Unsupervised Semantic Field Analysis
by example of the Heat Equation

Special Topic on NETWORKS

Candidate Number: 1072462

Abstract

This work will attempt to

1 Motivation

For non-periodic problem settings, Chebyshev series are a fantastic choice (Biemann 2006).

2 Introduction

Let $\mathbb{N}$ denote the positive integers, so $0 \notin \mathbb{N}$.

2.1 Definition: Undirected Graph

A graph $G = (V, E)$ with vertices V and edges $E \subseteq V \times V$ is undirected if and only if $(v_i, v_j) \in E \Rightarrow (v_j, v_i) \in E \quad \forall\, v_i, v_j \in V$.

2.2 Definition: Adjacency Matrix

Let $A \in \{0, 1\}^{n \times n}$ denote the symmetric adjacency matrix of an undirected graph $G = (V, E)$. Its entries are given by $a_{ij} = \{A\}_{ij} = \mathbb{1}_{(v_i, v_j) \in E}$, so $a_{ij} = 1$ if vertex v_i is connected to v_j and 0 otherwise. By construction, $A = A^T$.

Further let $m := |E|$ and $n = |V|$ denote the number of edges and vertices, respectively. The degree d_i of a vertex $v_i \in V$ is defined by the number of edges connecting to it, so

$$d_i := \deg(v_i) = \Big| \{(v_j, v_k) \in E \mid v_j = v_i\} \Big| ,$$

for an undirected graph $G = (V, E)$. By the Handshaking lemma,

$$\sum_{i=1}^{n} d_i = \sum_{v \in V} \deg(v) = 2m \, .$$

2.3 Definition: Graph Clustering

Let $C = \{C_i \subseteq V\}_{i=1 \ldots n_C}$ denote a clustering of $G = (V, E)$ into $n_C \in \mathbb{N}$ clusters where $C_i \cap C_j = \{\} \; \forall\, i, j \in \{1, ..., n_C\}$ and $\bigcup_{i=1}^{n_C} C_i = V$. Let $s_i \in \{1, ..., n_C\}$ denote the assigned cluster of vertex $v_i \in V$.

2.4 Definition: Modularity

For a given undirected graph $G = (V, E)$ and clustering C, let

$$Q := \frac{1}{2m} \sum_{i=1}^{n} \sum_{j=1}^{n} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \delta(s_i, s_j) \, ,$$

with $\delta(\cdot, \cdot)$ the Kronecker delta indicating whether two vertices v_i and v_j belong to the same cluster (Grindrod and Lambiotte 2022).

Modularity is a measure of the quality of a clustering (also referred to as a partitioning) of G. It can also be written as

$$Q = \frac{1}{2m} \sum_{c=1}^{n_C} \left[\sum_{v_i, v_j \in C_i} \left(A_{ij} - \frac{d_i d_j}{2m} \right) \right] .$$

3 Clustering Algorithms

3.1 Chinese Whispers

Biemann 2006.

3.2 Watset

Ustalov et al. 2019.

3.3 Fiedler

Fortunato 2010.

3.4 Louvain

Grindrod and Lambiotte 2022.

4 Discussion and Outlook

Bla

References

Biemann, Chris (June 2006). 'Chinese whispers: an efficient graph clustering algorithm and its application to natural language processing problems'. In: *TextGraphs: Proceedings of the First Workshop on Graph Based Methods for Natural Language Processing*. Association for Computational Linguistics, pp. 73–80. DOI: 10.5555/1654758.1654774.

Fortunato, Santo (2010). 'Community detection in graphs'. In: *Physics Reports* 486.3, pp. 75–174. ISSN: 0370-1573. DOI: https://doi.org/10.1016/j.physrep.2009.11.002.

Grindrod, Peter and Renaud Lambiotte (5th Nov. 2022). *C5.4 Networks*. Lecture Notes for the course.

Ustalov, Dmitry, Alexander Panchenko, Chris Biemann and Simone Paolo Ponzetto (Sept. 2019). 'Watset: Local-Global Graph Clustering with Applications in Sense and Frame Induction'. In: *Computational Linguistics* 45.3, pp. 423–479. ISSN: 0891-2017. DOI: 10.1162/coli_a_00354.

A Appendix Things?