive set of tools developed for analysis on graphs. An important way techniques from TDA can contribute in these contexts is to generalize the analysis from pair wise interactions to higher order interactions using simplicial complexes, and subsequent analysis of topological properties. Another example where TDA can be useful is to discover community structure in protein interaction networks.

Cyber-Physical Systems:
Real-time classification of events is a very important topic in cyber-physical systems where I believe TDA has a major role to play. The underlying dynamics of the power-grids depend on the slow changing topology of the network, and the much quicker changes in factors such as load/demand ans supply. This complexity is driving the research in this field towards data driven approaches. The approach of converting PMU data, a time signal, into a point cloud and classifying/detecting events by analyzing the topology of the resulting point cloud using TDA techniques, will make the process more robust to missing data, and will likely ease the requirement of perfect synchronization across PMUs resulting in lower cost of these units.

Another big way my research can contribute to this field is the development of distributed algorithms for analysis and control of the power grid. The current approaches are predominantly based on collecting the PMU data at a central location. This imposes a communication burden and transmission delays. In network processing on the other hand can lead to much quicker response. I will work on developing distributed signal processing algorithms for fast real time event detection in power grids.

Some other subjects which fascinate me, and which I think can benefit greatly from TDA, are linguistics and natural language processing. Both these topics are seeing increasing use of techniques from graph theory recently, and TDA in many ways generalizes these techniques.

In the long term, my research philosophy compels me to work on applications which bring engineering, mathematical and scientific principles together to have real impact on our society. Topological data analysis, which combines aspects from all these disciplines, will soon become one of the major branches of machine learning and big data analysis.

References

[1] H. Chintakunta and H. Krim. Distributed localization of coverage holes using topological persistence. *Signal Processing, IEEE Transactions on*, 62(10):2531–2541, May 2014.

[2] H. Chintakunta and H. Krim. Distributed tracking of systematic failures in sensor networks. Submitted to Journal of Complex Networks, 2014.

[3] S. Emrani, H. Chintakunta, and H. Krim. Real time detection of harmonic structure: A case for topological signal analysis. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 3445–3449, May 2014.

[4] J. Gamble, H. Chintakunta, and H. Krim. Emergence of core-periphery structure from local node dominance in social networks. Submitted to International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2015. Preprint: http://www.harishchintakunta.com/publications.html.

[5] J. Gamble, H. Chintakunta, and H. Krim. Coordinate-free quantification of coverage in dynamic sensor networks. Submitted to Signal Processing. Preprint: http://arxiv.org/abs/1411.7337, 2014.

[6] A.C. Wilkerson, H. Chintakunta, H. Krim, T.J. Moore, and A. Swami. A distributed collapse of a network's dimensionality. In *Global Conference on Signal and Information Processing (GlobalSIP), 2013 IEEE*, pages 595–598, Dec 2013.