Spatial Network Activity Summarization for Disaster Response
Dev Oliver, oliver@cs.umn.edu

Research overview and Potential Significance: My research explores novel computational techniques for enhancing situational awareness for disaster response, which is a national imperative [12]. Beyond the unquantifiable costs of injury and loss of life from disasters, economic damages from Hurricane Sandy alone in the United States exceeded $60 billion [13]. Disaster response during emergency management includes action taken immediately after a disastrous event with the aim of saving life, protecting property, and dealing with immediate disruption, damage, or other effects caused by the disaster (e.g., hurricanes, earthquakes, terrorist attacks, etc.) [1]. The 2010 Haiti earthquake saw numerous requests for relief including requests for food, water, and medicine. Emergency managers are tasked with making resource allocation decisions based on the locations of the affected population. Figure 1(a) shows the locations of a subset of various relief requests in Port-au-Prince, Haiti from January to March 2010. An important aspect of disaster response is to understand where people are in a given area (e.g., situational awareness). In other words it is important to be able to summarize people’s location on a map and place them into different groups depending on their location. In order to make informed, critical resource allocation decisions given such location information, emergency managers need tools to assist them with situational awareness, which I conceptualize in my work as the spatial network activity summarization problem (SNAS).

Figure 1: (a) Sample emergency requests (dots) occurring in Port-au-Prince, Haiti after the 2010 Earthquake [2] (b) Three summary routes produced by KMR [3].

Given a spatial network such as a road network, a collection of activities (e.g., disaster relief reports), and $k$ (a desired number of routes), spatial network activity summarization (SNAS) finds $k$ routes to group activities or disaster relief incidents [3]. Intuitively, SNAS seeks to group activities by routes so that emergency managers can know which $k$ routes would best serve the affected population in terms of number of activities covered. An activity is an object of interest, associated with only one edge of the spatial network (or one street in the road network). For example, in disaster response-related applications, an activity might be the location of a request for relief supplies. The objective of SNAS is to maximize coverage of the activities covered by each route and the constraint is that each route is a shortest route between its end-points. Shortest routes are necessary because immediate response is needed after a disastrous event [1] and shortest routes allow for quick delivery of relief supplies.

However, finding a set of $k$ shortest routes that maximizes the number of activities on selected routes is computationally challenging. This is due to the fact that if $k$ subsets
are selected from all shortest routes in a spatial network, there are a large number of possibilities to consider due to the overlapping of shortest routes. For disjoint routes, the problem would be relatively less computationally challenging. However, due to overlapping routes, the general problem of SNAS is very computationally challenging.

**Context of Existing Research:** My investigation of previous work has revealed that the summaries provided to emergency managers have been either geometry or network-based. In geometry-based summarization, spatial points are grouped using Euclidean distance (as the crow flies). Such techniques focus on the discovery of the geometry (e.g., circle, ellipse) of high density regions and include K-Means [6], K-medoid [7], and P-median [8]. These methods do not consider the underlying spatial network; they group spatial objects that are close in terms of Euclidean distance but not close in terms of network distance. Thus, they may fail to group activities that are on the same street. In network-based summarization, spatial objects are grouped using network (e.g., road) distance. Existing network-based summarization methods such as Mean Streets [9], Maximal Subgraph Finding (MSGF) [10], and Clumping [11] group activities over multiple routes, a single route/subgraph, or no routes at all. Mean Streets [9] finds anomalous streets or routes with unusually high activity levels. It is not designed to summarize activities over \( k \) routes because the number of high emergency-request streets returned is always relatively small. MSGF [10] identifies the maximal subgraph (e.g., a single route, \( k = 1 \)) under the constraint of a user specified length and cannot summarize activities when \( k > 1 \). The Network-Based Variable-Distance Clumping Method (NT-VCM) [11] groups activities that are within a certain shortest route distance of each other on the network but is not designed to summarize activities using \( k \) routes because the grouping of activities is based on the given distance threshold.

**Contributions:** In developing a solution to improve situational awareness for disaster response based on the data from Haiti, I developed an algorithm called K-Main Routes (KMR) that finds a set of \( k \) routes to summarize activities (Figure 1(b)). KMR aims to maximize the number of activities covered on each \( k \) route. I added several performance-tuning techniques to improve KMR’s performance without affecting solution quality. I evaluated KMR on both synthetic and real-world data and showed that KMR with performance-tuning decisions yielded substantial computational savings without reducing summary route coverage, meaning the tool performed faster and better without losing accuracy. This software was successfully transferred to the National Geospatial-Intelligence Agency (NGA) after being implemented in the Oracle database and is now available for use by NGA analysts. The results have been published in ICDM Workshops 2010 and have also been accepted with major revisions in the prestigious IEEE Transactions on Knowledge and Data Engineering (TKDE) journal [3].

However, several outstanding issues remain while attempting to further improve situational awareness for emergency managers, which will shape my research going forward. For example, I plan to extend our approach to account for different types of activities. For example, a building collapsing on someone might require more immediate attention than other types of activities and should be summarized accordingly. Incorporating time into summarizing activities will also be investigated to account for activities occurring at different times. I have done some initial work [4] on summarizing spatio-temporal data but new challenges remain due to the increased data size and complex inter-relationships when time is added. Additionally, determining KMR’s
usefulness to domain professionals will also be explored. I plan to explore other types of data that may not be associated with a single point in a street (e.g., aggregated crime data at the zip code level, trajectory data [5], etc.). Finally, I plan to investigate a distance-based rather than coverage-based objective function where the distance between activities and summary routes is minimized to reduce how far the affected population has to walk.

**Professional Goals:** I attended graduate school because I wanted to work on exciting challenges while being in an environment of learning and discovery. After graduate school, I would like to pursue a research career where I can work on real-world problems and mentor students. My goal is to serve society and work with organizations active in disaster management. A doctoral dissertation fellowship will afford me the opportunity to focus intently on my research, graduate on time, and make a significant real-world impact.

**References**


