Article

The Spatial-Comprehensiveness (S-COM) Index: Identifying Optimal Spatial Extents in Volunteered Geographic Information Point Datasets

Haydn Lawrence ^{1,*} , Colin Robertson ², Rob Feick ¹ and Trisalyn Nelson ³

¹ School of Planning, University of Waterloo, Waterloo, ON N2L 3G1, Canada; rdfeick@uwaterloo.ca

² Department of Geography and Environmental Studies, Wilfrid Laurier University, Waterloo, ON N2L 3C5, Canada; crobertson@wlu.ca

³ School of Geographical Sciences and Urban Planning, Arizona State University, Tempe, AZ 85281, USA; trisalyn.nelson@asu.edu

* Correspondence: hlawrenc@uwaterloo.ca

Received: 25 July 2020; Accepted: 19 August 2020; Published: 21 August 2020

Abstract: Social media and other forms of volunteered geographic information (VGI) are used frequently as a source of fine-grained big data for research. While employing geographically referenced social media data for a wide array of purposes has become commonplace, the relevant scales over which these data apply to is typically unknown. For researchers to use VGI appropriately (e.g., aggregated to areal units (e.g., neighbourhoods) to elicit key trend or demographic information), general methods for assessing the quality are required, particularly, the explicit linkage of data quality and relevant spatial scales, as there are no accepted standards or sampling controls. We present a data quality metric, the Spatial-comprehensiveness Index (S-COM), which can delineate feasible study areas or spatial extents based on the quality of uneven and dynamic geographically referenced VGI. This scale-sensitive approach to analyzing VGI is demonstrated over different grains with data from two citizen science initiatives. The S-COM index can be used both to assess feasible study extents based on coverage, user-heterogeneity, and density and to find feasible sub-study areas from a larger, indefinite area. The results identified sub-study areas of VGI for focused analysis, allowing for a larger adoption of a similar methodology in multi-scale analyses of VGI.

Keywords: citizen science; user-generated content; volunteered geographic information; spatial scale; data quality; Voronoi diagrams; quadtree; optimized extents; RinkWatch; FrogWatch

1. Introduction

Volunteered geographic information (VGI) [1], produced in full or in part by citizens, has been shown to have scientific, social, and cultural value. Methods have been developed to unlock this value through machine learning [2], data mining [3], and other computational methods [4,5]. Data that users contribute purposely or passively to Web sites and services (e.g., YouTube, Facebook, Twitter, Wikipedia) are often repurposed for a variety of applications, including some in which geographical information is important (e.g., Tweet maps). A persistent challenge is that the quality of these data varies from place-to-place since individuals choose where they author data and because non-experts differ in expertise, motivation, and data collection methods [6,7]. In particular, the irregular nature of informal data-authoring processes creates sampling patterns that often do not correspond to well-defined natural or administrative study areas (extents) for analysis. We explore this issue through the lens of a new data quality property termed “spatial data comprehensiveness” which we define as a “suitably even distribution of data observations in terms of coverage and density to fit the analytical

needs within a geographic area". Spatial data comprehensiveness is of particular importance for task/context-based analysis of VGI as it pertains to individuals.

Data quality has always been an important dimension in spatial data and GIS studies, with significant research to advance the understanding of spatial data accuracy [8–10], error propagation [11], data documentation standards [12], and the importance of semantics and fitness for use concepts [13,14]. In this paper, we present a new index that is designed to reduce the uncertainties implicit in the inability to control sampling by identifying feasible study areas within VGI datasets based on evaluation of a composite index of spatial data comprehensiveness (S-COM) without reference to external data. The key goal with the S-COM index is to delineate a feasible sub-area, or areas, utilizing only the data in the VGI dataset, lacking other authoritative or meta information sources, and allowing for a quick exploratory understanding of the key underlying spatial qualities of the dataset (e.g., optimized scale and extent for a data-rich analysis). When little is known about the data-authoring process giving rise to a VGI dataset, evaluating the evenness of contributions is critical to identify the relevant spatial scale for study. This is an especially important consideration when the geographic phenomena being investigated are in themselves fuzzy and uncertain such as what individuals consider as a city's "downtown". We have previously shown how the choice of spatial grain for analysis can be optimized with respect to the general data quality parameters for VGI [15]. Here, we build on that work to demonstrate how spatial extent can also be optimized based on a predefined index consistent with a study's overarching goals. As VGI becomes more prominent in research, tools to aid researchers in quickly gauging feasible grains and extents for study will be needed, especially given data with temporal characteristics where grain size may differ based on time slices (i.e., spatial data streams). Additionally, multi-scale patterns in VGI and sensitivity to zonal configurations (MAUP) can be investigated through the algorithms presented here.

Data from two citizen science projects, RinkWatch and FrogWatch, are used to illustrate different contexts where finding feasible extents and aggregation units is a required first step of VGI point dataset analysis. These two projects employ web-based maps and interfaces in order to solicit citizen-reporting of individual observations of ice rink conditions (RinkWatch) and frog or toad sightings (FrogWatch). While these initiatives have distinct goals, they are both used to gauge local variations of environmental variables, such as temperature and habitat quality, respectively.

2. Background

In terms of VGI with no reference dataset with which to compare it to, data quality evaluation takes on a meaning dependent upon the researchers' goals. With VGI [1], data quality is often heterogeneous since a single dataset is usually composed of the contributions from many users who often differ in skill sets, motivations, and areas of interest [16–19]. For researchers interested in using these data for understanding spatial processes, a greater limitation is that they lack control and knowledge over VGI data collection processes. While most data analysis methods assume data are representative and collected using a study design, volunteers typically decide where and when they want to collect data, what methods or tools are used, and whether their individual observations will correspond to what a researcher may require in a sampling design.

When faced with these new or relatively unknown datasets, such as in the context of citizen science, Dickenson [20] cites several data quality aspects of user-contributed data that can be problematic including: knowledge gaps between expert and amateur data collectors, limited demographic representation, and the challenge of maintaining volunteers' interest over extended periods of time. In many cases, analysis of VGI can even be considered representative of outliers rather than a representative sample [21,22]. Most VGI is created by an active minority of users, a phenomenon sometimes described as the 90-9-1 rule [23]. The motivations, aspirations, and characteristics of the users involved in the creation of these datasets may differ significantly from the wider population [16], passively excluding the critical voices bereft of access to the technology and resources required to participate [24]. These properties of VGI data analyses are similar to what has been reported in the

big data literature where research has been characterized as being overly data-driven in terms of problem focus, methods, and geographic locale [25–28]. For example, it is not accidental that most commercial activity and academic research related to VGI, especially geosocial forms (e.g., Twitter) and the near ubiquitous OpenStreetMap (OSM), are centred in the cores of major cities and near popular landmarks since these are typically the areas with the largest data volumes [23,29,30]. While many studies have analyzed the quality of these data and often found their quality to be comparable to expert-generated alternatives [10,23,31], rarely is the “site selection” or study extent aspect of these studies taken into consideration explicitly. Instead, analyses tend to be framed with reference to existing administrative areas (e.g., city boundaries) or otherwise pre-defined study extents (e.g., ecological zones) and exploratory approaches are used to assess the suitability of VGI resources in terms of geographic data coverage and distribution.

The need to study this implicit choice of pre-defined study extents for VGI data coverage and distribution is most apparent when the issue of spatial scale in VGI analysis is considered [32–35]. Point-based VGI, such as georeferenced photographs, and tweets, are often aggregated to areal units (e.g., neighbourhoods, cities, regions) and summaries of their content are produced to discern dominant perceptions of events, places, or other forms of citizen sensing [36,37]. Lacking direct control of VGI data collection, however, introduces uncertainties related to what geographic units of analysis are appropriate, the types of analyses the data are fit for, the role of mobility and place of residence [38], and what study areas are feasible for preselected analysis methods.

When a study area is chosen with non-uniform coverage, algorithms to find a specific subset of data will help to delineate new feasible sub-areas (i.e., data-rich extents/grains based on research parameters and criteria) for a more focused study. Unfortunately, the majority of research has focused on the assessment of VGI in comparison to extrinsic datasets (e.g., authoritative reference datasets, website metadata collections). This has led to a dearth of research utilizing the dataset itself (i.e., intrinsically) to gauge quality. A call by researchers into forming a new paradigm of intrinsic quality evaluation [6,39] has begun to take shape.

Senaratne et al. [6] built upon Goodchild and Li’s [40] proposal of the three approaches to VGI quality by adding “data mining” to the list. Their meta study of VGI quality assessment research showed that few researchers were actively working on an assessment of VGI without the use of authoritative reference datasets. These studies still addressed quality through comparisons to metadata supplied through the medium in which the data were collected (e.g., OSM user rank data, number of edits). None of the ~50 studies they reported upon assessed the quality of a VGI dataset through the use of the dataset itself—they all referenced other data for comparison. While image-based VGI quality assessment is in the process of moving to implicit-based quality measures [6,41,42], there is a definite need to move point-based VGI quality assessment from traditional reference comparison methodologies to respond to the multitude of different configurations of data that VGI may produce.

3. Methods

In this paper, we apply the term spatial comprehensiveness to a composite measure of three elements of VGI data quality that are pertinent for delineating feasible study areas: coverage, user-heterogeneity, and density. These components are expressed on a scale of 0 to 1 and combined multiplicatively to yield a final overall S-COM index. The relative weightings of the components can be varied to reflect differences in study parameters [15], such as a lack of unique user data. The first component of the S-COM index, coverage, assesses candidate extents and grains of the overall study area for desired spatial qualities in terms of observations, such as clustering of similar data or dispersion of attribute-less points. Coverage is important as VGI is inherently uneven due to users’ interests, unconcerned of research designs or technological specs. These candidate sub-areas can be chosen in the preliminary stages of the study and evaluated by the S-COM index or can be subunits defined by the process itself in an unsupervised fashion. The definition of “coverage” can be varied, allowing screening of potential sites based on application-specific minimum thresholds of observations.

For example, given a VGI dataset of the downtown of a large city in which we wish to gauge sentiment of visitors such as tourists, we would want to minimize coverage as tourist activity would be highly clustered and not dispersed throughout the entire city. In this paper, coverage is calculated as the ratio of data aggregated to areal units (e.g., rectangular grids, census tracts), labeled as cells for the purposes of this paper, with the specified minimum number of observations to the total number of cells in the study area. These cells could be grid-like or irregular Voronoi tessellations, or authoritative geographic units such as a census tract. General approaches to determining the minimum threshold parameter can be statistical in nature: examining the distribution through a histogram or excluding a certain percentage from the lower tail of the distribution. The minimum threshold can also be based on other factors, such as preserving the geoprivacy of participants.

The second component measure of S-COM is user-heterogeneity, which measures the ratio of observations to the number of unique contributors within each unit of analysis. This ratio is calculated by taking the total number of user submissions and the number of unique users within each cell divided by the total number of cells (and therefore can be thought of as a user-density measure). In the context of thematic analysis of geosocial data, high user-heterogeneity may indicate more certainty in an identified pattern or trend (i.e., Linus's Law) [40,43].

The final component is a spatial density measure which evaluates similarity among the aggregated data counts, visualized in a Moran's scatterplot. An estimate of how evenly distributed values are in a study area is made by evaluating the variance in the quadrants of the scatterplot. This is defined by the measured values on the X-axis and the spatially lagged values on the Y-axis, taking the largest and smallest quadrants by area and calculating a range where 0 represents perfect similarity. In this way, density is used to find areas of similarity both within geographic space and within data space (see Figure 1). The circled data points on the Moran's scatterplot (Y-axis) correspond to the shaded areas on the map (left). This sub-area of the map was found to be "similar" in terms of S-COM properties, with an index value of 0.816. The left pane (Voronoi diagram) is zoomed in to better highlight the polygons selected compared to the right pane with the complete set of data points. Density is formulated such that values near 1 indicate evenness in the Moran's scatterplot quadrants and therefore similarity of data distribution (i.e., near the mean). Values near zero indicate high variance in the Moran's scatterplot quadrants and are considered either spatial or aspatial outliers. Since this approach can be sensitive to outliers, tail thresholds can be used to reduce their effect on the measure of spatial density.

Figure 1. Voronoi tessellation created from generic test data.

To delineate feasible sub-areas within an overall study area, both regular and irregular tessellations are used. The regular tessellation method uses a quadtree approach and a recursive algorithm that determines whether adjacent polygons should be merged to create a more “comprehensive” parent polygon. The irregular tessellation method uses only the recursive algorithm, testing each adjacent polygon for improvement in the S-COM index. Both algorithms use the S-COM index to assess candidate sub-areas and return a value between 0 and 1 for each configuration tested.

The quadtree gridded approach was chosen for its ability to create variable resolutions recursively. The quadtree is a regular tessellation that continuously subdivides areas into four equal areas until a threshold of homogeneity is reached (Figure 2) [44–46]. The first iteration (computed metric value of 3) is divided into four nodes. The top right and bottom left nodes’ metric values (5 and 7) are greater than the parent node’s metric value (3) and are therefore further subdivided. In the third iteration, the top right quadrant has no nodes with metric values higher than the parent node, so it stops dividing. However, the bottom left continues until it also reaches the condition that all child nodes have lower metric values than the parent. The algorithm then takes the nodes from the lowest branch and merges them. It finds the metric value for them (8 in this case) and checks that value with the parent node’s value (7). If the value is higher, the merged polygon is sent up the tree instead of the complete parent polygon. This example returned two polygons with metric values 8 and 5. Bereuter and Weibel [46] explained that this type of data structure can increase query speed by continuously subdividing a heterogeneous study area until homogeneous data values are found. Examples of the use of quadtrees can be found in image retrieval [47,48] and environmental modelling [49,50].

Figure 2. (a) Quadtree branch creation NOTE: metric values are shown as whole numbers for illustration purposes. The metric values normally fall between 0 and 1; (b) quadtree return values.

The irregular tessellation approach uses an algorithm that iteratively merges adjacent polygons from a seed polygon (Figure 3). These Voronoi tessellations are created from the unit of analysis, in the case of RinkWatch, the individual rinks with the readings localized at the rinks. The algorithm iterates over every polygon present in the area of interest, treating each as a seed. The seed polygon’s neighbours are evaluated to find the one neighbouring polygon that creates the highest overall S-COM index value when combined with the seed’s value. If this value is higher than that of the current seed, the seed is replaced with the new merged polygon. This process is repeated until no adjacent polygons are found to create a higher S-COM index value. Adjacency is determined non-recursively

using a global adjacency matrix similar to the AMEOBA algorithm [51]. The algorithm checks each neighbour (green) around the merged polygon (red).

Figure 3. The polygon adjacency algorithm.

4. Data Sources and Study Area

Citizen science initiatives have grown in popularity in recent years, as developing and deploying data collection apps via mobile friendly websites and mobile phones has increased significantly. Projects such as RinkWatch, FrogWatch, and OakMapper.org [52] illustrate how simple geographic data applications can harness citizen observations on a variety of socially relevant topics. The RinkWatch project was designed to exploit the popularity of outdoor skating by recruiting citizens to contribute to a web-based log about the quality and availability of ice on their homemade rinks [53]. More importantly, observed changes in the availability of outdoor skating could motivate greater public engagement in climate change and climate science. FrogWatch aims to engage citizens in the collection of frog siting observations. Similar to RinkWatch, FrogWatch has both conservation science and educational objectives.

The case study of evaluating RinkWatch data to identify feasible study areas was focused on the Kitchener/Waterloo (KW) area of southern Ontario, Canada during the 2013 and 2014 winters. A feasible study for RinkWatch data would be slightly unique in comparison to other citizen science projects. OSM, for example, often has several users contributing data in each location as some will add new features while others may concentrate on improving existing data. In contrast, RinkWatch focuses on the monitoring of ice rinks, most of which are created and maintained by private homeowners. A total of 985 rinks (unique users) were registered over the two-year study period overall, with 22,661 individual readings of ice conditions (data points) specified. The KW area was chosen for the case study and has an estimated population of 535,154 and an average winter temperature of -7.7 degrees Celsius over the last three years [54]. This area was chosen due to the high participation rates with the RinkWatch project in addition to the familiarity of the area to researchers. Twenty-one rinks in the KW area are used in this analysis based on a filter of each rink having more than 20 ice condition readings (three consecutive weeks). This helps to eliminate spurious or malicious data from the analysis [31].

FrogWatch data were collected for the Toronto Census Metropolitan Area (CMA) which includes the City of Toronto as well as adjacent cities (e.g., Mississauga, Brampton, Markham) and a number of more sparsely populated towns and townships and had a population of approximately 6 million in 2014 [55]. The S-COM index was used in the Toronto CMA to identify areas where conservation efforts could be focused to reduce the number of frogs killed by roadway traffic. All frog sightings from FrogWatch within 500 m of any street, road, avenue, highway, or expressway were chosen for the analysis which resulted in 1108 frog or toad sightings over the 1999 to 2014 period, with 2015 onward showing a large drop in participation.

5. Case Study

Three types of aggregation methods were applied to the RinkWatch data: grids (quadrees), census tracts, and Voronoi polygons (see Figure 3). The quadtree implementation used in this research was recursive such that each cell was subdivided until all four child nodes' index values were less than the parent node's index value plus a small recursion tolerance (0.05 in the case of this implementation, where the index value is between 0 and 1). If any child node's value was greater than its parent's node index value plus the tolerance, the recursion would move to a deeper level. The algorithm finds the smallest feasible units across the study area in question and merges adjacent grid cells except when an isolated (i.e., island) grid cell is encountered. At the end of the algorithm's run, the polygon with the largest area and highest S-COM index value is identified as a feasible area (Figure 3), though it could be modified to return several feasible areas.

FrogWatch data were first aggregated to polygonal data before running the analysis, a common requirement for citizen science data, especially when potentially endangered species data are concerned. A grid was created with each cell measuring 10×10 km in area and frog sightings within 500 m of a road, street, avenue, highway, or expressway were found and aggregated within the grid cell in which they fell. This was done to simulate a study assessing the focus of resources for frog conservation efforts in the City of Toronto by looking at areas where frogs are often observed near heavily trafficked roads.

All data points within each polygon for both case studies were aggregated for S-COM index values determination with the index normalized by polygon area. The authoritative polygon delineations used were census tracts and dissemination areas, both chosen as census data disseminated at these geographic units and often used in the social sciences. Voronoi polygons for RinkWatch and a grid for FrogWatch were chosen for their non-standard properties. Index component weightings for analysis were varied for RinkWatch data: 50% density and coverage, and 33% density, coverage, and user-heterogeneity. Hereafter, these will be referred to as 505 weighting and 333 weighting, respectively. The 505 weighting was used for the analysis as user-heterogeneity was considered a less useful component. This is due to the fact that user-heterogeneity measures the ratio of data points to unique users while the case study will use a minimum of 20 data points per user for inclusion, causing a small user-heterogeneity S-COM component value. In addition, as the FrogWatch data did not include user information, the only weighting used was 505 (50% density and 50% coverage). The FrogWatch dataset was used to help understand the efficacy and limitations of the SCOM when utilized on a sparser dataset.

6. Results

6.1. RinkWatch

There were several differences between the regular tessellations when compared to the irregular tessellations. The irregular tessellations generally obtained higher S-COM index values and more homogenous data counts compared to the gridded areas which incorporated several extreme outliers (Figure 4). The regular tessellation algorithm returned feasible areas with very low S-COM index values (0.148/0.216 for the 333/505 index inputs) compared to both the irregular tessellation (census tracts) and geo-index tessellation (Voronoi polygons), both of which showed very similar values (0.968/0.9997 and 0.656/0.680, respectively) (Figure 4).

Figure 4. Three types of tessellations: (a) quadtree gridded polygons are used left, (b) census tracts centre, and (c) Voronoi polygons right.

The gridded area (Figure 4a) found an area of like counts of ice condition readings even though this same area contains much larger counts which do not meet the criteria needed for the skateability analysis described earlier. The S-COM index value is also very low for both the 505 and 333 weightings. Coverage would play a large role in the lower values as the grid area is rectangular by design, leaving a lot of empty space. In addition to this, the addition of the large outliers to the list of counts would decrease user-heterogeneity.

The irregular polygon analyses found higher S-COM index values overall (Figure 4b,c). The census tract analysis identified four polygons with similar numbers of readings. The individual polygonal index values were normalized by the area of the polygons. The Voronoi polygon analysis identified the same region. User-heterogeneity decreased the index values of both of these analyses as each aggregated point has just one unique user (rink). Dissemination areas were tested but proved to have too fine of a resolution for an analysis of this dataset. Overall, the adjacency algorithm provided higher S-COM index values than the quadtree algorithm.

The spline interpolation analysis shows a much smoother spline surface created than the surface created from the entire dataset (number of readings ≥ 20) (Figures 5 and 6), allowing us to better gauge the variability in the local area for skateability. While local variation would show differences from the official weather data stations (located at the airport in the KW region) and among different areas of the KW region (e.g., rural vs. urban), the variation shown using the entire dataset for the study areas is unusable (Figures 5b and 6b). Although there are many levels of users, from people who have been building rinks for decades to people learning for the first time, there should not be the amount of variability in the skateability of rinks as shown in the left images. The feasible areas found using both weightings (505 and 333) demonstrate much smoother surfaces than the overall area with similar values of skateability. The major outliers were not included in the feasible area. Figure 5 shows a 505 weighting with the 333 weighting (Figure 6) showing similar results. It should be noted that a spline creates the best surface based on fitting a smooth surface to the data, causing the values under 0% and over 100%.

Figure 5. Using Voronoi polygons, a spline interpolation surface fit based on rink skateability. The left surfaces (a) are the optimal areas with the right surfaces (b) showing the complete study area. Index component weightings are coverage 33%, user-heterogeneity 33%, and density 33%.

Figure 6. Using Voronoi polygons, a spline interpolation surface fit based on rink skateability. The left surfaces (a) are the optimal areas with the right surfaces (b) showing the complete study area. Index component weightings are coverage 50%, user-heterogeneity 0%, and density 50%.

6.2. FrogWatch

The FrogWatch data analysis was more limited in scope than the RinkWatch data analysis due to a lack of user/collector information, limiting the quality assessment such as checking if the data found in an area were collected by one user or many users. The polygon adjacency algorithm used with the FrogWatch data located large areas of similar data counts of frogs within the overall study area (10 × 10 km/cell). Figure 7 presents the most feasible study area based on the criteria of large coverage and data similarity. No minimum number of counts was applied to this trial. The two Moran's scatterplots found in Figure 7c illustrate the heterogeneity of the feasible area calculated by the algorithm. Coverage was found to be low at 0.056, while density was found to be high (0.884),

with an overall S-COM index value of 0.470. The coverage was influenced by a large number of cells with low counts. These low count cells were attached to the final feasible study area polygon due to the adjacency limitation of the algorithm. A larger lag value could allow the algorithm to calculate adjacency at different orders (i.e., allowing polygons to be included even if they were separated by n empty polygons), though this would create too large of a polygon for the small overall study area used. Buffer values could be used to combine several areas of like S-COM index values that fall within ranges of values, allowing for larger feasible areas to be found. The cell with 36 frog sightings was included in a feasible study area of approximately the same S-COM index value as the one in Figure 7 and would be included in the overall “best” one if ranges were implemented. One of the cells was found with a frog count of 142 sightings (Figure 7a). These were found to be all located at the same position and were probably due to a single bulk upload of all the sightings in the area. This highlights one of the problems when working with VGI. The algorithm did not use this cell in finding the highest S-COM index value as the density component was set to find outliers of large values in similar spaces but not in only one cell. The middle Moran’s scatterplot (b) presents the polygon in relation to the entire study area, while (c) illustrates the heterogeneity of the feasible area calculated by the algorithm (Y-axis = spatial lag, X-axis = frog counts).

Figure 7. The most feasible study area based on the criteria of large coverage and data similarity; (a) optimal extent; (b) Moran’s I scatterplot with optimal grids circled; (c) Moran’s I scatterplot focused on optimal grids.

7. Discussion

Previous research has demonstrated that analysis of VGI at different grains can result in different outcomes [29,56]. Here, we aimed to use the data quality index described in Lawrence et al. [15] to define a study extent based on the spatial characteristics inherent in VGI point patterns. The results showed that the methods were able to find sub-areas that matched the predefined criteria for the VGI case study of skateability of RinkWatch and for conservation efforts for FrogWatch. These criteria included similarity in data counts by each unique contributor in the case of RinkWatch, and spatial adjacency/similarity and a minimum amount of data counts in both case studies. Filtering out outliers, whether for exclusion, or for further study, and preserving spatial connectedness in spatial analysis of VGI point patterns may be an important pre-processing step for future studies of user-contributed data.

The S-COM index values showed that the differences between different spatial extents can be quite extreme. The first optimization method used was the quadtree gridded algorithm. While the

study area result encapsulated a reasonable subset of the overall data matching the criteria needed for a skateability analysis, it could not overcome the inherent deficiency of standard grids being used. The feasible area found for the RinkWatch data contained several aggregated counts of readings that were much higher than the surrounding counts creating a bias in the skateability analysis as a rink with 394 contributions would be much more accurate than the average rink with 73 contributions. While there could be uses for finding the most prolific contributors, such as an analysis to assess socio-economic factors that may explain strong participation rates [34,40,57–60], this was not the focus of this paper. The S-COM index, as a measure of spatial comprehensiveness, is extendable in the sense that component weightings can be tailored to study objectives. The analysis of FrogWatch differed in that the overall area was analyzed for outliers, or large concentrations of frogs, allowing for the demarcation of feasible areas for conservation efforts.

The irregular polygon study areas (census tracts, dissemination areas, and Voronoi polygons) were found to provide better S-COM index values. The census tract and Voronoi polygon study areas were much more conducive to spatial interpolation analysis. The census tract and Voronoi study areas proved to capture the best areas of all the different types of study areas tested, showing good compatibility with the criteria needed for the final skateability analysis. There was very little difference between using an index component weighting of 333 compared to the 505 weighting. The skateability analysis using spline interpolation helps to visualize the difference between using the complete study area versus the feasible Voronoi area. Figures 5 and 6 show that skateability within the extent shown is not as high as the complete set of points would have us believe. The value of 100% (bottom right of Figure 5) is a highly suspect value which should be further analyzed and most likely omitted from the analysis.

The FrogWatch analysis further helps to illustrate the S-COM index's efficacy in the delineation of sub-study areas within an initial study area. While a cursory visual examination of the study area might prove to elucidate the areas of large numbers of frog counts, the index used in this paper further quantifies the choice of sub-study area. However, the FrogWatch dataset was different in the unavailability of user information or meta knowledge creating difficulties in the use of the data [61–64], though the S-COM index was still able to find a suitable sub-area based on the conditions of the study (large concentrations of frogs).

The three measures of the S-COM metric can be replicated using many different approaches. DBSCAN is a well-known clustering technique that would be able to find areas of similar coverage. Mirahsan et al. [65] calculated user-heterogeneity of wireless cellular users utilizing the coefficient of variation of Voronoi cell areas and the resulting Delaunay cell edge lengths; however, this study used OSM data. Similar ideas are presented in Feng et al. [66] dealing with highly heterogeneous but noisy datasets, yet again using external data to verify heterogeneity. The S-COM's density and coverage measures approximate the Geospatial Data ISO guidelines for Completeness, which has been extensively studied [10,13,31,67] in terms of quality assessments. The S-COM metric allows for its three measures to be combined in various algorithms (e.g., Voronoi, census tracts, quad-tree) as demonstrated in this paper. This approach has two advantages over current methods. First and foremost, it is calculated solely through the dataset itself. Secondly, it utilizes all three measures in the calculations at each step, as opposed to calculating each measure separately and then attempting to combine them into one optimal extent. Regardless of whether geolocated VGI is being used to analyze daily activity patterns or sentiment, or citizen science data are being used to investigate the abundance of an amphibian, the spatial characteristics of the data need to be interrogated, and the method developed here is robust to different forms of subject-object studies. The three components of the metric used to identify candidate sub-areas for further analysis are important irrespective of application, but can be specifically tailored using the weighting parameters when combined into the final S-COM.

The need for methods to find study extents and identifying specific spatial scales are not limited to VGI. For example, Galpern [68] noted that the optimal spatial grain is usually unknown before

a landscape genetic analysis and gives the example of looking at various grains to filter out unimportant variation in the landscape of a wide-ranging organism gene flow. As the understanding of these various sources of VGI are still in their infancy, an approach to the utilization of a VGI dataset without a multiscale focus may leave important information hidden or ignored by the grain or extent chosen. Sester et al. [69] discussed how many of the delineations that may be found in VGI are not official and may have semi-permanent boundaries or fuzzy borders, as in the case of colloquial characterizations of urban neighbourhoods as “uptown”, “the college district”, or other culturally defined delineations. Modern clustering algorithms (e.g., DBScan, KMeans) could also offer additional patterns in the data that these algorithms would not find; however, many unsupervised clustering algorithms do not offer the ability to tailor sensitivity to initial centers of clusters and are not as likely to find optimal local minimums with several more subjective parameters with outlier data [70,71].

Overall, the use of the S-COM index was successful in identifying sub-study areas that showed similarity, were spatially connected, and had a minimum number of data. Finding the right study area without predefined borders can be challenging [72,73]. VGI data are created by users based on personal or project-related criteria which may not correspond to researchers’ needs. The results from the RinkWatch and FrogWatch case studies have shown that feasible sub-areas can be found through simple algorithms using a predefined index, allowing for an easier and faster study of data or comparison of different extents based on predefined conditions. Through these algorithms, we can reach a better understanding of how VGI can be used for analysis, in particular its use for public consumption of real-time data collection and immediate dissemination [69]. While this paper focused on the most feasible extent using different grains and feasibility approaches, altering the weights of the index components will allow its adoption in other types of studies. For example, the algorithm could have a minimum threshold for the index values it returns, allowing for several study areas of interest (extents) to be further analyzed by researchers. We anticipate that these findings will facilitate a wider adoption of multi-scale analysis of VGI and promote the data quality assessment of VGI in geography and other disciplines.

Author Contributions: Colin Robertson and Haydn Lawrence conceptualized the study. Haydn Lawrence performed the analysis and developed algorithms. Rob Feick and Trisalyn Nelson contributed to the writing and methodology development. All authors have read and agreed to the published version of the manuscript.

Funding: Social Sciences and Humanities Research Council of Canada (SSHRC).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Goodchild, M.F. Citizens as sensors: The world of volunteered geography. *GeoJournal* **2007**, *69*, 211–221. [\[CrossRef\]](#)
2. Li, W.; Goodchild, M.F.; Church, R.L.; Zhou, B. Geospatial Data Mining on the Web: Discovering Locations of Emergency Service Facilities. In Proceedings of the Advanced Data Mining and Applications, Beijing, China, 17–19 December 2011; Zhou, S., Zhang, S., Karypis, G., Eds.; Springer: Berlin/Heidelberg, Germany, 2012; pp. 552–563.
3. Elwood, S. Geographic information science: Visualization, visual methods, and the geoweb. *Prog. Hum. Geogr.* **2011**, *35*, 401–408. [\[CrossRef\]](#)
4. Goetz, M.; Zipf, A. Towards defining a framework for the automatic derivation of 3D CityGML models from volunteered geographic information. *Int. J. 3 Inf. Model. IJ3DIM* **2012**, *1*, 1–16. [\[CrossRef\]](#)
5. Tenney, M.; Hall, G.B.; Sieber, R.E. A crowd sensing system identifying geotopics and community interests from user-generated content. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 1497–1519. [\[CrossRef\]](#)
6. Senaratne, H.; Mobasher, A.; Ali, A.L.; Capineri, C.; Haklay, M. A review of volunteered geographic information quality assessment methods. *Int. J. Geogr. Inf. Sci.* **2017**, *31*, 139–167. [\[CrossRef\]](#)
7. Brown, G.; Rhodes, J.; Lunney, D.; Goldingay, R.; Fielding, K.; Garofano, N.; Hetherington, S.; Hopkins, M.; Green, J.; McNamara, S. The influence of sampling design on spatial data quality in a geographic citizen science project. *Trans. GIS* **2019**, *23*, 1184–1203. [\[CrossRef\]](#)

8. Barron, C.; Neis, P.; Zipf, A. A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Trans. GIS* **2014**, *18*, 877–895. [[CrossRef](#)]
9. Mooney, P.; Corcoran, P.; Winstanley, A.C. Towards quality metrics for OpenStreetMap. In Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems, San Jose, CA, USA, 2–5 November 2010; Association for Computing Machinery: San Jose, CA, USA, 2010; pp. 514–517.
10. Girres, J.-F.; Touya, G. Quality assessment of the French OpenStreetMap dataset. *Trans. GIS* **2010**, *14*, 435–459. [[CrossRef](#)]
11. Devillers, R.; Stein, A.; Bédard, Y.; Chrisman, N.; Fisher, P.; Shi, W. Thirty years of research on spatial data quality: Achievements, failures, and opportunities. *Trans. GIS* **2010**, *14*, 387–400. [[CrossRef](#)]
12. Jeansoulin, R.; Devillers, R. Chapter 2. Spatial Data Quality: Concepts. In *Fundamentals of Spatial Data Quality*; Devillers, R., Jeansoulin, R., Eds.; ISTE: London, UK, 2006; pp. 31–42.
13. Basiri, A.; Haklay, M.; Foody, G.; Mooney, P. Crowdsourced geospatial data quality: Challenges and future directions. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 1588–1593. [[CrossRef](#)]
14. Lansley, G.; de Smith, M.; Goodchild, M.; Longley, P. Big Data and Geospatial Analysis. *arXiv* **2019**, arXiv:1902.06672.
15. Lawrence, H.; Robertson, C.; Feick, R.; Nelson, T. Identifying Optimal Study Areas and Spatial Aggregation Units for Point-Based VGI from Multiple Sources. In *Advances in Spatial Data Handling and Analysis: Select Papers from the 16th IGU Spatial Data Handling Symposium*; Harvey, F., Leung, Y., Eds.; Advances in Geographic Information Science; Springer International Publishing: Cham, Switzerland, 2015; pp. 65–84. ISBN 978-3-319-19950-4.
16. Coleman, D.; Georgiadou, Y.; Labonte, J. Volunteered geographic information: The nature and motivation of producers. *Int. J. Spat. Data Infrastruct. Res.* **2009**, *4*, 332–358.
17. Coleman, D.J. Potential contributions and challenges of VGI for conventional topographic base-mapping programs. In *Crowdsourcing Geographic Knowledge*; Sui, D., Elwood, S., Goodchild, M., Eds.; Springer: Dordrecht, The Netherlands, 2013; pp. 245–263.
18. Budhathoki, N.R.; Haythornthwaite, C. Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap. *Am. Behav. Sci.* **2013**, *57*, 548–575. [[CrossRef](#)]
19. Robertson, C.; Feick, R. Defining local experts: Geographical expertise as a basis for geographic information quality. In Proceedings of the 13th International Conference on Spatial Information Theory (COSIT 2017), L'Aquila, Italy, 4–8 September 2017; Clementini, E., Donnelly, M., Yuan, M., Kray, C., Fogliaroni, P., Ballatore, A., Eds.; Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik: Dagstuhl, Germany, 2017; Volume 86, pp. 22:1–22:14.
20. Dickinson, J.L.; Zuckerberg, B.; Bonter, D.N. Citizen science as an ecological research tool: Challenges and benefits. *Annu. Rev. Ecol. Evol. Syst.* **2010**, *41*, 149–172. [[CrossRef](#)]
21. Haklay, M. Citizen science and volunteered geographic information: Overview and typology of participation. In *Crowdsourcing Geographic Knowledge*; Sui, D., Elwood, S., Goodchild, M., Eds.; Springer: Dordrecht, The Netherlands, 2013; pp. 105–122.
22. Antoniou, V.; Capineri, C.; Haklay, M.E. VGI and beyond: From data to mapping. In *Routledge Handbook of Mapping and Cartography*; Kent, A., Vujakovic, P., Eds.; Routledge: Abingdon, UK, 2017; pp. 475–488.
23. Neis, P.; Zielstra, D.; Zipf, A. The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. *Future Internet* **2012**, *4*, 1–21. [[CrossRef](#)]
24. Czepkiewicz, M.; Jankowski, P.; Młodkowski, M. Geo-questionnaires in urban planning: Recruitment methods, participant engagement, and data quality. *Cartogr. Geogr. Inf. Sci.* **2017**, *44*, 551–567. [[CrossRef](#)]
25. Hay, S.I.; George, D.B.; Moyes, C.L.; Brownstein, J.S. Big data opportunities for global infectious disease surveillance. *PLoS Med.* **2013**, *10*, e1001413. [[CrossRef](#)]
26. Kitchin, R. Big data and human geography: Opportunities, challenges and risks. *Dialogues Hum. Geogr.* **2013**, *3*, 262–267. [[CrossRef](#)]
27. Goodchild, M.F. The quality of big (geo) data. *Dialogues Hum. Geogr.* **2013**, *3*, 280–284. [[CrossRef](#)]
28. Lazer, D.; Kennedy, R.; King, G.; Vespignani, A. The parable of google flu: Traps in big data analysis. *Science* **2014**, *343*, 1203–1205. [[CrossRef](#)]

29. Li, L.; Goodchild, M.F.; Xu, B. Spatial, temporal, and socioeconomic patterns in the use of Twitter and Flickr. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 61–77. [\[CrossRef\]](#)
30. Hecht, B.; Stephens, M. A tale of cities: Urban biases in volunteered geographic information. In Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media, Ann Arbor, MI, USA, 1–4 June 2014; pp. 197–205.
31. Haklay, M. How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environ. Plan. B Plan. Des.* **2010**, *37*, 682–703. [\[CrossRef\]](#)
32. Neis, P.; Zielstra, D. Recent developments and future trends in volunteered geographic information research: The case of OpenStreetMap. *Future Internet* **2014**, *6*, 76–106. [\[CrossRef\]](#)
33. Jeffery, C.; Ozonoff, A.; Pagano, M. The effect of spatial aggregation on performance when mapping a risk of disease. *Int. J. Health Geogr.* **2014**, *13*, 9. [\[CrossRef\]](#) [\[PubMed\]](#)
34. Hollenstein, L.; Purves, R. Exploring place through user-generated content: Using Flickr tags to describe city cores. *J. Spat. Inf. Sci.* **2010**, *2010*, 21–48.
35. Feick, R.; Robertson, C. A multi-scale approach to exploring urban places in geotagged photographs. *Comput. Environ. Urban Syst.* **2015**, *53*, 96–109. [\[CrossRef\]](#)
36. Derungs, C.; Purves, R.S. From text to landscape: Locating, identifying and mapping the use of landscape features in a Swiss Alpine corpus. *Int. J. Geogr. Inf. Sci.* **2014**, *28*, 1272–1293. [\[CrossRef\]](#)
37. Shelton, T.; Poorthuis, A.; Graham, M.; Zook, M. Mapping the data shadows of Hurricane Sandy: Uncovering the sociospatial dimensions of ‘big data’. *Geoforum* **2014**, *52*, 167–179. [\[CrossRef\]](#)
38. Robertson, C.; Feick, R. Inference and analysis across spatial supports in the big data era: Uncertain point observations and geographic contexts. *Trans. GIS* **2018**, *22*, 455–476. [\[CrossRef\]](#)
39. Antoniou, V.; Skopeliti, A. Measures and indicators of VGI quality: An overview. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci. Present. ISPRS Geospatial Week* **2015**, *II-3/W5*, 345–351. [\[CrossRef\]](#)
40. Goodchild, M.F.; Li, L. Assuring the quality of volunteered geographic information. *Spat. Stat.* **2012**, *1*, 110–120. [\[CrossRef\]](#)
41. Fuchs, G.; Andrienko, N.; Andrienko, G.; Bothe, S.; Stange, H. Tracing the German centennial flood in the stream of tweets: First lessons learned. In Proceedings of the Second ACM SIGSPATIAL International Workshop on Crowdsourced and Volunteered Geographic Information, Orlando, FL, USA, 5 November 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 31–38.
42. Foody, G.M.; See, L.; Fritz, S.; Van der Velde, M.; Perger, C.; Schill, C.; Boyd, D.S.; Comber, A. Accurate attribute mapping from volunteered geographic information: Issues of volunteer quantity and quality. *Cartogr. J.* **2015**, *52*, 336–344. [\[CrossRef\]](#)
43. Haklay, M.; Basiouka, S.; Antoniou, V.; Ather, A. How many volunteers does it take to map an area well? The validity of Linus’ law to volunteered geographic information. *Cartogr. J.* **2010**, *47*, 315–322. [\[CrossRef\]](#)
44. Brisaboa, N.R.; Cerdeira-Pena, A.; de Bernardo, G.; Navarro, G.; Pedreira, Ó. Extending general compact queryable representations to GIS applications. *Inf. Sci.* **2020**, *506*, 196–216. [\[CrossRef\]](#)
45. Gage, T.; González-Nova, J.I.; Ladra, S.; Navarro, G.; Seco, D. Faster compressed quadrees. In Proceedings of the 2015 Data Compression Conference, Snowbird, UT, USA, 7–9 April 2015; IEEE Computer Society: Washington, DC, USA, 2015; pp. 93–102.
46. Bereuter, P.; Weibel, R. Real-time generalization of point data in mobile and web mapping using quadrees. *Cartogr. Geogr. Inf. Sci.* **2013**, *40*, 271–281. [\[CrossRef\]](#)
47. Ramanathan, V.; Mishra, S.; Mitra, P. Quadtree decomposition based extended vector space model for image retrieval. In Proceedings of the 2011 IEEE Workshop on Applications of Computer Vision (WACV), Kona, HI, USA, 5–7 January 2011; IEEE Computer Society: Washington, DC, USA, 2011; pp. 139–144.
48. Zhang, C.; Chen, R.; Zhu, L.; Liu, A.; Lin, Y.; Huang, F. Hierarchical information quadtree: Efficient spatial temporal image search for multimedia stream. *Multimed. Tools Appl.* **2019**, *78*, 30561–30583. [\[CrossRef\]](#)
49. Popinet, S. Quadtree-adaptive tsunami modelling. *Ocean Dyn.* **2011**, *61*, 1261–1285. [\[CrossRef\]](#)
50. Qiu, L.; Du, Z.; Zhu, Q.; Fan, Y. An integrated flood management system based on linking environmental models and disaster-related data. *Environ. Model. Softw.* **2017**, *91*, 111–126. [\[CrossRef\]](#)
51. Valles, G. *AMOEBa: A Multidirectional Optimum Ecotope-Based Algorithm*; R package; CRAN: Toronto, ON, Canada, 2014.
52. Connors, J.P.; Lei, S.; Kelly, M. Citizen science in the age of neogeography: Utilizing volunteered geographic information for environmental monitoring. *Ann. Assoc. Am. Geogr.* **2012**, *102*, 1267–1289. [\[CrossRef\]](#)

53. Robertson, C.; McLeman, R.; Lawrence, H. Winters too warm to skate? Citizen-science reported variability in availability of outdoor skating in Canada. *Can. Geogr. Géographie Can.* **2015**, *59*, 383–390. [\[CrossRef\]](#)
54. Statistics Canada Historic Climate Data. Available online: <http://climate.weather.gc.ca/> (accessed on 12 December 2019).
55. Statistics Canada Population of Census Metropolitan Areas. Available online: <http://www.statcan.gc.ca/tables-tableaux/sum-som/101/cst01/demo05a-eng.htm> (accessed on 26 March 2017).
56. Mackaness, W.A.; Chaudhry, O. Assessing the Veracity of Methods for Extracting Place Semantics from Flickr Tags. *Trans. GIS* **2013**, *17*, 544–562. [\[CrossRef\]](#)
57. Truong, Q.T.; De Runz, C.; Touya, G. Analysis of collaboration networks in OpenStreetMap through weighted social multigraph mining. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 1651–1682. [\[CrossRef\]](#)
58. Purves, R.; Edwardes, A.; Wood, J. Describing place through user generated content. *First Monday* **2011**, *16*. [\[CrossRef\]](#)
59. Elwood, S.; Goodchild, M.; Sui, D. Prospects for VGI research and the emerging fourth paradigm. In *Crowdsourcing Geographic Knowledge*; Sui, D., Elwood, S., Goodchild, M., Eds.; Springer: Dordrecht, The Netherlands, 2013; pp. 361–375.
60. Mooney, P. Understanding the activity of contributors to VGI projects. How, why, where, and when do they contribute geographic information? In Proceedings of the 26th International Cartographic Conference (ICC), Dresden, Germany, 25–30 August 2013.
61. Gardner, Z.; Mooney, P.; De Sabbata, S.; Dowthwaite, L. Quantifying gendered participation in OpenStreetMap: Responding to theories of female (under) representation in crowdsourced mapping. *GeoJournal* **2019**, 1–18. [\[CrossRef\]](#)
62. McLafferty, S.; Schneider, D.; Abelt, K. Placing volunteered geographic health information: Socio-spatial bias in 311 bed bug report data for New York City. *Health Place* **2020**, *62*, 102282. [\[CrossRef\]](#)
63. Bastin, L.; Schade, S.; Schill, C. Data and metadata management for better VGI reusability. In *Mapping and the Citizen Sensor*; Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C., Antoniou, V., Eds.; Ubiquity Press: London, UK, 2017; pp. 249–272.
64. Brown, G. A review of sampling effects and response bias in internet participatory mapping (PPGIS/PGIS/VGI). *Trans. GIS* **2017**, *21*, 39–56. [\[CrossRef\]](#)
65. Mirahsan, M.; Schoenen, R.; Szyszkowicz, S.S.; Yanikomeroğlu, H. Measuring the spatial heterogeneity of outdoor users in wireless cellular networks based on open urban maps. In Proceedings of the 2015 IEEE International Conference on Communications (ICC), London, UK, 8–12 June 2015; IEEE Computer Society: Washington, DC, USA, 2015; pp. 2834–2838.
66. Feng, J.; Zhang, M.; Wang, H.; Yang, Z.; Zhang, C.; Li, Y.; Jin, D. Dplink: User identity linkage via deep neural network from heterogeneous mobility data. In Proceedings of the The World Wide Web Conference, San Francisco, CA, USA, 13–17 May 2019; pp. 459–469.
67. Zielstra, D.; Zipf, A. A comparative study of proprietary geodata and volunteered geographic information for Germany. In Proceedings of the 13th AGILE International Conference on Geographic Information Science, Guimarães, Portugal, 11–14 May 2010; Volume 2010.
68. Galpern, P.; Manseau, M.; Wilson, P. Grains of connectivity: Analysis at multiple spatial scales in landscape genetics. *Mol. Ecol.* **2012**, *21*, 3996–4009. [\[CrossRef\]](#)
69. Sester, M.; Jokar Arsanjani, J.; Klammer, R.; Burghardt, D.; Haunert, J.-H. Integrating and Generalising Volunteered Geographic Information. In *Abstracting Geographic Information in a Data Rich World: Methodologies and Applications of Map Generalisation*; Burghardt, D., Duchêne, C., Mackaness, W., Eds.; Lecture Notes in Geoinformation and Cartography; Springer International Publishing: Cham, Switzerland, 2014; pp. 119–155.
70. Nayini, S.E.Y.; Geravand, S.; Maroosi, A. A novel threshold-based clustering method to solve K-means weaknesses. In Proceedings of the 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS), Chennai, Tamil Nadu, India, 1–2 August 2017; IEEE Computer Society: Washington, DC, USA, 2017; pp. 47–52.
71. Patel, K.M.A.; Thakral, P. The best clustering algorithms in data mining. In Proceedings of the 2016 International Conference on Communication and Signal Processing (ICCSP), Melmaruvathur, Tamil Nadu, India, 6–8 April 2016; IEEE Computer Society: Washington, DC, USA, 2016; pp. 2042–2046.

72. Brown, G.; Kelly, M.; Whittall, D. Which 'public'? Sampling effects in public participation GIS (PPGIS) and volunteered geographic information (VGI) systems for public lands management. *J. Environ. Plan. Manag.* **2014**, *57*, 190–214. [[CrossRef](#)]
73. Fonte, C.C.; Antoniou, V.; Bastin, L.; Estima, J.; Arsanjani, J.J.; Bayas, J.-C.L.; See, L.; Vatsava, R. Assessing VGI data quality. In *Mapping and the Citizen Sensor*; Foody, G., See, L., Fritz, S., Mooney, P., Olteanu-Raimond, A.-M., Fonte, C., Antoniou, V., Eds.; Ubiquity Press: London, UK, 2017; pp. 137–163.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).