

Research Statement

Jie Gao¹

My research is on applied computational geometry and distributed algorithm design with applications in wireless sensor networks. Embedded networked sensing devices are becoming ubiquitous across many activities that are important to our economy and life. In these applications, the physical locations of the sensor nodes greatly impact the system design in all aspects from low-level networking and organization to high-level information processing and applications. From a networking point of view, node placement clearly influences network connectivity and sensing coverage, which subsequently affects basic network organization such as naming and routing in the network. From an application point of view, sensor readings exhibit spatial correlations that can be exploited for data compression, approximation and validation. The physical embedding of sensor nodes brings a unique character that differentiates sensor networks from traditional networks, and enables a whole new set of properties that can be exploited for design efficiency.

1 Sensor Network Topology

When a sensor network grows large in scale it is unrealistic to assume that nodes are placed uniformly in a regular shaped region. Terrain variations and obstacles such as buildings often forbid sensor placement; random sensor deployment also likely leads to deployment ‘holes’ or ‘voids’. The presence of holes and irregular sensor network shape bring new challenges in fundamental network components such as sensor coverage, routing and network localization, as well as information processing and data management. The word topology here refers to algebraic topology and the homologous features of the sensor field, rather than the network connectivity alone.

1.1 Extraction and Representation of sensor network topology.

We are the first to study the extraction of nodes on hole boundaries [10, 11] using distributed operations when the node locations are available. In [47] we use path homotopy and cut locus to detect network boundaries without node locations. See Figure 1 for an example. This algorithm is lightweight in communication costs and is the state-of-the-art solution for extracting network boundary cycles from graph connectivity. We extended the idea in [56] to maintain topologically faithful sensor signal contours in a dynamic signal field. Communication costs are made minimum and confined to only nodes with changed values.

To find a compact representation of the network topology, we developed the combinatorial Delaunay complex in [12]. In particular, some sensor nodes are selected as *landmarks* and flood the network. All the nodes closest to the same landmark are grouped into the same Voronoi cell, forming the *landmark Voronoi diagram*. We extract the *combinatorial Delaunay complex* as the dual complex of the landmark Voronoi diagram – there is a Delaunay edge (or in general a k -simplex) between two (or k) landmarks if their Voronoi cells share some common nodes. We show that in a continuous domain \mathcal{R} , the combinatorial Delaunay complex defined with geodesic distances is homotopy equivalent to \mathcal{R} , i.e., capturing the same topological features of \mathcal{R} , under sufficient landmark sampling conditions [18]. Thus the combinatorial Delaunay complex is a sparse representation of the network or domain topology and has many applications in network routing and localization, as shown below.

Another way to represent the network topology is through network segmentation. In [53, 54], we use flow complex to segment the sensor field into pieces, where in each piece all nodes following directions to move away fastest from boundaries arrive at the same sink. This naturally partitions the sensor field along narrow necks. The automatic grouping of the sensor nodes into segments with simple shape, provides a generic approach to handle sensor field with complex shapes. This enables the re-use of existing protocols on an irregular network and makes the development of new protocols transparent to the specifics of the shape of a sensor field.

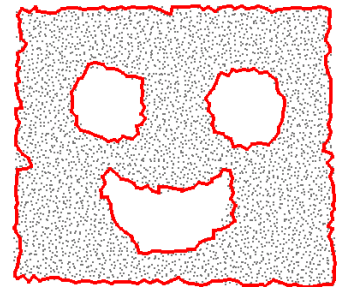


Figure 1. Detect the sensor field boundaries.

¹Department of Computer Science, Stony Brook University, Stony Brook, NY 11794. E-mail: jgao@cs.sunysb.edu.

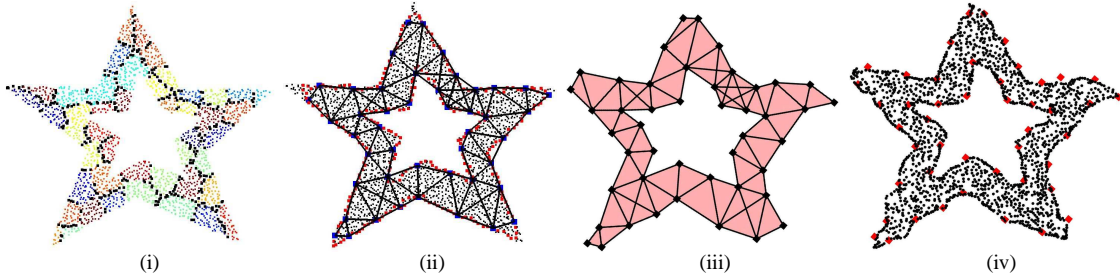


Figure 2. Anchor-free localization from network connectivity, on a double star shape. The number of nodes is 2171. The connectivity follows a unit disk graph model with average node degree 10. (i) The Voronoi cells of the landmarks (black nodes are on the Voronoi edges); (ii) The Delaunay edges extracted from the Voronoi cells of the landmarks; (iii) Our embedding result of the extracted Delaunay complex; (iv) Our localization result of the entire network.

1.2 Network Localization

Since wireless communications in sensor networks are short ranged, two nodes are connected by a link only when they are in close proximity. The problem of network localization is to discover sensor node positions from the graph connectivity, i.e., laying down the network in 2D recovering the network geometry.

Localization for Networks with Holes. Network localization has been extensively studied, but the performance of existing algorithms deteriorates rapidly when the network has holes.

In our recent work [30, 31, 48] we make use of the landmark Delaunay complex to successfully tackle a major challenge in anchor-free² network localization: graph rigidity [27]. A graph with fixed edge length is called rigid if it cannot be deformed continuously. It is globally rigid in 2D if it has a unique realization in the plane. However, even globally rigid graph do not admit an efficient algorithm as the realization problem remains NP-hard. In practice, many optimization algorithms run into local minimum and get stuck at a realization far from the ground truth. In contrast with the previous rigidity work on *graphs*, we focus on the global rigidity property of the *combinatorial Delaunay complex*, that has high-order topological structures (such as Delaunay triangles).

We established in [30, 31] the local sampling condition for landmark selection such that the Delaunay complex is globally rigid and admits a unique realization in the plane. This leads to an algorithm to put together the Delaunay triangles in an incremental manner which helps localize the rest of the nodes in the network. See Figure 2 for an example. In [48] we improved the landmark selection algorithm by developing local conditions and an incremental method that automatically selects landmarks according to the local geometric complexity of the network. As this approach does not explicitly require that the network to be embedded is globally rigid, the algorithm works well when nodes are deployed with low density (even uniformly sparse but non-rigid graphs such as a grid-like network with punched holes). Our method is the state-of-the-art solution for anchor-free network localization using only graph connectivity.

Localization using Angle Information. We also investigate how to make use of angle information for network localization. Previous work exclusively depends on measuring inter-distances between wireless nodes, by using received signal strength and Time-of-Arrival for example. On the other hand, one can measure the direction of incoming signals by using antenna array, directional antenna or laser transmissions. We systematically investigate what can be done with such angle measurements. In [5, 7], we show that for a unit disk graph (modeling the sensor network connectivity), the problem of network localization using angle information remains NP-hard, unfortunately. In [4], we show efficient, gossip algorithms that use both noisy angle and noisy distance information to help derive sensor locations.

1.3 Scalable Routing

In a dense network with a regular shape, scalable routing is easy as simple geographical greedy forwarding, i.e., sending the message to the neighbor closest to the destination in terms of Euclidean distance, is able to

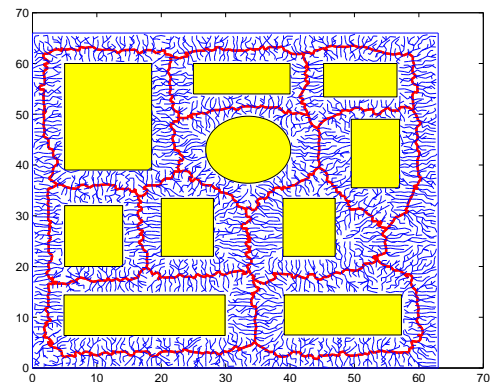


Figure 3. The medial axis and shortest path for-est rooted on the medial axis that define a road map.

²No nodes have known locations and the problem is to discover node positions in a relative coordinate system.

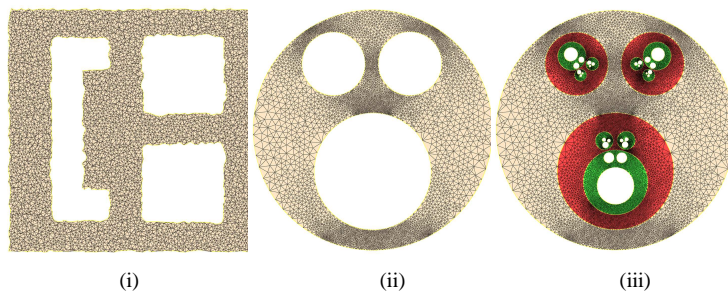


Figure 4. By using Ricci flow the network shown in (i) is mapped to the unit disk with circular holes in (ii). (iii) shows the 3-level circular reflections.

deliver the message with near shortest paths. However, when the network is irregular or has holes, greedy forwarding easily gets stuck at, for example, nodes on a concave hole boundary. Holes in a sensor field mostly reflect the underlying structure of the environment (e.g. obstacles, buildings, etc), and therefore are typically only a few in number and likely to remain stable. In most cases isolated or sporadic link failures will not change these large topological features. In my recent work, I developed a number of schemes that address efficient and scalable routing in large complex networks with holes.

Topology-enabled Routing. Since the topological features are stable, we can afford to explicitly compute an abstraction of these large features (e.g., holes) and store this compact structure at each node to facilitate routing or information processing in such a sensor field. For example, we can carry out proactive protocols at this abstract level, which is stable and compact, such that this high-level combinatorial guidance can be realized with localized and reactive protocols at sensor nodes with link volatility.

The gradient landmark-based routing (GLIDER) [12] uses the combinatorial Delaunay graph to capture the global network topology. The medial axis based routing [6] uses the *medial axis* [8, 2] of the network, which is defined as the set of nodes with at least two closest neighbors on network boundaries (Figure 3). In both cases a global compact structure is built to aid local greedy routing in a virtual road map. And the message is actually routed using local coordinate system in a greedy manner. The storage of the routing information at each node is merely the compact representation of the network global topology (the combinatorial Delaunay graph and the medial axis respectively).

Routing with Bounded Stretch. The routing methods mentioned above guarantee delivery but do not guarantee path quality. Thus we explore other geometric routing approaches that achieve bounded stretch. In [37] we investigated landmark selection rules such that greedily ‘routing to a landmark’ gets to the destination with a constant bounded stretch. In [44] we build size $\tilde{O}(\sqrt{n})$ routing tables for pairs selected with a spatial distribution to get $1 + \epsilon$ stretch routing.

Greedy Routing by Deforming Network Geometry. As we have seen that a network with a ‘bad’ shape can fail simple routing schemes such as greedy routing, we ask whether we can deform the network to a ‘good’ shape such that greedy routing works. In a recent work [40] we produce an embedding for the sensor nodes such that all the holes are mapped to circular disks (including the outer boundary), by using conformal mapping and discrete Ricci flow. An example is shown in Figure 4. Since all the holes are maximally convex, greedy routing with the virtual coordinates will never get stuck. The network deformation is computed once at the network initialization using a distributed gossip-style computation, and used throughout the network lifetime for point-to-point routing.

This technique of regulating the network shape can be carried further by reflecting the network using a Möbius transform with respect to each hole [41]. Doing this recursively will ‘fill up’ the holes. This defines a *covering space* of the sensor network, i.e., a tiling of the space with transformed copies of the sensor networks. See Figure 4. One does not need to pre-compute any of these mapping and is able to generate the reflections on the fly when necessary. This regulated network shape can be used as a generic solution for any information storage scheme that prefers a regular shape for load balancing. It also helps for load balanced greedy routing, as the message encountering nodes on the holes can choose to ‘cut through’ the hole (i.e., reflecting away from the hole in the original network), reducing the high load concentration on the boundaries.

We also applied conformal geometry to explore the path homotopy class³ with greedy routing [50]. This time we first cut the holes open and embed the network as a convex piece in the hyperbolic plane. Again we can find the universal

³Two paths in a Euclidean domain are *homotopy equivalent* if one can smoothly deform one to the other. Paths that are pairwise homotopy equivalent are said to have the same homotopy type.

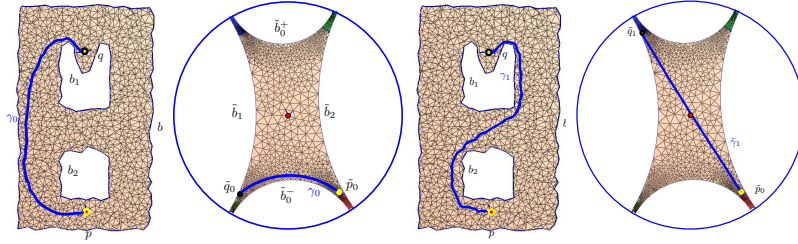


Figure 5. Two paths of different homotopy types are shown in the original network and as the greedy path computed in the hyperbolic universal covering space.

covering space by gluing congruent copies of the network to tile up the Poincaré disk. Now greedy routing to the images at different copies will generate paths of different homotopy types. See Figure 5 for an example.

Load Balanced Routing. A desirable property for routing in a network is load balancing – that is, nodes carry roughly the same amount of traffic and no nodes are over loaded. In wireless sensor networks, overloading a node is particularly dangerous as it may deplete the node battery and disconnect the network immaturely. We developed a greedy routing algorithm that guarantees constant path stretch for nodes deployed inside a strip such that the highest traffic load is at most a constant times the maximum load in the optimal solution [23]. We also show that in general the path stretch and the load balancing ratio are conflicting objectives and provide a tight tradeoff bound [24].

In a recent work [49], we consider the embedding of a network in 3D as the skeleton of a convex polytope. It is well known by Thurston’s Theorem that any 3-connected planar graph admits such an embedding where all faces of the convex polytope are tangent to a unit sphere. In particular, greedy routing using the embedded distances guarantees delivery [38]. We use Ricci flow to *compute* the embedding and further use the Möbius transform to optimize the sensor density for better load balancing.

Resilient Routing. In a wireless network various events may temporarily abrupt a communication link, such as jamming attacks, interference, battery outage, etc. We consider how to quickly recover from a failure event during the delivery of a packet. The first idea we use is to explore tree-based routing. Routing in a tree is easy as there is a unique path for any two nodes. But routing in a tree is not resilient to link failures. In [26] we consider routing using two trees extracted from a graph. When a link fails there is possibly another path in the second tree. We show that with two trees one can get both constant stretch and good path resilience. For the sensor network setting, we use conformal geometry to explore the path space for quick recovery in the case of a link/node failure. In particular, the embedding of a sensor network into a circular domain is not unique and different such embeddings differ by a Möbius transform, which can be exploited for quickly finding an alternative path [28].

1.3.1 Detecting Network Anomalies

We investigate two problems regarding malicious actions in a sensor network, related to routing and the changes of network topologies.

Location Certification. We investigate the problem of ‘securing’ sensor location claims. Since locations are important attribute tags on sensor data, a compromised node claiming an incorrect location may issue bogus data reports to the data sink. In [22], data routed through the network is tagged by participating nodes with belief ratings, collaboratively assessing the probability that the claimed source location is indeed correct. Authorized parties (e.g. final data sinks) now can evaluate a metric of trust for the claimed location of such reports. Belief ratings are derived from a data model of observed past routing activity. The solution is shown to feature a strong ability to detect false location claims and compromised nodes. For example, incorrect claims as small as 2 hops (from the actual location) are detected with over 90% accuracy.

Detecting Wormhole Attacks. In a wormhole attack, the adversary connects two distant points in the network using a direct low-latency link called the wormhole link. The end-points of this link (wormhole nodes) are equipped with radio transceivers that capture wireless transmissions on one end, send them through the wormhole link and replay them at the other end. A long wormhole link can significantly shorten the paths in the network and attract a lot of traffic through the wormhole link. Easy to deploy with no need to break the cryptographic protection, a wormhole attack is often launched in the first stage towards a series of more dangerous attacks on traffic analysis, selectively dropping messages, etc. In our work we developed wormhole detection test based on purely local network connectivity. In the first scheme [33], we

look for “forbidden structures”, i.e., abnormal connectivity features created only by a placement of wormhole attacks. This test is shown to have a high detection rate when the network is reasonably dense. It is a lenient test, i.e., any forbidden structures detected will surely indicate the existence of a wormhole attack, but a network passing such tests may still possibly suffer from wormhole attacks. In a recent work [3], we develop an aggressive test that is guaranteed to detect all wormhole attacks, at the cost of possibly detecting other legal network ‘bridges’.

2 Sensor Data Processing, Discovery and Brokerage

Traditional sensor network applications for scientific data collection gather sensor data routinely at a base station (also called a data sink). Emerging applications in which sensors are ubiquitously deployed in spaces where people live and work envision a smart environment where situations are continuously assessed and responded upon in a real time manner. The sensors serve two purposes: discovering/detecting the events of interest; and forming a supporting infrastructure for distributed resources/users to act on the detected events.

Sensors that detect interesting data are named as *information producers*. Users in the same space, searching for such sensor data are called *information consumers*. Unfortunately, neither information producers nor consumers are aware of each other, unless some *information brokerage* schemes bring them together. We develop various information processing, discovery and brokerage schemes, for different application requirements (query latency, storage, structures in the data, user query patterns), that best serve users’ requests in a resource-constrained network. We look at both discrete, isolated events such as target detections, as well as continuous physical data fields exhibiting strong spatial and temporal correlations.

2.1 Information Brokerage for Discrete Events

Information Discovery and Brokerage. We developed *double rulings* schemes for users to query for discrete events detected by sensors. In particular, an information producer propagates its data or pointers along a *producer curve* and an information consumer propagates its query along a *consumer curve*. We ensure that every consumer curve intersects with every producer curve, which guarantees successful data discovery. In [42] we use stereographic projection to project the sensor network on a sphere, and use the great circle that goes through the producer and a hash location $h(k)$ as the producer curve for data of type k . The nice properties of projective geometry allow a variety of efficient data queries on aggregated data or any subset of data in the network. For example, a consumer can issue a query along a latitude circle assuming the producer as a longitude curve (by rotating the sphere), which guarantees that data can be found in a ‘distance sensitive’ way (i.e., the query cost is proportional to the distance between producer and consumer). When the network has holes, we combined the double rulings scheme with the combinatorial Delaunay graph to incorporate nontrivial network topology [13]. Using conformal geometry to regulate network shape, as in [41], allows the double rulings scheme to be extended to all network topologies.

When the frequency of queries is high and the response time is critical, we developed an information diffusion scheme by using *harmonic functions* to build and maintain at sensor nodes compact information descriptors on data availability [32]. Using more aggressive information storage and processing, we made information discovery easy. A query following the information gradient in a greedy manner will arrive at the data source. The algebraic properties of harmonic functions also allow gradient aggregation, range queries, etc.

Sparse Data Aggregation and Distributed Matching. When there are multiple discrete events that appear in the network, how to efficiently collect such data is a non-trivial problem. The standard routine of building an aggregation tree spanning all sensors is non-optimal as many sensors, without any interesting data, are also involved. We investigated in [15] a sparse aggregation framework such that locally detected events sparsely spread in the network autonomously discover each other in a distributed fashion and form an aggregation structure that can be used to compute cumulants, moments, or other statistical summaries on the events. The communication cost is at most $O(\log n)|\text{MST}|$ where $|\text{MST}|$ is the weight of the minimum spanning tree on the events.

When the events are detections of mobile targets, we developed communication efficient algorithm to maintain a $O(\log n)$ approximate minimum Steiner tree for the mobile agents [52]. The agents can freely move around and each hop move for an agent incurs only amortized $O(\log n)$ message cost for updating the tree.

In [17] we also explore the resource management problem and investigate distributed online and offline matching algorithms that match spontaneous events (e.g., vehicles looking for on-street parking) with sparsely spread resources

(e.g., empty parking spots) in a distributed manner. Our scheme has low communication cost and provides polylogarithmic approximation factor solution to the minimum cost matching. These are the best known distributed algorithms on these problems.

Range Queries and Multi-resolution Data Representation. We propose to process and store data in a multi-resolution format following the principle of *fractional cascading* that states: “a sensor knows a fraction of the information from distant parts of the network, in an exponentially decaying fashion by distance” [14]. Naturally information relevant to each node is decaying with the distance to this node. Important information (e.g., the aggregates of larger regions) are naturally replicated more widely. These partial aggregates are useful for data validation, improve data survivability, and support efficient range queries as the aggregated value inside a geographical region can be answered by combining the pre-computed partial aggregates.

We achieve the fractional cascaded information representation with the following two methods. In [14], the sensor field is recursively partitioned by a standard quad-tree. Aggregates from each quad in the tree are computed and stored at all sensor nodes in the quad. Each node has the values of itself and aggregates of all the quads in which it resides. In [43], we use *spatial gossip* algorithm, such that each node x picks a node y with probability proportional to $1/d^3$, where d is the distance between x and y , and aggregate the data they carry. Using spatial distribution, information travels and mixes with each other very fast. Since most of the gossiping pairs are close, the total communication cost for all nodes to get a multi-resolution representation is low, nearly linear in the network size. This algorithm works for any order and duplicate insensitive synopsis (ODI-synopsis) [36, 9], where the same data can be aggregated multiple times but it is counted only once.

Another application of spatial gossip is to help with in-network coding by using its fast mixing properties and low communication cost. We use spatial gossip to spread linearly coded codewords to facilitate data collection and improve data robustness to node failures [1]. If the network has k original data pieces, any k codewords can be used to recover the sensor network data. The cost for establishing the codewords is only roughly linear in the number of nodes.

In a recent work [39], we construct an alternative, ‘flat’ scheme for efficient target tracking and range queries. We build differential 1-form on a planar graph (extracted from the communication graph, e.g., by using distributed algorithms [16]). Each edge is given a weight such that the integration along the boundary cycle of any region R gives the number of targets within R . When a target moves from one face to an adjacent face, we simply subtract the weight of the target from the weight of the edge just crossed. Thus target motion only trigger local updates to the differential forms.

2.2 Information Brokerage for Continuous Data Fields

Iso-contour Queries and Data-Guided Navigation. When sensors are used to monitor spatial signals, the sensor data naturally has strong correlations. A user can use the sensor data to perform data-driven routing and navigation — e.g., return the sensors with readings within a range (iso-contour query), or find a path from s to t with value below a threshold (low-value routing) — we need to understand and extract the relationship of the iso-contours of the signal field, which can be captured by the *contour tree* [46] on all the critical points, describing how the connected components of the iso-contours merge/split as we increase/decrease the isovalue. In [45] we developed a distributed sweep algorithm to advance iso-contours and identify saddles as two iso-contours merge or split. At the same time, we discover and store the contour tree in a distributed manner, using which gradient routing schemes can successfully solve iso-contour queries or low-value routing.

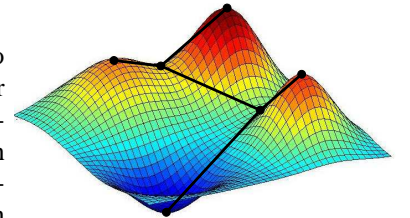


Figure 6. Contour tree.

One limitation of the contour-tree construction is that it does not handle network holes very well. A contour might be broken into multiple disconnected pieces that advance by themselves, causing it difficult to synchronize with each other. The competition resolution mechanism adopted to resolve this might incur a high communication cost by delivering messages between these pieces back and forth. To address these issues, we handle the topological structures of the signal field (in terms of critical points) and the topological structures of the sensor field (in terms of holes) simultaneously. We apply Morse theory [35] in sensor network setting and propose a communication-efficient distributed algorithm to decompose a sensor network to cells and constructing a coordinate system that allows for easy navigation in the domain with respect to the sensor signal [55].

Queries for Mobility Data. We investigated mechanisms to efficiently query for motion trajectories, which have naturally one-dimensional continuity [51]. We let sensor nodes store time-stamped local target detection and when

activated by a target, disseminates its knowledge of past seen targets to neighboring nodes. Thus a target traversing in a sensor field will activate the sensor nodes on its trajectory. And nodes on such a trajectory also exchange their knowledge of what has been detected in the network. This passive, minimal storage scheme uses little beyond what is necessary for a standard target detection system. A query for a target is randomized, issued as following straight lines of random directions searching for nodes with information regarding the target. Intuitively, the longer a target travels in the field and the ‘older’ it is (i.e., more targets appear after it), the more likely the target can be discovered.

3 Data Structures, Computational Geometry, and Algorithms

I work on data structures and geometric algorithms that have applications in peer-to-peer systems and social networks. Below are a couple of recent results and on-going projects in these directions.

Emergence of Distributed Spanners. A spanner graph on a set of points in \mathbb{R}^d provides shortest paths between any pair of points with lengths at most a constant factor of their Euclidean distance. A spanner with a sparse set of edges is thus a good candidate for network backbones, as desired in many practical scenarios such as the transportation network and peer-to-peer network overlays. In [25] we investigate new models and aim to interpret why good spanners ‘emerge’ in reality, when they are clearly built in pieces by agents with their own interests and the construction is not coordinated. Our main result is to show that the following algorithm generates a $(1 + \varepsilon)$ -spanner with a linear number of edges, constant average degree, and the total edge length a small logarithmic factor of the cost of the minimum spanning tree. In our algorithm, the points build edges at an *arbitrary* order. When a point p checks on whether the edge to a point q should be built, it will build this edge only if there is no existing edge $p'q'$ with p' and q' at distances no more than $\frac{1}{4(1+\varepsilon)} \cdot |p'q'|$ from p, q respectively. Eventually when all points have finished checking edges to all other points, the resulted collection of edges forms a sparse spanner as desired. This new spanner construction algorithm has applications in the construction of and local routing on nice network topologies for peer-to-peer systems, when peers join and leave the network and have only limited information about the rest of the network.

Navigation in Social Networks. Milgram’s experiment [34] is the earliest to verify ‘small-world phenomenon’ (aka ‘six degree of separation’) that there exists a short path between almost any pair of individuals in the world. In addition, Milgram’s experiments also show that the short path can be found by using purely local information, although forwarding decision making was not systematically recorded. This has motivated theoretical study on social network models that allow ‘local navigation’, the most prominent one being Kleinberg’s small world model [29]. However, given a real network, it is still unclear how to navigate in it. We take the approach of finding an embedding of the network in some space such that greedy routing using the distance in this space can guide navigation. This space could be a transformed space of attributes, named the *Blau space* in sociology. We started with a naive method of simply using the off-the-shelf embedding techniques (such as multi-dimensional scaling) and surprisingly most of the messages can be delivered with only 6 hops, for a number of real networks we have tested. We are working on a better understanding of the graph topology as well as finding better embedding methods.

IP Geolocation for Internet Hosts. IP Geolocation automatically estimates the coarse grained geographic location of an arbitrary computer on the Internet, using end-to-end delay measurements to landmarks with known locations. Existing methods use either fast but less accurate solutions based on simple triangulation methods, or slow but more accurate solutions using computationally expensive optimization routines. We work on efficient algorithms that run in real time and have accuracy that match or improve the best yet time consuming algorithm so far. Our approach is to carry out detailed experimental study and error analysis using real data sets. The most accurate algorithm uses semi-definite programming and throws in all possible constraints. Having too many constraints is not necessarily a beneficial thing, in particular, when the constraints might be self-conflicting with each other. We use rigidity-based constraint pruning to eliminate potentially conflicting constraints. In our on-going work, we demonstrate the performance improvement by using rigidity based constraint selection and also the benefit of including addition passive landmarks.

Clustering Lines. Clustering is a fundamental data analysis tool. Numerous clustering schemes have been developed for points in Euclidean spaces but not much work is done on clustering lines. I worked on this problem and developed near linear running time algorithms for lines in high dimensional spaces [20, 21, 19]. In particular, for n lines L in \mathbb{R}^d , we look for a ball with minimum radius that intersects with all lines. We show first that a Helly-type theorem holds, i.e., if any three lines can be intersected by a ball of radius r , then all the lines can be intersected by a ball of radius $2r$. We also find a coreset of size $O(1/\varepsilon)$ such that the minimum radius ball intersecting the coreset is at least $1 - \varepsilon$

that of L . This leads to an algorithm of near linear running time (i.e., $O(nd)$) to find a 1-center of a set of lines with approximation ratio $1 + \varepsilon$. In [19], we propose $2 + \varepsilon$ approximation algorithms for 2-center and 3-center problem of lines in \mathbb{R}^d with $O(nd)$ running time. To do this, we develop a new Helly theorem for crosses: if any n crosses have no common intersection, then there is a subset of $O(1/\varepsilon)$ crosses, each contracted by a factor of $1 - \varepsilon$, that have no common intersection.

References

- [1] M. Albano and J. Gao. In-network coding for resilient sensor data storage and efficient data mule collection. In *Proc. of the 6th International Workshop on Algorithms for Sensor Systems, Wireless Ad Hoc Networks and Autonomous Mobile Entities (ALGOSENSOR'10)*, pages 105–117, July 2010.
- [2] D. Attali, J.-D. Boissonnat, and H. Edelsbrunner. Stability and computation of the medial axis — a state-of-the-art report. In *Mathematical Foundations of Scientific Visualization, Computer Graphics, and Massive Data Exploration*. Springer-Verlag, 2004.
- [3] X. Ban, R. Sarkar, and J. Gao. Local connectivity tests to identify wormholes in wireless networks. manuscript, under submission, 2010.
- [4] A. Basu, J. Gao, J. S. B. Mitchell, and G. Sabhnani. Distributed localization using noisy distance and angle information. In *Proc. of the Seventh ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'06)*, pages 262–273, May 2006.
- [5] J. Bruck, J. Gao, and A. Jiang. Localization and routing in sensor networks by local angle information. In *Proc. of the Sixth ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc'05)*, pages 181–192, May 2005.
- [6] J. Bruck, J. Gao, and A. Jiang. MAP: Medial axis based geometric routing in sensor networks. *Wireless Networks*, 13(6):835–853, 2007.
- [7] J. Bruck, J. Gao, and A. Jiang. Localization and routing in sensor networks by local angle information. volume 5, February 2009.
- [8] H. I. Choi, S. W. Choi, and H. P. Moon. Mathematical theory of medial axis transform. *Pacific Journal of Mathematics*, 181(1):57–88, 1997.
- [9] J. Considine, F. Li, G. Kollios, and J. Byers. Approximate aggregation techniques for sensor databases. In *ICDE '04: Proceedings of the 20th International Conference on Data Engineering*, pages 449–460, 2004.
- [10] Q. Fang, J. Gao, and L. Guibas. Locating and bypassing routing holes in sensor networks. In *The 23rd Conference of the IEEE Communications Society (INFOCOM)*, volume 23, pages 2458–2468, March 2004.
- [11] Q. Fang, J. Gao, and L. Guibas. Locating and bypassing routing holes in sensor networks. In *Mobile Networks and Applications*, volume 11, pages 187–200, 2006.
- [12] Q. Fang, J. Gao, L. Guibas, V. de Silva, and L. Zhang. GLIDER: Gradient landmark-based distributed routing for sensor networks. In *Proc. of the 24th Conference of the IEEE Communication Society (INFOCOM)*, volume 1, pages 339–350, March 2005.
- [13] Q. Fang, J. Gao, and L. J. Guibas. Landmark-based information storage and retrieval in sensor networks. In *The 25th Conference of the IEEE Communication Society (INFOCOM'06)*, volume 1, pages 339–350, April 2006.
- [14] J. Gao, L. Guibas, J. Hershberger, and L. Zhang. Fractionally cascaded information in a sensor network. In *Proc. of the 3rd International Symposium on Information Processing in Sensor Networks (IPSN'04)*, pages 311–319, April 2004.
- [15] J. Gao, L. J. Guibas, J. Hershberger, and N. Milosavljevic. Sparse data aggregation in sensor networks. In *Proc. of the International Conference on Information Processing in Sensor Networks (IPSN'07)*, pages 430–439, April 2007.
- [16] J. Gao, L. J. Guibas, J. Hershberger, L. Zhang, and A. Zhu. Geometric spanners for routing in mobile networks. *IEEE Journal on Selected Areas in Communications Special issue on Wireless Ad Hoc Networks*, 23(1):174–185, 2005.
- [17] J. Gao, L. J. Guibas, N. Milosavljevic, and D. Zhou. Distributed resource management and matching in sensor networks. In *Proc. of the 8th International Symposium on Information Processing in Sensor Networks (IPSN'09)*, pages 121–132, April 2009.
- [18] J. Gao, L. J. Guibas, S. Y. Oudot, and Y. Wang. Geodesic delaunay triangulation and witness complex in the plane. In *Proc. of ACM-SIAM Symposium on Discrete Algorithms (SODA'08)*, pages 571–580, January 2008.
- [19] J. Gao, M. Langberg, and L. Schulman. Clustering lines in high dimensional space: Classification of incomplete data. *ACM Transactions on Algorithms*, 2009.
- [20] J. Gao, M. Langberg, and L. J. Schulman. Analysis of incomplete data and an intrinsic-dimension helly theorem. In *ACM-SIAM Symposium on Discrete Algorithms (SODA'06)*, pages 464–473, January 2006.

- [21] J. Gao, M. Langberg, and L. J. Schulman. Analysis of incomplete data and an intrinsic-dimension helly theorem. *Discrete and Computational Geometry*, 40(4):537–560, 2008.
- [22] J. Gao, R. Sion, and S. Lederer. Collaborative location certification for sensor networks. *ACM Transactions on Sensor Networks*, 6(4), 2010.
- [23] J. Gao and L. Zhang. Load balanced short path routing in wireless networks. *IEEE Transactions on Parallel and Distributed Systems, Special Issue on Localized Communications*, 17(4):377–388, April 2006.
- [24] J. Gao and L. Zhang. Tradeoffs between stretch factor and load balancing ratio in routing on growth restricted graphs. *IEEE Transactions on Parallel and Distributed Computing*, 20(2):171–179, February 2009.
- [25] J. Gao and D. Zhou. The emergence of sparse spanners and greedy well-separated pair decomposition. In *Proc. of the the 12th Scandinavian Symposium and Workshops on Algorithm Theory (SWAT'10)*, pages 50–61, June 2010.
- [26] J. Gao and D. Zhou. Resilient and low stretch routing through embedding into tree metrics. In *Proc. of the 12th Algorithms and Data Structures Symposium (WADS'11)*, pages 438–450, August 2011.
- [27] J. E. Graver, B. Servatius, and H. Servatius. *Combinatorial Rigidity*. Graduate Studies in Math., AMS, 1993.
- [28] R. Jiang, X. Ban, M. Goswami, W. Z. adn Jie Gao, and X. D. Gu. Exploration of path space using sensor network geometry. In *Proc. of the 10th International Symposium on Information Processing in Sensor Networks (IPSN'11)*, pages 49–60, April 2011.
- [29] J. Kleinberg. The small-world phenomenon: an algorithm perspective. In *STOC '00: Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 163–170, 2000.
- [30] S. Lederer, Y. Wang, and J. Gao. Connectivity-based localization of large scale sensor networks with complex shape. In *Proc. of the 27th Annual IEEE Conference on Computer Communications (INFOCOM'08)*, pages 789–797, May 2008.
- [31] S. Lederer, Y. Wang, and J. Gao. Connectivity-based localization of large scale sensor networks with complex shape. *ACM Transactions on Sensor Networks*, 5(4), November 2009.
- [32] H. Lin, M. Lu, N. Milosavljević, J. Gao, and L. Guibas. Composable information gradients in wireless sensor networks. In *Proc. of the International Conference on Information Processing in Sensor Networks (IPSN'08)*, pages 121–132, April 2008.
- [33] R. Maheshwari, J. Gao, and S. R. Das. Detecting wormhole attacks in wireless networks using connectivity information. In *Proceedings of the 26th Conference of the IEEE Communications Society (INFOCOM'07)*, pages 107–115, May 2007.
- [34] S. Milgram. The small world problem. *Psychology Today*, (1), 1967.
- [35] J. W. Milnor. *Morse Theory*. Princeton University Press, Princeton NJ, 1963.
- [36] S. Nath, P. B. Gibbons, S. Seshan, and Z. R. Anderson. Synopsis diffusion for robust aggregation in sensor networks. In *SenSys '04: Proceedings of the 2nd international conference on Embedded networked sensor systems*, pages 250–262, 2004.
- [37] A. Nguyen, N. Milosavljevic, Q. Fang, J. Gao, and L. J. Guibas. Landmark selection and greedy landmark-descent routing for sensor networks. In *Proceedings of the 26th Conference of the IEEE Communications Society (INFOCOM'07)*, pages 661–669, May 2007.
- [38] C. H. Papadimitriou and D. Ratajczak. On a conjecture related to geometric routing. *Theor. Comput. Sci.*, 344(1):3–14, 2005.
- [39] R. Sarkar and J. Gao. Differential forms for target tracking and aggregate queries in distributed networks. In *Proc. of the 16th Annual International Conference on Mobile Computing and Networking (MobiCom'10)*, pages 377–388, September 2010.
- [40] R. Sarkar, X. Yin, J. Gao, F. Luo, and X. D. Gu. Greedy routing with guaranteed delivery using ricci flows. In *Proc. of the 8th International Symposium on Information Processing in Sensor Networks (IPSN'09)*, pages 97–108, April 2009.
- [41] R. Sarkar, W. Zeng, J. Gao, and X. D. Gu. Covering space for in-network sensor data storage. In *Proc. of the 9th International Symposium on Information Processing in Sensor Networks (IPSN'10)*, pages 232–243, April 2010.
- [42] R. Sarkar, X. Zhu, and J. Gao. Double rulings for information brokerage in sensor networks. In *Proc. of the ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, pages 286–297, September 2006.
- [43] R. Sarkar, X. Zhu, and J. Gao. Hierarchical spatial gossip for multi-resolution representations in sensor networks. In *Proc. of the International Conference on Information Processing in Sensor Networks (IPSN'07)*, pages 420–429, April 2007.
- [44] R. Sarkar, X. Zhu, and J. Gao. Spatial distribution in routing table design for sensor networks. In *Proc. of the 28th Annual IEEE Conference on Computer Communications (INFOCOM'09), mini-conference*, pages 2766–2770, April 2009.
- [45] R. Sarkar, X. Zhu, J. Gao, L. J. Guibas, and J. S. B. Mitchell. Iso-contour queries and gradient descent with guaranteed delivery in sensor networks. In *Proc. of the 27th Annual IEEE Conference on Computer Communications (INFOCOM'08)*, pages 1175–1183, May 2008.
- [46] M. van Kreveld, R. van Oostrum, C. Bajaj, V. Pascucci, and D. Schikore. Contour trees and small seed sets for isosurface traversal. In *Proc. 13th Annu. ACM Sympos. Comput. Geom.*, pages 212–220, 1997.

- [47] Y. Wang, J. Gao, and J. S. B. Mitchell. Boundary recognition in sensor networks by topological methods. In *Proc. of the ACM/IEEE International Conference on Mobile Computing and Networking (MobiCom)*, pages 122–133, September 2006.
- [48] Y. Wang, S. Lederer, and J. Gao. Connectivity-based sensor network localization with incremental delaunay refinement method. In *Proc. of the 28th Annual IEEE Conference on Computer Communications (INFOCOM'09)*, pages 2401–2409, April 2009.
- [49] X. Yu, X. Ban, R. Sarkar, W. Zeng, X. D. Gu, and J. Gao. Spherical representation and polyhedron routing for load balancing in wireless sensor networks. In *Proc. of 30th Annual IEEE Conference on Computer Communications (INFOCOM'11)*, April 2011.
- [50] W. Zeng, R. Sarkar, F. Luo, X. D. Gu, and J. Gao. Resilient routing for sensor networks using hyperbolic embedding of universal covering space. In *Proc. of the 29th Annual IEEE Conference on Computer Communications (INFOCOM'10)*, pages 1694–1702, March 2010.
- [51] D. Zhou and J. Gao. Opportunistic processing and query of motion trajectories in wireless sensor networks. In *Proceedings of the 28th Annual IEEE Conference on Computer Communications (INFOCOM'09)*, pages 1197–1205, April 2009.
- [52] D. Zhou and J. Gao. Maintaining approximate minimum steiner tree and k-center for mobile agents in a sensor network. In *Proc. of the 29th Annual IEEE Conference on Computer Communications (INFOCOM'10)*, pages 511–515, March 2010.
- [53] X. Zhu, R. Sarkar, and J. Gao. Shape segmentation and applications in sensor networks. In *Proceedings of the 26th Conference of the IEEE Communications Society (INFOCOM'07)*, pages 1838–1846, May 2007.
- [54] X. Zhu, R. Sarkar, and J. Gao. Segmenting a sensor field: Algorithms and applications in network design. *ACM Transactions on Sensor Networks*, 2009.
- [55] X. Zhu, R. Sarkar, and J. Gao. Topological data processing for distributed sensor networks with morse-smale decomposition. In *Proc. of the 28th Annual IEEE Conference on Computer Communications (INFOCOM'09), mini-conference*, pages 2911–2915, April 2009.
- [56] X. Zhu, R. Sarkar, J. Gao, and J. S. B. Mitchell. Light-weight contour tracking in wireless sensor networks. In *Proceedings of the 27th Annual IEEE Conference on Computer Communications (INFOCOM'08)*, pages 960–967, May 2008.