Form binary synonymy to near synonymy by optimal proxemy of lexical resources

Benoît Gaillard[⋆], *Emmanuel Navarro*[‡], Bruno Gaume[⋆]

[⋆] : CLLE-ERSS, Université de Toulouse II

[‡] : IRIT, Université de Toulouse III

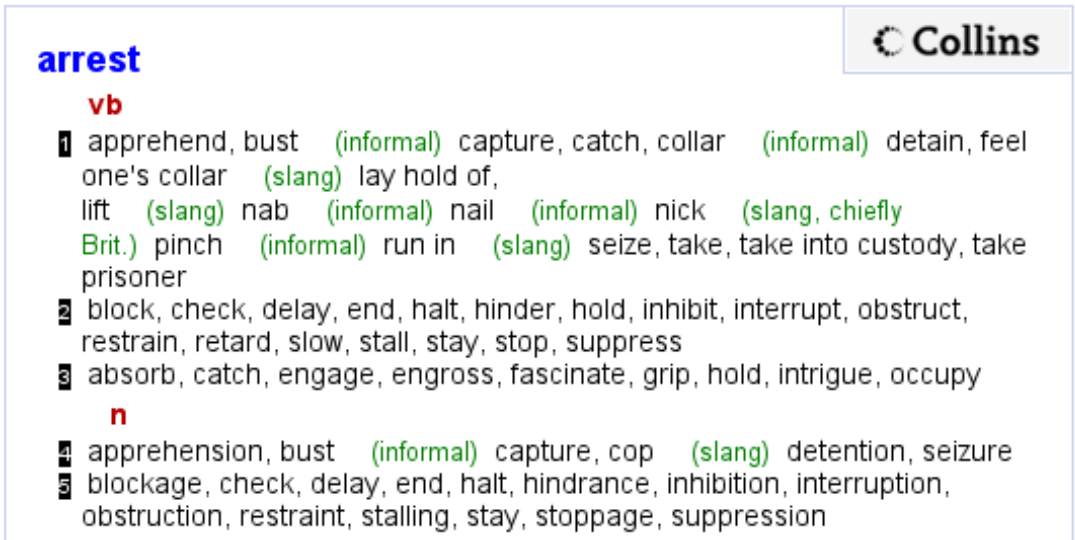
October 28, 2010

Synonymy as a binary relation

Traditional Synonymy Dictionaries

$\Leftrightarrow$

Binary Relation between Words



arrest

Collins

vb

1 apprehend, bust (informal) capture, catch, collar (informal) detain, feel one's collar (slang) lay hold of, lift (slang) nab (informal) nail (informal) nick (slang, chiefly Brit.) pinch (informal) run in (slang) seize, take, take into custody, take prisoner

2 block, check, delay, end, halt, hinder, hold, inhibit, interrupt, obstruct, restrain, retard, slow, stall, stay, stop, suppress

3 absorb, catch, engage, engross, fascinate, grip, hold, intrigue, occupy

n

4 apprehension, bust (informal) capture, cop (slang) detention, seizure

5 blockage, check, delay, end, halt, hindrance, inhibition, interruption, obstruction, restraint, stalling, stay, stoppage, suppression

from an on-line version of the *English Collins Dictionary*
`http://dictionary.reverso.net/english-synonyms/`

- may have POS and sub-sense divisions,
- may have additional information (register, ...)
- BUT:

Two words are synonyms or not!

utique

Toutes les boutiques

Livres en français (622)

Fournitures de bureau (4)

Livres anglais et étrangers (53)

Photo, Image, Son et GPS (8)

Informatique (1)

otion

expédition (Qu'est-ce?)

Toutes les options d'expédition

Livraison gratuite

Listmania!

Créer une liste thématique Listmania!

Rechercher Listmania!

GO!

" Dictionnaire des synonymes"

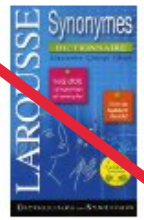
Recherches connexes: dictionnaire des synonymes larousse.

Résultats 1 - 16 sur 687

ver le tri

Dictionnaire des synonymes

Résultats 1 - 16 sur 687

1. 

Dictionnaire des synonymes de Emile Genouvrier, Claude Désirat, Tristan Hordé, et Dominique Désirillat-Leblanc (**Broché** - 13 juin 2007)

Acheter neuf: ~~EUR 7,90~~ **EUR 7,50**

16 neufs à pa

Recevez vo et choisissez la livraison Éclair.

Livraison

Livres en

2. 

Dictionnaires des synonymes, nuances et contraires de Collectif (**Relie** - 15 mars 2006)

Acheter neuf: ~~EUR 22,90~~ **EUR 21,75**

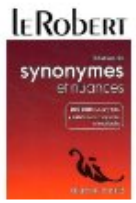
14 neufs à partir de EUR 21,75 2 d'occasion à partir de EUR 19,90

Recevez votre article au plus tard le **mardi 19 octobre**, si vous commandez d'ici **4 heures** et choisissez la livraison Éclair.

Livraison gratuite possible (voir fiche produit).

(2)

Livres en français: Voir l'ensemble des 622 articles

3. 

Dictionnaire des synonymes et nuances de Collectif (**Broché** - 27 juin 2006)

Acheter neuf: ~~EUR 11,90~~ **EUR 11,31**

12 neufs à partir de EUR 11,30 2 d'occasion à partir de EUR 11,29

Recevez votre article au plus tard le **mardi 19 octobre**, si vous commandez d'ici **4 heures** et choisissez la livraison Éclair.

Livraison gratuite possible (voir fiche produit).

(3)

Livres en français: Voir l'ensemble des 622 articles

4. 

Le dictionnaire des synonymes et des contraires de Larousse (**Broché** - 16 septembre 2009)

Acheter neuf: ~~EUR 19,90~~ **EUR 18,91**

synonym dictionary

ew Releases
Any New Release
Last 90 days (1)

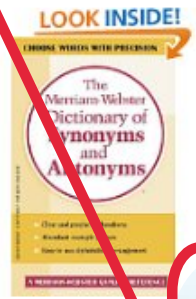
epartment
Any Department
Books
Reference (1,231)
Professional & Technical (1,220)
Nonfiction (1,215)
Children's Books (116)
Literature & Fiction (63)
Science (31)
Medicine (15)
Home & Garden (3)
History (26)
Entertainment (22)
Business & Investing (11)
Religion & Spirituality (18)
Arts & Photography (8)
Health, Mind & Body (7)
Computers & Internet (17)
Teens (26)
Biographies & Memoirs (3)
Travel (5)
Cooking, Food & Wine (1)
Outdoors & Nature (5)
Sports (1)
Law (2)
Gay & Lesbian (1)

ormat
Any Format

Books › "synonym dictionary"

Showing 1 - 12 of 1,644 Results

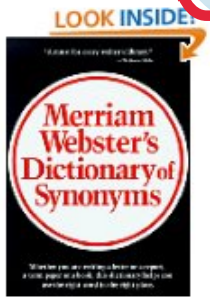
1 - 12 of 1,644 Results

1. 

The Merriam-Webster Dictionary of Synonyms and Antonyms by Merriam-Webster (May 1994)

(15 customer reviews)

Formats	Buy new	New from	Used from
Mass Market Paperback Order in the next 13 hours to get it by Tuesday, Oct 19.	$4.99	$2.12	$1.68

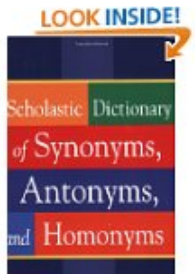
2. 

Merriam Webster's Dictionary of Synonyms: A Dictionary of Discriminated Synonyms With Antonyms and Analogous and Contrasted Words by Merriam-Webster (Jun 1984)

(15 customer reviews)

Formats	Buy new	New from	Used from	Collectible from
Hardcover Order in the next 13 hours to get it by Tuesday, Oct 19.	~~$22.95~~ $17.90	$13.90	$3.19	$15.00

Some formats eligible for **FREE** Super Saver Shipping.

3. 

Scholastic Dictionary of Synonyms, Antonyms, and Homonyms by Scholastic Inc. (Apr 1, 2001)

(11 customer reviews)

Formats	Buy new	New from	Used from
Paperback Order in the next 13 hours to get it by Tuesday, Oct 19.	$5.99	$3.19	$2.00

Also useful in NLP, and computer science...

- Query Expansion in I.R.,
- Disambiguation,
- Machine translation,
- ...

Many *digital* lexical resources are based on binary synonymy (at least WordNet(s), ...).

Computer scientist's issues:

- Not available, or very expensive !
- Small coverage !
- Accuracy ?

We came with another issue:

⇒ **Do different ressources indicate same synonymy links ?**

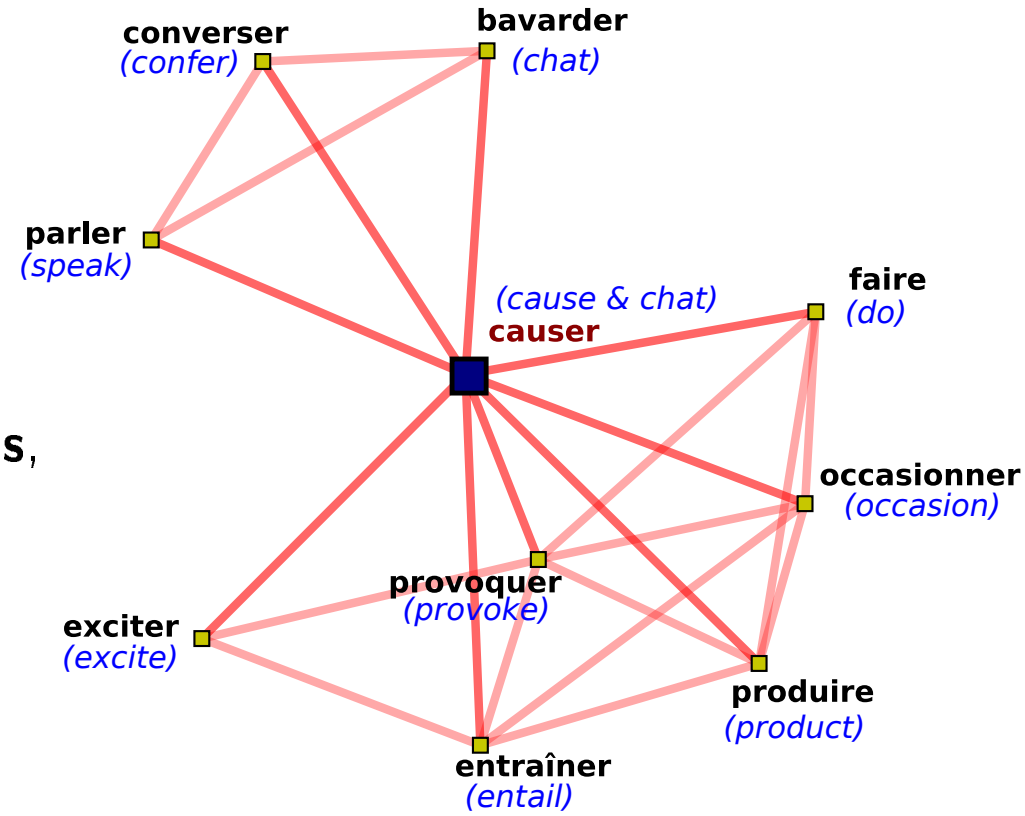
⇒ **What is the consistency between ressources ?**

Studying the consistency of 7 dictionaries of synonyms

Coverage of 7 dictionaries of French synonyms:

	Bailly	Benac	Bertaud	Guizot	Lafaye	Larousse	Robert	∪
$\|V\|$	12 739	21 231	50 130	3 160	3 119	25 441	52 429	83 244
$\|E\|$	14 212	32 428	126 260	2 199	2 500	78 909	116 289	246 667

- 7 french **general purpose** dictionaries of synonymy, electronic version[1] of paper dictionaries,
- Naturaly encoded as **un-weighted un-directed graphs**,
- No sub-sense,
- No POS separation.



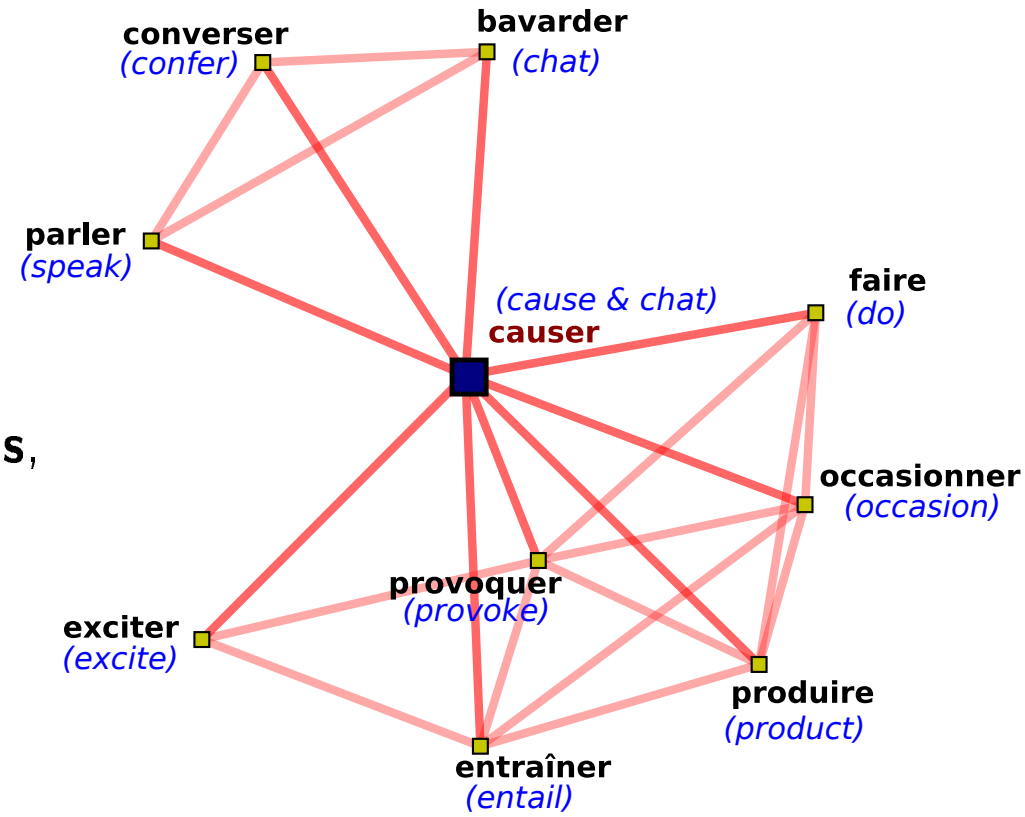
[1] Extracted from dictionaries by INALF (ATILF today), revised in CRISCO

Studying the consistency of 7 dictionaries of synonyms

Coverage of 7 dictionaries of French synonyms:

	Bailly	Benac	Bertaud	Guizot	Lafaye	Larousse	Robert	∪
$\|V\|$	12 739	21 231	50 130	3 160	3 119	25 441	52 429	83 244
$\|E\|$	14 212	32 428	126 260	2 199	2 500	78 909	116 289	246 667

- 7 french **general purpose** dictionaries of synonymy, electronic version[1] of paper dictionaries,
- Naturaly encoded as **un-weighted un-directed graphs**,
- **No sub-sense**,
- No POS separation.



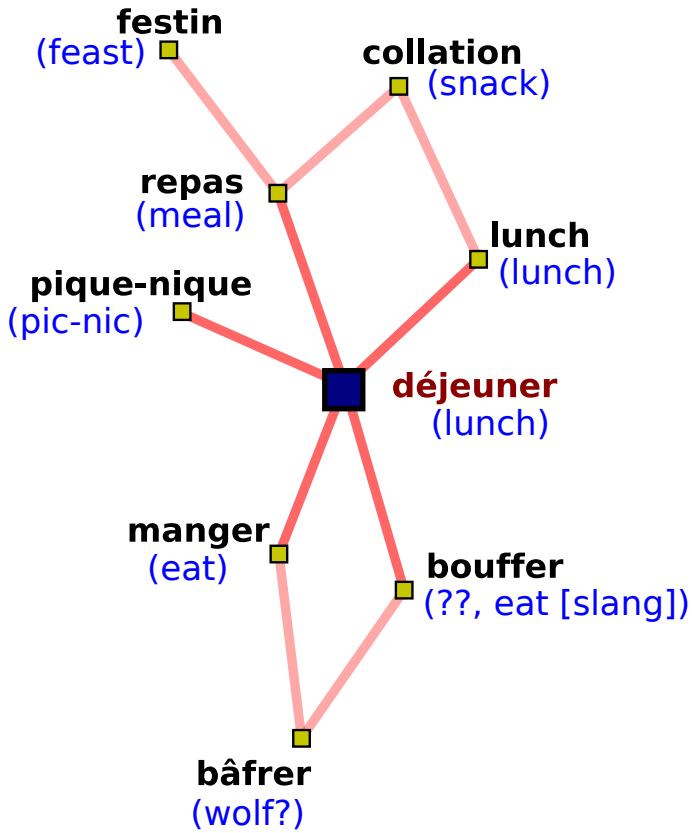
[1] Extracted from dictionaries by INALF (ATILF today), revised in CRISCO

Studying the consistency of 7 dictionaries of synonyms

Coverage of 7 dictionaries of French synonyms:

	Bailly	Benac	Bertaud	Guizot	Lafaye	Larousse	Robert	∪
$\|V\|$	12 739	21 231	50 130	3 160	3 119	25 441	52 429	83 244
$\|E\|$	14 212	32 428	126 260	2 199	2 500	78 909	116 289	246 667

- 7 french **general purpose** dictionaries of synonymy, electronic version[1] of paper dictionaries,
- Naturaly encoded as **un-weighted un-directed graphs**,
- **No sub-sense**,
- **No POS separation**.

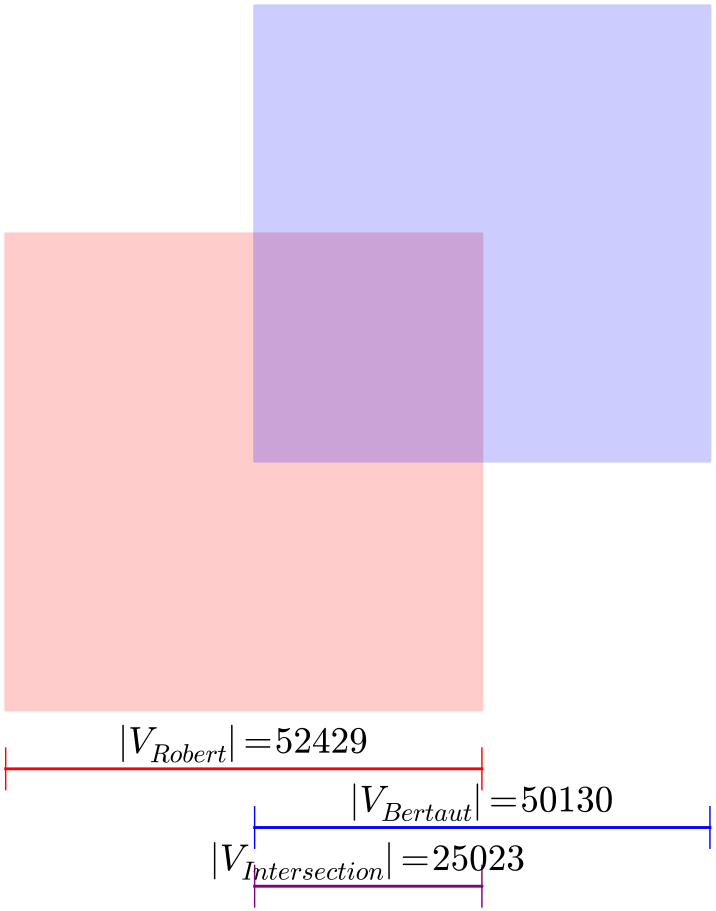


[1] Extracted from dictionaries by INALF (ATILF today), revised in CRISCO

How to compare two *binary* graphs ?

$$G_1 = (V_1, E_1), \quad G_2 = (V_2, E_2)$$

Robert vs. Bertaud



Lexical coverage :

$$R_\bullet = \frac{|V_1 \cap V_2|}{|V_2|} \qquad P_\bullet = \frac{|V_1 \cap V_2|}{|V_1|}$$

$$F_\bullet = 2.\frac{R_\bullet.P_\bullet}{R_\bullet + P_\bullet}$$

Synonymy links agreement :
(reduce graphs to common vertices)

$$R_\updownarrow = \frac{|E_{1\langle v_1 \cap v_2 \rangle} \cap E_{2\langle v_1 \cap v_2 \rangle}|}{|E_{2\langle v_1 \cap v_2 \rangle}|}$$

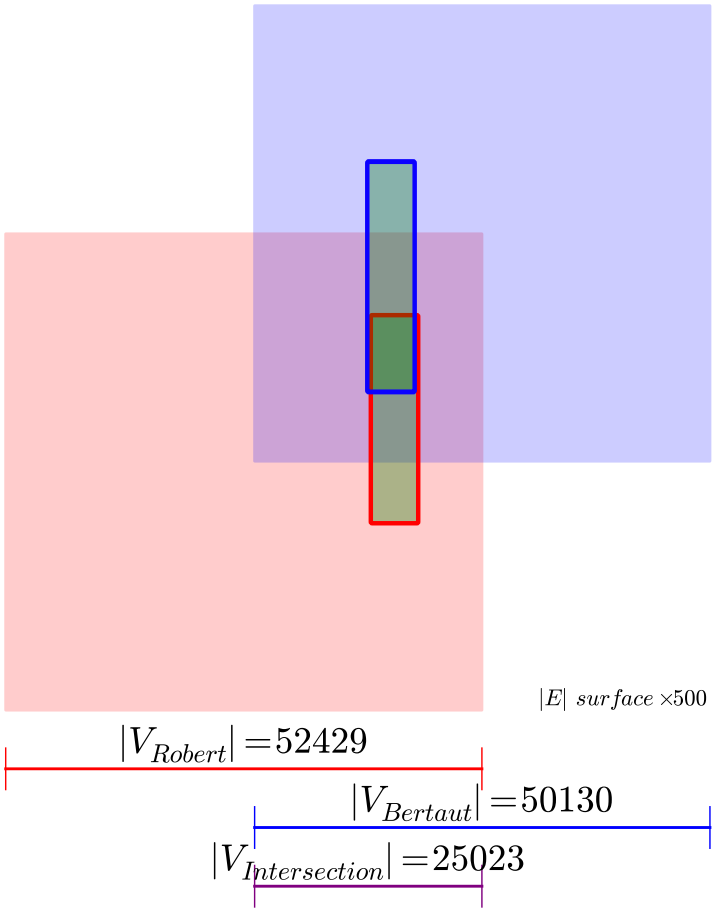
$$P_\updownarrow = \frac{|E_{1\langle v_1 \cap v_2 \rangle} \cap E_{2\langle v_1 \cap v_2 \rangle}|}{|E_{1\langle v_1 \cap v_2 \rangle}|}$$

$$F_\updownarrow = 2.\frac{R_\updownarrow.P_\updownarrow}{R_\updownarrow + P_\updownarrow}$$

How to compare two *binary* graphs ?

$$G_1 = (V_1, E_1), \quad G_2 = (V_2, E_2)$$

Robert vs. Bertaud



Lexical coverage :

$$R_\bullet = \frac{|V_1 \cap V_2|}{|V_2|} \qquad P_\bullet = \frac{|V_1 \cap V_2|}{|V_1|}$$

$$F_\bullet = 2.\frac{R_\bullet.P_\bullet}{R_\bullet + P_\bullet}$$

Synonymy links agreement :
(reduce graphs to common vertices)

$$R_\updownarrow = \frac{|E_{1\langle v_1 \cap v_2 \rangle} \cap E_{2\langle v_1 \cap v_2 \rangle}|}{|E_{2\langle v_1 \cap v_2 \rangle}|}$$

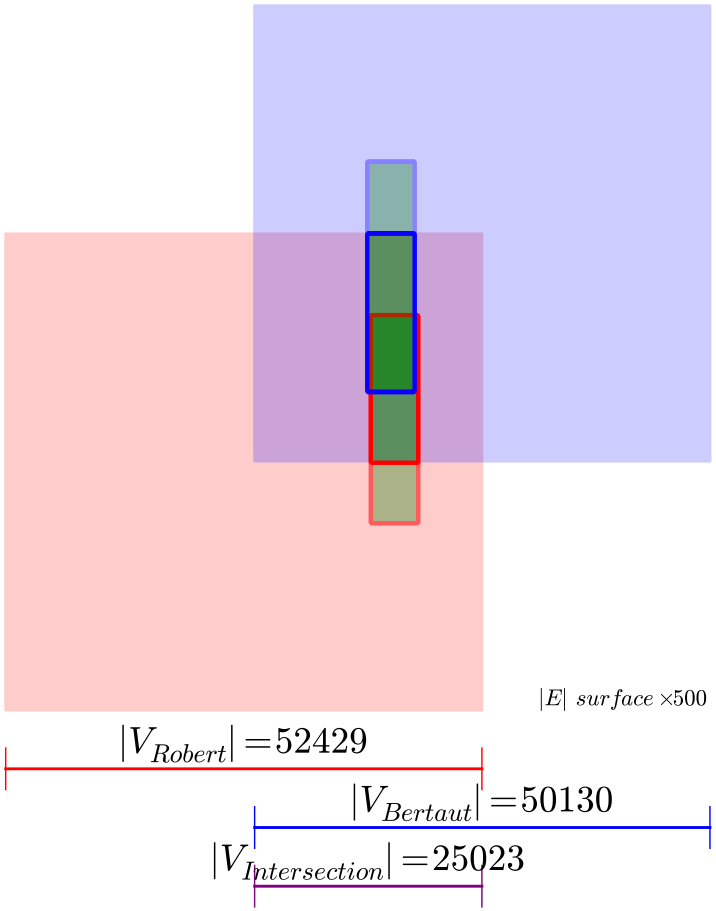
$$P_\updownarrow = \frac{|E_{1\langle v_1 \cap v_2 \rangle} \cap E_{2\langle v_1 \cap v_2 \rangle}|}{|E_{1\langle v_1 \cap v_2 \rangle}|}$$

$$F_\updownarrow = 2.\frac{R_\updownarrow.P_\updownarrow}{R_\updownarrow + P_\updownarrow}$$

How to compare two *binary* graphs ?

$$G_1 = (V_1, E_1), \quad G_2 = (V_2, E_2)$$

Robert vs. Bertaud



Lexical coverage :

$$R_\bullet = \frac{|V_1 \cap V_2|}{|V_2|} \qquad P_\bullet = \frac{|V_1 \cap V_2|}{|V_1|}$$

$$F_\bullet = 2.\frac{R_\bullet.P_\bullet}{R_\bullet + P_\bullet}$$

Synonymy links agreement :

(reduce graphs to common vertices)

$$R_\updownarrow = \frac{|E_{1\langle v_1 \cap v_2 \rangle} \cap E_{2\langle v_1 \cap v_2 \rangle}|}{|E_{2\langle v_1 \cap v_2 \rangle}|}$$

$$P_\updownarrow = \frac{|E_{1\langle v_1 \cap v_2 \rangle} \cap E_{2\langle v_1 \cap v_2 \rangle}|}{|E_{1\langle v_1 \cap v_2 \rangle}|}$$

$$F_\updownarrow = 2.\frac{R_\updownarrow.P_\updownarrow}{R_\updownarrow + P_\updownarrow}$$

A Weak Agreement ...

		Benac		Bertaud		Guizot		Lafaye		Larousse		Robert		∪	
		(•)	(↕)	(•)	(↕)	(•)	(↕)	(•)	(↕)	(•)	(↕)	(•)	(↕)	(•)	(↕)
Bail.	R	0.49	0.57	0.23	0.20	0.84	0.60	0.90	0.61	0.40	0.18	0.22	0.20	0.15	**0.17**
	P	0.82	0.60	0.91	0.73	0.21	0.50	0.22	0.52	0.81	0.62	0.89	0.69	1.00	1.00
	F	0.62	**0.58**	0.37	**0.31**	0.33	**0.54**	0.36	**0.56**	0.54	**0.28**	0.35	**0.30**	0.27	0.29
Ben.	R			0.37	0.31	0.84	0.58	0.90	0.68	0.52	0.18	0.28	0.18	0.26	**0.26**
	P			0.87	0.71	0.13	0.43	0.13	0.51	0.62	0.60	0.69	0.63	1.00	1.00
	F			0.52	**0.43**	0.22	**0.49**	0.23	**0.59**	0.57	**0.27**	0.40	**0.28**	0.41	0.42
Bert.	R					0.92	0.74	0.96	0.78	0.74	0.42	0.48	0.51	0.60	**0.63**
	P					0.06	0.17	0.06	0.18	0.38	0.38	0.50	0.49	1.00	1.00
	F					0.11	**0.27**	0.11	**0.29**	0.50	**0.40**	0.49	**0.50**	0.75	0.77
Guiz.	R							0.79	0.68	0.11	0.19	0.06	0.18	0.04	**0.15**
	P							0.78	0.69	0.88	0.72	0.91	0.82	1.00	1.00
	F							0.78	**0.69**	0.19	**0.30**	0.10	**0.29**	0.07	0.26
Laf.	R									0.11	0.18	0.06	0.17	0.04	**0.15**
	P									0.92	0.65	0.95	0.76	1.00	1.00
	F									0.20	**0.28**	0.11	**0.28**	0.07	0.26
Lar.	R											0.41	0.51	0.31	**0.53**
	P											0.84	0.53	1.00	1.00
	F											0.55	**0.52**	0.47	0.69
Rob.	R													0.63	**0.61**
	P													1.00	1.00
	F													0.77	0.75

A Weak Agreement ...

		Benac		Bertaud		Guizot		Lafaye		Larousse		Robert		∪	
		(•)	(↕)	(•)	(↕)	(•)	(↕)	(•)	(↕)	(•)	(↕)	(•)	(↕)	(•)	(↕)
Bail.	R	0.49	0.57	0.23	0.20	0.84	0.60	0.90	0.61	0.40	0.18	0.22	0.20	0.15	**0.17**
	P	0.82	0.60	0.91	0.73	0.21	0.50	0.22	0.52	0.81	0.62	0.89	0.69	1.00	1.00
	F	0.62	**0.58**	0.37	**0.31**	0.33	**0.54**	0.36	**0.56**	0.54	**0.28**	0.35	**0.30**	0.27	0.29
Ben.	R			0.37	0.31	0.84	0.58	0.90	0.68	0.52	0.18	0.28	0.18	0.26	**0.26**
	P			0.87	0.71	0.13	0.43	0.13	0.51	0.62	0.60	0.69	0.63	1.00	1.00
	F			0.52	**0.43**	0.22	**0.49**	0.23	**0.59**	0.57	**0.27**	0.40	**0.28**	0.41	0.42
Bert.	R					0.92	0.74	0.96	0.78	0.74	0.42	0.48	0.51	0.60	**0.63**
	P					0.06	0.17	0.06	0.18	0.38	0.38	0.50	0.49	1.00	1.00
	F					0.11	**0.27**	0.11	**0.29**	0.50	**0.40**	0.49	**0.50**	0.75	0.77
Guiz.	R							0.79	0.68	0.11	0.19	0.06	0.18	0.04	**0.15**
	P							0.78	0.69	0.88	0.72	0.91	0.82	1.00	1.00
	F							0.78	**0.69**	0.19	**0.30**	0.10	**0.29**	0.07	0.26
Laf.	R									0.11	0.18	0.06	0.17	0.04	**0.15**
	P									0.92	0.65	0.95	0.76	1.00	1.00
	F									0.20	**0.28**	0.11	**0.28**	0.07	0.26
Lar.	R											0.41	0.51	0.31	**0.53**
	P											0.84	0.53	1.00	1.00
	F											0.55	**0.52**	0.47	0.69
Rob.	R													0.63	**0.61**
	P													1.00	1.00
	F													0.77	0.75

What's wrong ??

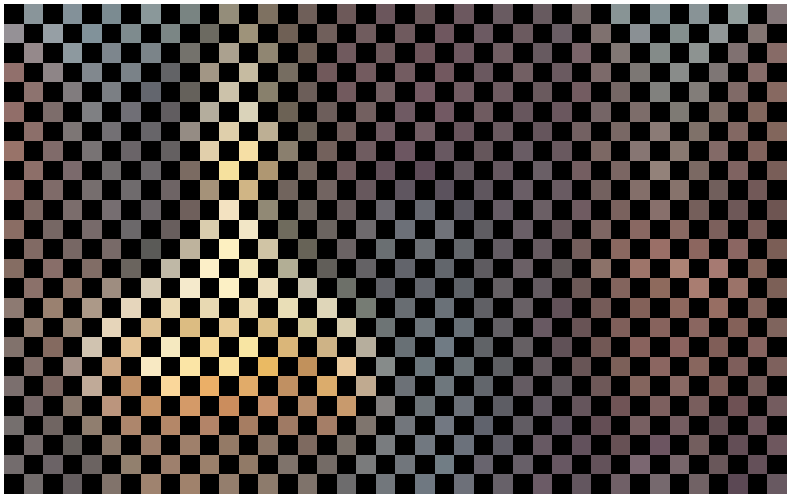
- Why resources describing the same lexicon appear so different ?
- *Maybe we look at them too much "in detail" ...*

What's wrong ??

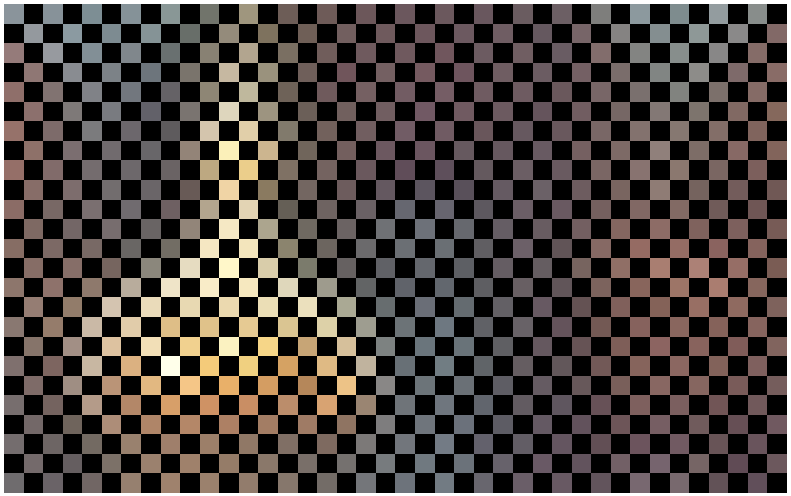
- Why resources describing the same lexicon appear so different ?
- *Maybe we look at them too much "in detail" ...*

The picture metaphor (1/2)

Uspenski Cathedral, Helsinki

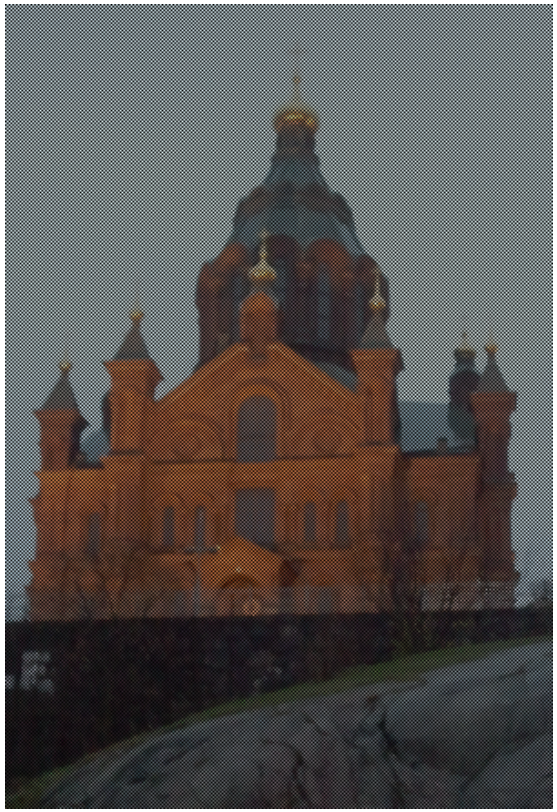


Each **even** pixel painted in black...

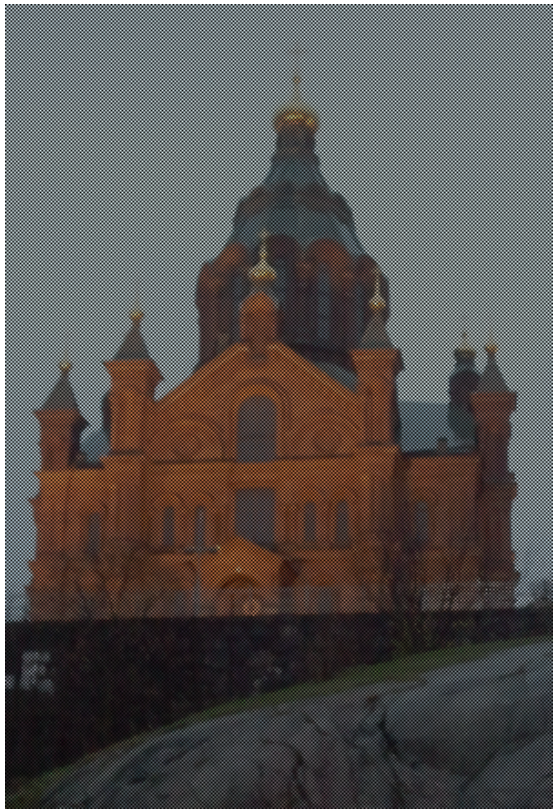


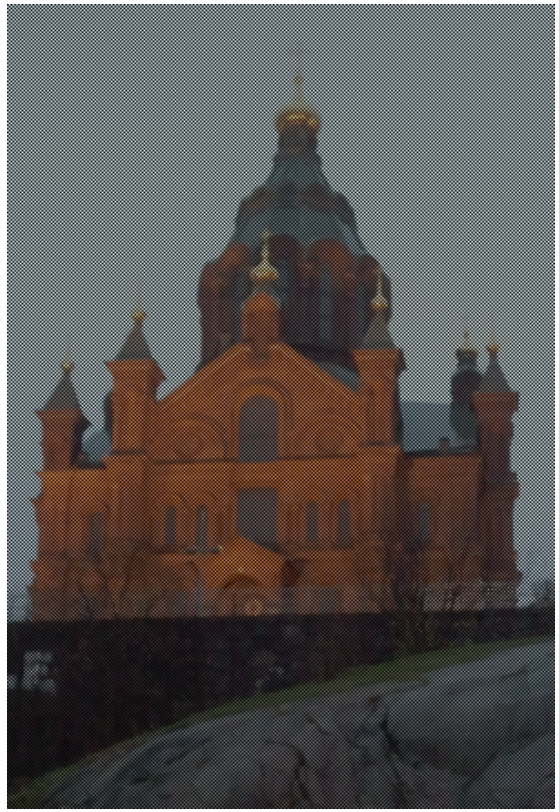
Each **odd** pixel painted in black...

The picture metaphor (2/2)

$Even =$ 

The picture metaphor (2/2)

$Even =$ 

$Odd =$ 

Can you see a difference ?

however... $sim(A, B) \approx 0$, when computed at **pixel level**.

Similarly on graphs : take a step back !

Weak agreement at the **edge level**

but

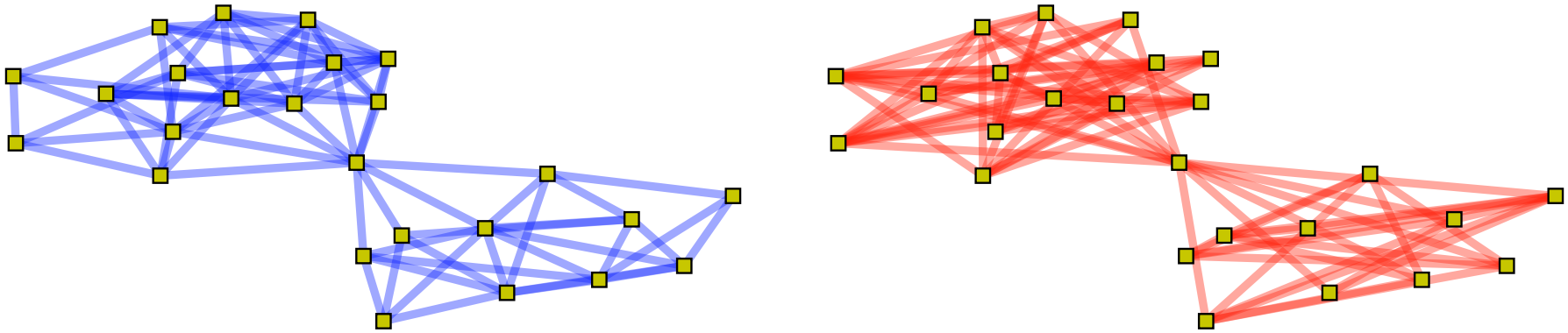
a stronger agreement at a **coarser grain level**...

Similarly on graphs : take a step back !

Weak agreement at the **edge level**

but

a stronger agreement at a **coarser grain level**...

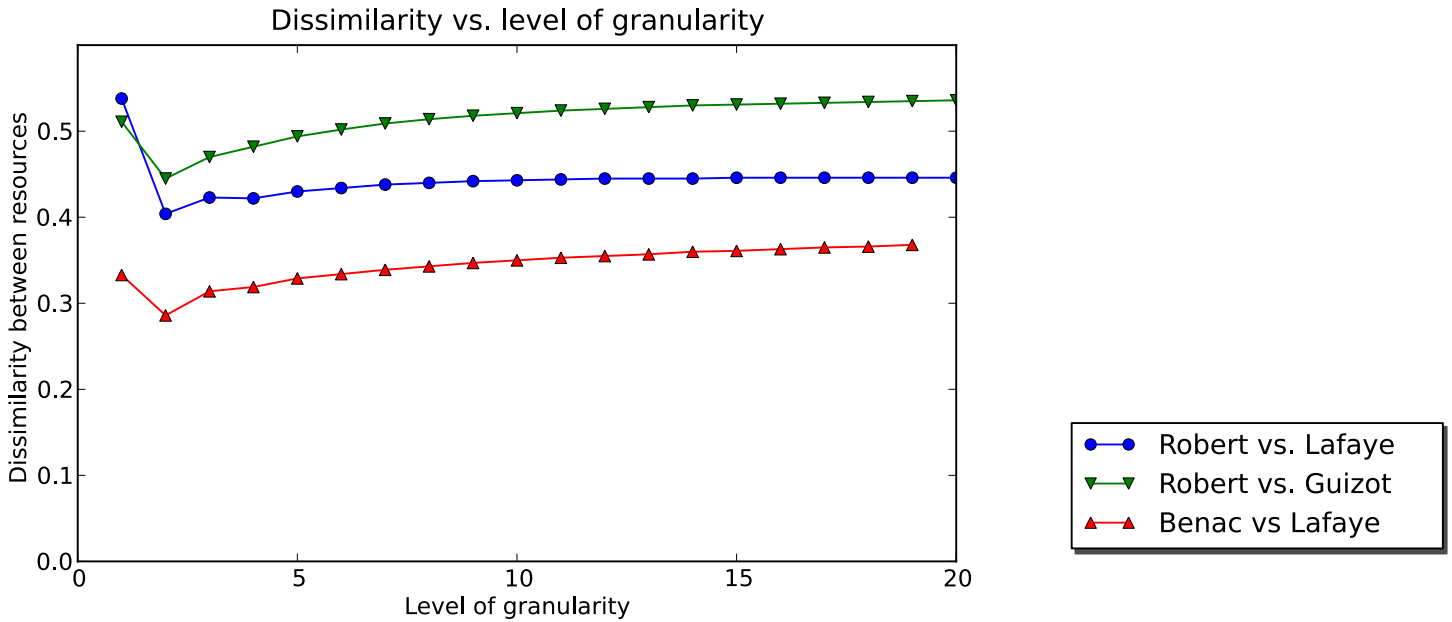


As for the cat picture: there is **no edge in common** between these two graphs...

How to validate this hypothesis ? (1/2)

Given a method to represent graphs at different levels of granularity,

- Distance/Similarity between graphs at different levels of granularity...
- Is there an optimal level ?



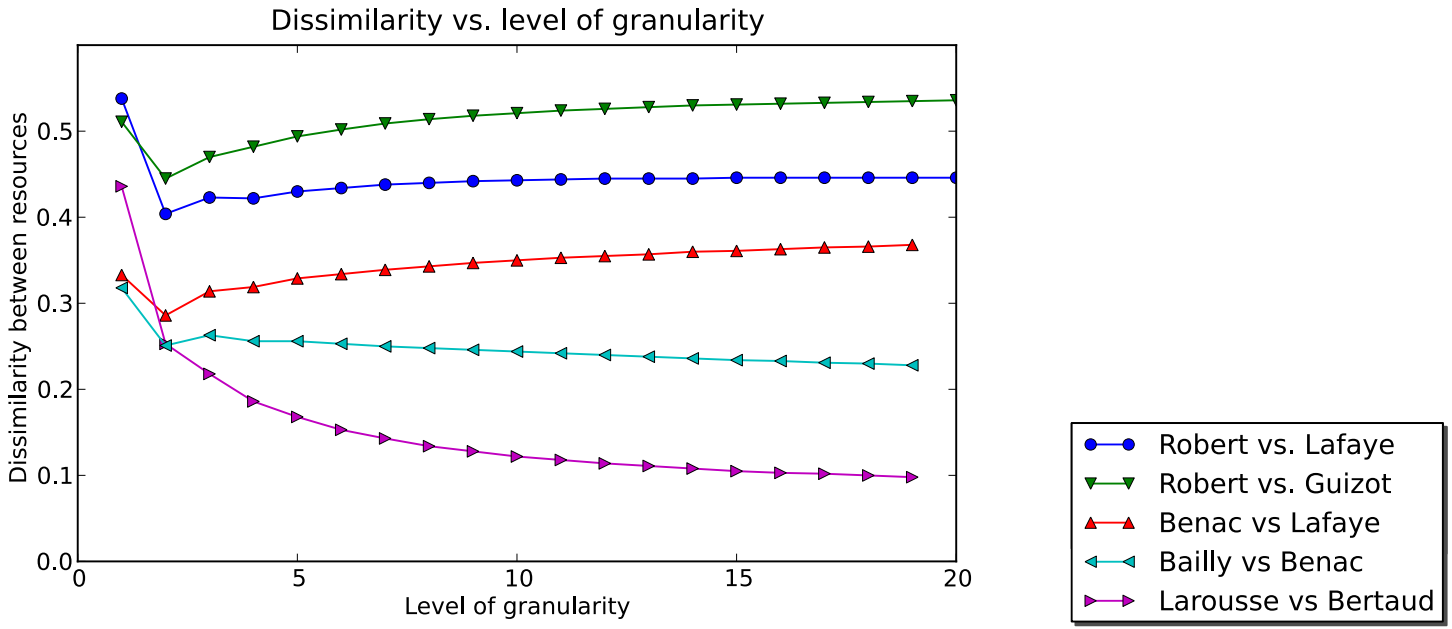
But:

- Pb 1: some graphs tend to be very similar at the *coarsest grain level*...
- Pb 2: "**level of granularity**" is given by the method, is it correct ?

How to validate this hypothesis ? (1/2)

Given a method to represent graphs at different levels of granularity,

- Distance/Similarity between graphs at different levels of granularity...
- Is there an optimal level ?



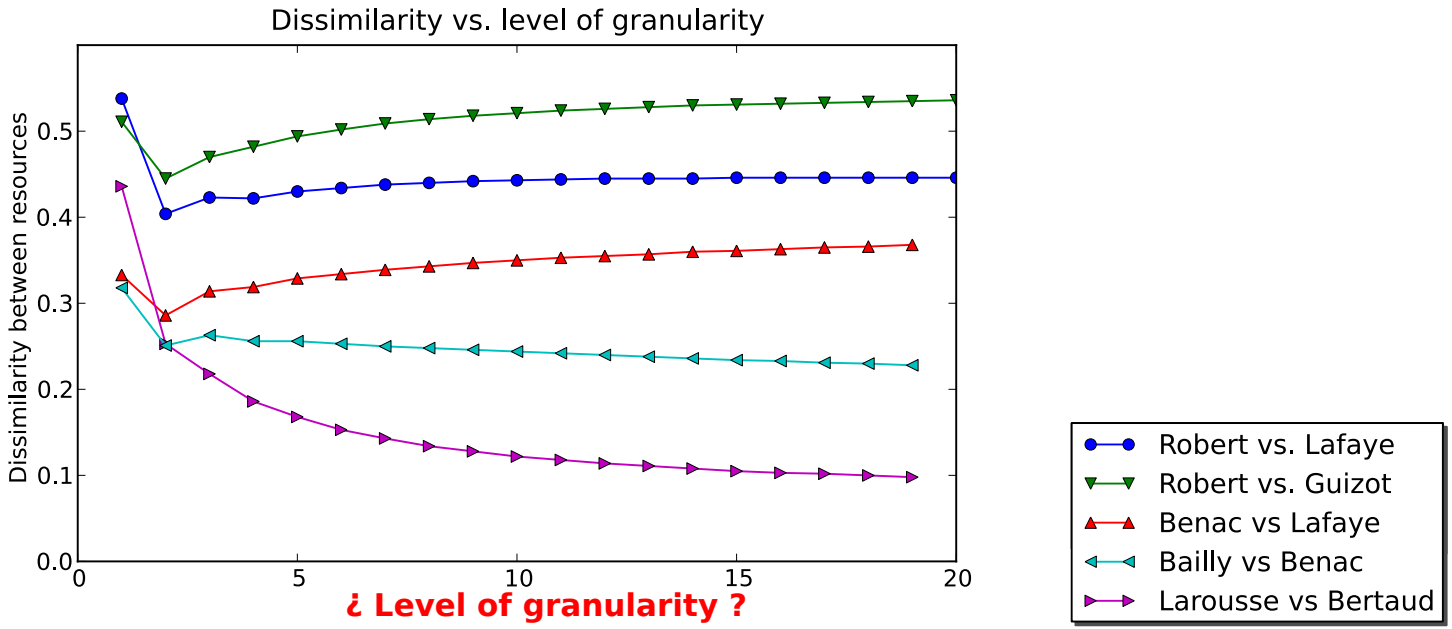
But:

- Pb 1: some graphs tend to be very similar at the *coarsest grain level*...
- Pb 2: "**level of granularity**" is given by the method, is it correct ?

How to validate this hypothesis ? (1/2)

Given a method to represent graphs at different levels of granularity,

- Distance/Similarity between graphs at different levels of granularity...
- Is there an optimal level ?



But:

- Pb 1: some graphs tend to be very similar at the *coarsest grain level*...
- Pb 2: "**level of granularity**" is given by the method, is it correct ?

How to validate this hypothesis ? (2/2)

1. Enrich a graph G with **edges that appear at a higher level of granularity**,
2. Compare it to the union graph ($G_\cup$).

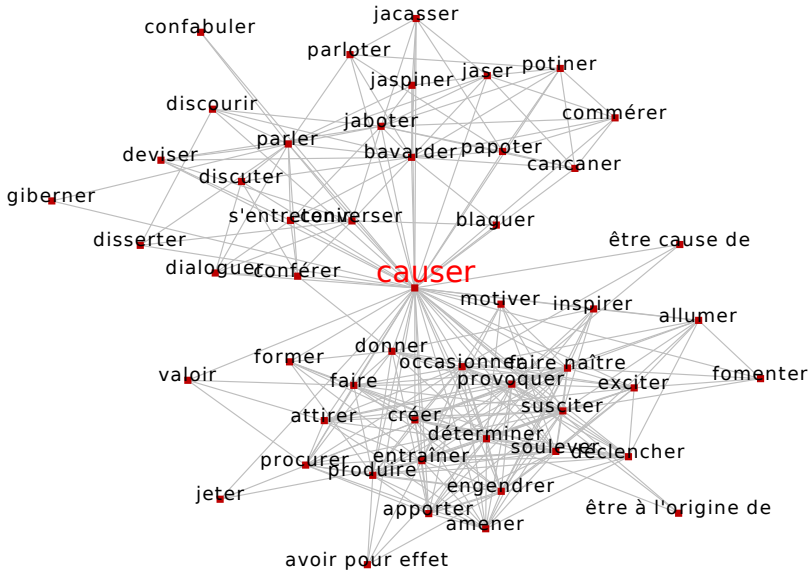
Can we find **edges missing** in G when we look at it at a coarser grain level ?

- **Missing edges**: edges present in $G_\cup$ but not in G
  ($G_\cup$ restricted to vertices of G)

How can we look at our graph at different grain levels ?

Graph Clustering

- how detect **mesoscopique structures** ?
- Currently a lot of research...
- but not (yet!) good method handling **overlapping** communities.



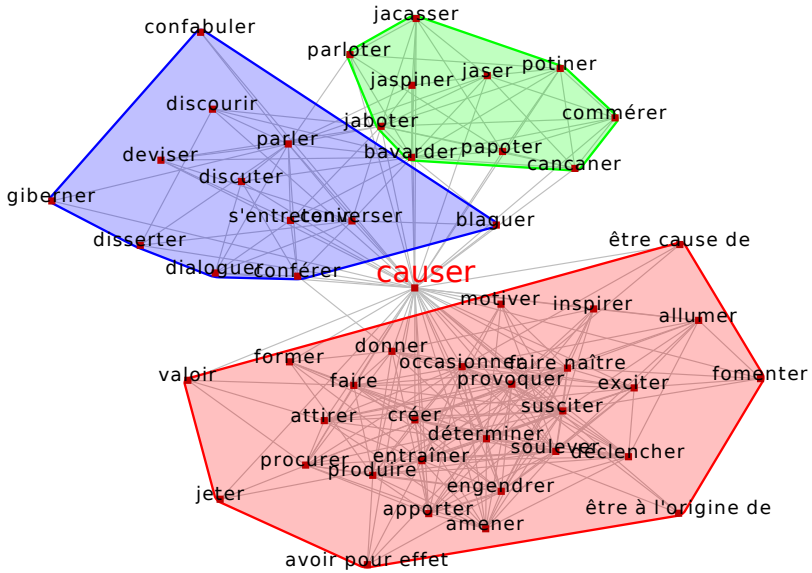
Random Walks

- the idea: if (u, v) are **in the same *"cluster"***, they may **not** be **adjacent**, but **going from** u **to** v should be **very likely** with a **short lenght random walk**.
- Random walkers tends to be trapped into clusters,
- Note: idea used by many graph clustering methods,
- See next slide...

How can we look at our graph at different grain levels ?

Graph Clustering

- how detect **mesoscopique structures** ?
- Currently a lot of research...
- but not (yet!) good method handling **overlapping** communities.



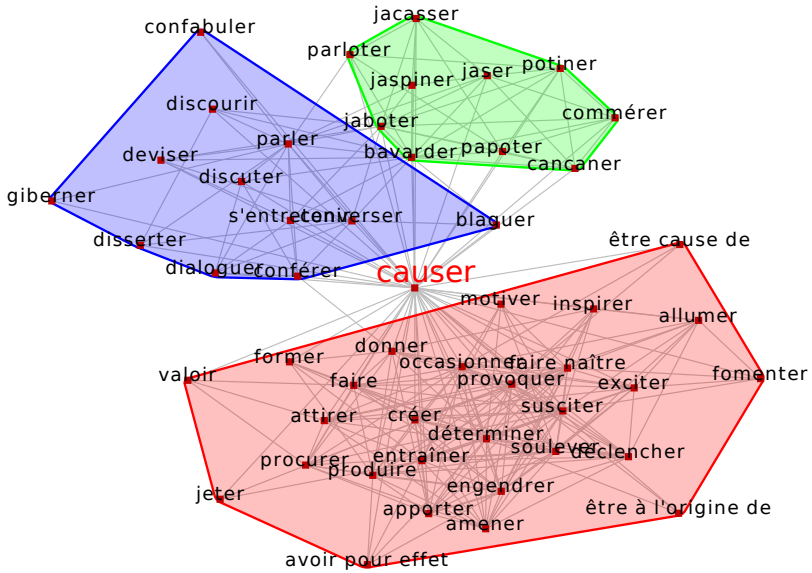
Random Walks

- the idea: if (u, v) are **in the same *"cluster"***, they may **not** be **adjacent**, but **going from u to v** should be **very likely** with a **short lenght random walk**.
- Random walkers tends to be trapped into clusters,
- Note: idea used by many graph clustering methods,
- See next slide...

How can we look at our graph at different grain levels ?

Graph Clustering

- how detect **mesoscopique structures** ?
- Currently a lot of research...
- but not (yet!) good method handling **overlapping** communities.



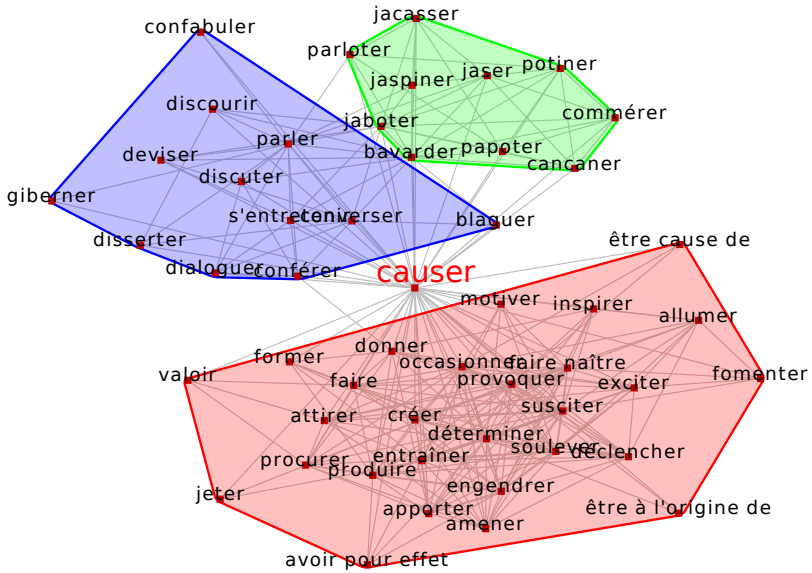
Random Walks

- the idea: if (u, v) are **in the same *"cluster"***, they may **not** be **adjacent**, but **going from u to v** should be **very likely** with a **short lenght random walk**.
- Random walkers tends to be trapped into clusters,
- Note: idea used by many graph clustering methods,
- See next slide...

How can we look at our graph at different grain levels ?

Graph Clustering

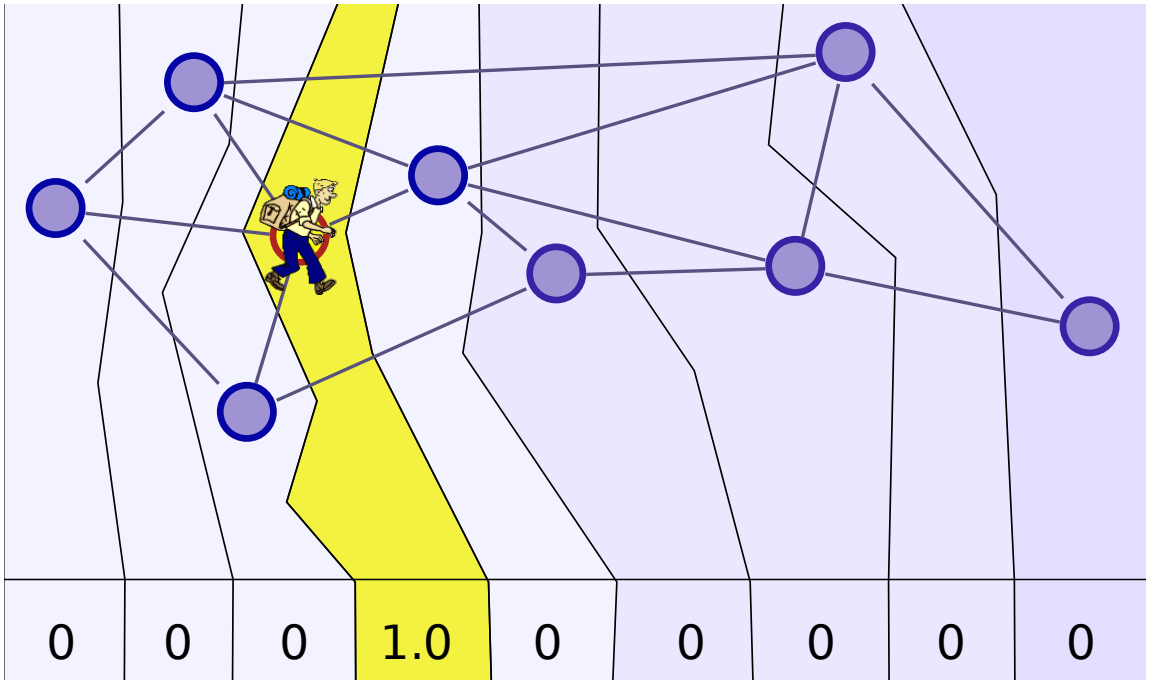
- how detect **mesoscopique structures** ?
- Currently a lot of research...
- but not (yet!) good method handling **overlapping** communities.



Random Walks

- the idea: if (u, v) are **in the same *"cluster"***, they may **not** be **adjacent**, but **going from** u **to** v should be **very likely** with a **short lenght random walk**.
- Random walkers tends to be trapped into clusters,
- Note: idea used by many graph clustering methods,
- See next slide...

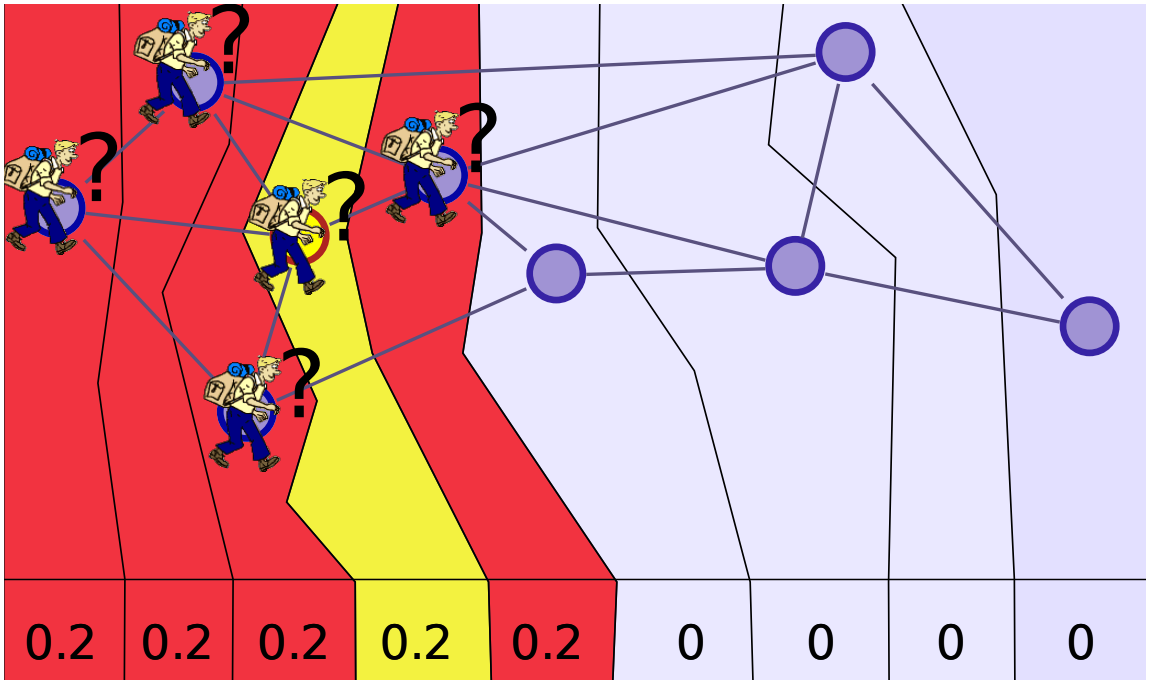
Random walks on graph...



1. start from a node u,
2. walk to a neighbour with equal probability,
3. walk to a neighbour with equal probability,
4. etc...

$$t = 0 \ , \quad P^t(u, *) = [0, 0, 0, 1.0, 0, 0, 0, 0, 0]$$

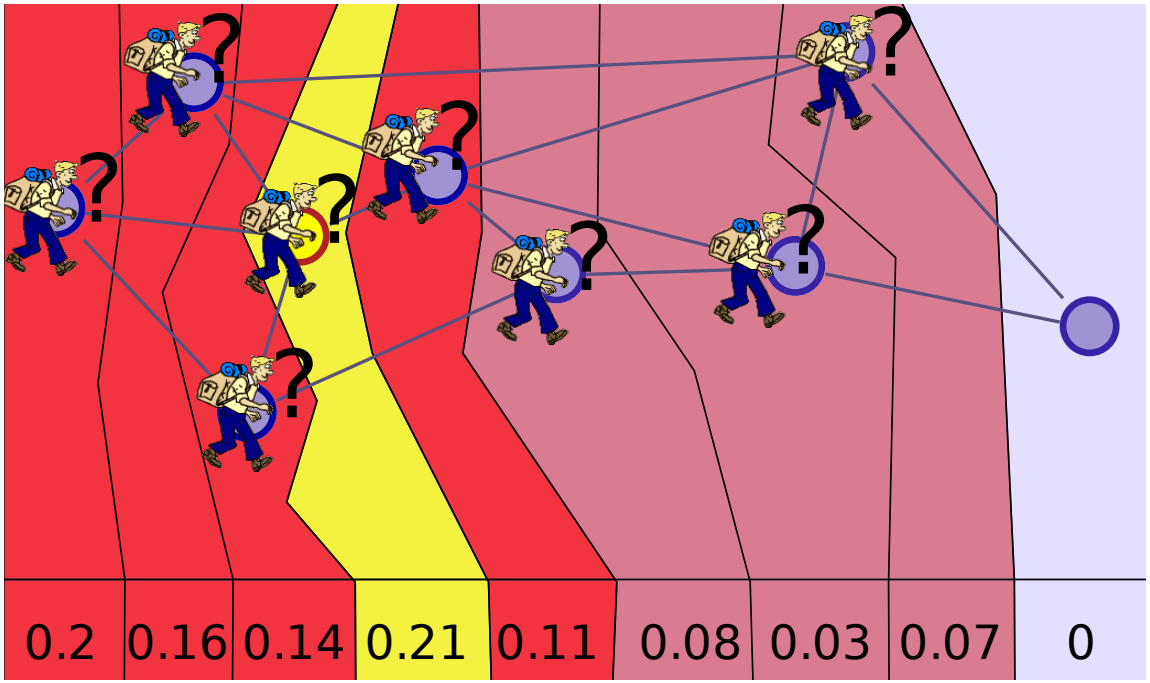
Random walks on graph...



1. start from a node u,
2. walk to a neighbour with equal probability,
3. walk to a neighbour with equal probability,
4. etc...

$$t = 1 \ , \quad P^t(u, *) = [0.2, 0.2, 0.2, 0.2, 0.2, 0, 0, 0, 0, 0]$$

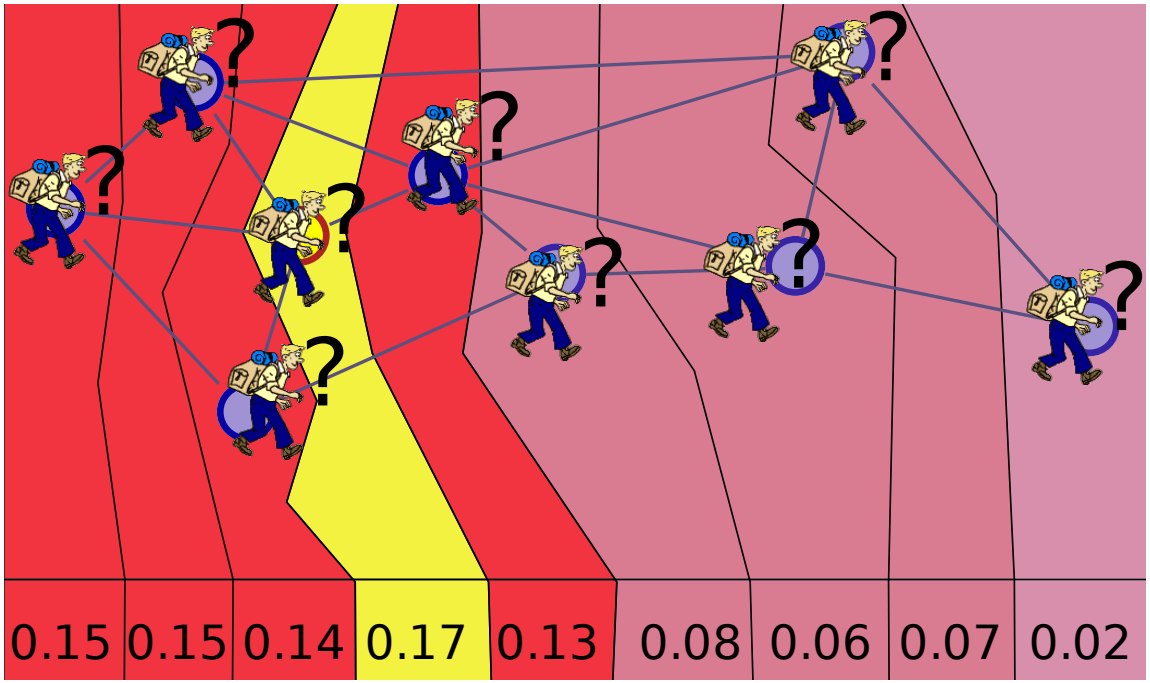
Random walks on graph...



1. start from a node u,
2. walk to a neighbour with equal probability,
3. walk to a neighbour with equal probability,
4. etc...

$$t = 2\ , \quad P^t(u, *) = [0.2, 0.16, 0.14, 0.21, 0.11, 0.08, 0.03, 0.07, 0]$$

Random walks on graph...



1. start from a node u,
2. walk to a neighbour with equal probability,
3. walk to a neighbour with equal probability,
4. etc...

$$t = 3\ , \quad P^t(u, *) = [0.15, 0.15, 0.14, 0.17, 0.13, 0.08, 0.06, 0.07, 0.02]$$

Build coarse grain graphs with *random walks*

- $P^t(u,v)$: **probability to reach** v **from** u **in** t **steps** (that we called *proxemy*).
- $P^\infty(u,v) = P^\infty(v) = \frac{\deg(v)}{D}$: no more depends on u[2]
  (simplest version of *pagerank*)

For a given t and a number M of edges to add:

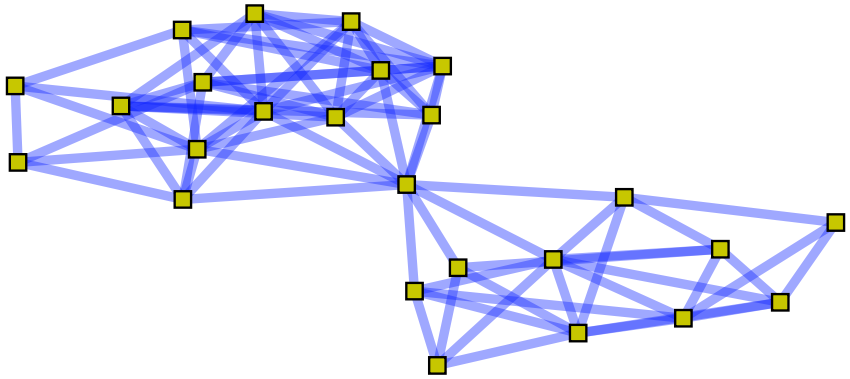
1. Compute all $P^t(u,v)$,
2. Add edges between the M pairs (u,v) that have the largest :

$$\frac{P^t(u,v)}{P^\infty(v)}$$

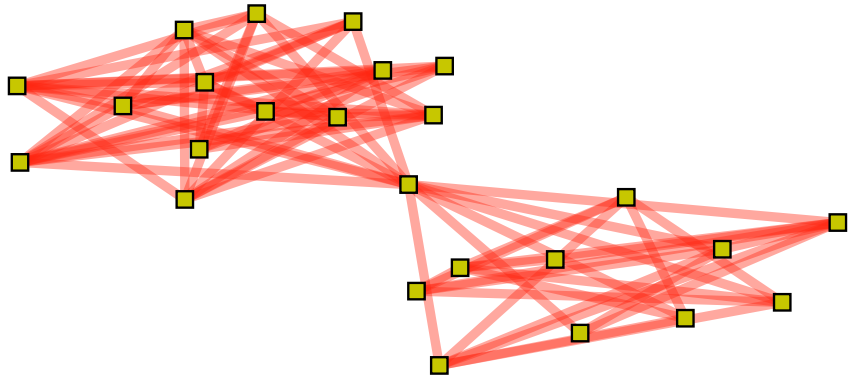
(we choose $t = 3$, experiments with $t = 2, 5, 10, 20$ give similar results)

[2] If the graph is not directed and reflexive

Toy example

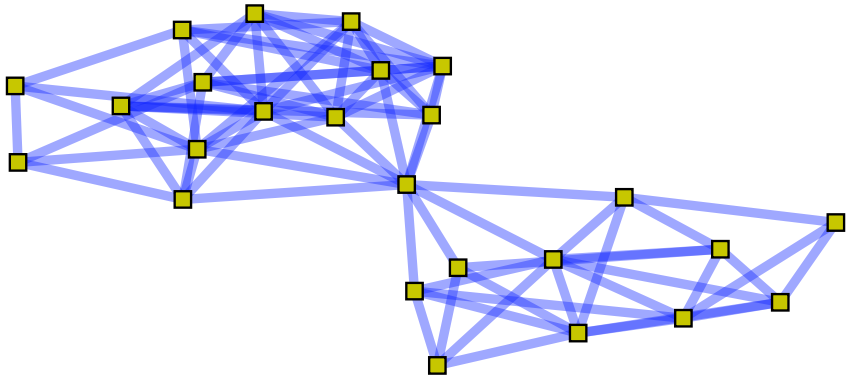


$R_{\updownarrow} = 0.51,\ P_{\updownarrow} = 1,\ F_{\updownarrow} = 0.67$

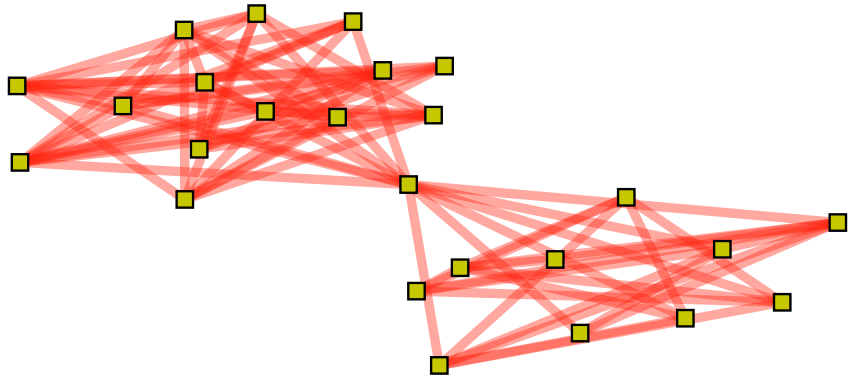


$R_{\updownarrow} = 0.49,\ P_{\updownarrow} = 1,\ F_{\updownarrow} = 0.66$

Toy example

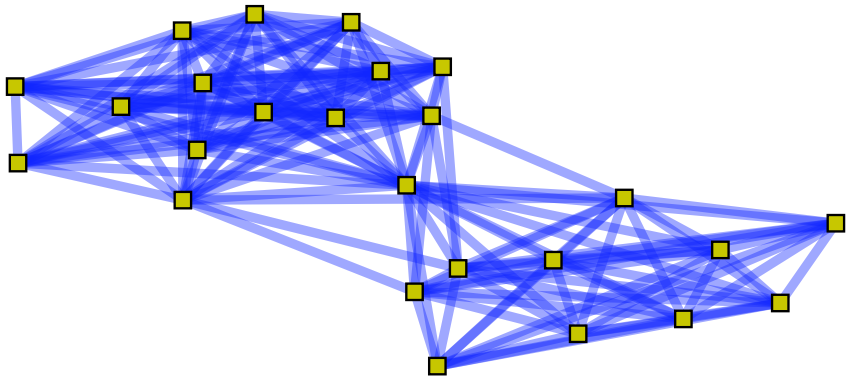


$R_{\updownarrow} = 0.51,\ P_{\updownarrow} = 1,\ F_{\updownarrow} = 0.67$

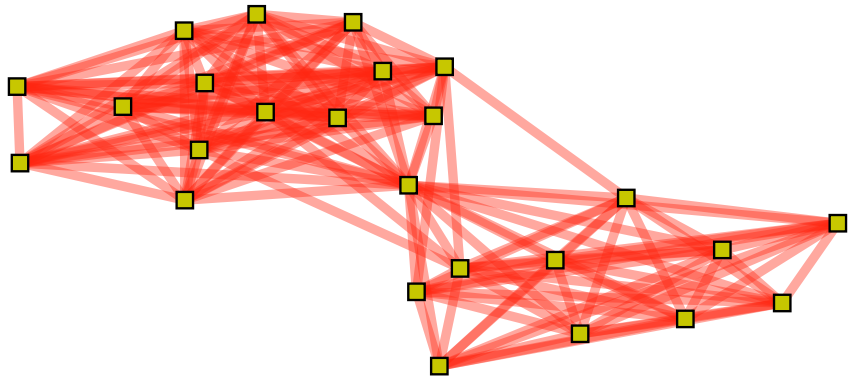


$R_{\updownarrow} = 0.49,\ P_{\updownarrow} = 1,\ F_{\updownarrow} = 0.66$

With 86 edges added :

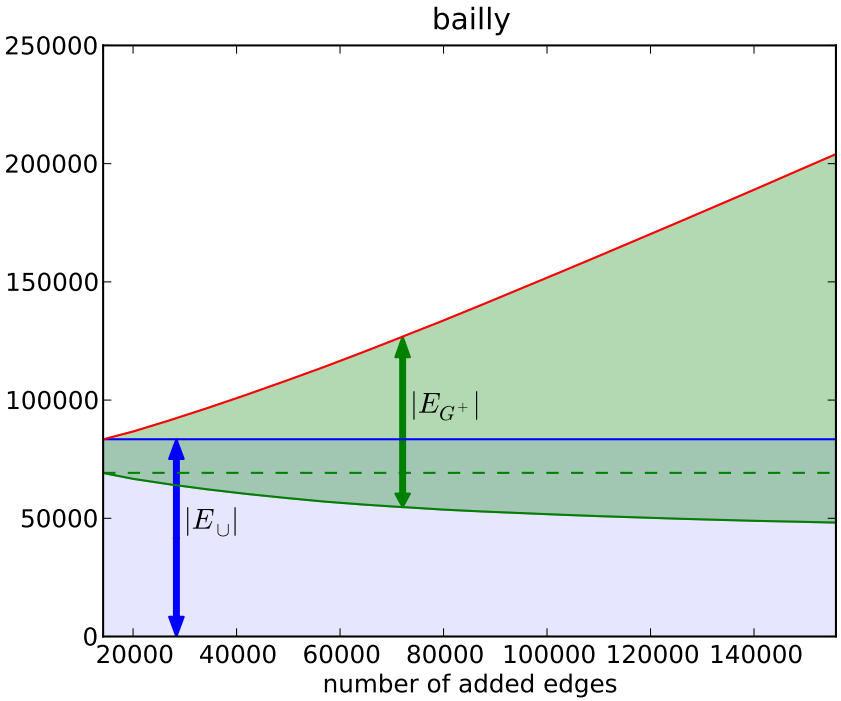


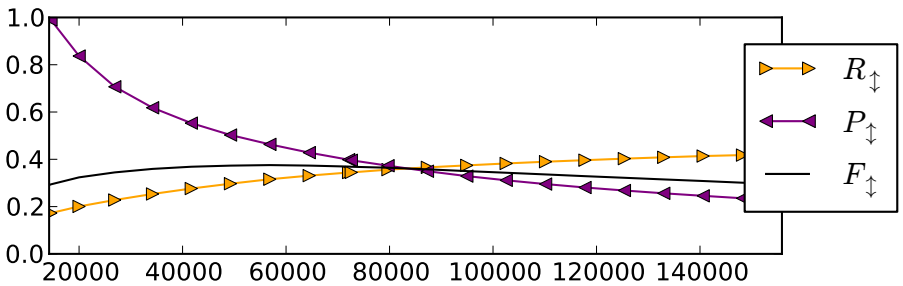
$R_{\updownarrow} = 1,\ P_{\updownarrow} = 0.95,\ F_{\updownarrow} = 0.98$



$R_{\updownarrow} = 1,\ P_{\updownarrow} = 0.96,\ F_{\updownarrow} = 0.98$

Enriched graph vs. union graph (Bailly vs Union Graph)





(1) Edges in $G_\cup$ that are in G or in G^+: 17% to 42%: not so inconsistent !

(2) **not** in $G_\cup$ **and** added to G: what is it?

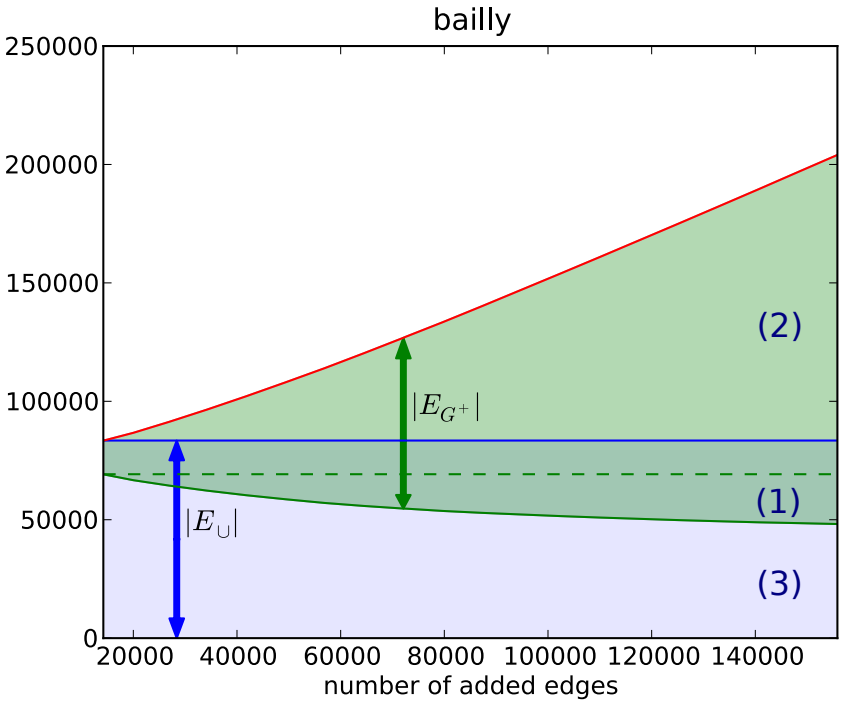
(3) in $G_\cup$ **but not** added to G: still inconsistent...

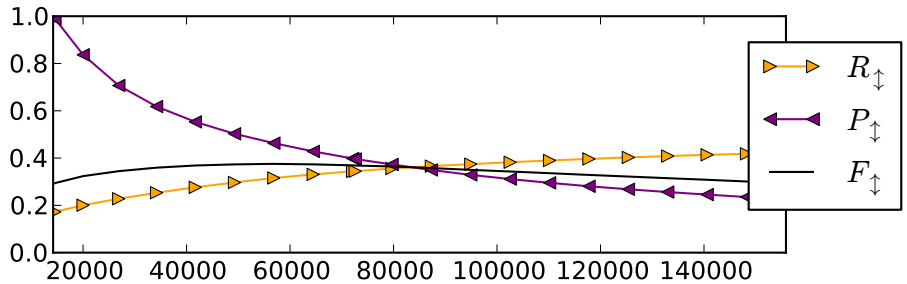
Probability to find an edge of $G_\cup$ by random :

$$\frac{2.|E_{\cup \cap V_{bailly}}|}{|V_{bailly}|.|V_{bailly}|} = \frac{2 * 83430}{12739^2} \approx 0.001$$

if we add 120 000 edges: about 120 in $G_\cup$...

Enriched graph vs. union graph (Bailly vs Union Graph)





(1) Edges in $G_\cup$ that are in G or in G^+:
17% to 42%: not so inconsistent !

(2) **not** in $G_\cup$ **and** added to G:
what is it?

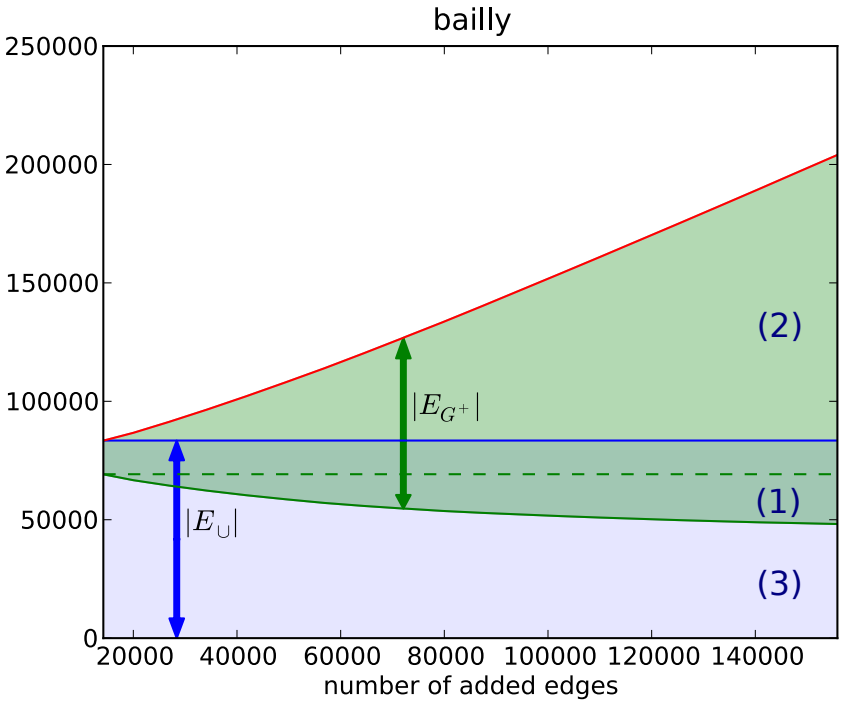
(3) in $G_\cup$ **but not** added to G:
still inconsistent...

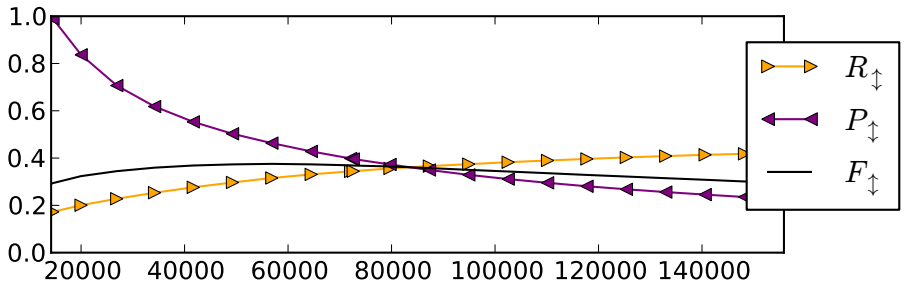
Probability to find an edge of $G_\cup$ by random :

$$\frac{2.|E_{\cup \cap V_{bailly}}|}{|V_{bailly}|.|V_{bailly}|} = \frac{2 * 83430}{12739^2} \approx 0.001$$

if we add 120 000 edges:
about 120 in $G_\cup$...

Enriched graph vs. union graph (Bailly vs Union Graph)





(1) Edges in $G_\cup$ that are in G or in G^+: 17% to 42%: not so inconsistent !

(2) not in $G_\cup$ **and** added to G: what is it?

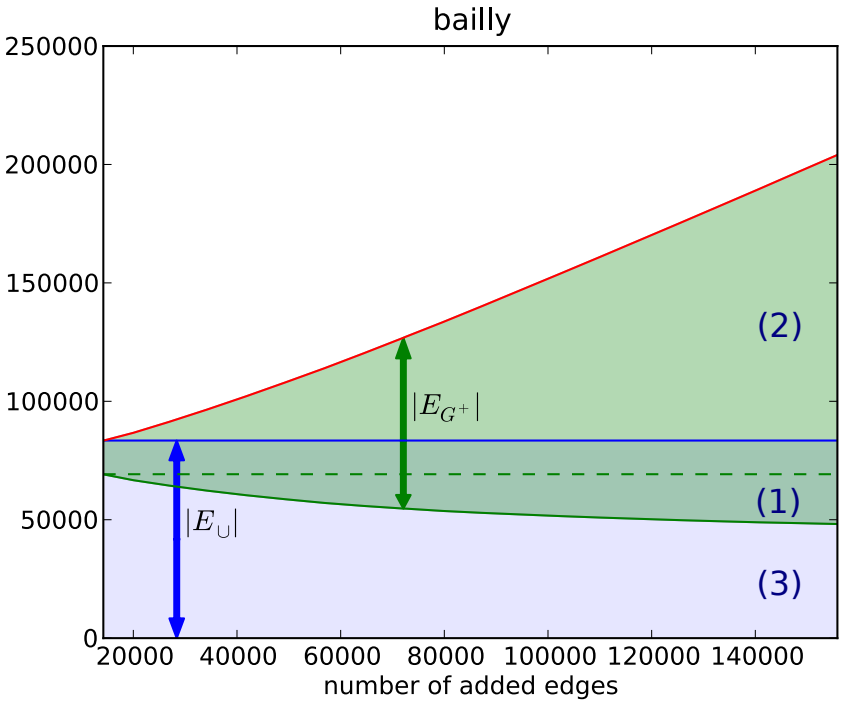
(3) in $G_\cup$ **but not** added to G: still inconsistent...

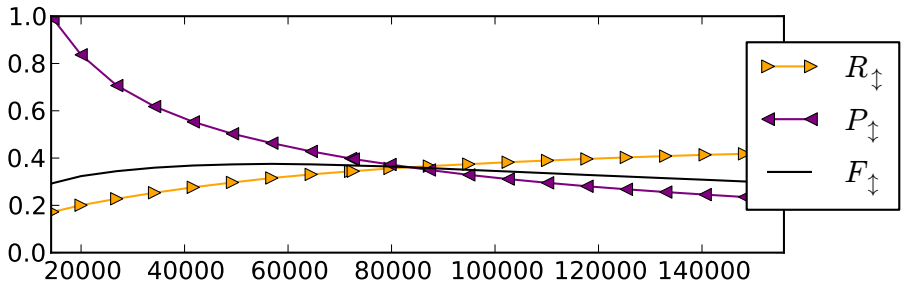
Probability to find an edge of $G_\cup$ by random :

$$\frac{2.|E_{\cup \cap V_{bailly}}|}{|V_{bailly}|.|V_{bailly}|} = \frac{2 * 83430}{12739^2} \approx 0.001$$

if we add 120 000 edges: about 120 in $G_\cup$...

Enriched graph vs. union graph (Bailly vs Union Graph)





(1) Edges in $G_\cup$ that are in G or in G^+:
17% to 42%: not so inconsistent !

(2) **not** in $G_\cup$ **and** added to G:
what is it?

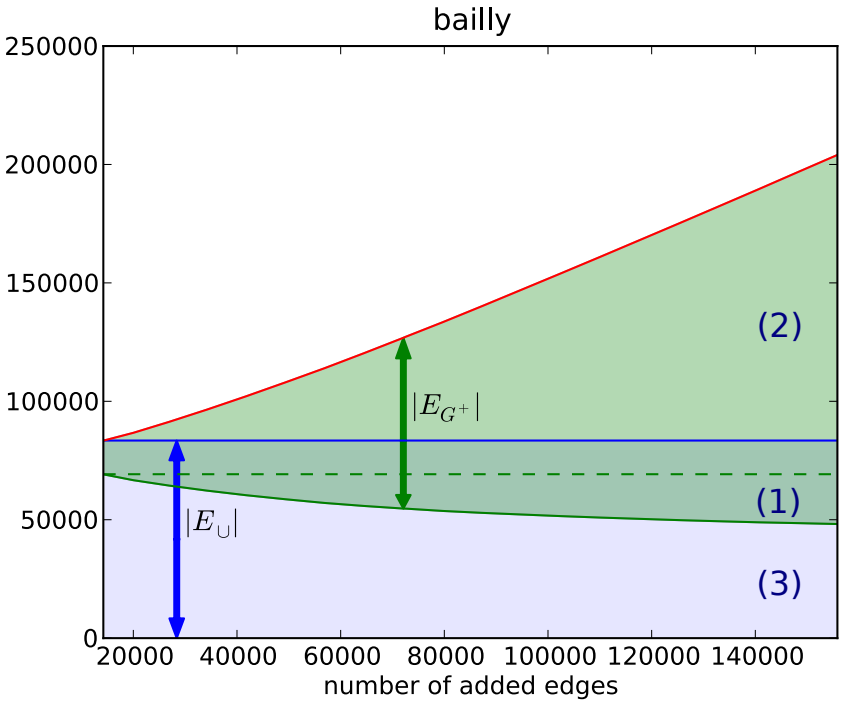
(3) in $G_\cup$ **but not** added to G:
still inconsistent...

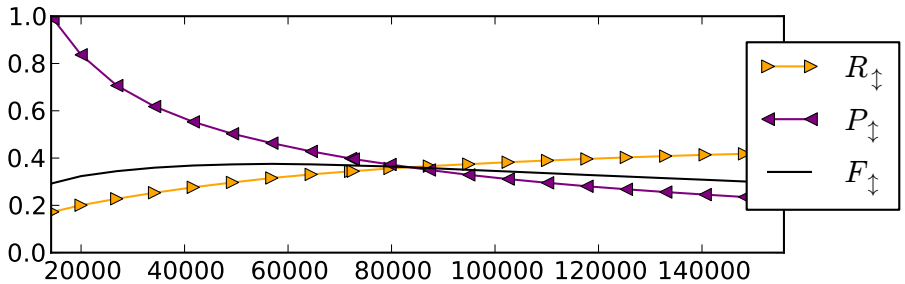
Probability to find an edge of $G_\cup$ by random :

$$\frac{2.|E_{\cup \cap V_{bailly}}|}{|V_{bailly}|.|V_{bailly}|} = \frac{2 * 83430}{12739^2} \approx 0.001$$

if we add 120 000 edges:
about 120 in $G_\cup$...

Enriched graph vs. union graph (Bailly vs Union Graph)





(1) Edges in $G_\cup$ that are in G or in G^+: 17% to 42%: not so inconsistent !

(2) **not** in $G_\cup$ **and** added to G: what is it?

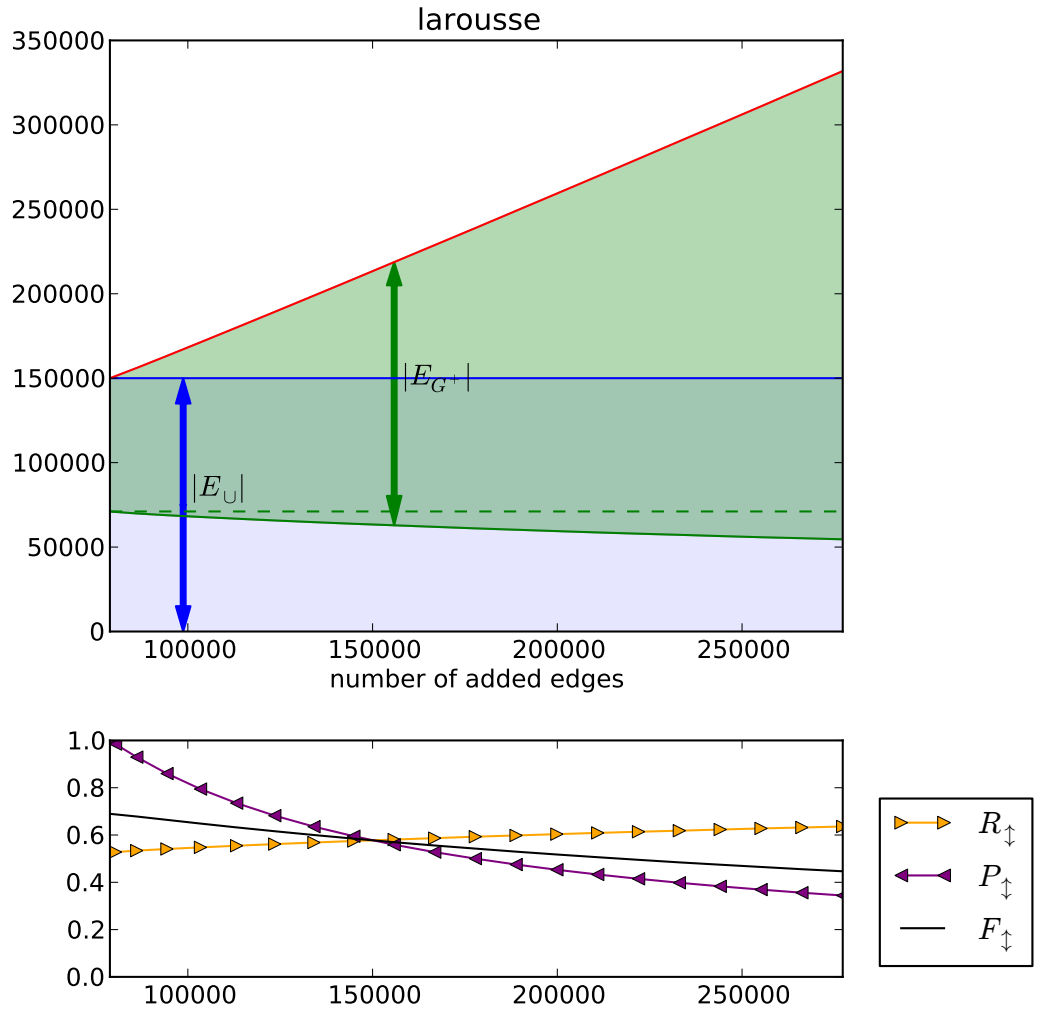
(3) in $G_\cup$ **but not** added to G: still inconsistent...

Probability to find an edge of $G_\cup$ by random :

$$\frac{2.|E_{\cup \cap V_{bailly}}|}{|V_{bailly}|.|V_{bailly}|} = \frac{2 * 83430}{12739^2} \approx 0.001$$

if we add 120 000 edges: about 120 in $G_\cup$...

Enriched graph vs. union graph (Larousse vs Union Graph)



Conclusion and Perspectives

- Binary synonymy is not a good model !
- **however** they may be good proxy to compute a **graded semantic similarity.**
- Taking a step back: a more consistent image,
- Still some inconsistencies :
  - experts may have left out some word senses ie. problem of coverage... (?)
- Graph clustering with overlaps may help expliciting word senses...

Application to semi-automatic enrichment of Wiktionary

http://en.wiktionary.org/wiki/manuscript

manuscript - Wiktionary

తెలుగు
Tiếng Việt
中文

Feedback

Submit anonymous feedback about Wiktionary:

Good
Bad
Messy
Mistake in definition
Confusing
Could not find the word I want
Incomplete
Entry has inaccurate information
Definition is too complicated
If you have time, leave us a note.

Noun

manuscript (*plural* **manuscripts**)

1. A book, composition or any other document, (type)written by hand, not mechanically
2. A single, original copy of a book, article, composition etc, written by hand or even printed, submitted as original for (copy-editing and) reproductive publication.

Derived terms [edit]

- manuscriptal
- manuscription

Synonyms [edit]

- handwrit [-]
- autograph [-]
- handwriting [-]

Wisigoth: candidate synonyms for *manuscript* (N) *[More synonyms]*

script [+] [x]	*copy* [+] [x]	*journal* [+] [x]	*register* [+] [x]
writing [+] [x]	*scenario* [+] [x]	*screenplay* [+] [x]	*inscription* [+] [x]
signature [+] [x]	*original* [+] [x]		

Propose your own synonym here: add

Related terms [edit]

- script

WISIGOTH Project : `http://redac.univ-tlse2.fr/wisigoth/`

Thank you !

Download slides: http://www.irit.fr/~Emmanuel.Navarro/publications.php

Edmonds, P. and Hirst, G. (2002).

Near-synonymy and lexical choice.

Computational Linguistics, 28(2):105—144.

Gaume, B., Mathieu, F., and Navarro, E. (2010).

Building real-world complex networks by wandering on random graphs.

Information - Interaction – Intelligence, 10(1).

Navarro, E., Sajous, F., Gaume, B., Prévot, L., Hsieh, S., Kuo, I., Magistry, P., and Huang, C.-R. (2009).

Wiktionary and NLP: Improving synonymy networks.

In Proceedings of the 2009 ACL-IJCNLP Workshop on The People's Web Meets NLP: Collaboratively Constructed Semantic Resources, pages 19–27, Suntec, Singapore. Association for Computational Linguistics.

Sajous, F., Navarro, E., Gaume, B., Prévot, L., and Chudy, Y. (2010).

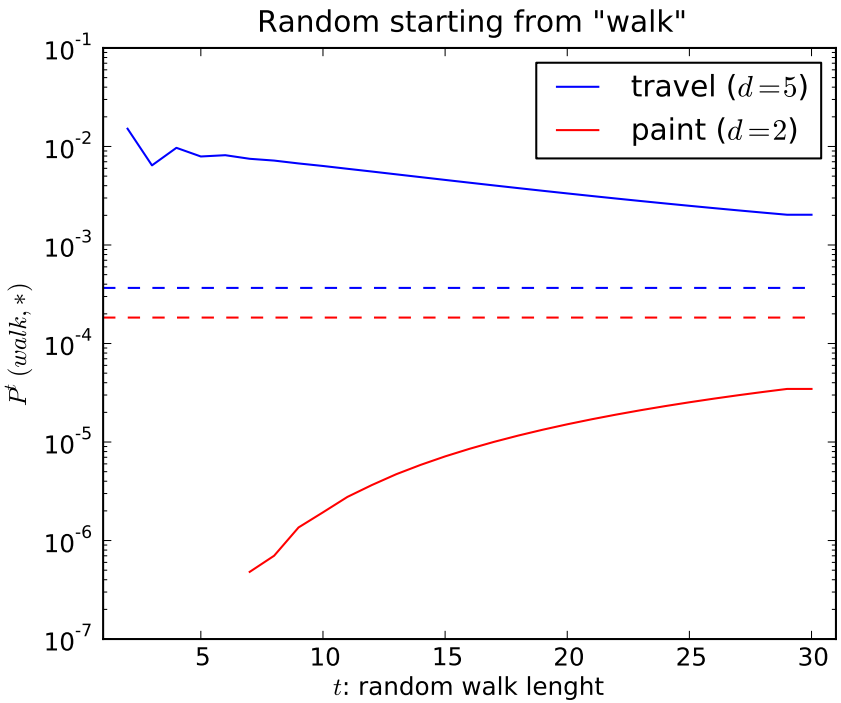
Semi-automatic Endogenous Enrichment of Collaboratively Constructed Lexical Resources: Piggybacking onto Wiktionary.

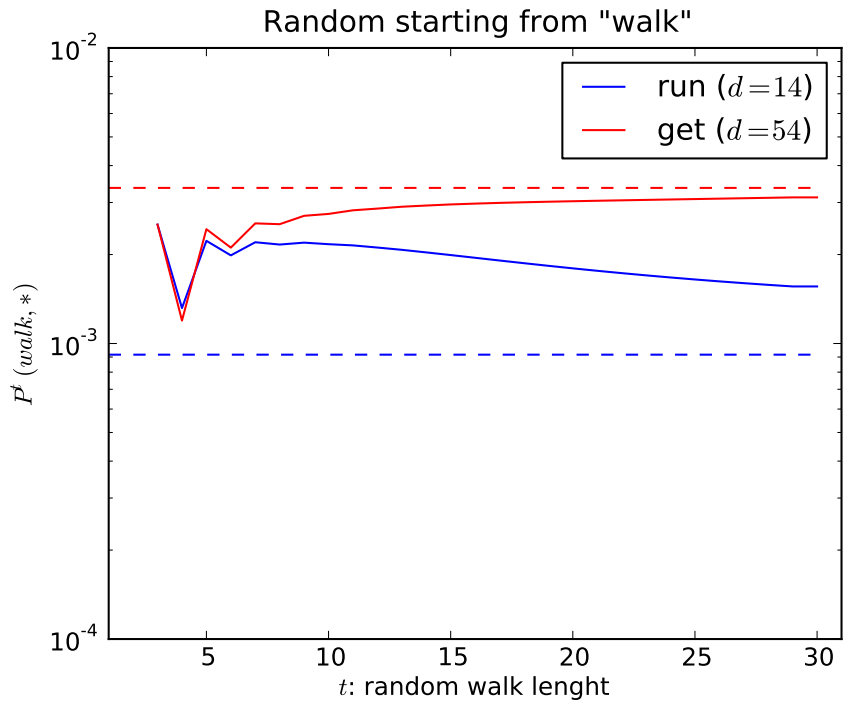
In Loftsson, H., Rögnvaldsson, E., and Helgadóttir, S., editors, Advances in Natural Language Processing, volume 6233 of Lecture Notes in Computer Science, pages 332–344. Springer Berlin / Heidelberg.

Appendix

Proxemy convergence examples

Starting from *walk* in synonymy graph of english *wiktionary* :





- The ratio of $P^t(u, v)$ to the limit $P^\infty(v)$ is more informative !