

Projet EVOLEX

Bruno Gaume

CLLE, CNRS & Univ. Toulouse

Projet EVOLEX

Séminaire du master LITL,
Toulouse, 18 Décembre 2017

- 1 Structure des réseaux lexicaux
- 2 Mesurer la similarité structurelle entre réseaux lexicaux
- 3 Hypothèses

Applications à EVOLEX

- 4 La ressource Jeux de mots
- 5 Résultats

- ▶ Gaume, B., Duvignau, K., Navarro, (2016). Skillex: a graph-based lexical score for measuring the semantic efficiency of used verbs by human subjects describing actions. In Revue TAL : 55, 3; Numéro spécial sur Traitement Automatique des Langues et Sciences Cognitives.
- ▶ Navarro, E., Chudy, Y., Magistry, P., Gaume, B. (2014). BACANAL : Balades Aléatoires Courtes pour ANALyses Lexicales, Application à la substitution lexicale. In Proceedings of TALN-20 2014 : Atelier SEMDIS.

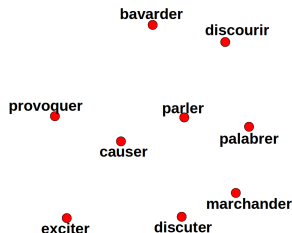
Structure des réseaux lexicaux

$\mathbf{G} = (\mathbf{V}, \mathbf{E})$ où $E \subseteq P_2^V$: un graphe de $n = |V|$ sommets et $m = |E|$ arêtes

Structure des réseaux lexicaux

$\mathbf{G} = (\mathbf{V}, \mathbf{E})$ où $E \subseteq P_2^V$: un graphe de $n = |V|$ sommets et $m = |E|$ arêtes

$\mathbf{V} = \{\text{bavarder, causer, discourir, discuter, exciter, palabrer, parler, p  rer, provoquer}\}$

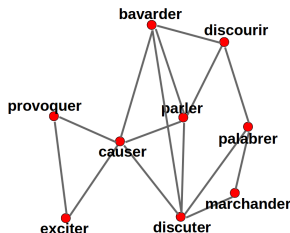


Structure des réseaux lexicaux

$G = (V, E)$ où $E \subseteq P_2^V$: un graphe de $n = |V|$ sommets et $m = |E|$ arêtes

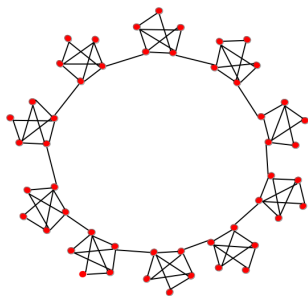
$V = \{\text{bavarder, causer, discourir, discuter, exciter, palabrer, parler, p  rer, provoquer}\}$

$E = \{\{\text{bavarder, parler}\}, \{\text{bavarder, discourir}\}, \{\text{bavarder, causer}\}, \{\text{causer, discuter}\}, \{\text{causer, parler}\}, \{\text{causer, provoquer}\}, \{\text{causer, exciter}\}, \{\text{discourir, parler}\}, \{\text{discourir, palabrer}\}, \{\text{discourir, p  rer}\}, \{\text{discuter, parler}\}, \{\text{discuter, palabrer}\}, \{\text{discuter, p  rer}\}, \{\text{exciter, provoquer}\}, \{\text{palabrer, parler}\}, \{\text{palabrer, p  rer}\}, \{\text{parler, p  rer}\}\}$

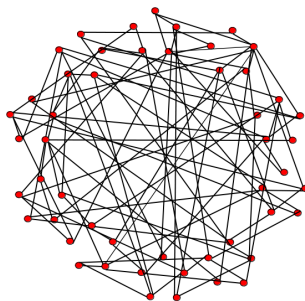
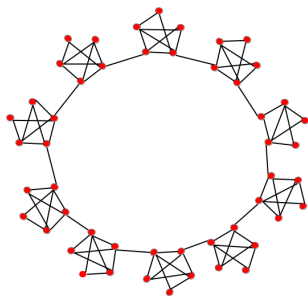


Graphes de laboratoire

Graphes de laboratoire



Graphes de laboratoire



Graphes de terrain

Graphes de terrain

- ▶ Les graphes d'acointance d'un groupe d'humains
- ▶ Le graphe des appels téléphoniques aux USA
- ▶ Le graphe du World Wide Web
- ▶ Le graphe des collaborations cinématographiques
- ▶ Le graphe de *Caenorhabditis elegans*
- ▶ Les graphes Lexicaux

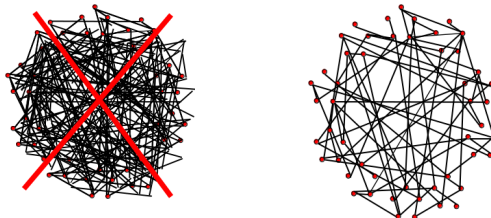
Graphes de terrain

- ▶ Les graphes d'acointance d'un groupe d'humains
- ▶ Le graphe des appels téléphoniques aux USA
- ▶ Le graphe du World Wide Web
- ▶ Le graphe des collaborations cinématographiques
- ▶ Le graphe de *Caenorhabditis elegans*
- ▶ Les graphes Lexicaux

Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

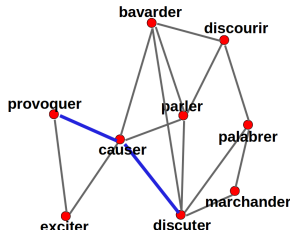
Graphes de terrain



Réseaux Communautaires Hiérarchiques

- ▶ **Faible densité**
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

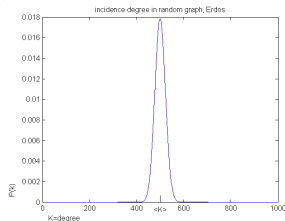
Graphes de terrain



Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ **Chemins courts**
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

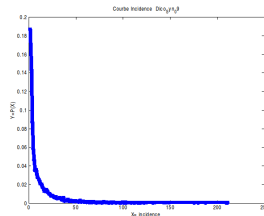
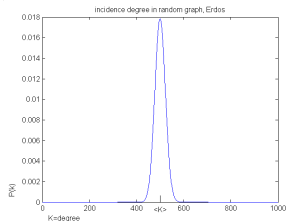
Graphes de terrain



Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ **Distribution des degrés à queue lourde (loi de puissance)**
- ▶ Fort coefficient de clustering : zones denses en arêtes

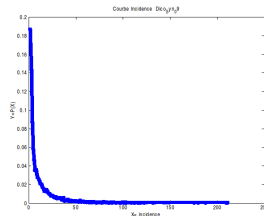
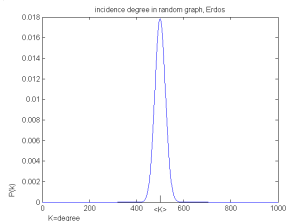
Graphes de terrain



Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ **Distribution des degrés à queue lourde (loi de puissance)**
- ▶ Fort coefficient de clustering : zones denses en arêtes

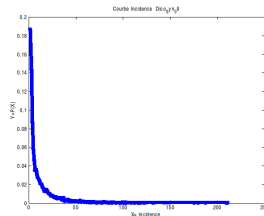
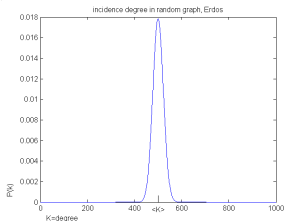
Graphes de terrain



Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ **Distribution des degrés à queue lourde (loi de puissance)**
- ▶ Fort coefficient de clustering : zones denses en arêtes

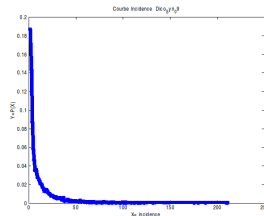
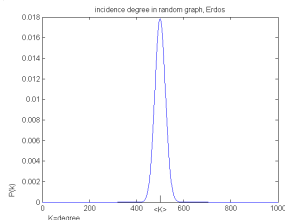
Graphes de terrain



Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ **Distribution des degrés à queue lourde (loi de puissance)**
- ▶ Fort coefficient de clustering : zones denses en arêtes

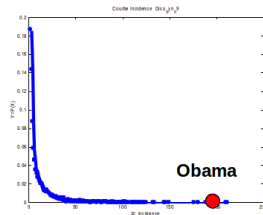
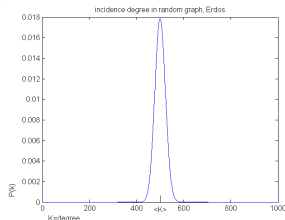
Graphes de terrain



Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ **Distribution des degrés à queue lourde (loi de puissance)**
- ▶ Fort coefficient de clustering : zones denses en arêtes

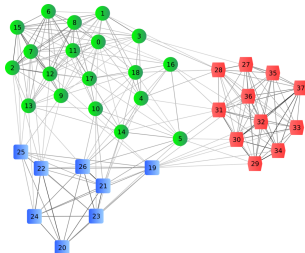
Graphes de terrain



Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ **Distribution des degrés à queue lourde (loi de puissance)**
- ▶ Fort coefficient de clustering : zones denses en arêtes

Graphes de terrain



Réseaux Communautaires Hiérarchiques

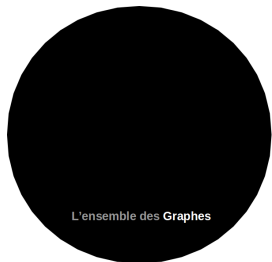
- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ **Fort coefficient de clustering : zones denses en arêtes**

Graphes de terrain

Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

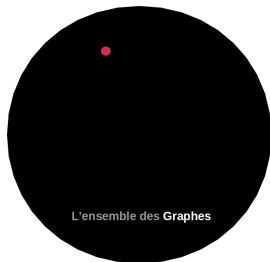
Graphes de terrain



Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

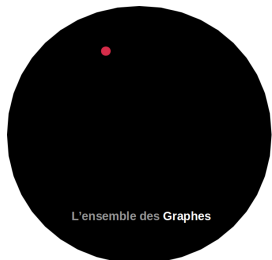
Graphes de terrain



Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

Graphes de terrain

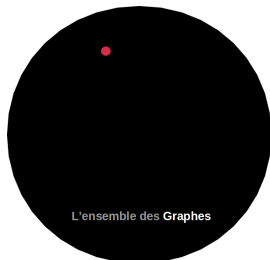


- Les graphes d'acointance d'un groupe d'humains
- Le graphe des appels téléphoniques aux USA
- Le graphe du World Wide Web
- Le graphe des collaborations cinématographiques
- Le graphe de *Caenorhabditis elegans*
- Les graphes Lexicaux

Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

Graphes de terrain



- Les graphes d'acointance d'un groupe d'humains
- Le graphe des appels téléphoniques aux USA
- Le graphe du World Wide Web
- Le graphe des collaborations cinématographiques
- Le graphe de Caenorhabditis elegans
- Les graphes Lexicaux

Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ **Distribution des degrés à queue lourde (loi de puissance)**
- ▶ **Fort coefficient de clustering : zones denses en arêtes**

1 Structure des réseaux lexicaux

2 Mesurer la similarité structurelle entre réseaux lexicaux

3 Hypothèses

Applications à EVOLEX

4 La ressource Jeux de mots

5 Résultats

Problème

Hypothèse : Les relations lexicales reflètent les structures sémantiques du lexique communes aux membres d'une même communauté linguistique

Hypothèse : Les relations lexicales reflètent les structures sémantiques du lexique communes aux membres d'une même communauté linguistique

Incompatible avec :

Constat : Il existe de nombreux désaccords entre les membres d'une même communauté linguistique :

- **EXEMPLE** : **Beaucoup** de paires de mots sont synonymes pour l'individu I_1 et pas pour l'individu I_2 et vice et versa



Hypothèse : Les relations lexicales reflètent les structures sémantiques du lexique communes aux membres d'une même communauté linguistique

Résoudre cette contradiction apparente

Constat : Il existe de nombreux désaccords entre les membres d'une même communauté linguistique :

- **EXEMPLE** : **Beaucoup** de paires de mots sont synonymes pour l'individu I_1 et pas pour l'individu I_2 et vice et versa



bavarder : parler, discourir, causer

causer : parler, discuter, bavarder, provoquer, exciter

discourir : parler, palabrer, pérorer, bavarder

discuter : parler, palabrer, causer, pérorer

exciter : causer, provoquer

palabrer : parler, discourir, pérorer, discuter

parler : palabrer, discourir, causer, pérorer, discuter, bavarder

pérorer : parler, palabrer, discourir, discuter

provoquer : causer, exciter

bavarder : parler, discourir, causer

causer : parler, discuter, bavarder, provoquer, exciter

discourir : parler, palabrer, pérorer, bavarder

discuter : parler, palabrer, causer, pérorer

exciter : causer, provoquer

palabrer : parler, discourir, pérorer, discuter

parler : palabrer, discourir, causer, pérorer, discuter, bavarder

pérorer : parler, palabrer, discourir, discuter

provoquer : causer, exciter

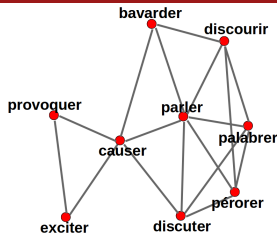
$$G = (V, E) \text{ où } E \subseteq P_2^V$$

$$V = \{\text{bavarder, causer, discourir, discuter, exciter, palabrer, parler, pérorer, provoquer}\}$$

$$E = \{\{\text{bavarder, parler}\}, \{\text{bavarder, discourir}\}, \{\text{bavarder, causer}\}, \{\text{causer, discuter}\}, \\ \{\text{causer, parler}\}, \{\text{causer, provoquer}\}, \{\text{causer, exciter}\}, \{\text{discourir, parler}\}, \\ \{\text{discourir, palabrer}\}, \{\text{discourir, pérorer}\}, \{\text{discuter, parler}\}, \{\text{discuter, palabrer}\}, \\ \{\text{discuter, pérorer}\}, \{\text{exciter, provoquer}\}, \{\text{palabrer, parler}\}, \{\text{palabrer, pérorer}\}, \\ \{\text{parler, pérorer}\}\}$$

Un graphe de $n = |V|$ sommets et $m = |E|$ arêtes

Réseaux lexicaux construits à partir de ressources électroniques

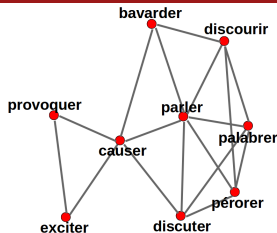


$G = (V, E)$ où $E \subseteq P_2^V$

$V = \{\text{bavarder, causer, discourir, discuter, exciter, palabrer, parler, péroter, provoquer}\}$

$E = \{\{\text{bavarder, parler}\}, \{\text{bavarder, discourir}\}, \{\text{bavarder, causer}\}, \{\text{causer, discuter}\}, \{\text{causer, parler}\}, \{\text{causer, provoquer}\}, \{\text{causer, exciter}\}, \{\text{discourir, parler}\}, \{\text{discourir, palabrer}\}, \{\text{discourir, péroter}\}, \{\text{discuter, parler}\}, \{\text{discuter, palabrer}\}, \{\text{discuter, péroter}\}, \{\text{exciter, provoquer}\}, \{\text{palabrer, parler}\}, \{\text{palabrer, péroter}\}, \{\text{parler, péroter}\}\}$

Un graphe de $n = |V|$ sommets et $m = |E|$ arêtes



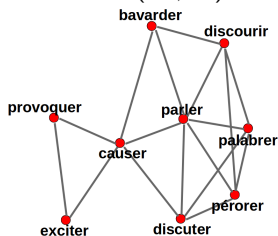
Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde (loi de puissance)
- ▶ Fort coefficient de clustering : zones denses en arêtes

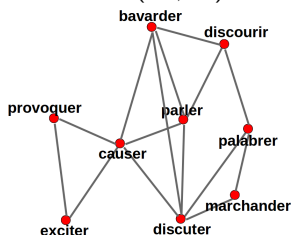
Comparer E_1 et E_2 comme des «sacs de liens»

Comparer E_1 et E_2 comme des «sacs de liens»

$$G_1 = (V_1, E_1)$$

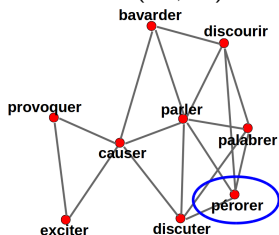


$$G_2 = (V_2, E_2)$$

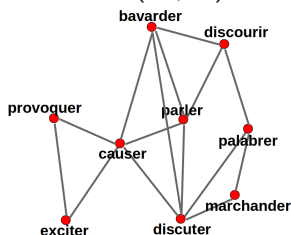


Comparer E_1 et E_2 comme des «sacs de liens»

$$G_1 = (V_1, E_1)$$

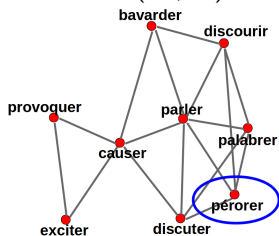


$$G_2 = (V_2, E_2)$$

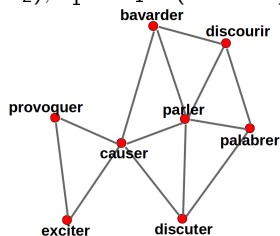


Comparer E_1 et E_2 comme des «sacs de liens»

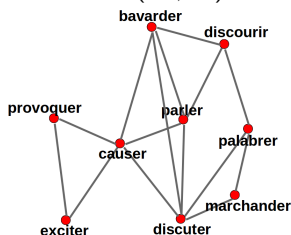
$$G_1 = (V_1, E_1)$$



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

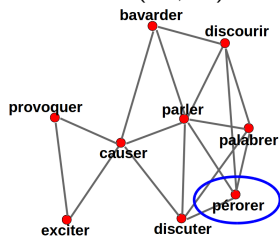


$$G_2 = (V_2, E_2)$$

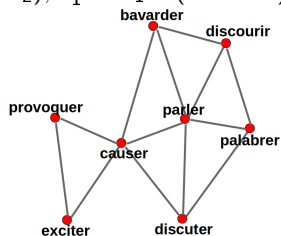


Comparer E_1 et E_2 comme des «sacs de liens»

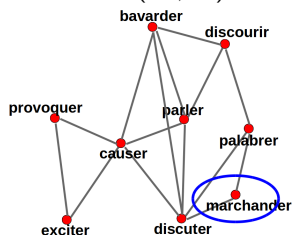
$$G_1 = (V_1, E_1)$$



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

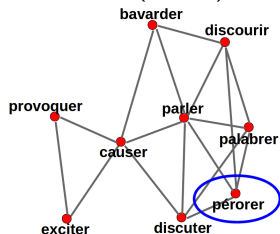


$$G_2 = (V_2, E_2)$$

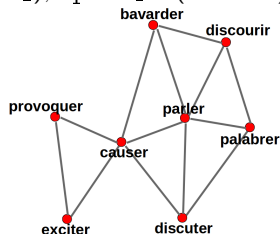


Comparer E_1 et E_2 comme des «sacs de liens»

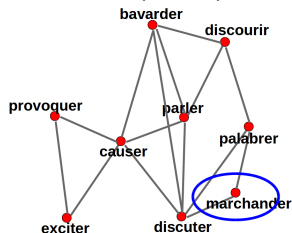
$$G_1 = (V_1, E_1)$$



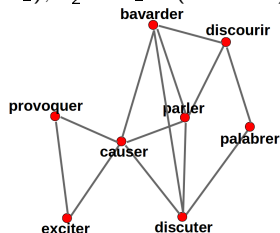
$$G'_1 = (V' = (V_1 \cap V_2), E'_1 = E_1 \cap (V' \times V'))$$



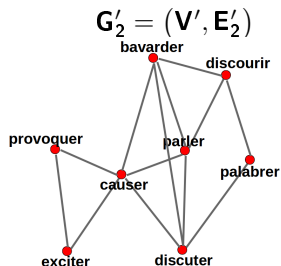
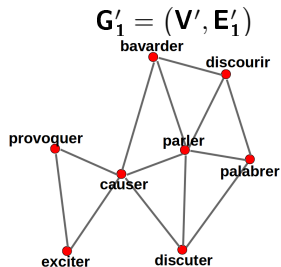
$$G_2 = (V_2, E_2)$$



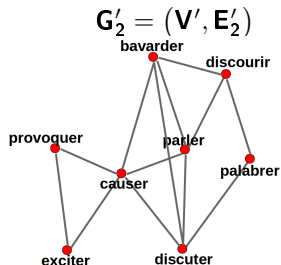
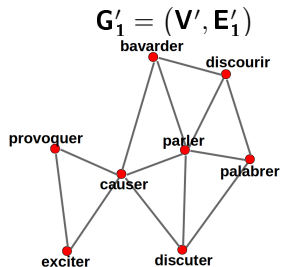
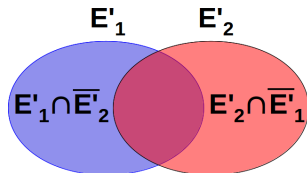
$$G'_2 = (V' = (V_1 \cap V_2), E'_2 = E_2 \cap (V' \times V'))$$



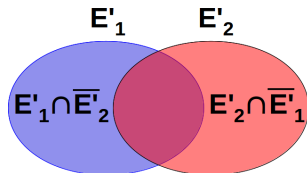
Comparer E_1 et E_2 comme des «sacs de liens»



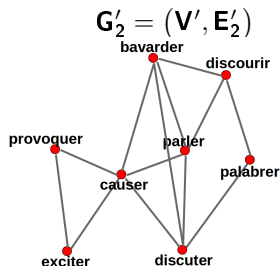
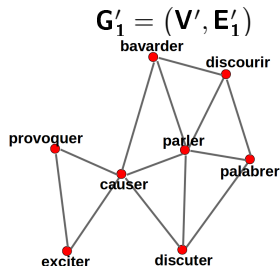
Comparer E_1 et E_2 comme des «sacs de liens»



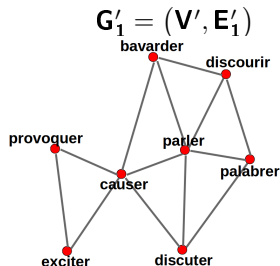
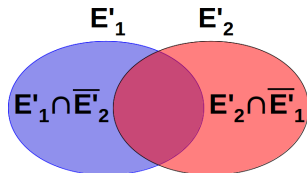
Comparer E_1 et E_2 comme des «sacs de liens»



$$GED(G'_1, G'_2) = \frac{|E'_1 \cap \overline{E'_2}| + |E'_2 \cap \overline{E'_1}|}{|E'_1| + |E'_2|} \in [0, 1]$$

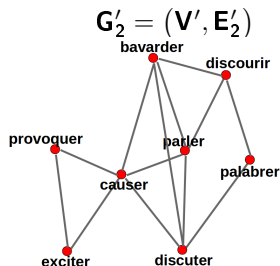


Comparer E_1 et E_2 comme des «sacs de liens»

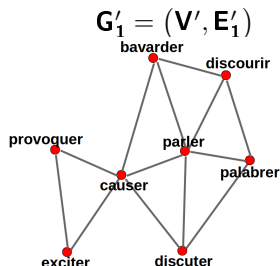
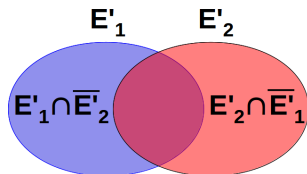


$$GED(G'_1, G'_2) = \frac{|E'_1 \cap \overline{E'_2}| + |E'_2 \cap \overline{E'_1}|}{|E'_1| + |E'_2|} \in [0, 1]$$

$$GED(Lar', Rob') = 0.47$$



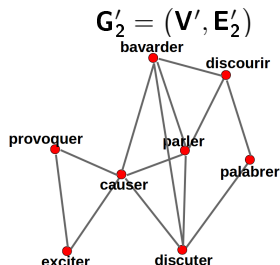
Comparer E_1 et E_2 comme des «sacs de liens»



$$GED(G'_1, G'_2) = \frac{|E'_1 \cap \overline{E'_2}| + |E'_2 \cap \overline{E'_1}|}{|E'_1| + |E'_2|} \in [0, 1]$$

$$GED(Lar', Rob') = 0.47$$

Nombreux désaccords entre Larousse et Robert

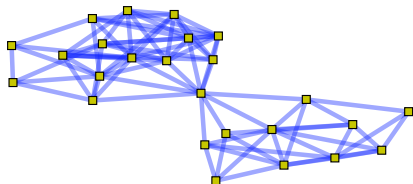


Hypothèse

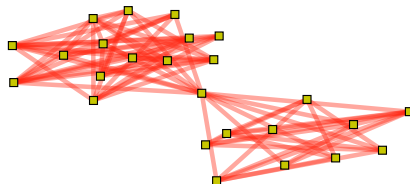
Nombreux désaccords **au niveau des arêtes**

Nombreux désaccords **au niveau des arêtes**

$$G_1 = (V, E_1)$$



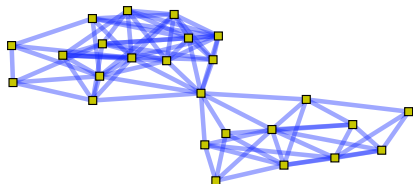
$$G_2 = (V, E_2)$$



Aucune arête commune entre ces deux graphes.

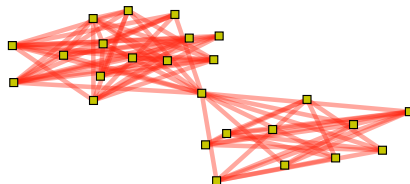
Nombreux désaccords **au niveau des arêtes**

$$G_1 = (V, E_1)$$



$$\text{GED}(G_1, G_2) = 1$$

$$G_2 = (V, E_2)$$



Aucune arête commune entre ces deux graphes.

Nombreux désaccords **au niveau des arêtes**

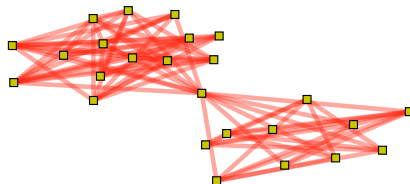
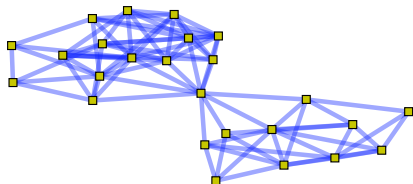
Mais

Bon accord **au niveau des clusters...**

$$G_1 = (V, E_1)$$

$$\text{GED}(G_1, G_2) = 1$$

$$G_2 = (V, E_2)$$



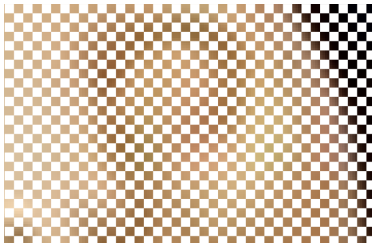
Aucune arête commune entre ces deux graphes.

Métaphore

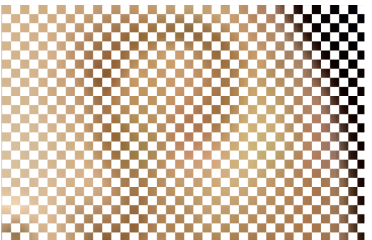
Métaphore



La photo d'un chat



Chaque pixel **impair** est peint en blanc



Chaque pixel **pair** est peint en blanc

Métaphore

Pair =

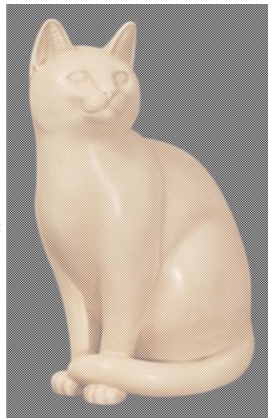


Métaphore

Pair =



Impair =



Métaphore

Pair =



Impair =



Aucun pixel commun

Images distantes (au niveau des pixels)

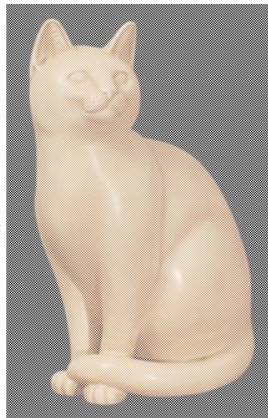
Cependant ces deux images se ressemblent

Métaphore

Pair =



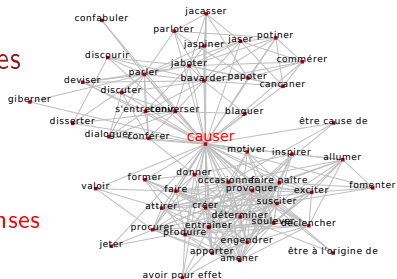
Impair =



Marches Aléatoires

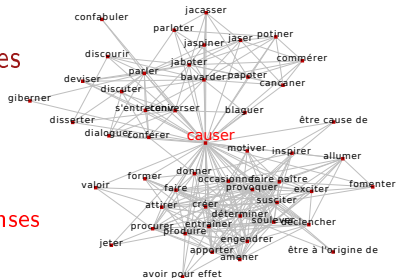
Réseaux Communautaires Hiérarchiques

- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde
- ▶ Fort coefficient de clustering : **zones denses**



Réseaux Communautaires Hiérarchiques

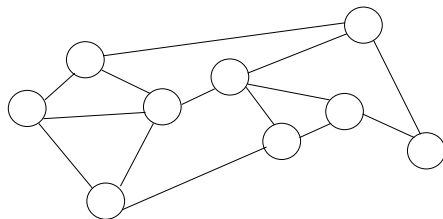
- ▶ Faible densité
- ▶ Chemins courts
- ▶ Distribution des degrés à queue lourde
- ▶ Fort coefficient de clustering : **zones denses**



Marches Aléatoires

- ▶ **Idee:** si (u, v) sont **dans un même “cluster”**, alors **beaucoup de chemins** mènent de u à v (même si ils ne sont pas adjacents).
- ▶ Un **marcheur aléatoire** a **tendance à être capturé par les “clusters”**

Marches Aléatoires



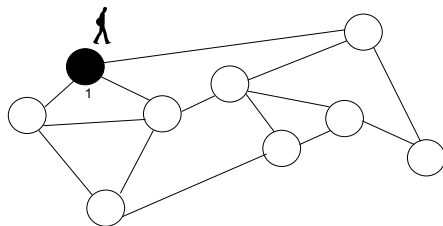
1 État initial ($t=0$)

2 $t = 1$

3 $t = 2$

4 $t = 3$

Marches Aléatoires



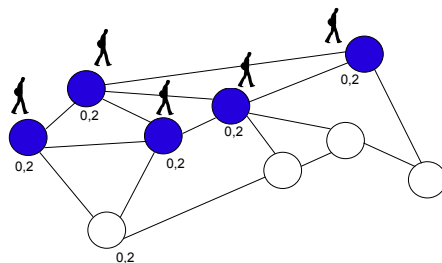
1 État initial ($t=0$)

2 $t = 1$

3 $t = 2$

4 $t = 3$

Marches Aléatoires



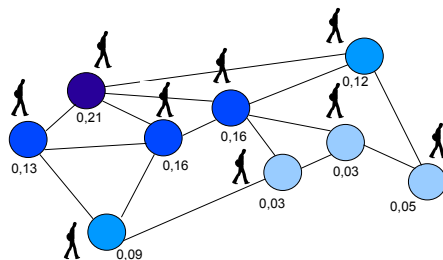
1 État initial ($t=0$)

2 $t = 1$

3 $t = 2$

4 $t = 3$

Marches Aléatoires



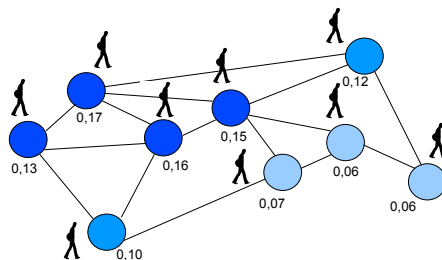
1 État initial ($t=0$)

2 $t = 1$

3 $t = 2$

4 $t = 3$

Marches Aléatoires



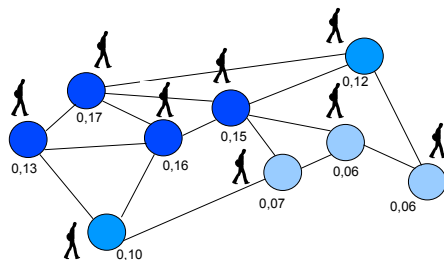
1 État initial ($t=0$)

2 $t = 1$

3 $t = 2$

4 $t = 3$

Marches Aléatoires



- 1 État initial ($t=0$)
- 2 $t = 1$
- 3 $t = 2$
- 4 $t = 3$

La Proxémie : $P_G^t(u, v)$

Dans G , probabilité d'être sur v à l'instant t en partant de u

Marches Aléatoires

Liste ordonnées des verbes selon leur score de proxémie à partir de *déshabiller* ($t = 4$)

1) dépouiller*	11) dévoiler	21) arracher	31) tailler	41) dire
2) défaire*	12) dégager	22) délacer*	32) développer	42) ouvrir
3) démunir*	13) étaler*	23) voler	33) déchirer	43) déployer
4) déshabiller*	14) ôter*	24) dénouer	34) déchausser	44) dépiauter
5) dévêtir*	15) enlever	25) desserrer	35) vider	45) déposséder
6) révéler	16) défrusquer	26) médire*	36) peler	46) frustrer
7) découvrir*	17) afficher	27) débarrasser	37) dépourvoir	47) indiquer
8) montrer*	18) désaffubler	28) exposer	38) défubler	48) démontrer
9) dénuder*	19) écorcher*	29) couper	39) priver	49) manifester
10) dégarnir*	20) démasquer*	30) exhiber*	40) trahir	50) détacher

Table : Les cinquante verbes qui ont la plus forte probabilité d'être atteints en 4 pas depuis *déshabiller* (sur 10000 verbes).

La Proxémie : $P_G^t(u, v)$

Dans G , probabilité d'être sur v à l'instant t en partant de u

Marches Aléatoires

Liste ordonnées des verbes selon leur score de proxémie à partir de *déshabiller*
($t = 4$)

1) dépouiller*	11) dévoiler	21) arracher	31) tailler	41) dire
2) défaire*	12) dégager	22) délacer*	32) développer	42) ouvrir
3) démunir*	13) étaler*	23) voler	33) déchirer	43) déployer
4) déshabiller*	14) ôter*	24) dénouer	34) déchausser	44) dépiauter
5) dévêtir*	15) enlever	25) desserrer	35) vider	45) déposséder
6) révéler	16) défrusquer	26) médire*	36) peler	46) frustrer
7) découvrir*	17) afficher	27) débarrasser	37) dépourvoir	47) indiquer
8) montrer*	18) désaffubler	28) exposer	38) défubler	48) démontrer
9) dénuder*	19) écorcher*	29) couper	39) priver	49) manifester
10) dégarnir*	20) démasquer*	30) exhiber*	40) trahir	50) détacher

Table : Les cinquante verbes qui ont la plus forte probabilité d'être atteints en 4 pas depuis *déshabiller* (sur 10000 verbes).

La Proxémie : $P_G^t(u, v)$

Dans G , probabilité d'être sur v à l'instant t en partant de u

Marches Aléatoires

Liste ordonnées des verbes selon leur score de proxémie à partir de *déshabiller* ($t = 4$)

1) dépouiller*	11) dévoiler	21) arracher	31) tailler	41) dire
2) défaire*	12) dégager	22) délacer*	32) développer	42) ouvrir
3) démunir*	13) étaler*	23) voler	33) déchirer	43) déployer
4) déshabiller*	14) ôter*	24) dénouer	34) déchausser	44) dépiauter
5) dévêtir*	15) enlever	25) desserrer	35) vider	45) déposséder
6) révéler	16) défrusquer	26) médire*	36) peler	46) frustrer
7) découvrir*	17) afficher	27) débarrasser	37) dépourvoir	47) indiquer
8) montrer*	18) désaffubler	28) exposer	38) défubler	48) démontrer
9) dénuder*	19) écorcher*	29) couper	39) priver	49) manifester
10) dégarnir*	20) démasquer*	30) exhiber*	40) trahir	50) détacher

Table : Les cinquante verbes qui ont la plus forte probabilité d'être atteints en 4 pas depuis *déshabiller* (sur 10000 verbes).

La Proxémie : $P_G^t(u, v)$

Dans G , probabilité d'être sur v à l'instant t en partant de u

Marches Aléatoires

Liste ordonnées des verbes selon leur score de proxémie à partir de *déshabiller* ($t = 4$)

1) dépouiller*	11) dévoiler	21) arracher	31) tailler	41) dire
2) défaire*	12) dégager	22) délacer*	32) développer	42) ouvrir
3) démunir*	13) étaler*	23) voler	33) déchirer	43) déployer
4) déshabiller*	14) ôter*	24) dénouer	34) déchausser	44) dépiauter
5) dévêtir*	15) enlever	25) desserrer	35) vider	45) déposséder
6) révéler	16) défrusquer	26) médire*	36) peler	46) frustrer
7) découvrir*	17) afficher	27) débarrasser	37) dépourvoir	47) indiquer
8) montrer*	18) désaffubler	28) exposer	38) défubler	48) démontrer
9) dénuder*	19) écorcher*	29) couper	39) priver	49) manifester
10) dégarnir*	20) démasquer*	30) exhiber*	40) trahir	50) détacher

Table : Les cinquante verbes qui ont la plus forte probabilité d'être atteints en 4 pas depuis *déshabiller* (sur 10000 verbes).

La Proxémie : $P_G^t(u, v)$

Dans G , probabilité d'être sur v à l'instant t en partant de u

$P_G^t(u, v)$ = Dans G , probabilité d'être sur v à l'instant t en partant de u .

- **Marches longues:** la probabilité d'atteindre un sommet v ne dépend que du degrés de v :

$$\lim_{t \rightarrow \infty} P_G^t(u, v) = \frac{\deg_G(v)}{\sum_{r \in V} \deg_G(r)} = \pi_G(v)$$

$P_G^t(u, v)$ = Dans G , probabilité d'être sur v à l'instant t en partant de u .

- **Marches longues:** la probabilité d'atteindre un sommet v ne dépend que du degrés de v :

$$\lim_{t \rightarrow \infty} P_G^t(u, v) = \frac{\deg_G(v)}{\sum_{r \in V} \deg_G(r)} = \pi_G(v)$$

- **Marches courtes:** forte probabilité de rester dans un cluster:

- $P_G^t(u, v) > \pi_G(v)$ si u et v sont dans un même cluster
- $P_G^t(u, v) < \pi_G(v)$ sinon

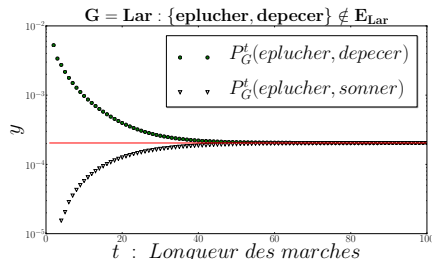
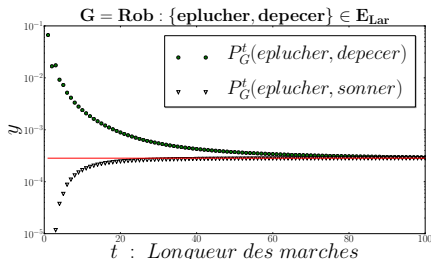
$P_G^t(u, v)$ = Dans G , probabilité d'être sur v à l'instant t en partant de u .

- **Marches longues:** la probabilité d'atteindre un sommet v ne dépend que du degrés de v :

$$\lim_{t \rightarrow \infty} P_G^t(u, v) = \frac{\deg_G(v)}{\sum_{r \in V} \deg_G(r)} = \pi_G(v)$$

- **Marches courtes:** forte probabilité de rester dans un cluster:

- $P_G^t(u, v) > \pi_G(v)$ si u et v sont dans un même cluster
- $P_G^t(u, v) < \pi_G(v)$ sinon



$P_G^t(u, v)$ = Dans G , probabilité d'être sur v à l'instant t en partant de u .

- **Marches longues:** la probabilité d'atteindre un sommet v ne dépend que du degrés de v :

$$\lim_{t \rightarrow \infty} P_G^t(u, v) = \frac{\deg_G(v)}{\sum_{r \in V} \deg_G(r)} = \pi_G(v)$$

- **Marches courtes:** forte probabilité de rester dans un cluster:

- $P_G^t(u, v) > \pi_G(v)$ si u et v sont dans un même cluster
- $P_G^t(u, v) < \pi_G(v)$ sinon

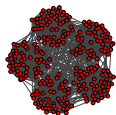
$$t = 5 : \quad \text{CONF}_G(u, v) = \frac{P_G^5(u, v)}{P_G^5(u, v) + \pi_G(v)}$$

CONF_G est une mesure de proximité normalisée basée sur les marches aléatoires courtes dans G :

- Beaucoup de chemins courts entre u et v : $\text{CONF}_G(u, v) > 0,5$;
- Peu de chemins courts entre u et v : $\text{CONF}_G(u, v) < 0,5$;

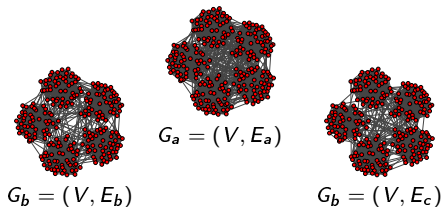
Confluence: Expérience contrôlée

Confluence: Expérience contrôlée

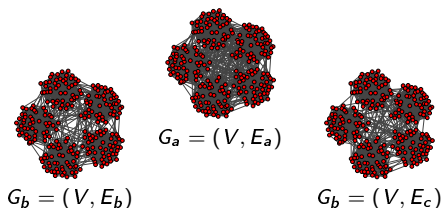


$$G_a = (V, E_a)$$

Confluence: Expérience contrôlée

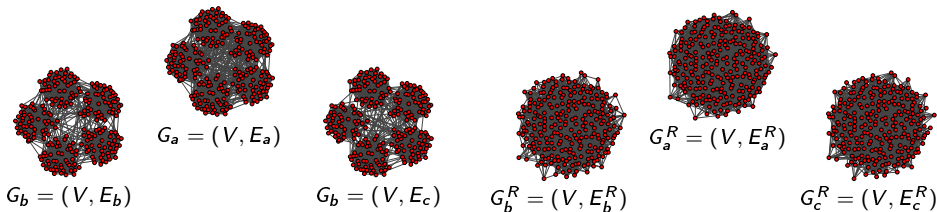


Confluence: Expérience contrôlée



$$E_b \cap E_c = \emptyset \Rightarrow$$
$$GED(G_b, G_c) = 1$$

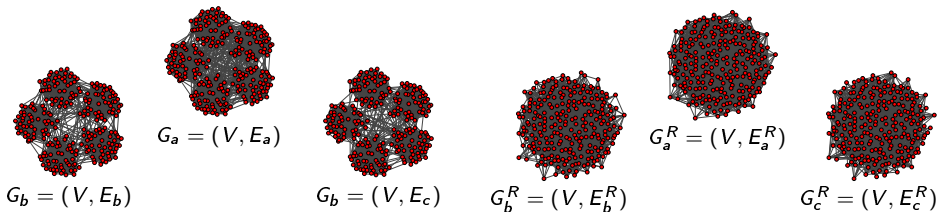
Confluence: Expérience contrôlée



$$E_b \cap E_c = \emptyset \Rightarrow \\ GED(G_b, G_c) = 1$$

$$E_b^R \cap E_c^R = \emptyset \Rightarrow \\ GED(G_b^R, G_c^R) = 1$$

Confluence: Expérience contrôlée

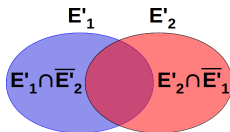


$$E_b \cap E_c = \emptyset \Rightarrow$$

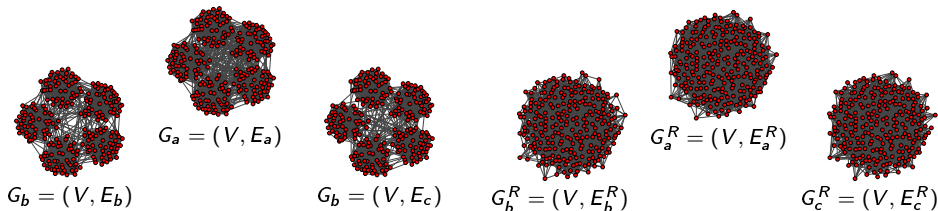
$$GED(G_b, G_c) = 1$$

$$E_b^R \cap E_c^R = \emptyset \Rightarrow$$

$$GED(G_b^R, G_c^R) = 1$$



Confluence: Expérience contrôlée

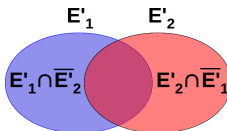
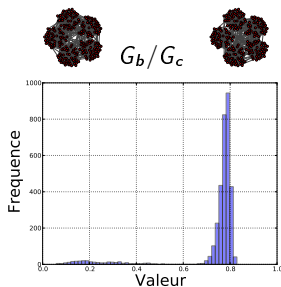


$$E_b \cap E_c = \emptyset \Rightarrow$$

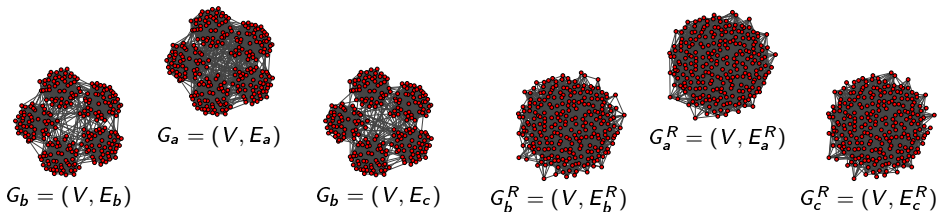
$$GED(G_b, G_c) = 1$$

$$E_b^R \cap E_c^R = \emptyset \Rightarrow$$

$$GED(G_b^R, G_c^R) = 1$$

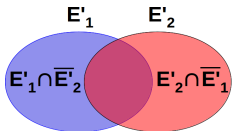
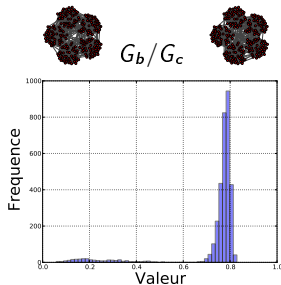


Confluence: Expérience contrôlée



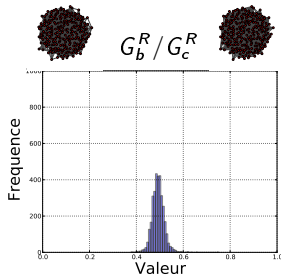
$$E_b \cap E_c = \emptyset \Rightarrow$$

$$GED(G_b, G_c) = 1$$



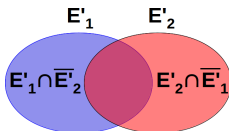
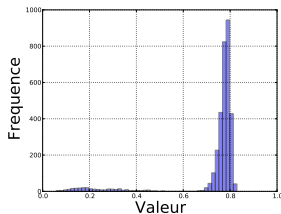
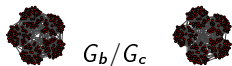
$$E_b^R \cap E_c^R = \emptyset \Rightarrow$$

$$GED(G_b^R, G_c^R) = 1$$

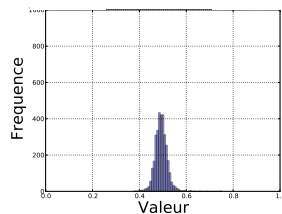
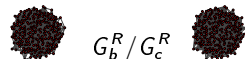


Confluence: Expérience contrôlée

$$GED(G_b, G_c) = 1$$



$$GED(G_b^R, G_c^R) = 1$$

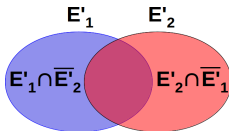
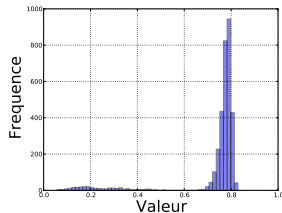
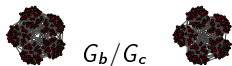


Confluence: Expérience contrôlée

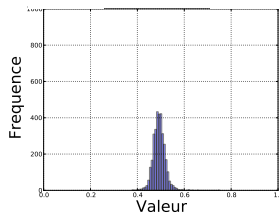
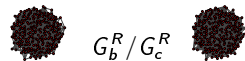
$$\mu(\mathbf{G}_1, \mathbf{G}_2) =$$

$$\frac{1}{|E_1 \cap \bar{E}_2| + |E_2 \cap \bar{E}_1|} \left(\sum_{\{u,v\} \in (E_2 \cap \bar{E}_1)} \text{CONF}_{G_1}(\{u,v\}) + \sum_{\{u,v\} \in (E_1 \cap \bar{E}_2)} \text{CONF}_{G_2}(\{u,v\}) \right)$$

$$\text{GED}(G_b, G_c) = 1$$



$$\text{GED}(G_b^R, G_c^R) = 1$$



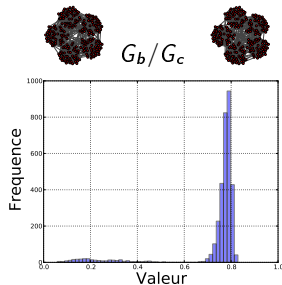
Confluence: Expérience contrôlée

$$\mu(\mathbf{G}_1, \mathbf{G}_2) =$$

$$\frac{1}{|E_1 \cap \bar{E}_2| + |E_2 \cap \bar{E}_1|} \left(\sum_{\{u,v\} \in (E_2 \cap \bar{E}_1)} \text{CONF}_{G_1}(\{u,v\}) + \sum_{\{u,v\} \in (E_1 \cap \bar{E}_2)} \text{CONF}_{G_2}(\{u,v\}) \right)$$

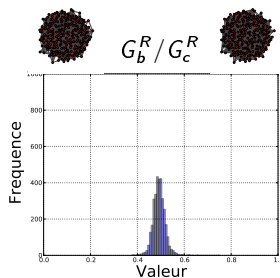
$$\mu(G_b, G_c) = 0.74$$

$$\text{GED}(G_b, G_c) = 1$$



$$\mu(G_b^R, G_c^R) = 0.49$$

$$\text{GED}(G_b^R, G_c^R) = 1$$



Comparer les réseaux lexicaux

Comparer les réseaux lexicaux

$$\mu(\mathbf{G}_1, \mathbf{G}_2) = \frac{1}{|E_1 \cap \overline{E_2}| + |\overline{E_2} \cap E_1|} \left(\sum_{\{u,v\} \in (E_2 \cap \overline{E_1})} \text{CONF}_{\mathbf{G}_1}(\{u,v\}) + \sum_{\{u,v\} \in (E_1 \cap \overline{E_2})} \text{CONF}_{\mathbf{G}_2}(\{u,v\}) \right)$$

Robert / Larousse

Comparer les réseaux lexicaux

$$\mu(\mathbf{G}_1, \mathbf{G}_2) = \frac{1}{|E_1 \cap \overline{E_2}| + |\overline{E_1} \cap E_2|} \left(\sum_{\{u,v\} \in (E_2 \cap \overline{E_1})} \text{CONF}_{G_1}(\{u,v\}) + \sum_{\{u,v\} \in (\overline{E_1} \cap E_2)} \text{CONF}_{G_2}(\{u,v\}) \right)$$

Robert / Larousse

$$\text{GED}(G'_{Rob}, G'_{Lar}) = 0.47$$



$$\text{GED}(G'^R_{Rob}, G'^R_{Lar}) = 0.47$$



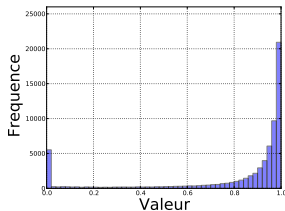
Comparer les réseaux lexicaux

$$\mu(\mathbf{G}_1, \mathbf{G}_2) =$$

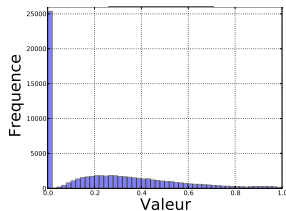
$$\frac{1}{|E_1 \cap \overline{E_2}| + |E_2 \cap \overline{E_1}|} \left(\sum_{\{u,v\} \in (E_2 \cap \overline{E_1})} \text{CONF}_{G_1}(\{u,v\}) + \sum_{\{u,v\} \in (E_1 \cap \overline{E_2})} \text{CONF}_{G_2}(\{u,v\}) \right)$$

Robert / Larousse

$$\text{GED}(G'_{\text{Rob}}, G'_{\text{Lar}}) = 0.47$$



$$\text{GED}(G'^R_{\text{Rob}}, G'^R_{\text{Lar}}) = 0.47$$



Comparer les réseaux lexicaux

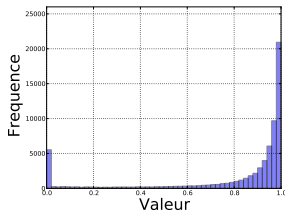
$$\mu(\mathbf{G}_1, \mathbf{G}_2) =$$

$$\frac{1}{|E_1 \cap \overline{E_2}| + |E_2 \cap \overline{E_1}|} \left(\sum_{\{u,v\} \in (E_2 \cap \overline{E_1})} \text{CONF}_{G_1}(\{u,v\}) + \sum_{\{u,v\} \in (E_1 \cap \overline{E_2})} \text{CONF}_{G_2}(\{u,v\}) \right)$$

Robert / Larousse

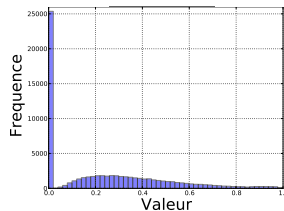
$$\mu(\mathbf{G}'_{\text{Rob}}, \mathbf{G}'_{\text{Lar}}) = 0.80$$

$$\text{GED}(\mathbf{G}'_{\text{Rob}}, \mathbf{G}'_{\text{Lar}}) = 0.47$$



$$\mu(\mathbf{G}'^{\text{R}}_{\text{Rob}}, \mathbf{G}'^{\text{R}}_{\text{Lar}}) = 0.24$$

$$\text{GED}(\mathbf{G}'^{\text{R}}_{\text{Rob}}, \mathbf{G}'^{\text{R}}_{\text{Lar}}) = 0.47$$



Comparer les réseaux lexicaux

Comparer les réseaux lexicaux

Table : Pédigrés de quelques graphes lexicaux

Graphes lexicaux		n	m	$\langle k \rangle$	C	L_{lcc}	λ (r^2)
Lar	Adjectifs	5510	21147	7,68	0,21	4,92	-2,06 (0,88)
	Noms	12159	31601	5,20	0,20	6,10	-2,39 (0,88)
	Verbes	5377	22042	8,20	0,17	4,61	-1,94 (0,88)
Rob	Adjectifs	7693	20011	5,20	0,14	5,26	-2,05 (0,94)
	Noms	24570	55418	4,51	0,11	6,08	-2,34 (0,94)
	Verbes	7357	26567	7,22	0,12	4,59	-2,01 (0,93)
Jdm	Adjectifs	9859	30087	6,10	0,16	5,44	-2,24 (0,90)
	Noms	29213	56381	3,86	0,14	6,48	-2,66 (0,93)
	Verbes	7658	22260	5,81	0,14	5,06	-2,08 (0,89)
Wik	Adjectifs	6960	6594	1,89	0,15	8,48	-2,46 (0,95)
	Noms	43206	37661	1,74	0,13	10,56	-2,51 (0,89)
	Verbes	7203	7497	2,08	0,25	9,22	-2,28 (0,92)

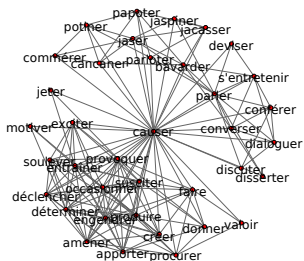
Comparer les réseaux lexicaux

Table : Mesurer la similarité structurelle entre réseaux lexicaux

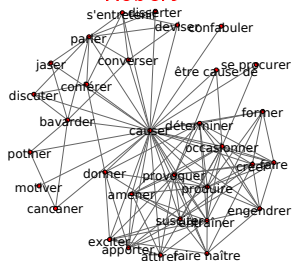
GED (μ) (μ^R) sur les paires de graphes			
G_1/G_2	Rob_A	Jdm_A	Wik_A
Lar_A	0,45 (0,76) (0,34)	0,47 (0,71) (0,38)	0,75 (0,41) (0,06)
Rob_A		0,51 (0,70) (0,29)	0,71 (0,42) (0,05)
Jdm_A			0,54 (0,43) (0,03)
G_1/G_2	Rob_N	Jdm_N	Wik_N
Lar_N	0,48 (0,70) (0,20)	0,48 (0,70) (0,19)	0,72 (0,31) (0,03)
Rob_N		0,47 (0,70) (0,13)	0,71 (0,29) (0,02)
Jdm_N			0,46 (0,32) (0,01)
G_1/G_2	Rob_V	Jdm_V	Wik_V
Lar_V	0,48 (0,73) (0,40)	0,46 (0,70) (0,39)	0,78 (0,25) (0,05)
Rob_V		0,47 (0,70) (0,37)	0,78 (0,25) (0,06)
Jdm_V			0,55 (0,31) (0,04)

Comparer les réseaux lexicaux

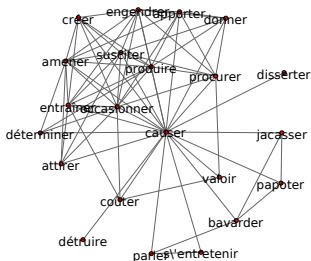
Larousse



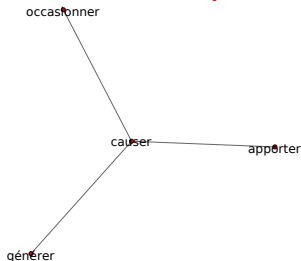
Robert



Jeux de mots



Wiktionary



Hypothèse : Les relations lexicales reflètent les structures sémantiques du lexique communes aux membres d'une même communauté linguistique

Constat : Il existe de nombreux désaccords entre membres d'une même communauté linguistique:

$$\text{GED}(\mathbf{G}_1, \mathbf{G}_2) \approx 0.50$$

- **EXEMPLE** : **Beaucoup** de paires de mots sont synonymes pour l'individu I_1 et pas pour l'individu I_2



Hypothèse : Les relations lexicales reflètent les structures sémantiques du lexique communes aux membres d'une même communauté linguistique

Incompatible ?

Constat : Il existe de nombreux désaccords entre membres d'une même communauté linguistique:

GED(G_1, G_2) ≈ 0.50

- **EXEMPLE** : **Beaucoup** de paires de mots sont synonymes pour l'individu I_1 et pas pour l'individu I_2



Hypothèse : Les relations lexicales reflètent les structures sémantiques du lexique communes aux membres d'une même communauté linguistique

Incompatible ?

Constat : Il existe de nombreux désaccords entre membres d'une même communauté linguistique:

GED(G_1, G_2) ≈ 0.50

- **EXEMPLE** : **Beaucoup** de paires de mots sont synonymes pour l'individu I_1 et pas pour l'individu I_2



Non

Parceque : Ces désaccords sont des désaccords faibles :

$\mu(G_1, G_2) \approx 0.70$ (accord au niveau structurel)

Hypothèse : Les relations lexicales reflètent les structures sémantiques du lexique communes aux membres d'une même communauté linguistique

L'analyse d'une relation lexicale doit être faite au niveau de la structure globale dessinée par la relation. C'est ce que la mesure μ peut faire: avec un niveau de représentation adéquat (la confluence), elle réconcilie les jugements portés par deux juges différents sur une même relation lexicale.

Hypothèses

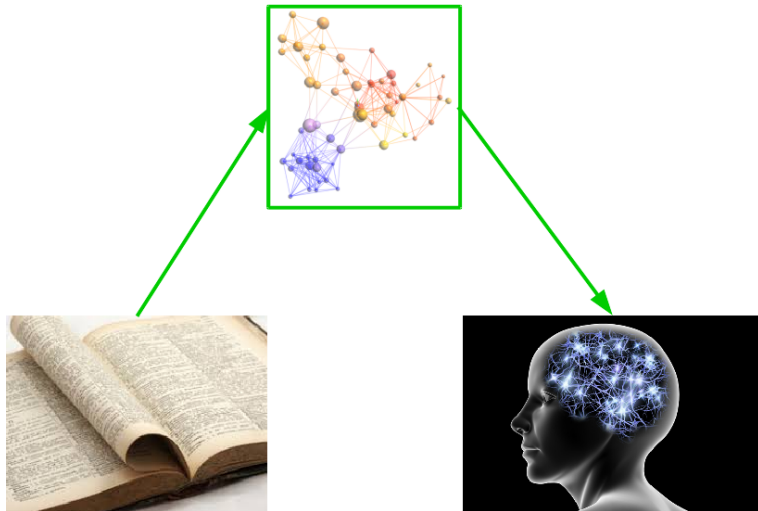
- 1 Structure des réseaux lexicaux
- 2 Mesurer la similarité structurelle entre réseaux lexicaux

3 Hypothèses

Applications à EVOLEX

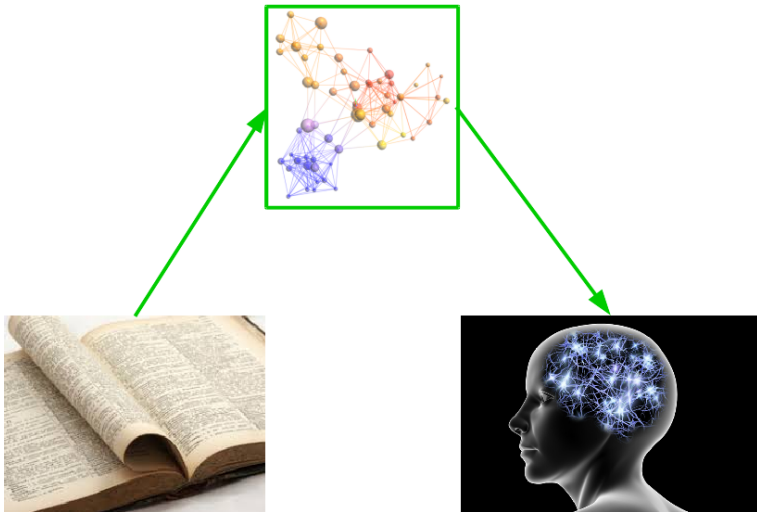
- 4 La ressource Jeux de mots
- 5 Résultats

Hypothèses



Hypothèses

Avec les bons outils !!



- 1 Structure des réseaux lexicaux
- 2 Mesurer la similarité structurelle entre réseaux lexicaux
- 3 Hypothèses

Applications à EVOLEX

- 4 La ressource Jeux de mots
- 5 Résultats

La ressource Jeux de mots (Lafourcade, 2007)

Trouver des mots en **REL**ation sémantique avec une **CIBLE** donnée
que le joueur concurrent n'a pas trouvés,
mais que les autres joueurs ont trouvées.

La ressource Jeux de mots (Lafourcade, 2007)

Trouver des mots en **REL**ation sémantique avec une **CIBLE** donnée
que le joueur concurrent n'a pas trouvés,
mais que les autres joueurs ont trouvées.

CIBLE = couverture, REL = synonymie

La ressource Jeux de mots (Lafourcade, 2007)

Trouver des mots en **REL**ation sémantique avec une **CIBLE** donnée
que le joueur concurrent n'a pas trouvés,
mais que les autres joueurs ont trouvées.

CIBLE = couverture, REL = synonymie

CIBLE = crocodile, REL = association libre

La ressource Jeux de mots (Lafourcade, 2007)

Trouver des mots en **REL**ation sémantique avec une **CIBLE** donnée
que le joueur concurrent n'a pas trouvés,
mais que les autres joueurs ont trouvées.

CIBLE = couverture, REL = synonymie

CIBLE = crocodile, REL = association libre

CIBLE = hirondelle, REL = hyperonymie

- 1 Structure des réseaux lexicaux
- 2 Mesurer la similarité structurelle entre réseaux lexicaux
- 3 Hypothèses

Applications à EVOLEX

- 4 La ressource Jeux de mots
- 5 Résultats

JDM	freq-JDM	Confluence
Synonymie	0.09	0.11
Association libre	0.34	0.46

Table : Corrélations avec la fréquence d'association d'EVOLX.