Semantic Propagation on Contextonyms using SentiWordNet.

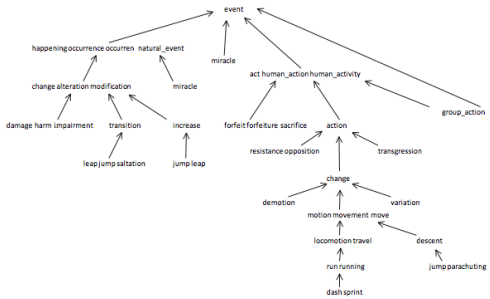
Ovidiu Şerban[⋆,⋆⋆]
Alexandre Pauchet[⋆], Alexandrina Rogozan[⋆], Jean-Pierre Pecuchet[⋆]

[⋆]LITIS Laboratory, INSA Rouen
Saint-Etienne-du-Rouvray, France
{forename.surname}@insa-rouen.fr

[⋆⋆]Department of Computer Science
Faculty of Mathematics and Computer Science
"Babeş-Bolyai" University,
Cluj-Napoca, Romania

15 November 2012

WordNet



WordNet Extensions

- WOLF (WordNet Libre du Français), MultiWordNet, EuroWordNet, BalkaNet
- WordNet++, ExtendedWordNet, eXtended WordNet
- **SentiWordNet**, WordNetAffect

WordNet

"WordNet is a lexical database for the English language.It groups English words into sets of synonyms called synsets, provides short, general definitions, and records the various semantic relations between these synonym sets."
(Source: Wikipedia)

SentiWordNet

"SentiWordNet is a lexical resource for opinion mining. SentiWordNet assigns to each synset of WordNet three sentiment scores: positivity, negativity, objectivity."

(Source:http://sentiwordnet.isti.cnr.it/**)**

- covers most of the synsets of WordNet 3.0
- maintained by (Baccianella, Esuli, Sebastiani) Group, at Istituto di Scienza e Tecnologie dell'Informazione, Consiglio Nazionale delle Ricerche, Italy
- generated using a set of manually annotated seeds, and using sentiment propagation on WordNet

The problem with SentiWN ?

The disagreement level:

Dictionary	Dis.	Obs.	Cnt.	Price[1]
MPQA	27%	+/-	8,221	30 $
Op. Lexicon	25%	+/-	6,789	-
Inquirer	23%	+/-	11,788	Free
LIWC	25%	Categ.	4,500	90 $

⋆ Christopher Potts, *Sentiment Tutorial on sentiment analysis*, 2011

⋆⋆ SentiWordNet (WordNet) Count: **117,659**

[1] For research purpose

SentiWordNet Example

heart **heart#1** bosom#5: the locus of feelings and intuitions; "in your heart you know it is true", -0.125

From context to contextonyms

Contextonyms

def. The contextonyms are maximal ***cliques*** extracted in a word co-occurrence graph

def. Maximal ***cliques*** are the largest complete sub-graphs that can be found in a certain graph

- Introduced by Ji et al. 2003, among with certain filtering techniques (children co-occurrences, final nodes co-occurrences)

Subtitles dataset

- OpenSubtitles.org dataset: 53,384 unique movie files
- 86,276 words maintained after filtering (freq $>$ 0.01 %)
- 3,948,359 co-occurrences maintained after filtering (freq $>$ 0.01 %)

Problems ?!?

DDMCE for Clique Extraction

- The Dynamic Distributable Maximal Clique Exploration Algorithm
- A new distributable version of the Bron-Kerbosch Algorithm, with (Tomita et al.) strategy for node selection, which can be used for dynamic data
- 354,109 conflictual clique found

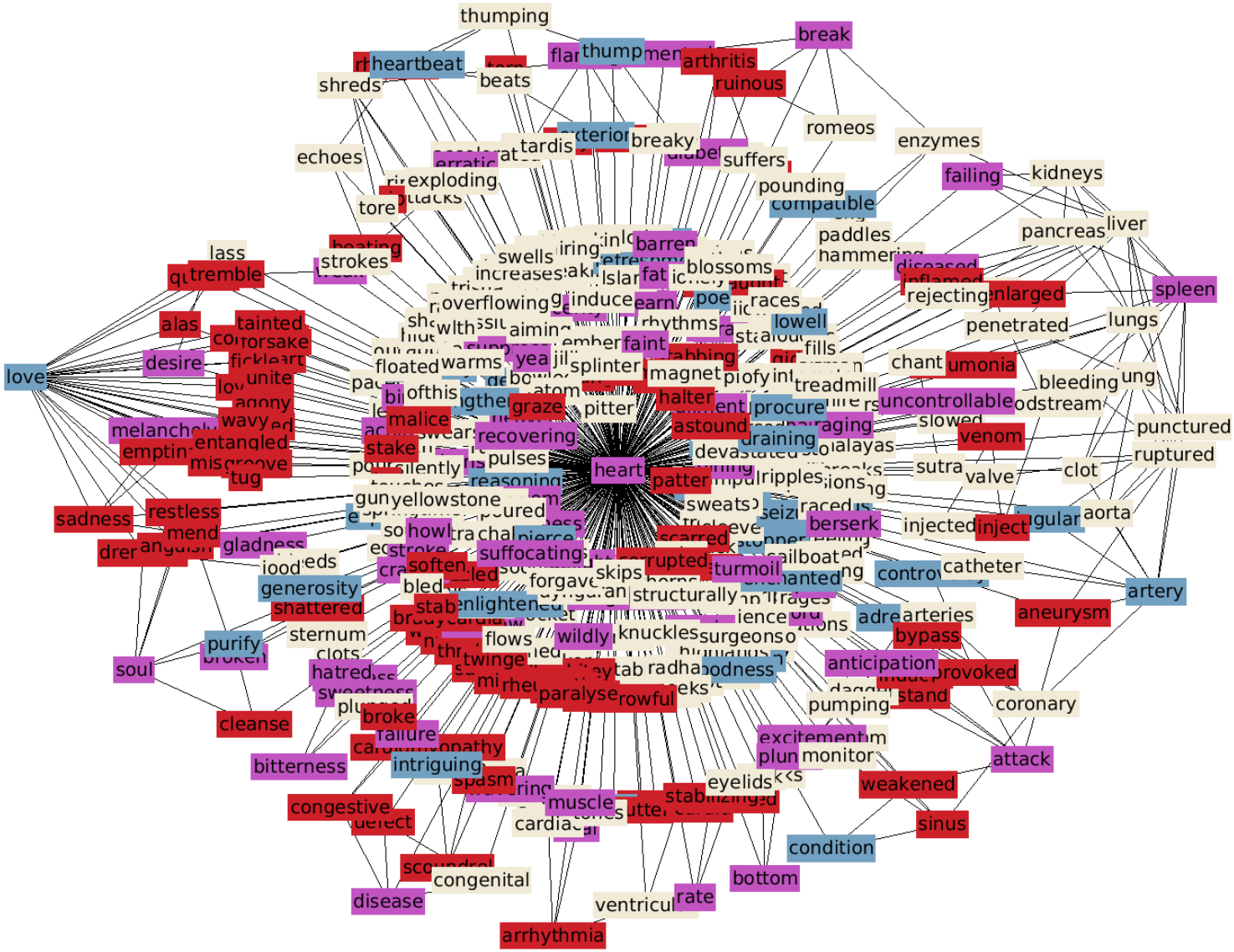
SentiWordNet Conflicts

heart spirit#8 **heart#6**: an inclination or tendency of a certain kind; "he had a change of heart", +0.5

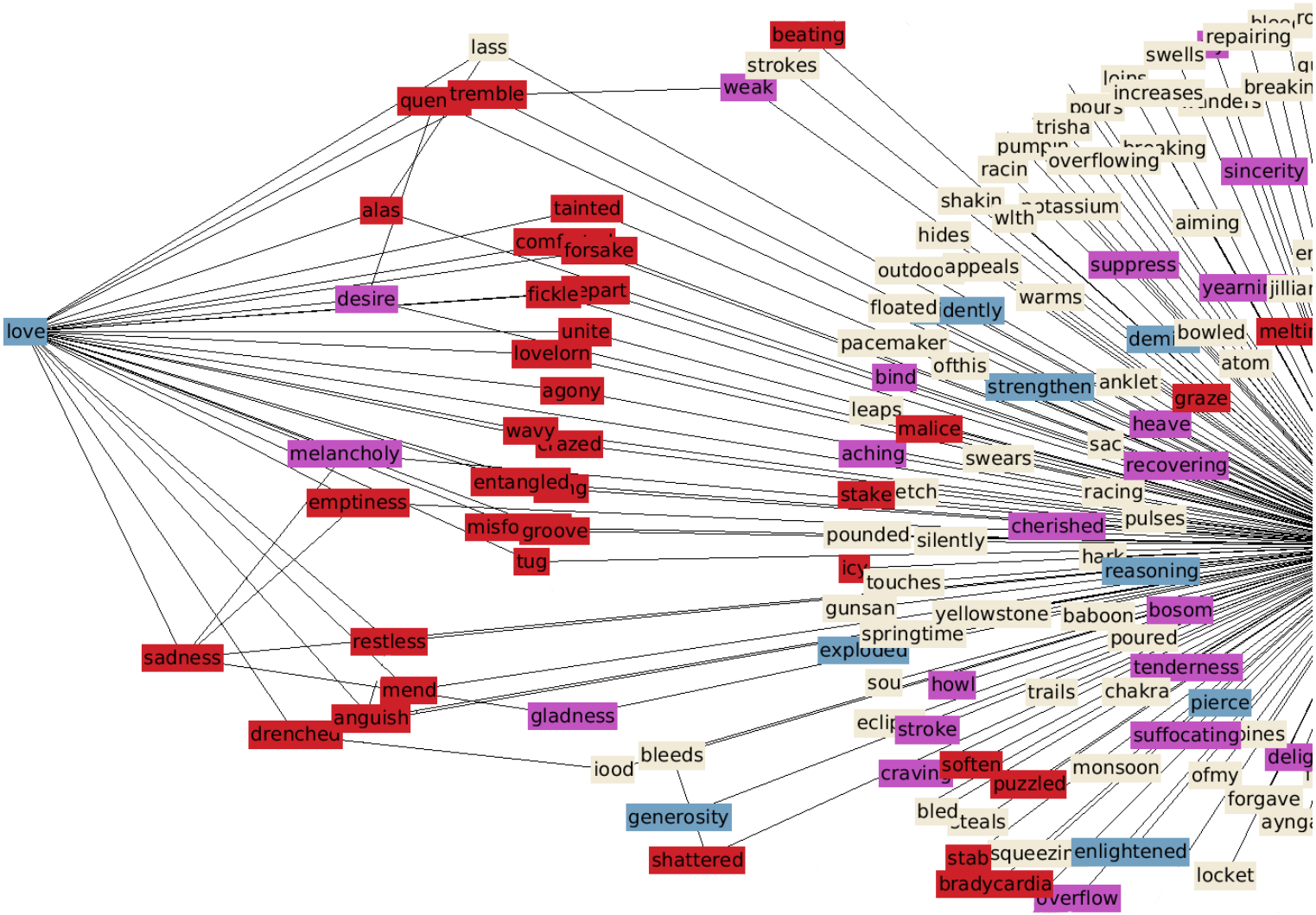
heart **heart#1** bosom#5: the locus of feelings and intuitions; "in your heart you know it is true", -0.125

heart spunk#2 nerve#2 mettle#1 **heart#3**: the courage to carry on; "you haven't got the heart for baseball", +0.25 -0.25

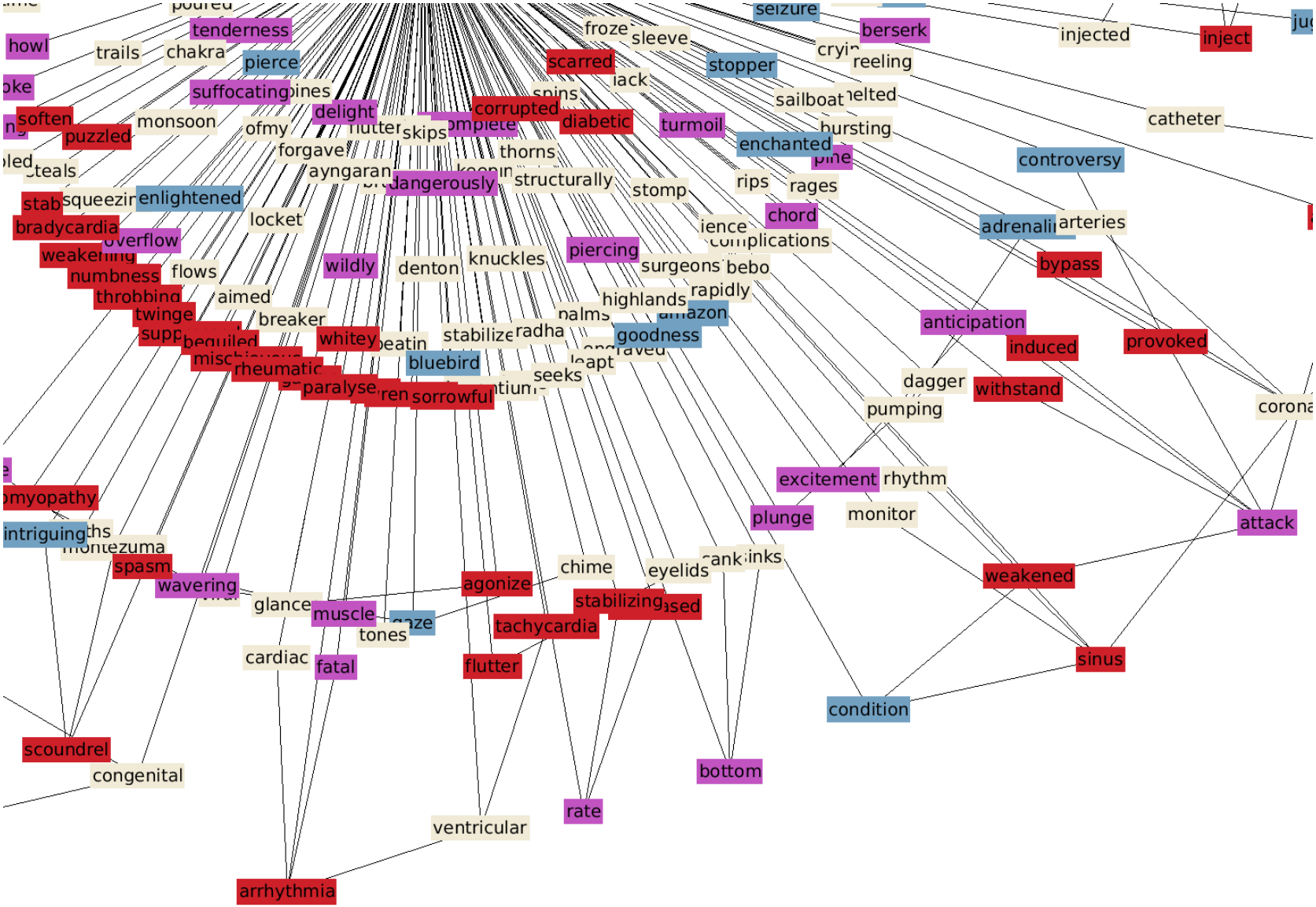
From context to contextonyms



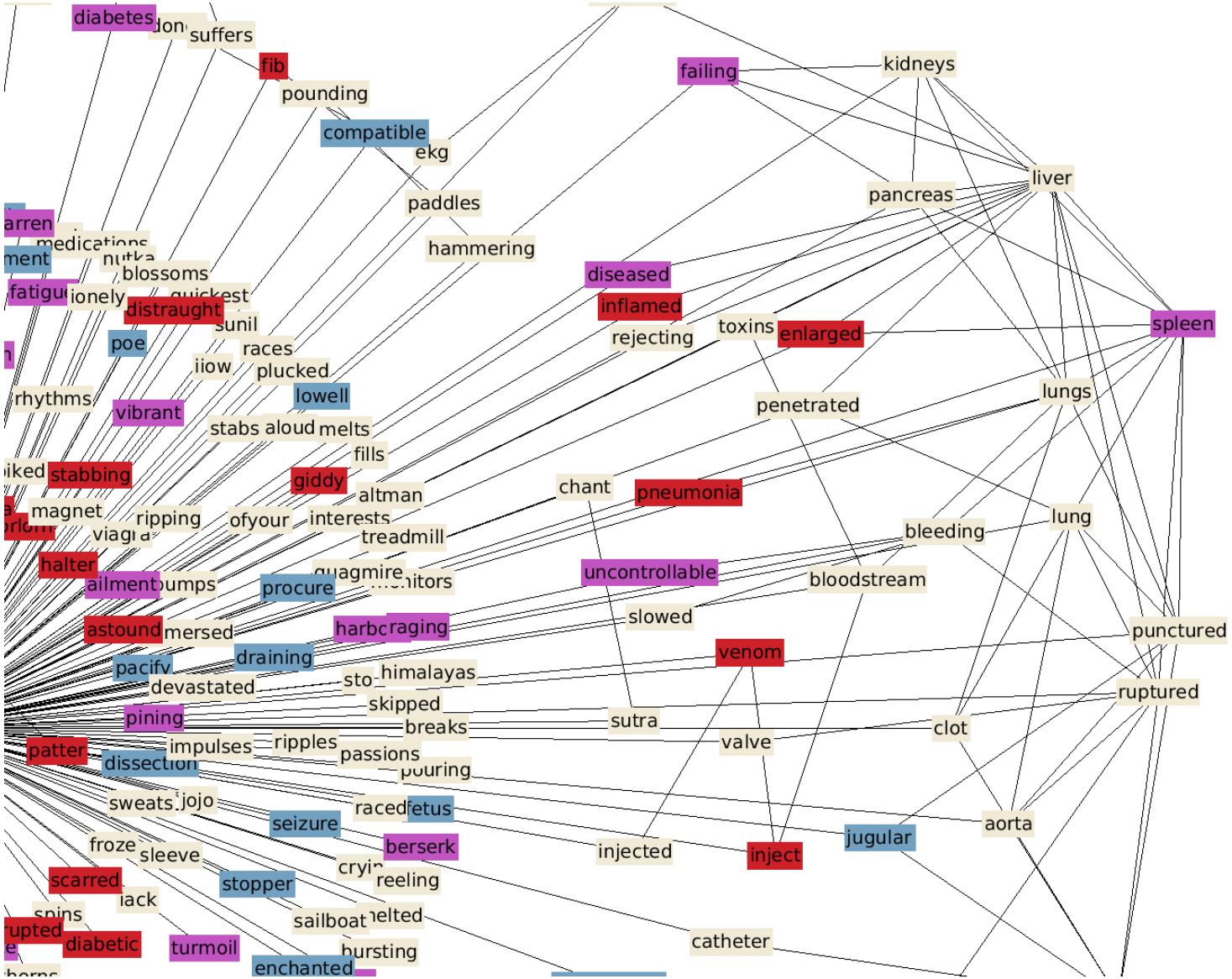
From context to contextonyms (more)



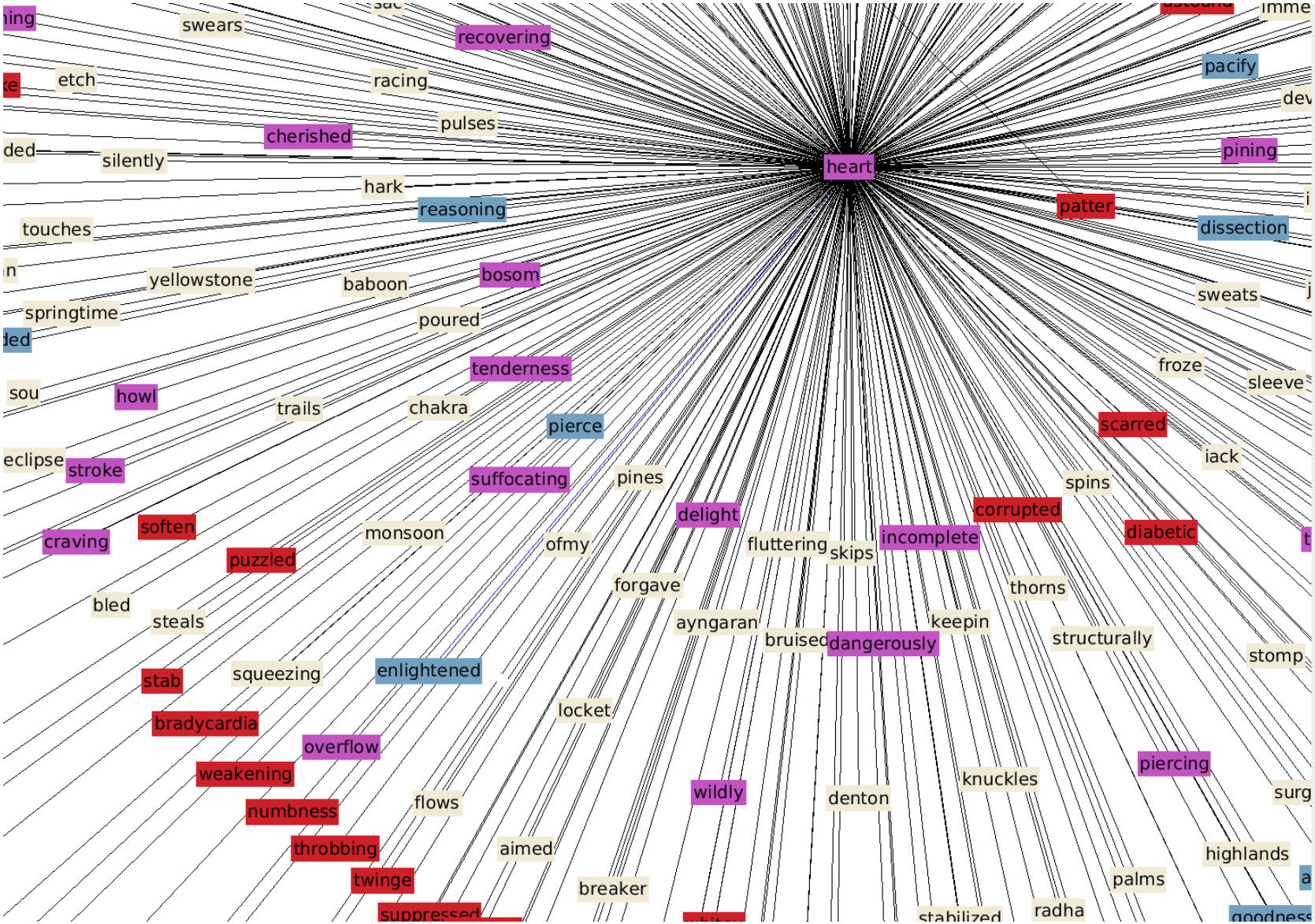
From context to contextonyms (more)



From context to contextonyms (more)



From context to contextonyms (more)

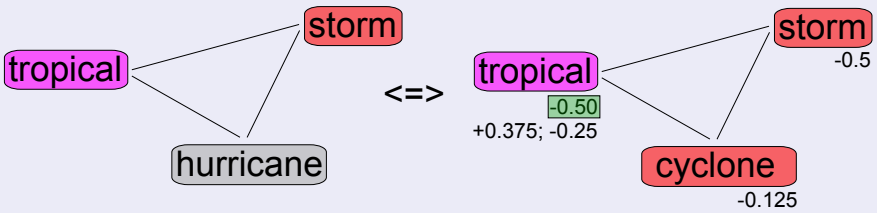


Usage example

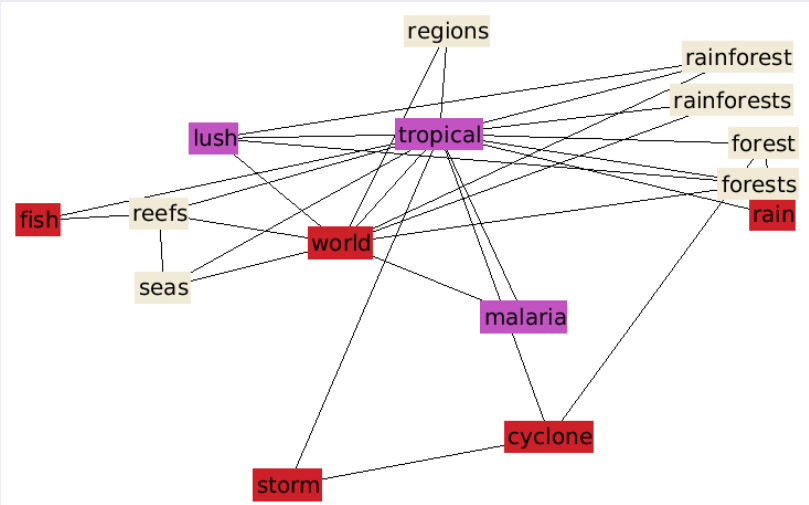
SemEval'07, task 14 Corpus

Hurricane Paul Weakens To Tropical Storm

hurricane paul weakens tropical storm



"Tropical" contextonyms



Clique Ranking

Given: ph **a set of words corresponding a phrase**

$$\forall q \in Q,\ R(q, ph) = \frac{f(q \cap ph) - f(q \setminus ph)}{f(ph)} \quad (1)$$

Where $f(X)$ **represents the combined frequency of the set X**

Conclusion and perspectives

Conclusion

- The contextonyms solve some of the conflicts found in SentiWordNet
- Preliminary validation shows encouraging results

Perspectives

- This approach needs to be validated on large data
- Context is domain/corpus dependent. A multi-domain model and validation needs to be conducted

Merci pour votre attention !