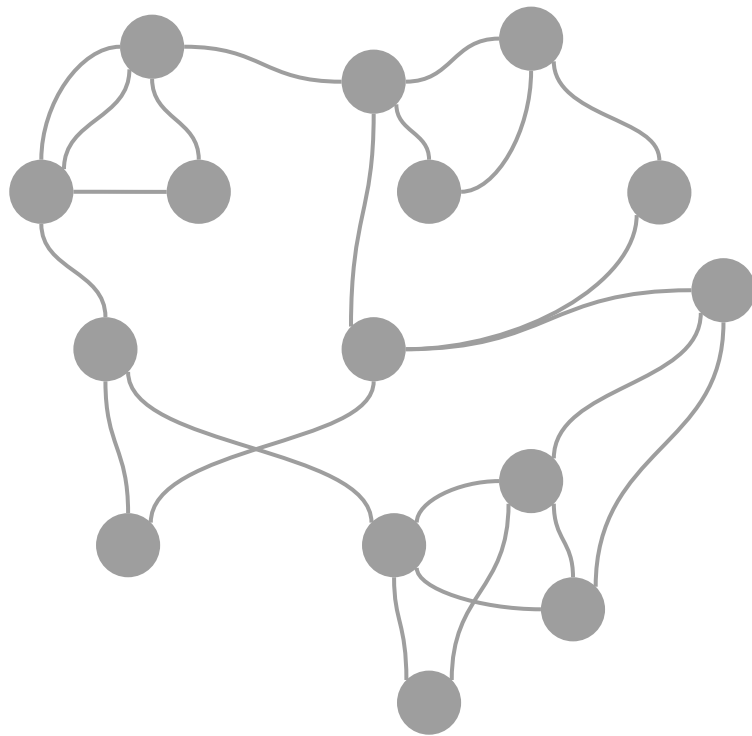# Enriched semantic graphs for extractive text summarization

Antonio F. G. Sevilla, Alberto Fernández-Isabel & Alberto Díaz
16 Sep. 2016

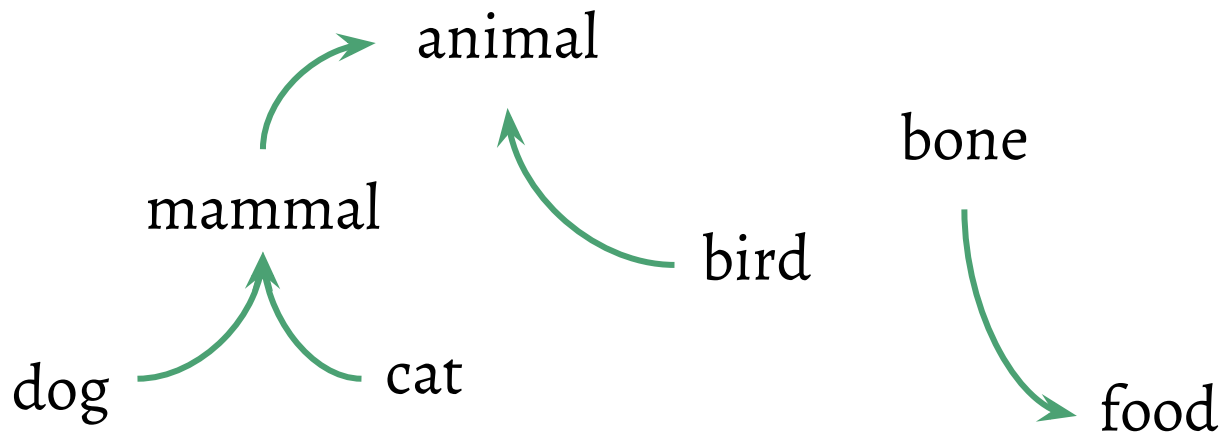# Representation of meaning

- AI processing requires understanding.
- This requires a good model for knowledge.
- In text processing, how to represent meaning?

# Semantic graph: Free-form, linked network of concepts

# Basic Semantic Graph

- ❏ Nodes = concepts (content words)
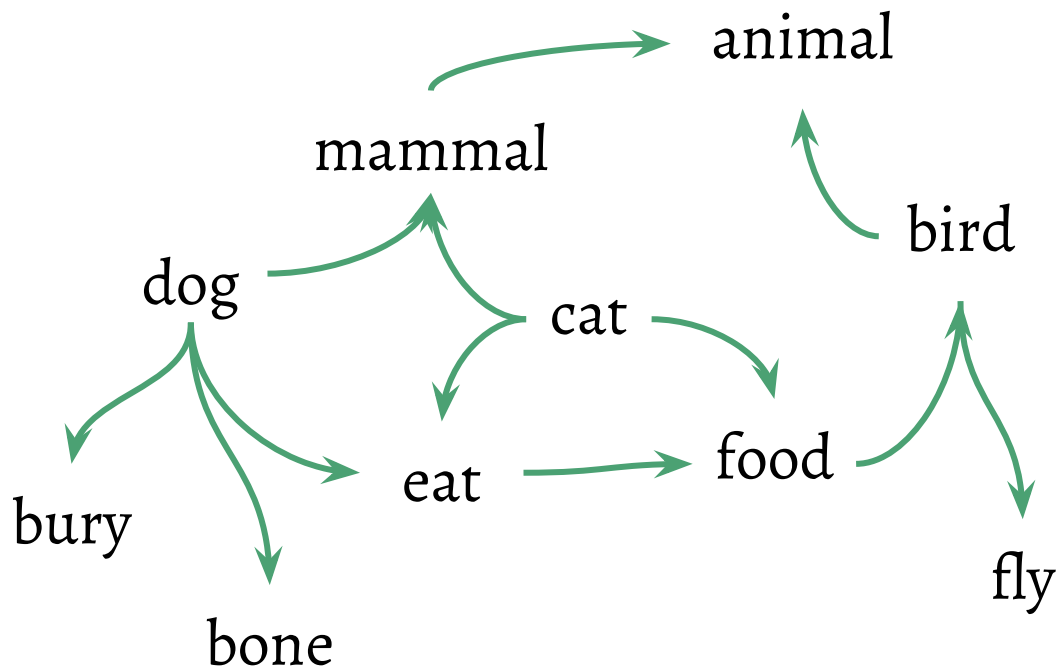- ❏ Edges = lexical relations (from a database)
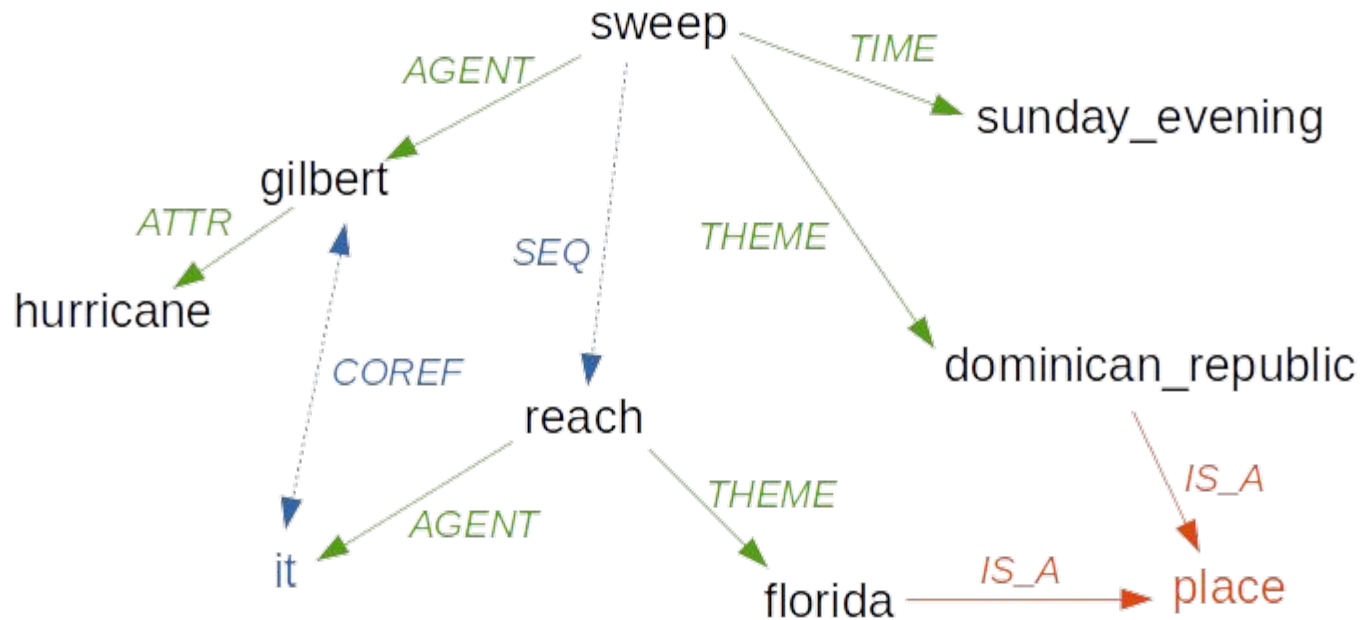
# Enriched Semantic Graph

- ❏ Grammar edges (subject, object)
- ❏ Similarity edges
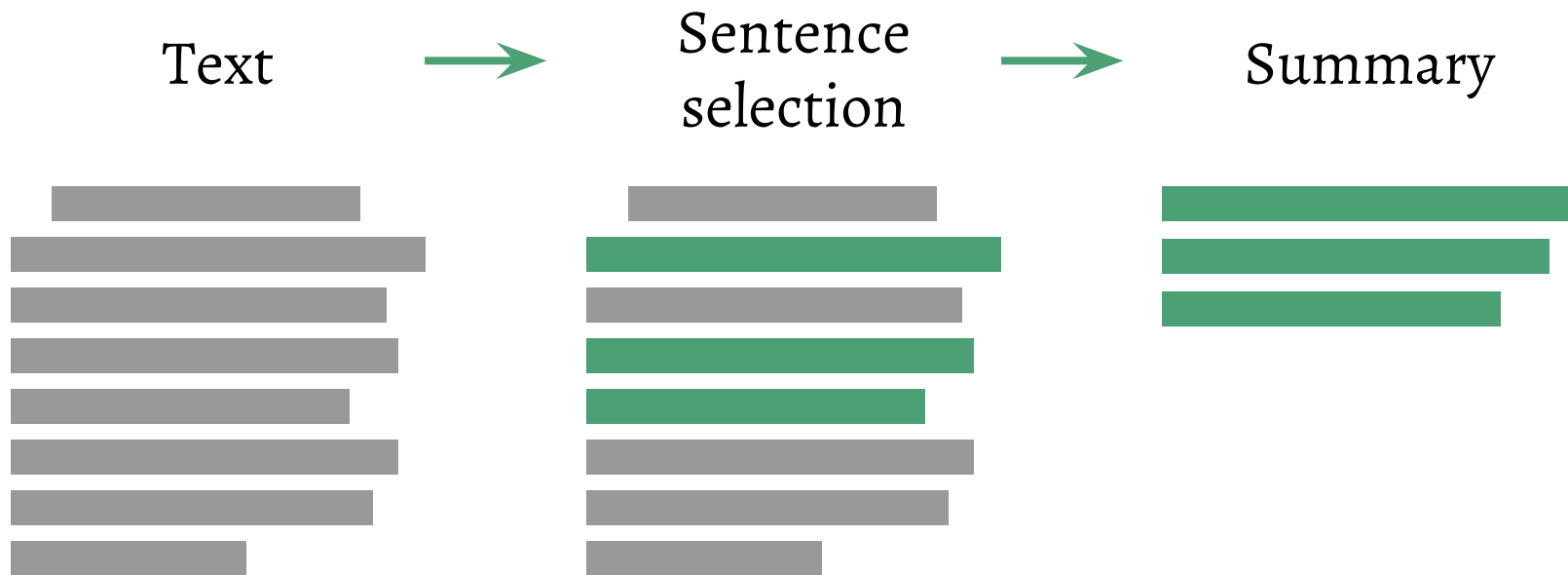- ❏ Node attributes (gram. features)

Expert knowledge:
- ❏ Lexical databases
- ❏ analysis tools

*"Hurricane Gilbert swept the Dominican republic Sunday evening. It then reached Florida."*

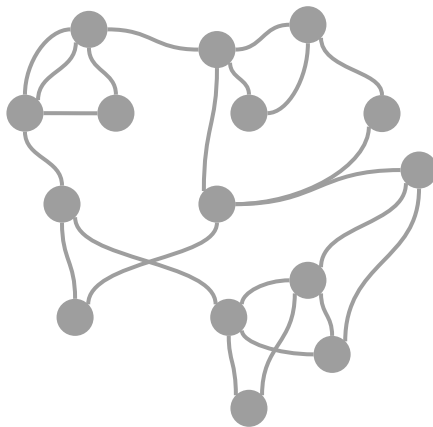# Use case: Extractive Summarization

Text $\rightarrow$ Sentence selection $\rightarrow$ Summary
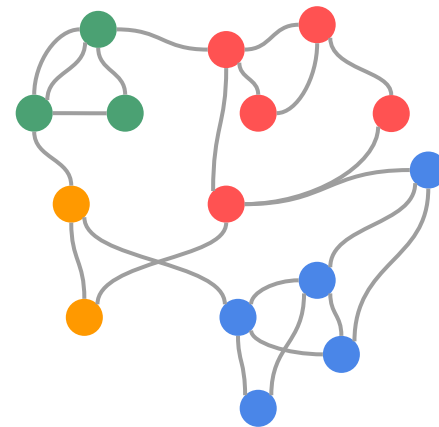
# Topic extraction

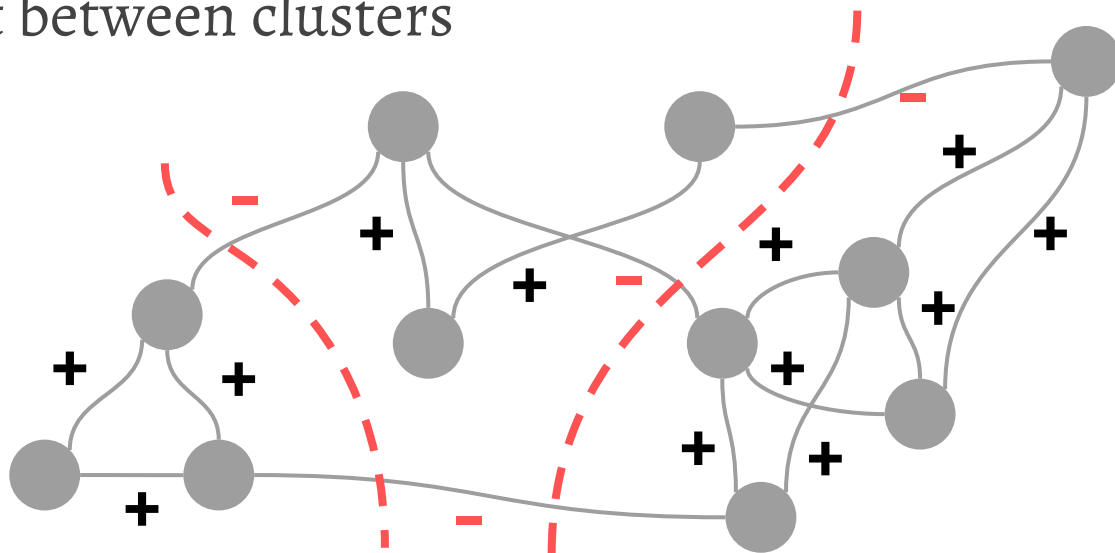Text $\rightarrow$ Semantic graph $\rightarrow$ Clusters = topics

Connections between nodes thanks to the enrichment process tie together related words, creating clusters of topics in the text.
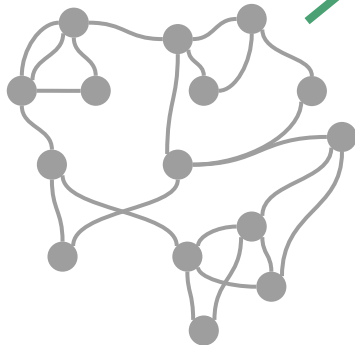
# Clustering

- Degree based method
- Maximize edge weight inside each cluster
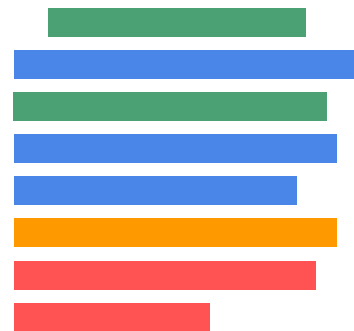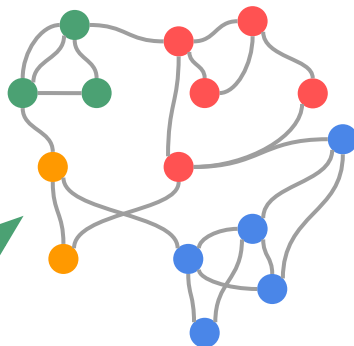- Minimize weight between clusters

# Full process

Text

Clusters

Summary

Graph

Scores

# Evaluation



Text

Human summary

Automatic summary

0.67

ROUGE measure (sequence of words recall)
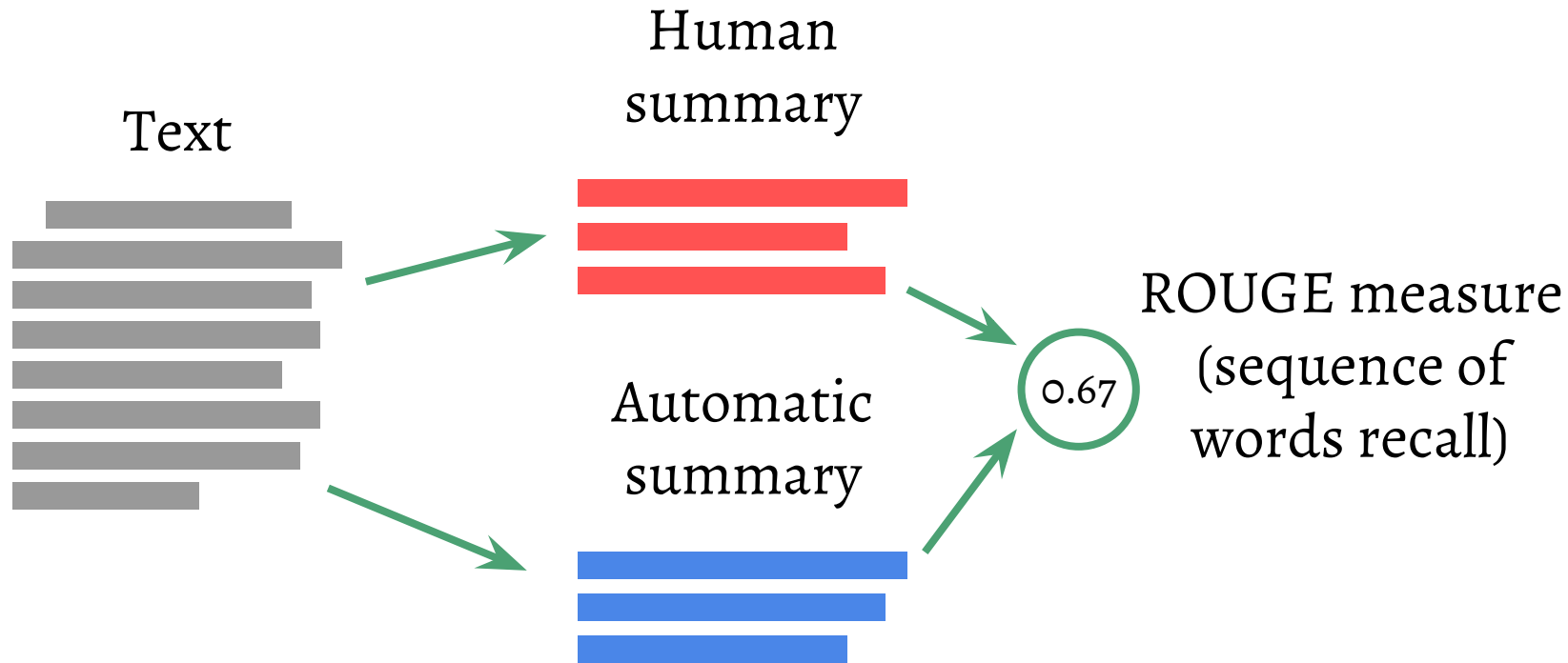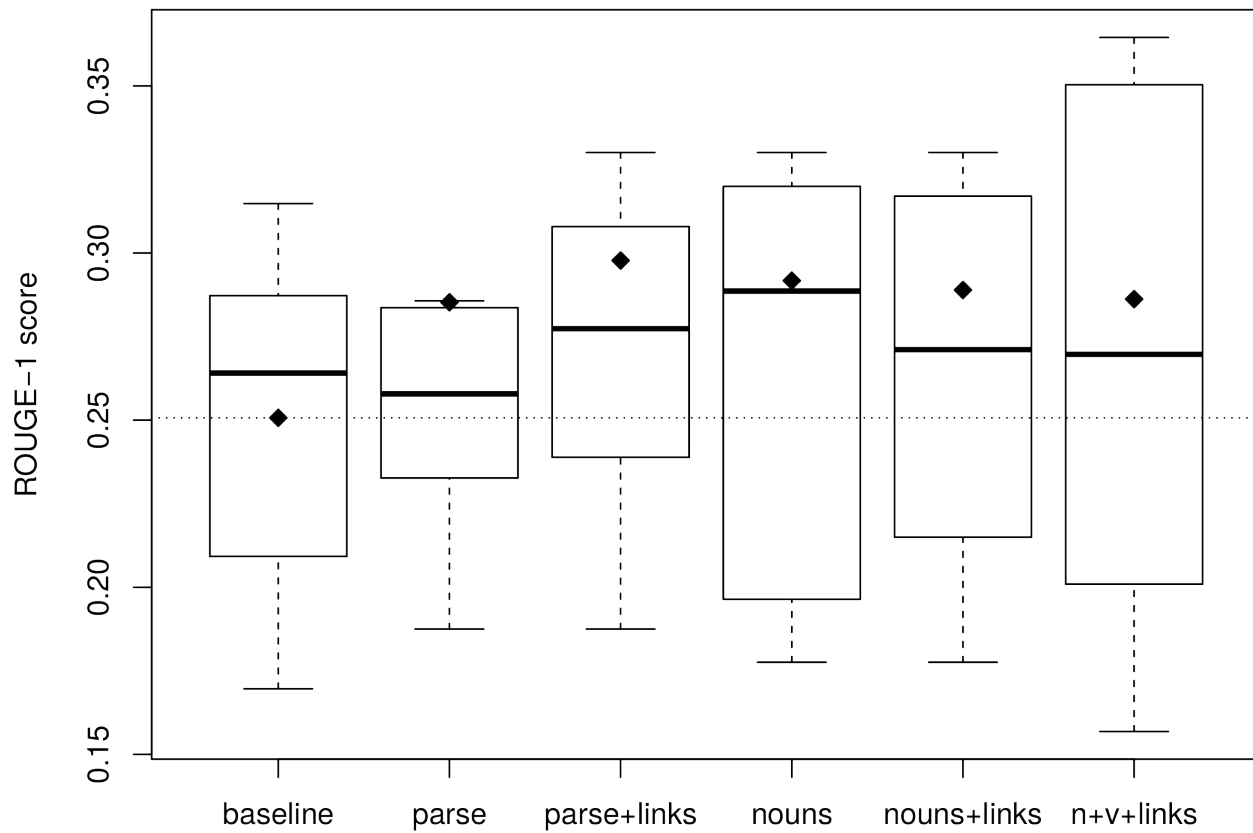
# Results

# Discussion
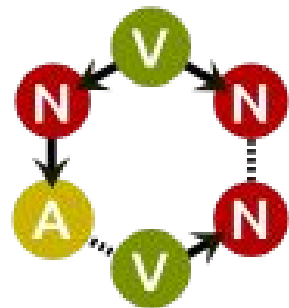
- More information -> More connectedness -> Better clustering
- But asymptotic improvement
- News text baseline is hard
- Extractive approach doesn't benefit much
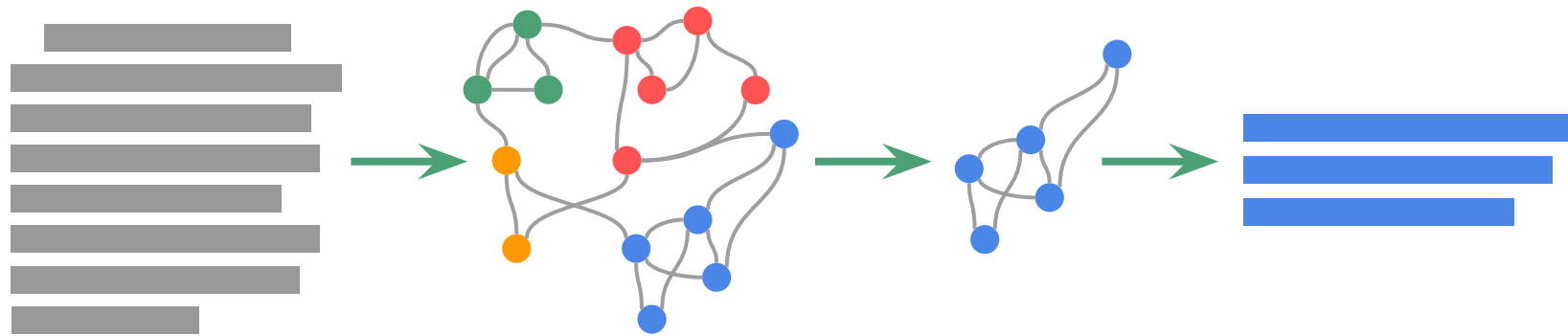- Heterogeneity is not so good for clustering

# Another Result: Grafeno

- ❏ Semantic graph library
- ❏ Python tools for creation and processing of semantic graphs
- ❏ Open-source
- ❏ To be presented in ICDIM, Porto, 19-21 Sept.

# Future Work

- ❏ Better graphs, better clustering
- ❏ Abstractive summarization: natural language generation

# FIN

Antonio F. G. Sevilla
afgs@ucm.es

Alberto Fernández-Isabel
afernandezisabel@ucm.es

Alberto Díaz
albertodiaz@fdi.ucm.es

NiL - Natural Interaction based on Language
http://nil.fdi.ucm.es

Grafeno - semantic graph library
http://github.com/agarsev/grafeno

# Sentence selection

1. Topic extraction
2. Relevance scoring
3. Final score
   (e.g only the main topic)
4. Choose top sentences up to summary length

| Sentence | Topic 1 | Topic 2 | Topic 3 | Final | Use? |
|----------|---------|---------|---------|-------|------|
| 1 | 0.1 | 0.3 | 0.2 | 0.1 | |
| 2 | 0.3 | 0.0 | 0.1 | 0.3 | ✔ |
| 3 | 0.0 | 0.0 | 0.9 | 0.0 | |
| 4 | 0.7 | 0.4 | 0.2 | 0.7 | ✔ |
| 5 | 0.6 | 0.6 | 0.3 | 0.6 | ✔ |
| 6 | 0.0 | 0.1 | 0.2 | 0.0 | |