How “Alternative” are Alternative Facts? Towards Measuring Statement Coherence via Spatial Analysis

[Vision Paper]

Krzysztof Janowicz
STKO Lab
University of California, Santa Barbara
janowicz@ucsb.edu

Grant McKenzie
University of Maryland, College Park
gmck@umd.edu

ABSTRACT

Following the AAA principle by which anybody can say anything about any topic, the Web is no stranger to alternative facts. Nonetheless, with the increasing volume and velocity at which content is being published and difficulties to assess the credibility of information and the trustworthiness of sources, alternative facts are becoming a major challenge and an instrument for spreading disinformation. Interestingly, the diversity of today's data sources can also help us to counter alternative facts by measuring their coherence, i.e., the degree to which data from one source confirms or contradict data from another source. While a single dataset can be biased towards supporting or discrediting a statement, the diverse sources of data across media types that are publicly accessible today offer unique perspectives on which to assess a given statement. To give an intuitive example, a statement about the comparison of crowd sizes should align with photos of said crowds. However, these photos could be taken at different times, from different viewpoints, and could lead to different, sample-based estimations. Adding further data from heterogeneous sources, such as metro ridership, can either further support a statement or contradict it. We use three thought experiments to discuss the role of geographic data, knowledge graphs, and spatial analysis in approaching alternative facts from a novel angle, namely by studying their coherence, i.e., whether they align with other statements, instead of trying to falsify them. In doing so, we aim at increasing the costs for maintaining alternative facts.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous

Keywords

Alternative facts, spatial analysis, knowledge graphs

1. INTRODUCTION AND MOTIVATION

SIGSPATIAL'17, November 7–10, 2017, Los Angeles Area, CA, USA
ACM ISBN .
DOI:

While notions such as alternative facts, fake news, or bullshit [1] have different meanings, e.g., fake news producers consciously spread falsehoods, whereas the term bullshit characterizes expressions intended to persuade without regard for truth, they share the common idea that truth is either relative or of secondary importance. *Truth* itself is a more difficult concept than one would naively assume. In a non-scientific context, e.g., policy making, the *scientific method* [10] makes it difficult for researchers to assert observations such as climate change as true or to effectively disprove falsehoods. Falsehoods and fake news are, of course, not a new concept. In fact, several travel reports by famous early explorers and even entire expeditions are now being questioned, e.g., Amerigo Vespucci's first voyage. Phantom island, for example, were sometimes made up to generate funding for future expeditions. The problem today, however, is one of *scale*. In 2013, Brandolini summarized the problem by pointing to its asymmetric nature and stating that ‘the amount of energy needed to refute bullshit is an order of magnitude bigger than to produce it.’ Simply put, with an ever growing number of falsehoods that can be readily and very widely spread via social media, it is not possible anymore to catch up with them, e.g., by disproving statements or putting them into the right context, following traditional journalism or the scientific method.

There are at least five major theories of truth, namely coherence theory, correspondence theory, constructivist theory, consensus theory, and the pragmatic theory, as well as various other approaches that challenge those major theories and the notion of or need for truth. The coherence theory of truth, for instance, mandates that truth is a property of systems of propositions and can only be applied to individual properties based on their coherence to the other propositions. Consequently, in theory, contradictory proposition can be true as long as they cohere to some system of propositions. In contrast, the correspondence theory of truth defines individual propositions as being true if they accurately reflect (or explain) reality. Each of these theories can be criticized from different angles. Nonetheless, they still provide us with criteria to assert the truth of propositions in a certain *context*.

With regards to geographic information, the question of truth can be approached from a wide range of perspectives. To give a few examples, uncertainty is a related concept that has been extensively studied in GIScience from at least three different levels, namely conceptualization uncertainty, representational uncertainty, and analysis uncertainty [5]. The misuse of maps [9] and geo-visualization (and analysis more broadly) has also been studied extensively. Location is also a key element in fraud prevention and frequently used by credit card companies. While our work focuses on statements, fake news detection [13] is an related effort. Finally, and most related to our proposal, one can study the nature of geographic information to reason about the quality of data contributed via user-generated content.

For example, Goodchild and Li studied the imagined country of *Allestone* drawn by Thomas Williams Malkin by measuring whether it was adhering to three well-known geographic principles [2]. First they studied the fractal dimension of the imaginary coastline, then Horton's Law on the bifurcation ratios of streams, and finally Central Place Theory using Ripley's K for the places drawn on the map of Allestone. The results were inconclusive with regards to the ability of these statistics to expose Allestone as an imagined island. Simply put, the drawing of child prodigy Malkin followed geographic principles to a degree where differences remained within the expected margins of natural variation. From the viewpoint of the truth theories introduced above, one could state that the geographic characteristics of Allestone are *coherent* with our set of beliefs about geography.

One of the interesting decisions taken when creating Semantic Web technologies and Linked Data, was to accept the AAA(AA) slogan that anyone can say anything about any topic (at any place and at any time) as a fundamental characteristic of an open and distributed knowledge graph. Linked Data is about *statements*, whether these are true or not is explicitly left out of scope. This decision turned out to be a key strength of Linked Data, also in the realm of geographic information [4]. Contradictory statements, e.g., about the population and spatial extent of a region, can co-exist side-by-side. Decisions about their usage and importance can be made by the user at runtime. This follows the insight that in a graph of knowledge bases, global inconsistencies are acceptable as long as local branches, also called micro-theories or contexts, remain consistent [12].

We propose to apply the same concepts to alternative facts and regard them as sets of statements, irrespective of whether they are 'true' or not. From a knowledge representation and reasoning perspective, this can be modeled as a graph of micro-theories as described above with the alternative facts for any given situation in one micro-theory (af), a second micro-theory containing the competing facts (cf), and a third micro-theory (n) of statements about the world contributed by a variety of (neutral) sources. Following McCarty and Buvac [6], $ist(mt_{af}, p)$ signifies that p is true for micro-theory mt_{af} and $genlMt$ is the generalization relation between two micro-theories such that if $genlMt(mt_n, mt_{af})$, then everything that holds true in mt_n is also true in mt_{af}. Sibling theories can contain contradict, so we would not assert $genlMt(mt_{af}, mt_{cf})$. The question is now what happens if we construct a hierarchy such that both mt_{af} and mt_{cf} rely on mt_n (Figure 1). This is reasonable for two reasons. First, mt_n contains neutral statements about the state of the world, and, second, the key aspect of alternative facts (as opposed to fairy tales) is that they only dispute some statements but otherwise follow common sense. They question crowd sizes, not principles of geography.

Our core idea is as follows, instead of falling into the trap of trying to falsify alternative facts, we want to reverse Brandolini's asymmetry principle. From the viewpoint of mt_{cf}, adding statements from mt_n will result in a (largely) coherent set of statements, while this will not be the case for mt_{af}. In fact, one would have to add (or remove) statements to mt_{af} (or mt_n) to reach coherence. To give a simplified example, imagine that mt_{af} states that the crowd size at event A was larger than at event B and mt_{cf} states the opposite. Imagine further that mt_n contains a variety of statements from different sources such as that according to the institution operating the local metro system, there were more riders during event B than A. Hence, this statement is coherent with mt_{cf} (and other statements in mt_n, e.g., number of taxi rides), and one would have to add a statement to mt_{af} stating that the audience attracted to event A relied to a lesser degree on using the public metro system for traveling. This may, or may not be an actual fact, but given the volume and variety of data today, mt_n contains other statements about spatial patterns extracted from the location of twitter users, booked hotel rooms, and so forth, leading to the problem that one would have to add more and more statements to mt_{af} to keep it alive as an 'alternative'. This reverses Brandolini's asymmetry principle.

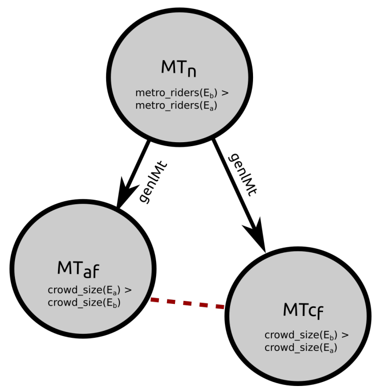

Figure 1: Micro-theories example.

This leads to the question of how one would arrive at statements that can be used in such a system and for our thought experiment, we argue that spatial analysis can play a role here, e.g., by measuring densities, point patterns, impedance, and so forth. Finally, there will be many ways to measure coherence, e.g., counting the number of (atomic) statements that have to be added or removed to keep the micro-theories coherent, computing the *self-information* of each statement, taking into account the changing entailment, i.e., what can be inferred before and after the statements have been added or removed, and so forth.

2. SCENARIOS

In this section we examine the role of these micro-theories and statements through a selection of scenarios. Each of these scenarios present a situation were spatial analysis can be employed to assess competing statements.

2.1 Crowd Size at a Political Event

Imagine a government official making a statement that contrasts crowd sizes at two major political events. If such a statement claims that one of the crowds was substantially larger than the other, one would then assume that this can be verified or falsified in a number of ways, e.g., analysis of photographs. The problem with such an approach is relying on the inexact art of people counting and adjusting for different times, view points, and so on. Consequently, one could rely on proxy information such as metro usage statistics for which exact data are readily available. Metro statistics are not the only source of such information, however. The impact of a large (or small) number of people in a relatively small area has a ripple effect on the amount and types of activities that occur within a city. Typical activity patterns such as recreational cycling or running will be affected by the density of people in a specific area within a city. Similarly, transportation patterns are influenced not only by road closures, but also by requests for taxis and ride-share services. The density of people within a certain region can have a substantial impact on these activities, and this impact can be quantified via analysis of spatial data.

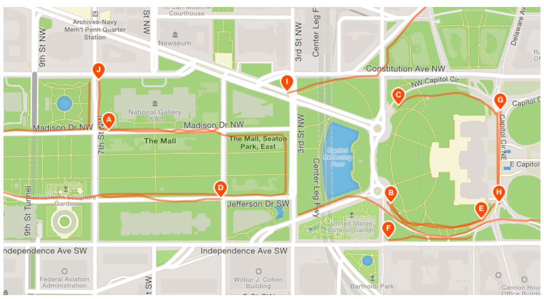

Figure 2: Sample Strava Segments on the National Mall in Washington, D.C.

For example, recreational cyclists or runners post their activities to social fitness tracking applications such as *Strava* (Figure 2) or *Garmin Connect*. A neutral micro-theory (mt_n) might state that an increase in person density within parts of a city would noticeably decrease movement activity within this region. Furthermore, one would expect recreational activities such as running or cycling to occur at a slower (than typical) average speeds within the region. In transportation network analysis, this would be referred to as an increase in impedance along a certain route. Given a micro-theory of competing fact (mt_{cf}) that states that crowd B was bigger than crowd A, this could be explained through a neutral statement that the impedance placed on recreational activities is bigger for a larger crowd than a smaller one and spatial data from social fitness tracking applications may either support or refute this. To support the alternative facts (mt_{af}) that state crowd A was bigger than B, however, one would need to add a new neutral statement, one option being that this increase in impedance was due to construction in the area or that people exercising during event B were slower on average than those during event A.

One could also posit that there would be an increase in *ride-share* pick-ups and drop-offs around a densely populated region with little to no vehicle traffic passing through. These data would lend themselves to convex hull analysis in order to determine the boundary of the region. The competing fact micro-theory could be explained through a neutral statement that the difference in convex hull size between two events reflects a difference in crowd size. The alternative fact micro-theory would need to provide a statement (other than crowd size) to explain why the convex hull was larger for crowd B than crowd A. That may present itself in a statement to the effect that fewer ride-sharing drivers were operating in the region during event A. Finally, social media posts have been used extensively in previous research as proxies for human mobility and activity patterns. Both geo-tagged *spatial* posts, such as tweets or Instagram photographs, and *place-based* check-ins, from applications such as Swarm can be used to approximate the number of individuals and their engagement with a region [3, 7]. These data, while biased in their socio-demographic representation, approximate the distribution of individuals throughout a crowded region with a high number of posts and check-ins coming from higher density parts of the crowd and fewer social media activity in lower density regions.

Returning to coherence theory, we find that as the heterogeneity of geospatial datasets increase and the range of spatial analysis applied to these datasets grows, a coherent view of the phenomena, namely the crowd, becomes apparent. Using spatial analysis to compare spatial data from two political events, we find that the analysis increasingly supports one claim over another. From a micro-theory perspective, the competing facts micro-theory is supported via statements in the neutral micro-theory. The alternative fact micro-theory, requires that additional statements be added to the neutral micro-theory, in order for coherence to exist between them. Furthermore, the complexity and number of statements necessary to accept the alternative fact would be much larger than that of the competing fact theory.

2.2 Tracking Military Equipment Movement

In this scenario, we imagine that ground-based military equipment from one region is transported across a border into another region, employed to disable an aircraft and then transported back across the border to the original region. Statements related to this hypothetical event range in their telling of how the equipment was transported, the region from which it originated and primary party responsible for the equipment. While photographs exist that show the equipment in different regions, various types of spatial analysis can be used to contribute to a better understanding of how the events unfolded. Provided data related to maximum travel velocity of such equipment as well as local and regional road networks, spatial trajectory analysis [11] and routing analysis can be used to calculate all potential routes between regions and the time it would take to transport such equipment. Similarly, existing spatiotemporal research into space-time prisms [8] shows the range of possible locations where the equipment could exist, given photographic evidence and time frame in question. Anecdotal and circumstantial evidence could also be confirmed or rejected based on spatial models constructed from previously confirmed sightings of the equipment. Neutral statements would point to the results of this spatial analysis, namely that given two locations and a known maximum speed, a vehicle's has a maximum area over which it can traverse. The mt_{cf} statement that a piece of equipment was at location X at time T can be confirmed through this neutral statement and the location of a vehicle sighting. A mt_{af} statement that the equipment could not possibly have been at location X would again need to rely on additional statements potentially pointing to alternative modes of transportation.

2.3 Nuclear Reactor Radiation Dissemination

A third scenario may be concerned with the environmental health and political impact of a nuclear reactor meltdown. Imagine a major incident involving a nuclear reactor power plant outside of a coastal city. The outcome of such an event would be felt in a number of ways, both environmentally and politically. It may be in the best interest of a representative of the power plant or government official to react politically, making a statement (mt_{af}) that reduces the perceived severity of such an event. The coherence of these statements could be assessed in a variety of ways, many of which involve analysis of geospatial information. Environment regulations set a threshold for the amount of radiation that is disseminated from a nuclear power plant under normal functioning conditions. One could verify an official's statement via marine buoys that test for radiation levels at specific points around the globe. Given models of macro and micro wind and water flow patterns, e.g., The National Weather Service's Global Real-time Ocean Forecast System, built through geospatial analysis, a neutral statement might indicate that a water parcel starting at location X would move to location Y after time interval T. Based on these models (e.g., Figure 3), a buoy in the water parcel at location Y after time interval T would either show an acceptable or unacceptable level of radiation. Should radiation be discovered at location Y, the alternative facts micro-theory may propose that a statement discrediting the method or equipment used for testing/reporting be added to the neutral fact theory. The competing fact micro-theory that a nuclear reactor meltdown on one side of the globe is responsible for increased radiation in a water parcel on the other side of the globe would point to the neutral statement concerning the geospatial dispersion and flow of water. On a more local scale, spatial analysis tools such as *Darcy Flow Analysis* or watershed delineation models can be used to understand how ground water flows within the vicinity of the nuclear reactor, allowing for prediction of where radioactive water may end up. These spatial analysis techniques are essential for determining neutral statements on which different micro-theory facts can be compared.

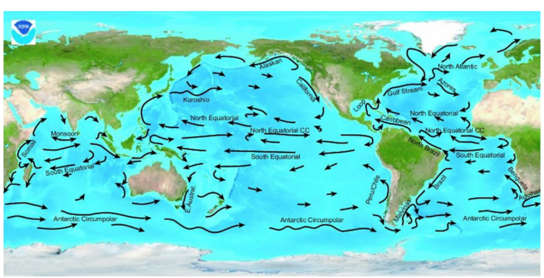

Figure 3: Visualization of major ocean currents throughout the globe. Credit: NOAA.

3. CONCLUSIONS

In this vision paper we proposed how one could apply spatial analysis to translate existing data from heterogeneous sources and across different media formats into quantifiable *statements*, e.g., by estimating the density and convex hull of a crowd from social media data. We argued that instead of trying to falsify alternative facts, which does not scale due to the asymmetry principle, we can make use of the observation that alternative facts, while being self-coherent, either become incoherent when compared to neutral statements about the state of the world or would require additional statements to justify why they are not in line with such neutral statements. This reverses the asymmetry principle. We are, of course, aware that the producers of alternative facts may not care about all this, our thought experiment rather assumes that we will be able to learn which quantifiable statements are valuable, similarly to the approach taken by IBM's Watson, and hence provide the public with a tool to raise question such as why there were less metro riders when there were supposedly more visitors.

4. REFERENCES

[1] Harry G Frankfurt. *On bullshit*. Princeton University Press, 2009.

[2] Michael F Goodchild and Linna Li. Assuring the quality of volunteered geographic information. *Spatial statistics*, 1:110–120, 2012.

[3] Bartosz Hawelka, Izabela Sitko, Euro Beinat, Stanislav Sobolevsky, Pavlos Kazakopoulos, and Carlo Ratti. Geo-located twitter as proxy for global mobility patterns. *Cartography and Geographic Information Science*, 41(3):260–271, 2014.

[4] Werner Kuhn, Tomi Kauppinen, and Krzysztof Janowicz. Linked data-a paradigm shift for geographic information science. In *Int. Conference on Geographic Information Science*, pages 173–186. Springer, 2014.

[5] Paul A Longley, Michael F Goodchild, David J Maguire, and David W Rhind. *Geographic information science and systems*. John Wiley & Sons, 2015.

[6] John McCarthy and Sasa Buvac. Formalizing context (expanded notes). 1997.

[7] Grant McKenzie and Krzysztof Janowicz. Where is also about time: A location-distortion model to improve reverse geocoding using behavior-driven temporal semantic signatures. *Computers, Environment and Urban Systems*, 54:1–13, 2015.

[8] Harvey J Miller. Modelling accessibility using space-time prism concepts within geographical information systems. *International Journal of Geographical Information System*, 5(3):287–301, 1991.

[9] Mark Monmonier. How to lie with maps, 2014.

[10] Karl Popper. *The logic of scientific discovery*. Hutchinson & Co., 1959.

[11] Katarzyna Siła-Nowicka, Jan Vandrol, Taylor Oshan, Jed A Long, Urška Demšar, and A Stewart Fotheringham. Analysis of human mobility patterns from GPS trajectories and contextual information. *International Journal of Geographical Information Science*, 30(5):881–906, 2016.

[12] Ipke Wachsmuth. The concept of intelligence in AI. *Prerational Intelli-gence-Adaptive Behavior and Intelligent Systems without Symbols and Logic*, 1:43–55, 2013.

[13] William Yang Wang. "liar, liar pants on fire": A new benchmark dataset for fake news detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL 2017)*, Vancouver, BC, Canada, July 2017. ACL.