Nodes, Paths, and Edges: Using Mental Maps to Augment Crime Data Analysis in Urban Spaces

Alex Godwin and John T. Stasko

School of Interactive Computing
Georgia Institute of Technology
Atlanta, GA

Abstract
Citizen participation can provide valuable insight into data that is not captured by official sources. In this paper, we propose a technique for using mental maps consisting of three fundamental elements: nodes, paths, and edges. These elements can be used to augment crime data analysis in urban spaces by incorporating the values and knowledge of citizens. We apply this technique to an analysis of property crime in three US cities: Baltimore, Atlanta, and Chicago. Subsequently, we find these cities have neighborhoods where the crime could be substantially higher—or perceived by citizens as higher—than is accounted for in the official public crime data. This analysis can be a vital first step for identifying hidden hotspots or better understanding public perceptions of high crime.

1. Introduction

Citizen participation in local government is vital to maintaining a functioning city, as paternalistic city programs that do not incorporate the values and knowledge of citizens are at risk for being misguided at best and manipulative at worst [Arn69]. For example, in an analysis of crime data and violence prevention programs in Cardiff, researchers determined that fewer than one-third of violent incidents requiring emergency treatment in the UK and Scandinavia appear in official police records [FSBS11]. The absence of these incidents from the official record points to a striking difference of perspective between the police and the inhabitants of the city, notably, victims of violence. Clearly, citizen participation is necessary to provide insight into the nature of urban crime data.

It is challenging to take something as qualitative and ephemeral as public perception and translate it into a tangible form. Local knowledge can be acquired through a variety of approaches based on Public Participation in Geographic Information Systems (PPGIS) [Sie06, WHC02] and Bottom-Up GIS (BUGIS) [Tal00], in which researchers typically conduct a series of workshops to document the values, ideas, and opinions of citizens. For example, Dennis interviewed local youths and asked them to create sketches of their perception of the qualitative aspects of the environment and how "bad" intersections affected their planned paths through the neighborhood [Den06]. If these maps were created based solely on authoritative data, it is unlikely they would capture the knowledge available to the youth in the workshop. Combining these sketches—or mental maps—with what data the police do have, however, has the potential to greatly increase the accuracy of that data with respect to unknown spatial distributions and the values of citizens living in those neighborhoods.

In this paper, we introduce a technique utilizing three elements of the "image of the city" identified by Lynch: nodes (areas of heightened activity and interaction), paths (channels that people take to move around the city), and edges (barriers that divide regions) [Lyn60]. These elements have been used extensively to analyze the structure of urban spaces and crime patterns that are governed by human movement [BB82, Arm13]. Brantingham and Brantingham have documented the ways in which Lynch's elements—notably nodes, paths, and edges—contribute to trends in criminal behavior [BB93]. For example, the highest concentrations of crime typically occur within the vicinity of nodes or the paths between them. Analysis of the path network can reveal the route that criminals are taking, potential sites for crimes that have not been recorded, or paths that contribute to crimes of opportunity. Finally, edges experience high crime rates as a function of both opportunity and criminal location preference. We also describe our implementation of this technique in a sketch-based system for capturing citizens' mental maps and highlighting the disparities between them and raw spatial data. Finally, we provide preliminary findings from the application of our technique to property crime data in three US cities: Atlanta, Baltimore, and Chicago.

2. Related Work

As a support mechanism for visualization, imageability is frequently used to determine the characteristics of a scene that will allow a user to navigate through a 2D or 3D environment and bet-

ter understand the data. An early attempt by Ingram and Benford set the stage for many later efforts in using imageability elements (e.g., nodes, paths, edges) to improve the legibility of a data visualization [IB95, IB96]. Many of these efforts seek to automatically detect notable elements, though, rather than allowing the user to explicitly define a mental map. For example, Chang et al. combined a building aggregation algorithm and a demographic data exploration panel to analyze the differences in census data between neighborhoods [CWK*07]. Similarly, van Wijk et al. support the creation of wayfinding maps through the simplified representations of urban networks given a focal origin node [vDHO16]. Glander and Döllner also use focal points and building aggregation, but incorporate a balanced tree of landmark elements to help navigate a 3D representation of the city at varying levels of abstraction [GD09].

Research to improve public participation in GIS (PPGIS) has found that creating rapid sketch-based representations improve communication during participatory planning workshops, and that many current GIS tools fail to facilitate this capability [AK02]. Moreover, many GIS applications remain inaccessible to a wider audience [HT03, Sie06]. Al-Kodmany avoided this dilemma by pairing an expert-operated GIS with an artist that took requests from participants during planning sessions in the Pilsen community [AK99]. Many systems allow for sketching in a GIS context, though more as a natural interaction technique for exploring data [TSH*14] or creating a query of existing spatial features [Ege97]. Rarer is the capability to express elements within a spatial context that are of importance to the user, such as the components of a mental map. This type of expression could be acquired implicitly, by tracking the areas of the city and spatial data items that a user inspects. This approach has been used quite successfully in other contexts, notably text document analysis, by generating a semantic model from user interaction at varying levels of detail [EFN12]. This type of interaction, however, often divides the user from understanding the internal mechanisms that are being used to generate a representation of the data. Instead, following Green's et al.'s guidance on process initiators [GRF09], our technique allows the user to explicitly capture her knowledge of the spatial environment by directly interacting with the interface.

3. Elements of the City

Though competent mental maps can be drawn from any of the five elements [WUS13], we have chosen to focus on nodes, paths, and edges for our technique due to previous research establishing a connection between those elements and criminal behavior. Brantingham and Brantingham demonstrated that these elements are necessary, if not sufficient, components in a framework for the analysis of the spatial distribution of crime [BB93]. Spatial data visualization represented along paths has been previously explored by Xie and Yan for traffic accidents [XY08, XY13], Wong et al. for power grids [WSM*09], and Kim et al. for crime [KMM*13]. Wood et al. have also demonstrated abstract hierarchical representation that depict the connectivity between regions on a map, though this is primarily for trajectory data rather than joining of spatial data to paths [WDS10, WSD11]. Nodes, or areas of high activity, can most directly be compared to the hotspots derived during more spatial analysis of point-based data. However, unlike Eu-

clidean [MRH*10] or grid-based [RMK12] approaches, our technique allows users to actively specify nodes rather than try to passively detect them. Our technique is novel in its incorporation of edges, which are not present in other approaches.

3.1. Paths

We utilize paths as the primary mental map element in our technique. Building from previous work in network-based Kernel Density Estimation (KDE), we utilize a path network consisting of roads [XY08]. This approach differs from standard planar KDE in that the distance between points on the map are not measured in Euclidean space, but based upon network distance. In our approach, we obtain road-level data from OpenStreetMap (OSM) and construct a coarse graph between intersections. As an open data source that can be modified by the public, OSM is ideal for constructing a backdrop for analysis of community-oriented data. This allows us to construct path networks that match real roads, rather than interpreting sketched paths on the map as passing through or above impassable areas (e.g., buildings).

We begin by dividing the road topology into *lixels*, or linear pixels [XY08]. Lixels consist of linear road segments of equal length. In terms of KDE, lixels are similar to selection of a pixel resolution for the planar space. The selection of lixel length in network KDE is, as with the selection of pixel resolution in planar KDE, an important consideration affecting the variation details of spatial patterns. Once road segments have been divided into lixels, we then assign each crime event in the set to the nearest lixel. Each lixel with one or more assigned data items is a *source lixel*, and serves as the point of origin for the network KDE within the network topology. We follow the approach of Kim et al. [KMM*13] rather than Xie and Yan, and assign scores to each lixel based on a weighted kernel function (Equation 1) and a minimum detection bandwidth rather than using a count of nearby crimes. This allows us to take into account the distance from the event to the lixel, because unlike the traffic accidents analyzed by Xie and Yan, crimes do not always occur directly on roads. For each of these events $e_1, e_2, \ldots, e_n$, we determine the minimum distance d_i from that event e_i to any part of the lixel. For this approach, the choice of kernel function does not affect the results as much as the choice of bandwidth h_l, which should be chosen carefully based upon the domain. We utilize the Epanechnikov kernel, depicted in Equation 2.

$$f(x) = \frac{1}{nh_l} \sum_{i=1}^{n} K(\frac{d_i}{h_l}) \qquad (1)$$

$$K(u) = \begin{cases} \frac{3}{4}(1-u^2), & \text{if } ||u|| \leq 1 \\ 0, & \text{otherwise} \end{cases} \qquad (2)$$

Once the spatial data items have been assigned to the correct lixel, we iterate through the source lixels to determine each of the other lixels that are reachable from their position given the current h_l using the Bellman-Ford algorithm. Each other lixel that a source lixel can "reach" receives a score adjustment relative to the network distance between them. The resulting score for any lixel rep-
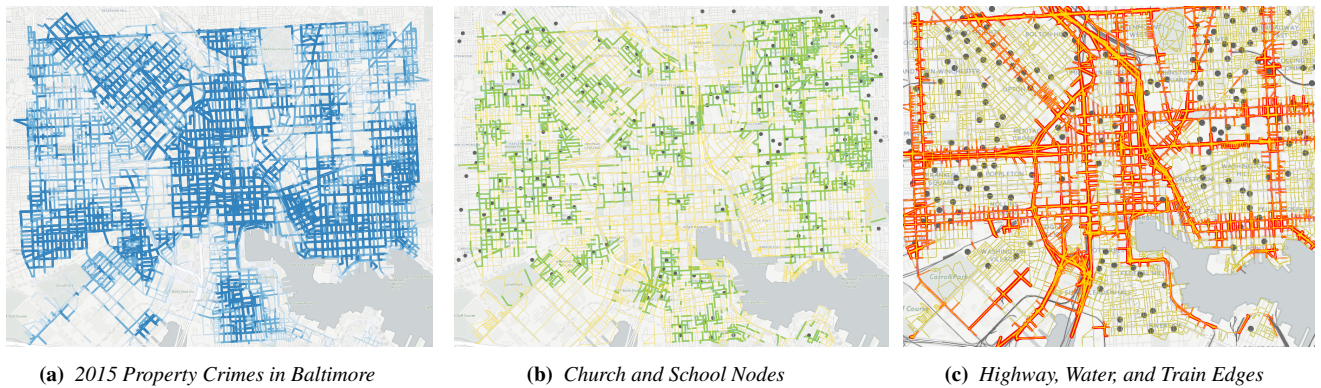
(a) *2015 Property Crimes in Baltimore* **(b)** *Church and School Nodes* **(c)** *Highway, Water, and Train Edges*

Figure 1: *Path KDE is rendered as line thickness and saturation in Figure 1a, with thicker and darker lines indicating a higher network score. In Figure 1b, the path network is shown in yellow, with green highlights indicating the lixels that have been amplified due to node proximity (church and school nodes are black circles). In Figure 1c, the red highlights indicate lixels that have been dampened due to edge proximity (highway, water, and train edges are black lines mostly occluded by the path highlights).*

resents the network KDE score aggregated from all of the source lixels that can reach it (Figure 1a). This score is aggregated using Equation 1, though we substitute a bisquare equation for the kernel (Equation 3). This equation utilizes a different bandwidth, h_g, than the preceding detection bandwidth, as we are calculating the density of the lixel with respect to the nearby lixels in the network topology rather than aggregating nearby crimes to their nearest source lixel.

$$K_i(u, h_g) = \begin{cases} [1 - \frac{d_i^2}{h_g^2}]^2, & \text{if } d_i < h_g \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

3.2. Nodes

To incorporate node elements into a mental map of the city, we update the existing path model to amplify the *reach* of a source lixel. This amplification provides an incremental boost to the h_g of all source lixels within close proximity to a node (Figure 1b). The effect of this amplification diminishes as the distance to the node increases, subject to the weighted distance obtained by Equation 3. We default nodes to an activity radius of half the current h_g, though other parameters would yield alternative results. For example, if $h_g = 100m$, then the activity radius would be 50m. A source lixel within this distance of the activity node would receive an improved reach in addition to the default reach. A source lixel that occurs in proximity to many activity nodes is further amplified.

3.3. Edges

In our technique, edge elements modify the lixel topology by artificially inflating lixel length, creating a dampening effect for source lixels that are nearby. When an edge is added to the map, it increases the artificial length of all lixels that are within its activity radius, subject to Equation 3. As with nodes, the default activity radius is half the current h_g. As source lixels spread to reachable nodes, they observe inflated lixels as longer than they actually are.

This reduces the propagation of source lixels across edges and creates a border effect of heightened scores near edges (Figure 1c).

3.4. Sketching Nodes, Paths, and Edges

We have implemented the capability to directly specify mental maps by sketching nodes, paths and edges. Mental maps can be built on top of a street network imported from OSM or created from scratch within a tile-based map viewer. To create a node on a map within the system, a user draws a circle around the area of activity. This indicates to the system that lixels within the area around the node should receive an amplified reach, or localized h_g. To create an edge, the user draws a line. This indicates to the system that it should dampen the reachability, or artificial lixel length, of lixels. Paths care drawn as straight lines, which are fitted to the closest path along available surface roads. Nodes, paths, and edges can also all be deleted using the eraser function. These capabilities are not used for obtaining the mental maps used in this paper, but are necessary for our planned investigations of crime data with participants through community workshops.

4. Mental Maps of Property Crime

To understand the potential effects of mental maps on quantitative spatial data, we analyzed property crime in three US cities using mental maps created from community-sourced OpenStreetMap (OSM) data. While these are not meant to substitute wholly for the actual mental maps sketched by citizens, they are a useful preliminary step for exploring the differences between community-sourced nodes, paths, and edges and the official crime data.

We obtained property crime data for Atlanta, Baltimore, and Chicago, three cities with significant differences in layout and navigability. We utilized road network topology data obtained from OSM to form the paths of the mental map constructed for each city, incorporating all path types accessible on foot or by car. We included a second set of OSM features as edges: major interstate highways, train tracks (excluding subway features), and water features (rivers and streams).
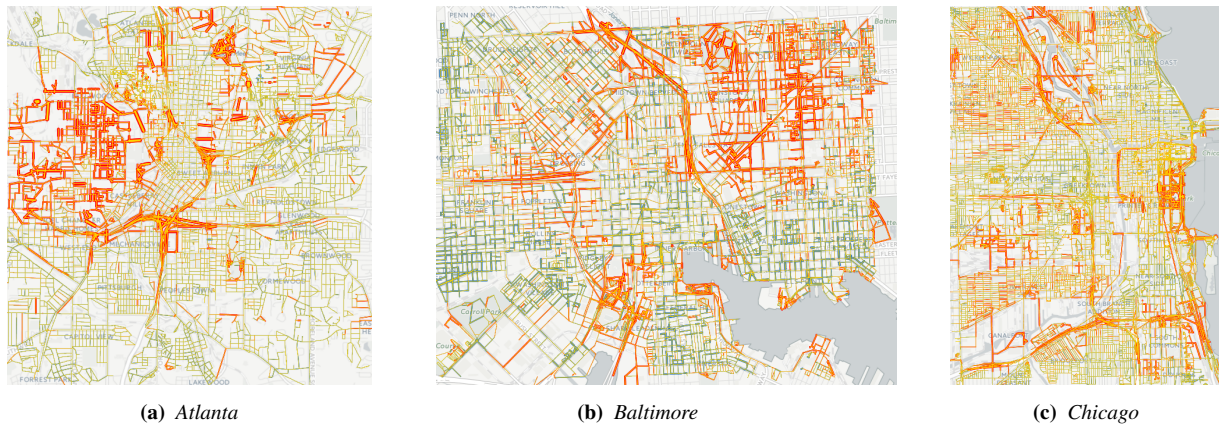
| **(a)** *Atlanta* | **(b)** *Baltimore* | **(c)** *Chicago* |

Figure 2: *The disparities between the mental maps and the original property crime distributions. All paths have a central yellow color to stand out against the disparity shading and background, so paths that are only yellow have no substantial difference. Disparity is calculated as difference in percentile, so red areas increased in value within the relative distribution of the mental map scores compared to their original score, blue values decreased in calculated percentile.*

The constructed mental maps for each city are used to interpret the distribution of property crimes: nonviolent (larceny, auto theft, and nonresidential burglary) and violent (pedestrian and residential robbery). The data sets consist of events that occurred in 2015 and were obtained through the Federal Uniform Crime Reporting (UCR) program. For each mental map, activity nodes are centered around church and school locations, as these represent important components of the community and family. Only the locations that OSM contributors thought were important enough to record and upload are included in the mental maps, and the subsequent mental maps reflect this. To limit the effects that clusters of relevant nodes might have on the model, we apply a hierarchical aggregation scheme to cluster together nodes that occur within h_l, the local bandwidth, of each other. Cities were analyzed with a lixel size of 25m to provide a resolution of approximately four lixels per city block. Local bandwidth was set at $h_l = 100m$ so crimes would be associated with a lixel at the nearest block but no further. Finally, the global bandwidth was set at $h_g = 200m$ to limit the propagation to a maximum distance of approximately two blocks.

Visual representations of the disparities between the original data distributions and the new distributions created using mental maps are shown in Figure 2. Red lines indicate neighborhoods where the distribution is more dense in the mental map, while blue lines indicate neighborhoods where the distribution is more dense in the original data. For the city of Atlanta, the greatest disparity between the mental map and the raw data occurs around West Midtown and the intersection between the major highways in the center of the map near downtown (Figure 2a). For the city of Baltimore, the largest disparity occurs with high mental map distributions in East Baltimore in the neighborhoods of Oliver, Dunbar-Broadway, and Middle East (Figure 2b). Many other neighborhoods, notably downtown and Fells Point, are lower, though Harlem Park to the west and the stadium area to the south also have higher property crime distributions. Finally, for the city of Chicago, most of the disparities in the mental map exist on the edge between the Loop and South Loop neighborhoods (Figure 2c). To the south, neigh-

borhoods along the lake and to the southwest along the highway also have an increase, as well as the northern neighborhoods around Goose Island. For the neighborhoods that are drawn in red, the presence of nodes and edges may indicate that the crime density is higher than what is captured in the official data source, or that inhabitants of those neighborhoods perceive crime as being higher because it occurs close to important nodes in their community. In either case, programs that are designed for reducing the amount of crime in those neighborhoods would be well-advised to include citizens of those neighborhoods in planning sessions.

5. Discussion and Future Work

In this paper, we proposed a novel technique for applying mental maps based on nodes, paths, and edges to spatial data. We described our implementation of this technique in a sketch-based system that will allow citizens to directly record their mental maps. Finally, we provided preliminary findings from the application of our technique to property crime data in three US cities. This technique facilitates the identification of disparities in mental maps and urban spatial data, providing insight into unknown or citizen-perceived hotspots.

There are several promising directions for future work. The most important would be to conduct a user study with citizens in these cities to collect individual mental maps and consolidate them for further analysis against the existing data as a first step for new crime prevention and public safety initiatives. Much like the Cardiff study described previously [FSBS11], additional sources of citizen data could be collected and compared to the distributions generated using mental maps to determine the degree to which they align or differ. Finally, mental maps and spatial crime data can be used together to explore citizens' qualitative perception of the city to better understand where they perceive crime to be higher or neighborhoods to be more dangerous when available data indicates crime is low. Changing the perception of these neighborhoods would be a first step to attracting new businesses and residents, which could be important for sustaining long-term community vitality.

References

[AK99]   AL-KODMANY K.: Using visualization techniques for enhancing public participation in planning and design: process, implementation, and evaluation. *Landscape and urban planning 45*, 1 (1999), 37–45. 2

[AK02]   AL-KODMANY K.: Visualization tools and methods in community planning: from freehand sketches to virtual reality. *Journal of planning Literature 17*, 2 (2002), 189–211. 2

[Arm13]   ARMITAGE R.: *Crime Prevention Through Housing Design: Policy and Practice*. Springer, 2013, ch. 6, pp. 123–143. 1

[Arn69]   ARNSTEIN S. R.: A ladder of citizen participation. *Journal of the American Institute of planners 35*, 4 (1969), 216–224. 1

[BB82]   BRANTINGHAM P. L., BRANTINGHAM P. J.: Mobility, notoriety, and crime: A study of crime patterns in urban nodal points. *Journal of Environmental Systems 11*, 1 (1982), 89–99. 1

[BB93]   BRANTINGHAM P. L., BRANTINGHAM P. J.: Nodes, paths and edges: Considerations on the complexity of crime and the physical environment. *Journal of Environmental Psychology 13*, 1 (1993), 3–28. 1, 2

[CWK*07]   CHANG R., WESSEL G., KOSARA R., SAUDA E., RIBARSKY W.: Legible cities: Focus-dependent multi-resolution visualization of urban relationships. *IEEE transactions on visualization and computer graphics 13*, 6 (2007), 1169–1175. 2

[Den06]   DENNIS S. F.: Prospects for qualitative gis at the intersection of youth development and participatory urban planning. *Environment and Planning A 38*, 11 (2006), 2039–2054. 1

[EFN12]   ENDERT A., FIAUX P., NORTH C.: Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (2012), ACM, pp. 473–482. 2

[Ege97]   EGENHOFER M. J.: Query processing in spatial-query-by-sketch. *Journal of Visual Languages & Computing 8*, 4 (1997), 403–424. 2

[FSBS11]   FLORENCE C., SHEPHERD J., BRENNAN I., SIMON T.: Effectiveness of anonymised information sharing and use in health service, police, and local government partnership for preventing violence related injury: experimental study and time series analysis. *BMJ 342* (2011), d3313. 1, 4

[GD09]   GLANDER T., DÖLLNER J.: Abstract representations for interactive visualization of virtual 3d city models. *Computers, Environment and Urban Systems 33*, 5 (2009), 375–387. 2

[GRF09]   GREEN T. M., RIBARSKY W., FISHER B.: Building and applying a human cognition model for visual analytics. *Information visualization 8*, 1 (2009), 1–13. 2

[HT03]   HAKLAY M., TOBÓN C.: Usability evaluation and PPGIS: towards a user-centred design approach. *International Journal of Geographical Information Science 17*, 6 (2003), 577–592. 2

[IB95]   INGRAM R., BENFORD S.: Legibility enhancement for information visualisation. In *Proceedings of the 6th conference on Visualization'95* (1995), IEEE Computer Society, p. 209. 2

[IB96]   INGRAM R., BENFORD S.: The application of legibility techniques to enhance information visualizations. *The Computer Journal 39*, 10 (1996), 819–836. 2

[KMM*13]   KIM S., MACIEJEWSKI R., MALIK A., JANG Y., EBERT D. S., ISENBERG T.: Bristle maps: A multivariate abstraction technique for geovisualization. *IEEE Transactions on Visualization and Computer Graphics 19*, 9 (2013), 1438–1454. 2

[Lyn60]   LYNCH K.: *The Image of the City*, vol. 11. MIT press, 1960. 1

[MRH*10]   MACIEJEWSKI R., RUDOLPH S., HAFEN R., ABUSALAH A. M., YAKOUT M., OUZZANI M., CLEVELAND W. S., GRANNIS S. J., EBERT D. S.: A visual analytics approach to understanding spatiotemporal hotspots. *IEEE Transactions on Visualization and Computer Graphics 16*, 2 (2010), 205–220. 2

[RMK12]   REY S. J., MACK E. A., KOSCHINSKY J.: Exploratory space–time analysis of burglary patterns. *Journal of Quantitative Criminology 28*, 3 (2012), 509–531. 2

[Sie06]   SIEBER R.: Public participation geographic information systems: A literature review and framework. *Annals of the Association of American Geographers 96*, 3 (2006), 491–507. 1, 2

[Tal00]   TALEN E.: Bottom-up GIS: A new tool for individual and group expression in participatory planning. *Journal of the American Planning Association 66*, 3 (2000), 279–294. 1

[TSH*14]   TURKAY C., SLINGSBY A., HAUSER H., WOOD J., DYKES J.: Attribute signatures: Dynamic visual summaries for analyzing multivariate geographical data. *IEEE Transactions on Visualization and Computer Graphics 20*, 12 (2014), 2033–2042. 2

[vDHO16]   VAN DIJK T. C., HAUNERT J.-H., OEHRLEIN J.: Location-dependent generalization of road networks based on equivalent destinations. In *Computer Graphics Forum* (2016), vol. 35, Wiley Online Library, pp. 451–460. 2

[WDS10]   WOOD J., DYKES J., SLINGSBY A.: Visualisation of origins, destinations and flows with OD maps. *The Cartographic Journal 47*, 2 (2010), 117–129. 2

[WHC02]   WEINER D., HARRIS T., CRAIG W. J.: Community participation and geographic information systems. In *Community participation and geographic information systems*, Craig W. J., Harris T., Weiner D., (Eds.). Taylor & Francis, London and New York, 2002, ch. 1, pp. 3–16. 1

[WSD11]   WOOD J., SLINGSBY A., DYKES J.: Visualizing the dynamics of London's bicycle-hire scheme. *Cartographica: The International Journal for Geographic Information and Geovisualization 46*, 4 (2011), 239–251. 2

[WSM*09]   WONG P. C., SCHNEIDER K., MACKEY P., FOOTE H., CHIN JR G., GUTTROMSON R., THOMAS J.: A novel visualization technique for electric power grid analytics. *IEEE Transactions on Visualization and Computer Graphics 15*, 3 (2009), 410–423. 2

[WUS13]   WESSEL G., UNRUH E., SAUDA E.: Heads up: using cognitive mapping to develop a baseline description for urban visualization. In *Proceedings of the Sixth International Conference on advances in Computer-Human Interactions (ACHI)* (2013), vol. 24, Citeseer. 2

[XY08]   XIE Z., YAN J.: Kernel density estimation of traffic accidents in a network space. *Computers, Environment and Urban Systems 32*, 5 (2008), 396–406. 2

[XY13]   XIE Z., YAN J.: Detecting traffic accident clusters with network kernel density estimation and local spatial statistics: an integrated approach. *Journal of Transport Geography 31* (2013), 64–71. 2