JOURNAL OF SPATIAL INFORMATION SCIENCE
Number N (YYYY), pp. xx–yy

doi:10.5311/JOSIS.YYYY.II.NNN

RESEARCH ARTICLE

Bananas and tangerines spilled on streets

Martin Fleischmann

Geographic Data Science Lab, Department of Geography and Planning, University of Liverpool, United Kingdom

Anastassia Vybornova

NEtworks, Data and Society (NERDS), Computer Science Department, IT University of Copenhagen, Denmark

Received: December 24, 2015; returned: February 25, 2016; revised: July 13, 2016; accepted: September 5, 2016.

Abstract: 1-2 sentences basic introduction into field. 2-3 sentences more detailed background. 1 sentence clearly stating the general problem being addressed by this particular study. 1 sentence summarizing the main result ("here we show"). 2-3 sentences explaining what the main result reveals/adds. 1-2 sentences to put results into more general context. Optional - if accessibility is enhanced by this: 2-3 sentences to provide broader perspective.

Keywords: street networks, blocks, urban form, shape analysis, urban morphology, urban morphometrics, routing

1 Introduction

- Importance of street networks for urban analysis - talk about availability of data, different use cases from transport to morphology to ... - try to illustrate the wide applicability so we can then base the claims about the importance of the issue on top of it - general motivation - framework of urban data science. why everyone would benefit from having this issue solved. cite arcaute on 'recent advances, lobo on 'urban science'. also: alessandretti 2020, louail 2015, barthelemy books (morphogenesis 2018; spatial networks 2022) - Data need to look different for transport than for morphology ans why it matters - Networks vs polygons enclosed by networks (blocks? negative space? we need to pin down the terminology we want to use) - Problem description - Each network comes with a different detail and generated "blocks" are not always what they seem to be but sometimes are an artifact of transport-focused geometry - cite cardillo, geisberger, morer (computational costs), maybe venerandi 2016?; vanegas paper on actually *simulating* these spaces - Examples - other authors complaining about the issue, without having solved it yet (e.g. best paper ever [**?**]); grippa 2018; peponis 2007 merges these into urban blocks (replacing by center lines) - include morphometric literature here - mention ‘momepy.Blocks‘ algorithm that attemps to go around the issue in a specific way (but does not solve it) - (fleischmann, porta, dibble, etc.) diet 2018 on planar map classification. sharifi on urban forms. - description/terminology: cf. hermosilla 2014 ’UBRSA’; see strano 2012 for power law of "land cells" (spaces surrounded by street segments); most recent: shpuza 2011, 2017, 2022 (how to get the PDF...). circular compactness - inspired by louf; see also more recent barthelemy 2017 with the same figures; - summary of what happens in this paper - ’towards an automated detection of bananas’; method inspired by louf and barthelemy; tried out on 150 cities across the world

INTRODUCTION OF THE INTRODUCTION

Many studies within urban science use the street network of a city as primary input. Recent examples are numerous and cover a wide range of applications: from transportation to urban morphologies [cite]. The feasibility and broad applicability of quantitative urban science studies has greatly increased with open source GIS data becoming available on platforms like OpenStreetMap [cite].

However, even when a data set of sufficient quality is available, the conversion of geospatial features into a street network, sometimes referred to as “network simplification”, still poses several unresolved methodological questions to the research community. The challenge, in a nutshell, is to reduce granularity of detail without loss of relevant information; in many cases, deciding which information to keep and which to aggregate is easy for a human, but challenging for an algorithm (see Figure 1 [ref]). This is true both for studies that are concerned with the street network itself, and for studies that look at the polygons enclosed by the street network. Further complexity is added by the fact that the requirements towards the input network vary greatly depending on the use case. For example, traffic routing applications require adequately represented directionality of edges (street segments), while urban morphology studies are based on the shape of the polygons between the edges [cite].

In this article, we work towards resolving one specific issue within the features-to-network conversion process, which we call “bananas”. In the next sections, we describe the problem and define our research question, followed by a literature and terminology review. We then briefly describe the methodology that we apply to 150 cities across the globe, and present on overview of our results. The fully documented workflow, all input data and all results are made available in open source format on github [link]. We conclude with a discussion of implications and potential further steps.

PROBLEM DESCRIPTION

Each geospatially encoded street network comes with a certain level of granularity. A close look at the polygons enclosed by the edges (street segments) of a given network often reveals artifacts of transport-focused geometry. An illustration is found in Figure 1 [ref].

[Description of Figure 1: A close-up look at [city, location]. The black lines are network edges (street segments). Grey polygons are “true urban blocks”; the red polygon is a “banana”.]

Unlike its grey-colored neighbors, the polygon colored in red is not an urban block enclosed by streets; rather, it appears in the network due to the representation of a bidirectional street as two separate edges, one in each direction of traffic flow. The "banana" in Figure 1 [ref] poses a twofold problem. First, for studies concerned with urban form, it introduces a false signal into the distribution of urban shapes and distorts the actual shape of its neighboring polygons. Second, for any study that is concerned with the street pattern rather than with routing, the "banana" introduces a superfluous edge, and it does so in a frequently inconsistent matter – not for all, but only some bidirectional and/or multilane streets. The extent to which the "banana" situation distorts results depends on the analysis conducted, and cannot be quantified without prior "banana" identification. Thus, no matter whether one is interested in the network or in shapes enclosed by the network, the "banana" should simply not be there, but rather be replaced by one network edge. Human manual processing would be unambiguous, but prohibitively costly. We therefore pose the following research question:

RESEARCH QUESTION How can "bananas" in an urban street network be computationally identified?

LITERATURE REVIEW AND TERMINOLOGY

Disclaimer: We do not claim the literature review to be complete. Rather, we are sure that there are many studies on this which we simply did not find due to the lack of a coherent terminology in the problem description. For this we apologize, and hope to contribute to a future homogenization of the problem discussion.

Chronological review of problem description:

Within the GIS community, a similar problem has been described already in the 1970ies as "sliver polygon".

For definition see (but dont cite) https://en.wikipedia.org/wiki/Sliver_polygon

Goodchild 1978 (in "Statistical aspects of the **polygon overlay** problem) – apparently first time it is described (but can't get a hold of the article)

Rybaczuk 1993: First description of "information based rules for sliver polygon removal in GIS"

Delafontaine et al. 2009: More recent foundational study on Sliver polygon removal

Sliver polygon removal in QGIS: https://docs.qgis.org/3.22/en/docs/user_manual/processing_algs/qgis/vectorgeometry.html#eliminate-selected-polygons

Sliver polygon removal in ArcGIS: https://pro.arcgis.com/en/pro-app/latest/help/data/validating-data/polygon-sliver.htm

TO CHECK: **what criteria** are used to identify sliver polygons? and are they replaced by something upon their removal from the data set? If so, **what criteria** are used to compute that replacing something?

While geometrically similar to sliver polygons, "bananas" stem from a context-dependent redundancy of mapped line features, rather than from mismatching boundaries in overlays of polygon features. In addition, "bananas" are, in line with our problem definition, confined to the specific context of urban street networks; hence our choice of vocabulary to call "bananas" STREET SLIVER POLYGONS.

Studies wrangling with sliver polygons in the context specifically of STREET NETWORKS:

Peponis 2007: conduct an analysis of urban spatial profiles and point out that their input data includes street center lines, but lacks street widths, hence street surfaces are merged into urban blocks.

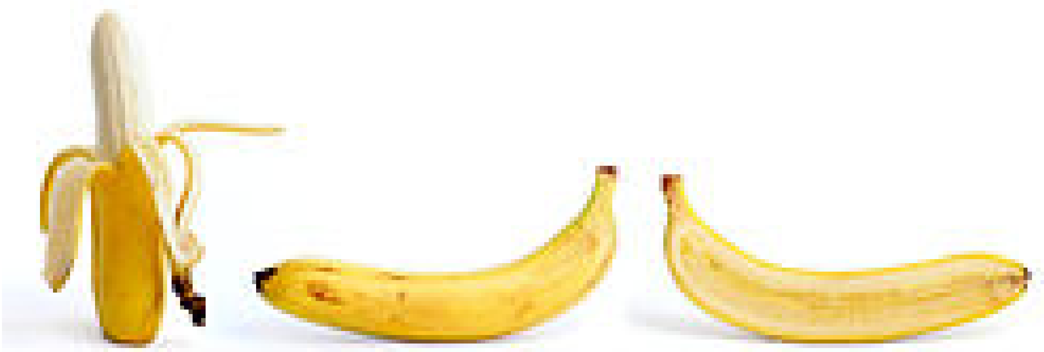

Figure 1: Banana

Strano et al. 2012: check if there is any reference to this issue? (or is the data they used not fine grained enough to encounter that problem?)

In a related, but methodologically different approach, Hermosilla et al. 2014 develop a method to derive so-called urban block related street areas (abbreviated by the authors as UBRSA), defined as the street area surrounding an urban block. This method, however, requires urban block boundaries as input.

Grippa 2018 et al: classify polygons derived from OpenStreetMap street network data into “urban blocks” and “sliver polygons” and present a semiautomated workflow, partially in PostGIS, for sliver polygon removal from the data set.

Shpuza 2022: [to be read]

Vybornova et al 2022: call them “parallel edges” and present a network path-based approach for identification (but no solution to how to remove these)

2 Method

This is method. And the figure **??** shows the banana.

- select sample of urban areas (FUA) - fetch the data from OSM - polygonize the network - measure shape charactersitics - TODO: measure initally more than Reock (get a sample from ESDA) - there is a conceptual backbone to this - we know that the artifacts are either small (small intersections) polygons or can be large but then they are very narrow (in between dual carriegaway) - we need a shape metric that captures this relationship - identify optimal measurements - plots that help us visually detect a cluster of artifacts - derivation of 1-dimensional index - from Roeck and area we can derive one value from which distribution we can identify a cut-off value for artifact/non-artifact polygons - cut-off value detection - exploration of geographical variation - differences between cities and continents - open tools, open data, open code with full reproducibility

3 Results

- area vs shape plots - use all cases together and show multiple shape indices - Reock as an optimal index (?) [I think it will be the optimal one but we need to verify that] - 1-dimensional index formula (if we use Reock it is the one from the banana notebook) - shape-index plots with cut-off values - plots based on geographical location - distributions, Reock-area scatters - describe the differences - formalise the detection workflow

4 Discussion

How could this be used?

how to move forward? (sneak preview of google summer of code) - the simplification problem can be seen as a problem of the elimination of banana

incorporate further data (ideas: directionality; street names; angles; land use; ...) use network formalism: on dual approach (intersections = edges): jiang 2004, yang 2022, rosvall/sneppen; barthelemy paper on shortest path shape

end with a call to action & 'towards open urban data science'

Acknowledgments

To be added. Remember to include ESRC/ATI funding covering initial experiments.