Socio-Semantic Knowledge Hypergraphs

Telmo Menezes[1] and Camille Roth[1,2]

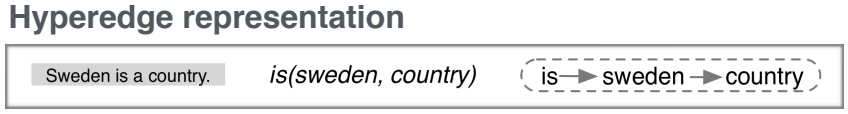
How to study beliefs or claims in digital media?

A recent stream of quantitative studies on social media data made extensive use of the joint analysis of interactions and content. This literature implicitly focuses on socio-semantic knowledge networks, where nodes can be either actors or concepts. While the identification of relationships between users is relatively straightforward, the automatic extraction and structuring of textual information comes however with a number of challenges.
Utterances and content are generally appraised as single words or *n-grams*, and sometimes represented as semantic networks, warranting further analysis using graph-based methods. In this context, traditional semantic networks typically reduce the original empirical data into simple binary relationships between atomic linguistic entities, thereby losing a large part of the underlying information.

Modern Machine Learning and NLP + Hypergraphs

Progress in machine learning approaches now provide us with reasonably **accurate and fast sentence parsers**. The richness of information provided by these parsers cannot be fully captured and analyzed using traditional graph-based network methods. On one hand, **natural language is recursive, allowing for statements about statements**, and on the other, it expresses **n-ary relationships**.
To address this, we propose a model based on the hypergraph formalism. Combined with **deep learning parsing techniques**, we contend that our model is capable of dealing with a richer array of concepts, taking into account the underlying actors, while remaining simple enough to lend itself to large-scale empirical analysis of human communication.

Hyperedge representation

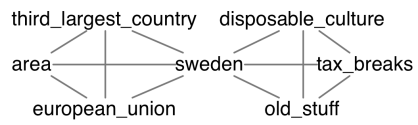


Example

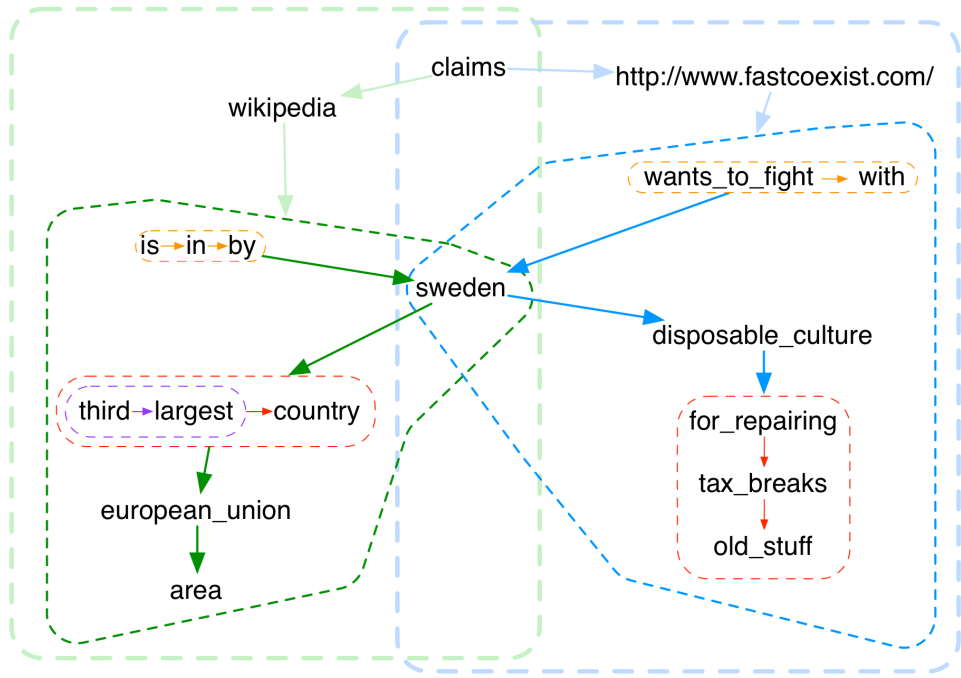
"Sweden wants to fight our disposable culture with tax breaks for repairing old stuff." *(blog post)*

"Sweden is the third-largest country in the European Union by area" *(Wikipedia)*

Co-occurrence graph



Hypergraph



Non-trivial socio-semantic similarities

The hypergraph can be explored to find non-trivial socio-semantic similarities between agents. The five building blocks presented here can be combined to discover commonalities that are beyond what is achievable by conventional n-gram-based approaches. To illustrate:

a) Find two agents *(e.g. agent1=Wikipedia; agent2=Reddit user)* that make the claim that **Sweden is a country** *[X=is→sweden→country]*.

b) Find **claims about Sweden** *[X=sweden; A, B are free variables (e.g. A=is→country, B=has→snow)]*.

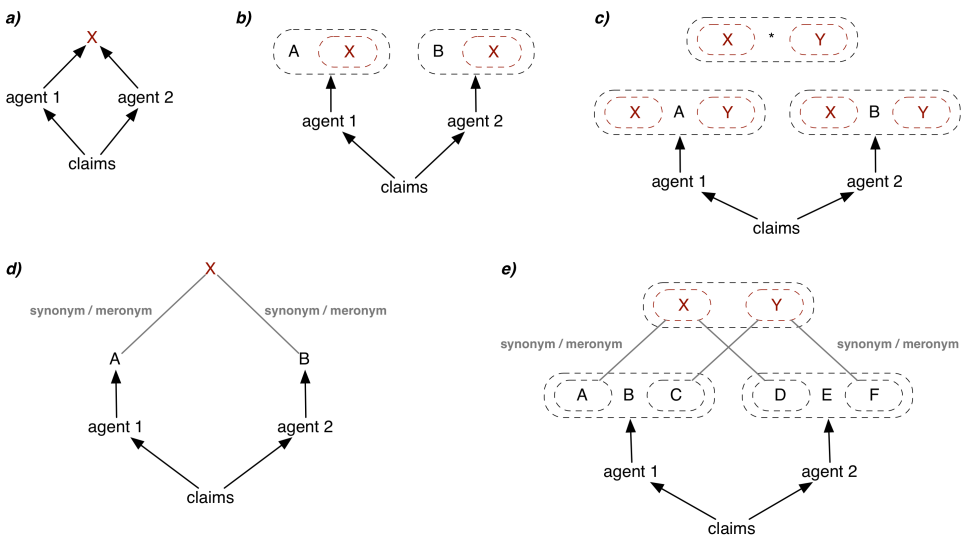
c) Find **claims about entities fighting disposable culture** *[X=fight; Y=disposable_culture; A,B are free variables (e.g. A=sweden, B=mary_poppins)]*.

d) By navigating the underlying ontology, we can connect concepts through abstractions. For example, agents that make **claims about countries in general** *[A, B have type X; X=country]*.

e) Combines patterns and abstraction, and can be used to find a communality of the type: **claims about things opposed by countries** *[A, D have type X; C, F have type Y; X=is_opposed_by; Y=country; B, E are free variables (e.g. B=disposable_culture, E=pollution)]*.

Notice that the five building blocks presented can be composed, leading to more complex patterns of similarity.

Building blocks of hypergraphic socio-semantic commonality



GraphBrain

http://graphbrain.org

GraphBrain is an **open-source research toolbox** created for the purpose of studying **collective human knowledge**. It consists of a number of tools for both social and semantic networks analysis, knowledge retrieval from varying sources, knowledge integration, natural language processing, exploration and visualization.
All of these tools are meant to be building blocks, organized around a central idea of the **recursive hypergraph as a knowledge base**.

[1]*Centre Marc Bloch Berlin (An-Institut der Humboldt Universität, UMIFRE CNRS-MAE), Friedrichstr. 191, 10117 Berlin, Germany*
[2]*Sciences Po, médialab, 84 rue de Grenelle, 75007 Paris, France*

{menezes, roth}@cmb.hu-berlin.de

Web: http://cmb.huma-num.fr