Mapping Rumours and Information Diffusion: Research update 20/08/2016

As described in the research proposal, this project aims to make progress in the mapping and tracking of rumours, both geographically and temporally, and to use the generated information in conjunction with information on stock market prices and trading volumes.

To do this, it is my intention to make use of technology developed by linguists to create early warning systems for the detection of disease outbreaks, earthquakes and other natural phenomena. In particular, I am exploring a system called BioCaster, which has successfully been deployed to detect infectious disease outbreaks. The BioCaster system is an automated system that mines texts from a large number of sources (such as news feeds, Twitter, blogs etc.) and categorises the relevant information in an easily usable format, e.g. by mapping it onto Google maps. The BioCaster system relies on the construction of a so-called ontology, i.e. a list of terms which are deemed relevant to the presence of disease in an individual or geographical. The ontology also allows the system to make intelligent inference based on the generated data.

To adopt this technology for the purpose of rumours in financial and real markets, a similar ontology must be constructed, but with terms relevant to financial and economic variables. I am currently exploring exactly how to construct the ontology and which terms to include. In doing this, I am currently in the process of determining whether textual analysis, already used in the finance literature, may be of help.

Once the relevant ontology has been created, I will have to determine whether the underlying BioCaster system (algorithms etc.) need modification or whether they can be employed largely as is.

Another issue I am working on is the details of the format in which information should be recorded and stored. In the original applications of the technology, the main aim is for policy makers to be able to swiftly detect outbreaks and so be able to deal with these in a timely manner. In the proposed application, the aim is somewhat different, as the idea is to gain insight into how the rumours themselves propagate (rather than to detect any underlying occurrences). It is therefore important to ensure, already at the design stage, exactly how the data will be collected and which data to generate (which is in fact part of the ontology production process). For this reason, I will liaise with a colleague (who may join me in conducting the research) who is an econometrician and has experience in the empirical analysis of financial market data.