

Invariants and Variability of Synonymy Networks: Self Mediated Agreement by Confluence

Benoît Gaillard
CLLE-ERSS, CNRS
University of Toulouse
Toulouse, France

benoit.gaillard@univ-tlse2.fr

Bruno Gaume
CLLE-ERSS, CNRS
University of Toulouse
Toulouse, France

bruno.gaume@univ-tlse2.fr

Emmanuel Navarro
IRIT
University of Toulouse
Toulouse, France

navarro@irit.fr

Abstract

Edges of graphs that model real data can be seen as judgements whether pairs of objects are in relation with each other or not. So, one can evaluate the similarity of two graphs with a measure of agreement between judges classifying pairs of vertices into two categories (connected or not connected). When applied to synonymy networks, such measures demonstrate a surprisingly low agreement between various resources of the same language. This seems to suggest that the judgements on synonymy of lexemes of the same lexicon radically differ from one dictionary editor to another. In fact, even a strong disagreement between edges does not necessarily mean that graphs model a completely different reality: although their edges seem to disagree, synonymy resources may, at a coarser grain level, outline similar semantics. To investigate this hypothesis, we relied on shared common properties of real world data networks to look at the graphs at a more global level by using random walks. They enabled us to reveal a much better agreement between dense zones than between edges of synonymy graphs. These results suggest that although synonymy resources may disagree at the level of judgements on single pairs of words, they may nevertheless convey an essentially similar semantic information.

1 Introduction

More and more resources exist, built with various approaches and methods and with many different aims and intended uses. A new issue raised by this

growth is that of comparing various resources. A lexical resource is usually based on semantic judgements about lexical elements (a human judgement performed by a lexicographer, or a machine-based judgement in the case of automatically built resources). Often, two independently built resources that describe the same linguistic reality only show a weak agreement even when based on human judgements under the same protocol (Murray and Green, 2004).

Many of such resources, such as WordNet (Fellbaum, 1998) or Wiktionary¹ (Zesch et al., 2008; Sajous et al., 2010) can be modelled as graphs. A graph encodes a binary relation on a set V of vertices. A graph $G = (V, E)$ is therefore defined by a finite, non empty set of $n = |V|$ vertices and by a set $E \subseteq V \times V$ of $m = |E|$ couples of vertices (edges). In the linguistic field, vertices can be various elements of the lexicon: lemmas, word senses, syntactic frames... and edges can describe various relations: synonymy, hyperonymy, translation, co-occurrence... Edges between two vertices can be seen as judgements that decide whether the considered relation applies to this pair. For example, in a synonymy graph, an edge exists between two words if they were judged to be synonyms by the lexicographer who was compiling the dictionary. So, different graphs that model dictionaries of synonyms are built according to the judgements of various “judges”.

We first illustrate, in section 2, how various standard synonymy resources of English and French share common structural properties: they all are Hierarchical Small Worlds (HSW). However, we then

¹<http://www.wiktionary.org/>

show that the synonymy judgements they describe seem to disagree: the Kappa (Cohen, 1960) between the edges of any two such resources remains surprisingly low. In the third section, we analyse this apparent disagreement and in section 4, we address it by proposing an alternative view of the networks, based on random walks. This more global view enables us to assess if disagreeing synonymy networks nevertheless concord at a more global level, because they model the same linguistic reality. Beyond the usual Kappa agreement measure, which is based on the local comparison of two category judgements (a pair is or is not a pair of synonyms), we can show that synonymy judgements do not essentially diverge on the lexical semantic structure that emerges from them. In the fifth section, we conclude by outlining possible applications and perspectives of this work.

2 Graph modelling of various synonymy resources

In order to study the similarities and variations of lexical resources, let us study a sample of graphs that model several standard synonymy resources. We analyse five standard, general purpose, paper dictionaries of French synonyms²: Bailly (*Bai*), Benac (*Ben*), Bertaud du Chazaut (*Ber*), Larousse (*Lar*), Robert (*Rob*). We also study synonymy relations extracted from the Princeton Word Net (*PWN*) and from the English Wiktionary (*Wik*). The *PWN* synonymy network was built according to the following rule: an edge is drawn between any two words that belong to the same synset. The Wiktionary synonymy network was extracted from Wiktionary dumps³ by methods exposed in (Sajous et al., 2010). Each of these resources is split⁴ by parts of speech (Nouns, Verbs, Adjectives) resulting in three different synonymy graphs, designated, for example for the Robert dictionary, as follows: *Rob_N*, *Rob_V*, *Rob_A*.

²Synonymy relations from each of these dictionaries were extracted by the INALF/ATILF Research Unit and corrected by the CRISCO Research Unit.

³<http://redac.univ-tlse2.fr/lexiques/wiktionaryx.html>

⁴Note that splitting is not necessary. The following work would apply similarly to whole resources.

2.1 Invariants : similar structural properties

Most lexical networks, as most field networks⁵, are Hierarchical Small World (HSW) Networks that share similar properties (Watts and Strogatz, 1998; Albert and Barabasi, 2002; Newman, 2003; Gaume et al., 2010; Steyvers and Tenenbaum, 2005). They exhibit a **low density** (not many edges), **short paths** (the average number of edges L on the shortest path between two vertices is low), a high **clustering rate** C (locally densely connected subgraphs can be found whereas the whole graph is globally sparse in edges), and the distribution of their degrees follows a **power law**. All graphs in our sample exhibit the HSW properties. For example, Table 1 shows the pedigrees of synonymy graphs of verbs (for space reasons we only show results for verbs, results are similar for the two other parts of speech). In this table, n and m are the number of vertices and edges, $\langle k \rangle$ is the average degree of vertices, and λ is the coefficient of the power law that fits the distribution of degrees, with a correlation coefficient r^2 . n_{lcc} and L_{lcc} are the number of vertices and the average path length measured on the largest connected component. Even if n and $\langle k \rangle$ vary across dictionaries, L_{lcc} is always small, C is always higher than for equivalent random graphs (Newman, 2003) and the distribution of degrees remains close to a power law with a good correlation coefficient.

Table 1: Pedigrees of seven synonymy graphs (verbs).

	n	m	$\langle k \rangle$	n_{lcc}	m_{lcc}	C	L_{lcc}	λ	r^2
<i>Bai_V</i>	3082	3648	2.46	2774	3417	0.04	8.24	-2.33	0.94
<i>Ben_V</i>	3549	4680	2.73	3318	4528	0.03	6.52	-2.10	0.96
<i>Ber_V</i>	6561	25177	7.71	6524	25149	0.13	4.52	-1.88	0.93
<i>Lar_V</i>	5377	22042	8.44	5193	21926	0.17	4.61	-1.94	0.88
<i>Rob_V</i>	7357	26567	7.48	7056	26401	0.12	4.59	-2.01	0.93
<i>PWN_V</i>	11529	23019	6.3	6534	20806	0.47	5.9	-2.4	0.90
<i>Wik_V</i>	7339	8353	2.8	4285	6093	0.11	8.9	-2.4	0.94

2.2 Variability : a low agreement between edges

Although all these graphs are HSW, Table 1 shows that the lexical coverage (n) and the number of synonymy links (m) significantly vary across graphs. Given two graphs $G_1 = (V_1, E_1)$ and $G_2 =$

⁵Field networks are networks that model real data gathered by field work, for example in sociology, linguistics or biology. They contrast with artificial networks (deterministic or random).

(V_2, E_2) , in order to compare their lexical coverages, we compute the Recall ($R_\bullet$), Precision ($P_\bullet$) and F-score ($F_\bullet$) of their vertex sets:

$$R_\bullet(G_1, G_2) = \frac{|V_1 \cap V_2|}{|V_2|} \quad P_\bullet(G_1, G_2) = \frac{|V_1 \cap V_2|}{|V_1|}$$

$$F_\bullet(G_1, G_2) = 2 \cdot \frac{R_\bullet(G_1, G_2) \cdot P_\bullet(G_1, G_2)}{R_\bullet(G_1, G_2) + P_\bullet(G_1, G_2)}$$

F-scores of pairs of comparable graphs (same language and same part of speech) of our sample remain moderate. Table 2 illustrates these measures on the eleven pairs of graphs involving the five French synonymy graphs (verbs) and the two English ones. It shows that the lexical coverages of the various synonymy graphs do not perfectly overlap.

Table 2: Precision, Recall and F-score of vertex sets of eleven pairs of graphs. G_1 in rows, G_2 in cols.

		<i>Ben_V</i>	<i>Ber_V</i>	<i>Lar_V</i>	<i>Rob_V</i>	<i>Wik_V</i>
<i>Bai_V</i>	$R_\bullet$	0.66	0.45	0.51	0.40	
	$P_\bullet$	0.76	0.96	0.90	0.95	
	$F_\bullet$	0.71	0.61	0.65	0.56	
<i>Ben_V</i>	$R_\bullet$		0.52	0.58	0.45	
	$P_\bullet$		0.96	0.88	0.93	
	$F_\bullet$		0.68	0.70	0.60	
<i>Ber_V</i>	$R_\bullet$			0.85	0.73	
	$P_\bullet$			0.70	0.82	
	$F_\bullet$			0.77	0.77	
<i>Lar_V</i>	$R_\bullet$				0.68	
	$P_\bullet$				0.92	
	$F_\bullet$				0.78	
<i>PWN_V</i>	$R_\bullet$					0.49
	$P_\bullet$					0.31
	$F_\bullet$					0.38

The value of $F_\bullet(G_1, G_2)$ measures the relative lexical coverage of G_1 and G_2 but does not evaluate the agreement between the synonymy judgements modelled by the graphs' edges. The Kappa of Cohen (Cohen, 1960) is a common measure of agreement between different judges who categorize the same set of objects. In the case of graphs, the judgements are not applied to simple entities but to relations between pairs of entities. Two synonymy graphs $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ give two judgements on pairs of vertices. For example, if a pair $(u, v) \in V_1 \times V_1$ is judged as synonymous then $(u, v) \in E_1$, else $(u, v) \in \overline{E_1}$. To measure the agreement between edges of G_1 and G_2 , one first has to reduce the two graphs to their common vertices:

- $G'_1 = (V' = (V_1 \cap V_2), E'_1 = E_1 \cap (V' \times V'))$;

- $G'_2 = (V' = (V_1 \cap V_2), E'_2 = E_2 \cap (V' \times V'))$;

For each pair of vertices $(a, b) \in (V' \times V')$, four cases are possible:

- $(a, b) \in E'_1 \cap E'_2$: agreement on pair (a, b) , (a, b) is synonymous for G'_1 and for G'_2 ;
- $(a, b) \in \overline{E'_1} \cap \overline{E'_2}$: agreement on pair (a, b) , (a, b) is neither synonymous for G'_1 nor for G'_2 ;
- $(a, b) \in E'_1 \cap \overline{E'_2}$: disagreement on pair (a, b) , (a, b) is synonymous for G'_1 but not for G'_2 ;
- $(a, b) \in \overline{E'_1} \cap E'_2$: disagreement on pair (a, b) , (a, b) is synonymous for G'_2 but not for G'_1 ;

The agreement between the two synonymy judgements of G_1 and G_2 is measured by $K_\uparrow(G'_1, G'_2)$, the Kappa between the two sets of edges E'_1 and E'_2 :

$$K_\uparrow(G'_1, G'_2) = \frac{(p_0 - p_e)}{(1 - p_e)} \quad (1)$$

where:

$$p_0 = \frac{1}{\omega} \cdot (|E'_1 \cap E'_2| + |\overline{E'_1} \cap \overline{E'_2}|) \quad (2)$$

is the relative observed agreement between vertex pairs of G'_1 and vertex pairs of G'_2 , where ω is the number of possible edges⁶ $\omega = \frac{1}{2} \cdot |V'| \cdot (|V'| - 1)$.

$$p_e = \frac{1}{\omega^2} \cdot (|E'_1| \cdot |E'_2| + |\overline{E'_1}| \cdot |\overline{E'_2}|) \quad (3)$$

is the hypothetical probability of chance agreement, assuming that judgements are independent⁷.

The value of agreement on synonymy judgements $K_\uparrow(G'_1, G'_2)$ varies significantly across comparable dictionary pairs of our sample, however it remains quite low. For example: $K_\uparrow(Rob'_V, Lar'_V) = 0.518$ and $K_\uparrow(PWN'_V, Wik'_V) = 0.247$ (cf. Table 3). On the whole sample studied in this work this agreement value ranges from 0.25 to 0.63 averaging to 0.39. This shows that, although standard dictionaries of synonyms show similar structural properties, they considerably disagree on which pairs of words are synonymous.

⁶Here, we do not consider reflexivity edges, that link vertices to themselves, as they are obviously in agreement across graphs and are not informative synonymy judgements.

⁷Note that $K_\uparrow(G'_1, G'_2) = K_\uparrow(G'_2, G'_1)$.

3 Analysis of the disagreement between synonymy networks

When comparing two lexical resources built by lexicographers, one can be surprised to find such a level of disagreement on synonymy relations. This divergence in judgements can be explained by editorial policies and choices (regarding, for example printed size constraints, targeted audiences...). Furthermore, lexicographers also have their subjectivity. Since synonymy is more a continuous gradient than a discrete choice (Edmonds and Hirst, 2002), an alternative limited to *synonym/not synonym* leaves ample room for subjective interpretation. However, these justifications do not account for such discrepancies between resources describing the semantic relations of words of the same language. Therefore, we expect that, if two words are deemed not synonyms in one resource G_1 , but synonyms in another G_2 , they will nevertheless share many neighbours in G_1 and G_2 . In other words they will belong to the same dense zones. Consequently the dense zones (or clusters) found in G_1 will be similar to those found in G_2 . Random walks are an efficient way to reveal these dense zones (Gaume et al., 2010). So, to evaluate the hypothesis, let us begin by studying the similarity of random walks on various synonymy networks.

3.1 Random walks on synonymy networks

If $G = (V, E)$ is a reflexive and undirected graph, let us define $d_G(u) = |\{v \in V / (u, v) \in E\}|$ the degree of vertex u in graph G , and let us imagine a walker wandering on the graph G :

- At a time $t \in \mathbb{N}$, the walker is on one vertex $u \in V$;
- At time $t + 1$, the walker can reach any neighbouring vertex of u , with uniform probability.

This process is called a simple random walk (Bollobas, 2002). It can be defined by a Markov chain on V with a $n \times n$ transition matrix $[G]$:

$$[G] = (g_{u,v})_{u,v \in V}$$

$$\text{with } g_{u,v} = \begin{cases} \frac{1}{d_G(u)} & \text{if } (u, v) \in E, \\ 0 & \text{else.} \end{cases}$$

Since G is reflexive, each vertex has at least one neighbour (itself) thus $[G]$ is well defined. Furthermore, by construction, $[G]$ is a stochastic matrix: $\forall u \in V, \sum_{v \in V} g_{u,v} = 1$.

The probability $P_G^t(u \rightsquigarrow v)$ of a walker starting on vertex u to reach a vertex v after t steps is:

$$P_G^t(u \rightsquigarrow v) = ([G]^t)_{u,v} \quad (4)$$

One can then prove (Gaume, 2004), with the Perron-Frobenius theorem (Stewart, 1994), that if G is connected⁸ (i.e. there is always at least one path between any two vertices), reflexive and undirected, then $\forall u, v \in V$:

$$\lim_{t \rightarrow \infty} P_G^t(u \rightsquigarrow v) = \lim_{t \rightarrow \infty} ([G]^t)_{u,v} = \frac{d_G(v)}{\sum_{x \in V} d_G(x)} \quad (5)$$

It means that when t tends to infinity, the probability of being on a vertex v at time t does not depend on the starting vertex but only on the degree of v . In the following we will refer to this limit as $\pi_G(v)$.

3.2 Confluence in synonymy networks

The dynamics of the convergence of random walks towards the limit (Eq. (5)) is heavily dependent on the starting node. Indeed, the trajectory of the random walker is completely governed by the topology of the graph: after t steps, any vertex v located at a distance of t links or less can be reached. The probability of this event depends on the number of paths between u and v , and on the structure of the graph around the intermediary vertices along those paths. The more interconnections between the vertices, the higher the probability of reaching v from u .

For example, if we take $G_1 = Rob_V$ and $G_2 = Lar_V$, and choose the three vertices $u = \text{éplucher (peel)}$, $r = \text{dépecer (tear apart)}$ and $s = \text{sonner (ring)}$, which are such that:

- $u = \text{éplucher (peel)}$ and $r = \text{dépecer (tear apart)}$ are synonymous in Rob_V : $(u, r) \in E_1$;
- $u = \text{éplucher (peel)}$ and $r = \text{dépecer (tear apart)}$ are not synonymous in Lar_V : $(u, r) \notin E_2$;
- $r = \text{dépecer (tear apart)}$ and $s = \text{sonner (ring)}$ have the same number of synonyms in G_1 : $d_{G_1}(r) = d_{G_1}(s) = d_1$;

⁸The graph needs to be connected for Eq. 5 to be valid but, in practice, the work presented here also holds on disconnected graphs.

- $r = \text{dépecer}$ (tear apart) and $s = \text{sonner}$ (ring) have the same number of synonyms in G_2 : $d_{G_2}(r) = d_{G_2}(s) = d_2$.

Then Equation (5) states that $(P_{G_1}^t(u \rightsquigarrow r))_{1 \leq t}$ and $(P_{G_1}^t(u \rightsquigarrow s))_{1 \leq t}$ converge to the same limit:

$$\pi_{G_1}(r) = \pi_{G_1}(s) = \frac{d_1}{\sum_{x \in V_1} d_{G_1}(x)}$$

as do $(P_{G_2}^t(u \rightsquigarrow r))_{1 \leq t}$ and $(P_{G_2}^t(u \rightsquigarrow s))_{1 \leq t}$:

$$\pi_{G_2}(r) = \pi_{G_2}(s) = \frac{d_2}{\sum_{x \in V_2} d_{G_2}(x)}$$

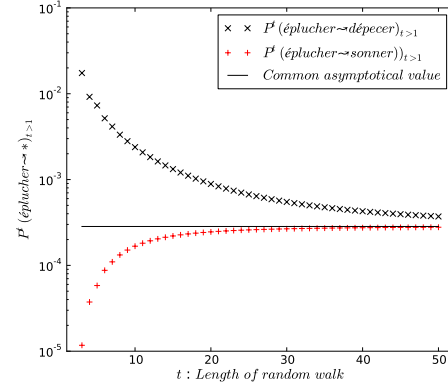
However the two series do not converge with the same dynamics. At the beginning of the walk, for t small, one can expect that $P_{G_1}^t(u \rightsquigarrow r) > P_{G_1}^t(u \rightsquigarrow s)$ and $P_{G_2}^t(u \rightsquigarrow r) > P_{G_2}^t(u \rightsquigarrow s)$ because *éplucher* is semantically closer to *dépecer* than to *sonner*. Indeed the number of short paths between *éplucher* and *dépecer* is much greater than between *éplucher* and *sonner*.

Figure 1(a) shows the values of $P_{G_1}^t(u \rightsquigarrow r)$ and $P_{G_1}^t(u \rightsquigarrow s)$ versus t , and compares them to their common limit. Figure 1(b) shows the values of $P_{G_2}^t(u \rightsquigarrow r)$ and $P_{G_2}^t(u \rightsquigarrow s)$ versus t , and compares them to their common limit. These figures confirm our intuition that, since *éplucher* (peel) and *dépecer* (tear apart) are semantically close, $P_{G_1}^t(u \rightsquigarrow r)$ and $P_{G_1}^t(u \rightsquigarrow s)$ decrease to their limit. We call this phenomenon *strong confluence*. It is worth noting that this remains true even if *éplucher* (peel) and *dépecer* (tear apart) are **not** synonyms in Lar_V . Conversely, since *éplucher* (peel) and *sonner* (ring) are semantically distant, $P_{G_1}^t(u \rightsquigarrow s)$ and $P_{G_2}^t(u \rightsquigarrow s)$ increase to their asymptotic value. We call this phenomenon *weak confluence*.

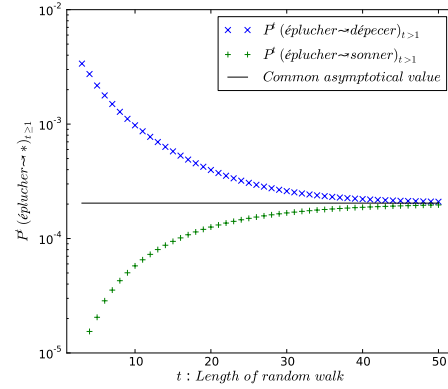
3.3 Correlation of the confluence of disagreeing synonymy pairs

When two graphs G_1 and G_2 disagree on a pair of vertices (a, b) (a is a neighbour of b in one graph but not in the other) there are three possible cases for the strength of the confluence between vertices a and b :

- (1) strong in both graphs (confluence agreement),



(a) $G_1 = Rob_V$



(b) $G_2 = Lar_V$

Figure 1: Confluences between *éplucher* (peel), *dépecer* (tear apart) and *éplucher* (peel), *sonner* (ring) in Rob_V and Lar_V .

- (2) weak in both graphs (confluence agreement),
- (3) strong in one graph, but weak in the other (confluence disagreement).

To contrast cases (1) and (2) from case (3) we measure the correlation between the confluences of disagreeing pairs of two synonymy networks G'_1 and G'_2 . We compare it to this same correlation on two reflexive and undirected *random* graphs $R_{G'_1} = (V', E_1^R)$ and $R_{G'_2} = (V', E_2^R)$ built such that:

$$|E_1^R \cap E_2^R| = |E'_1 \cap E'_2|,$$

$$|E_1^R \cap \overline{E_2^R}| = |E'_1 \cap \overline{E'_2}|,$$

$$|\overline{E_1^R} \cap E_2^R| = |\overline{E'_1} \cap E'_2|,$$

which means that the Kappa agreement between $R_{G'_1}$ and $R_{G'_2}$ is the same as between G'_1 and G'_2 .

For a given $t > 1$ and a set of vertex pairs $X \subseteq V' \times V'$, the correlation of confluences $\Gamma_X(G'_1, G'_2)$ is defined by the Pearson's linear correlation coefficient of the two value tables $(P_{G'_1}^t(u \rightsquigarrow v))_{(u,v) \in X}$ and $(P_{G'_2}^t(u \rightsquigarrow v))_{(u,v) \in X}$.

For all comparable pairs of our sample, we see that disagreeing pairs tend to have a much higher correlation of confluence than disagreeing pairs of equivalent random networks. As an example, for $G_1 = Rob_V$, $G_2 = Lar_V$ and $t = 3$, we have $\Gamma_{E'_1 \cap \overline{E'_2}}(G'_1, G'_2) = 0.41$ and $\Gamma_{\overline{E'_1} \cap E'_2}(G'_1, G'_2) = 0.38$, whereas in the case of the equivalent random graphs the same figures are close to zero.

This suggests that even if graphs disagree on the synonymy of a significant number of pairs, they nevertheless generally agree on the strength of their confluence. In other words, occurrences of cases (1) and (2) are the majority whereas occurrences of case (3) are rare. We propose in the next section an experiment to verify if we can rely on confluence to find a greater agreement between two graphs that disagree at the level of synonymy links.

4 Self mediated agreement by confluence

4.1 Hypothesis: Conciliation reveals structural similarity beyond disagreement of local synonymy

We saw in section 2.2 that the rate of agreement between edges of two standard synonymy networks G'_1 and G'_2 , $K_{\downarrow}(G'_1, G'_2)$, is usually low. However, we have noticed in Section 3.3 that the confluences of pairs on which synonymy graphs disagree are significantly more correlated ($\Gamma \approx 0.4$) than the confluence of equivalent random networks ($\Gamma \approx 0$). This suggests the following hypothesis: synonymy networks are in agreement at a level that is not taken into account by the Kappa measure on edges.

To verify this hypothesis, we try to make each pair of graphs conciliate on the basis of confluence values. We propose a conciliation process by which a graph can accept the addition of another's edges if they do not contradict its structure (i.e. there is a strong confluence value). We then assess if a

strong agreement is found between the two resulting graphs.

Let $G_1 = (V_1, E_1)$ and $G_2 = (V_2, E_2)$ be two synonymy networks, both reflexive, undirected, connected, and a given $t \in \mathbb{N}^*$. We define:

- $G'_1 = (V' = (V_1 \cap V_2), E'_1 = E_1 \cap (V' \times V'))$
- $G'_2 = (V' = (V_1 \cap V_2), E'_2 = E_2 \cap (V' \times V'))$
- $G_1^{(+G_2)} = (V', E_1^+ = E'_1 \cup C_1)$ where
$$C_1 = \{(u, r) \in \overline{E'_1} \cap E'_2 \mid P_{G'_1}^t(u \rightsquigarrow r) > \pi_{G'_1}(r)\} \quad (6)$$
- $G_2^{(+G_1)} = (V', E_2^+ = E'_2 \cup C_2)$ where
$$C_2 = \{(u, r) \in E'_1 \cap \overline{E'_2} \mid P_{G'_2}^t(u \rightsquigarrow r) > \pi_{G'_2}(r)\} \quad (7)$$

$G_1^{(+G_2)}$ and $G_2^{(+G_1)}$ are called *accommodating graphs*. The construction of the accommodating graphs may be metaphorically understood as a conciliation protocol by which two graphs accept proposals of the other that they can reconsider. For example, $G_1^{(+G_2)}$ is the graph G'_1 enriched by edges (u, r) of G'_2 such that there is a *strong confluence* between vertices u and r in G'_1 .

The following property is worth noticing:

Proposition 1. $\forall t \in \mathbb{N}^* :$

$$(E'_1 \cap E'_2) \subseteq (E_1^+ \cap E_2^+) \subseteq (E'_1 \cup E'_2) \quad (8)$$

Proof. By definition, $E_1^+ = E'_1 \cup C_1$ and $E_2^+ = E'_2 \cup C_2$, thus $(E'_1 \cap E'_2) \subseteq (E_1^+ \cap E_2^+)$, furthermore, by definition, $C_1 \subseteq \overline{E'_1} \cap E'_2$ and $C_2 \subseteq E'_1 \cap \overline{E'_2}$ thus $(E_1^+ \cap E_2^+) \subseteq (E'_1 \cup E'_2)$. $\square$

4.2 Experimental protocol

If, for any (G_1, G_2) synonymy resources of the same language, $K_{\downarrow}(G_1^{(+G_2)}, G_2^{(+G_1)})$ is significantly greater than $K_{\downarrow}(G'_1, G'_2)$, then the hypothesis is verified. The conciliation process depends on confluence measures that depend on a given t , the number of steps of the random walk. For $t = 1$, only vertices in the neighbourhood of the starting vertex are reachable. Consequently only pairs of vertices that are edges have a non null confluence. Thus $K_{\downarrow}(G_1^{(+G_2)}, G_2^{(+G_1)}) = K_{\downarrow}(G'_1, G'_2)$ which does not help us to contrast conciliated graphs from

initial binary synonymy graphs. So we fix $t = 2$ the shortest walk length that still yields informative results.

We propose a control experiment that consists in applying the conciliation process to random networks that have the same Kappa as the pairs of synonymy networks. The construction of these random graphs is described above, in section 3.3. We measure the agreement after conciliation of 20 different random graphs. With this control experiment we assess that the observed results are specific to graphs describing the same resource, and not a mere bias of the protocol (let us imagine a protocol whereby one would add all the disagreeing edges to the graphs: not only the Kappa of the pseudo accommodating synonymy graphs would be equal to one, but also the Kappa of pseudo accommodating random graphs, which would disqualify the protocol).

4.3 Results

Table 3 summarizes Kappa and conciliated Kappa values on the pairs of synonymy graphs of verbs. It shows a significant improvement of agreement after conciliation. For example, from a moderate Kappa (0.518) between graphs Rob'_V and Lar'_V (constructed by experts), the conciliation process leads to an excellent Kappa (0.852). Conversely the random networks only increase their agreement by 0.01 (with a very low standard deviation $\sigma < 0.001$). In English, from a poor (0.247) Kappa between PWN'_V (constructed by experts) and Wik'_V (constructed by the “crowds”), the conciliation process leads to a moderate Kappa (0.530), whereas the random networks only marginally increase their agreement (0.004).

Results are similar for other parts of speech. This means that the conciliation process significantly improves the agreement between resources, even if they are originally significantly diverge.

It is interesting to notice that the most similar pairs in terms of edge agreement do not necessarily produce the most agreeing pairs of accommodating graphs. For example, the pair (Bai_V, Rob_V) agrees more than the pair (Bai_V, Lar_V) , whereas for their accommodating graphs, the pair $(Bai_V^{(+Rob_V)}, Rob_V^{(+Bai_V)})$ agrees less than the pair $(Bai_V^{(+Lar_V)}, Lar_V^{(+Bai_V)})$.

Table 3: Kappa (ori.) and accommodating Kappa (**acc.**) values between French and English synonymy graphs (of verbs), compared with the Kappa values between pairs of equivalent random graphs (“ori. r.” and “acc. r.”).

	$K_{\downarrow}$	Ben_V	Ber_V	Lar_V	Rob_V	Wik_V
Bai_V	ori.	0.583	0.309	0.255	0.288	
	acc.	0.777	0.572	0.603	0.567	
	ori. r.	0.583	0.309	0.256	0.288	
	acc. r.	0.585	0.313	0.262	0.293	
Ben_V	ori.		0.389	0.276	0.293	
	acc.		0.657	0.689	0.636	
	ori. r.		0.390	0.276	0.294	
	acc. r.		0.392	0.283	0.301	
Ber_V	ori.			0.416	0.538	
	acc.			0.838	0.868	
	ori. r.			0.417	0.539	
	acc. r.			0.434	0.549	
Lar_V	ori.				0.518	
	acc.				0.852	
	ori. r.				0.518	
	acc. r.				0.529	
PWN_V	ori.					0.247
	acc.					0.540
	ori. r.					0.247
	acc. r.					0.251

So, when G_1 and G_2 are two synonymy graphs of a given language, then they are able to address their local synonymy disagreement and to reach a significantly better agreement. On the other hand, the agreement of random networks does not really improve after conciliation. This proves that the synonymy networks of the same language share specific similar structures that can be detected with the help of confluence measures.

5 Conclusion

Although graphs that encode synonymy judgements of standard semantic lexical resources share similar HSW properties they diverge on their synonymy judgements as measured by a low Kappa of edges. So, one could wonder whether the notion of synonymy is well defined, or if synonymy judgements are really independent. Without directly addressing this question, we nevertheless have shown that strong confluence measures help two synonymy graphs accommodate each others’ conflicting edges. They reach a much better agreement, whereas random graphs’ divergence is maintained. Since the graphs are HSW, they draw clusters of synonyms in which pairs of vertices have a strong confluence.

This suggests two conclusions. First, different synonymy resources that describe the same lexicon reveal dense zones that are much more similar across graphs than the binary synonymy categorisation (the synonym/not synonym alternative). These dense zones convey information about the semantic organisation of the lexicon. Second, random walks and confluence measures seem an appropriate technique to detect and compare the dense zones of various synonymy graphs.

This theoretical work validates the random walk/confluence approach as a potentially valid tool for detecting semantic similarities. This opens many perspectives for applications. For example, it can be used to enrich resources as was done for the Wisigoth project (Sajous et al., 2010). It may also help to merge, or aggregate, resources. If we apply the conciliation process to two graphs G_1 and G_2 , obtaining two accommodating graphs $G_1^{(+G_2)} = (V', E_1^+)$ and $G_2^{(+G_1)} = (V', E_2^+)$ then the graph $G = (V', E'' = (E_1^+ \cap E_2^+))$ could be a merged resource. Indeed, G 's set of edges, E'' seems like a good compromise because, according to the property 1, $(E_1' \cap E_2') \subseteq E'' \subseteq (E_1' \cup E_2')$. This new aggregation method would need to be validated by comparing the quality of the merged resource to the results of the union or intersection.

Furthermore, this work is a first step for defining a similarity measure between graphs, that could take into account the structural agreement rather than a simple edge-to-edge disagreement. Subsequent work should generalise the conciliation process along several axes:

- The number of steps t was chosen as the shortest possible for the confluence measures. It would be worthwhile to investigate the effect of the length of the walks on the agreement of the accommodating graphs.
- Another line of research would be to alter the conciliation ability of graphs, by increasing or decreasing the criterion for strong confluence. One can for example introduce a k parameter in the definition of C_1 (resp. C_2), in Equation 6:

$$P_{G_1'}^t(u \rightsquigarrow r) > k \cdot \pi_{G_1'}(r) \quad (9)$$

- The conciliation process seems unbalanced insofar as graphs only accept to *add* edges. It should be extended to a negotiating process where a graph could also accept to remove one edge if the other does not have it and its confluence is weak.
- The conciliation process could also be generalised to graphs that have different vertices, such as two synonymy networks of different languages. In that case the issue is not anymore to reveal a deeper similarity, beyond a local disagreement, because one can not compare the graphs vertex by vertex or edge by edge. However, questioning whether the semantic structures revealed by dense zones are similar from one lexicon to another is an interesting line of research. One approach to compare two synonymy graphs of two different languages would be to draw edges between vertices that are translations of each other. Random walks could then reach vertices of the two lexicons, so that the conciliation process could be generalised to accommodating two synonymy graphs via translation links.

Acknowledgements

The research presented in this paper was supported by the ANR-NSC (France-Taiwan) bilateral project M3 (Modeling and Measurement of Meaning). We would like to thank the reviewers for their insightful comments.

References

- [Albert and Barabasi2002] Réka Albert and Albert-László Barabasi. 2002. Statistical Mechanics of Complex Networks. *Reviews of Modern Physics*, 74:74–47.
- [Bollobas2002] Bela Bollobas. 2002. *Modern Graph Theory*. Springer-Verlag New York Inc., October.
- [Cohen1960] Jacob Cohen. 1960. A coefficient of agreement for nominal scales. *Educ. Psychol. Meas.*, (20):27–46.
- [Edmonds and Hirst2002] Philip Edmonds and Graeme Hirst. 2002. Near-Synonymy and Lexical Choice. *Computational Linguistics*, 28(2):105–144.

- [Fellbaum1998] Christiane Fellbaum, editor. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.
- [Gaume et al.2010] Bruno Gaume, Fabien Mathieu, and Emmanuel Navarro. 2010. Building Real-World Complex Networks by Wandering on Random Graphs. *I3: Information Interaction Intelligence*, 10(1).
- [Gaume2004] Bruno Gaume. 2004. Balades Aléatoires dans les Petits Mondes Lexicaux. *I3: Information Interaction Intelligence*, 4(2).
- [Murray and Green2004] G. Craig Murray and Rebecca Green. 2004. Lexical Knowledge and Human Disagreement on a WSD Task. *Computer Speech & Language*, 18(3):209–222.
- [Newman2003] M. E. J. Newman. 2003. The Structure and Function of Complex Networks. *SIAM Review*, 45:167–256.
- [Sajous et al.2010] Franck Sajous, Emmanuel Navarro, Bruno Gaume, Laurent Prévot, and Yannick Chudy. 2010. Semi-automatic endogenous enrichment of collaboratively constructed lexical resources: Piggybacking onto wiktionary. In Hrafn Loftsson, Eiríkur Rögnvaldsson, and Sigrún Helgadóttir, editors, *Advances in NLP*, volume 6233 of *LNCS*, pages 332–344. Springer Berlin / Heidelberg.
- [Stewart1994] G. W. Stewart. 1994. Perron-frobenius theory: a new proof of the basics. Technical report, College Park, MD, USA.
- [Steyvers and Tenenbaum2005] Mark Steyvers and Joshua B. Tenenbaum. 2005. The large-scale structure of semantic networks: Statistical analyses and a model of semantic growth. *Cognitive Science*, 29(1):41–78.
- [Watts and Strogatz1998] Duncan J. Watts and Steven H. Strogatz. 1998. Collective Dynamics of Small-World Networks. *Nature*, 393:440–442.
- [Zesch et al.2008] Torsten Zesch, Christof Müller, and Iryna Gurevych. 2008. Using wiktionary for computing semantic relatedness. In *Proceedings of the 23rd national conference on Artificial intelligence - Volume 2*, pages 861–866, Chicago, Illinois. AAAI Press.