Les capitalistes sociaux sur Twitter : détection, évolution et caractérisation

Nicolas Dugué[1] Vincent Labatut[2] Anthony Perez[1]

[1]Université d'Orléans, LIFO
{nicolas.dugue,anthony.perez}@univ-orleans.fr

[2]Université Galatasaray, Département d'informatique
vlabatut@gsu.edu.tr

Lip6
Complex Networks
7 novembre 2013

Plan


Vert d'Absinthe
@vertdabsinthe
La première boutique dédiée à l'absinthe The first shop dedicated to absinthe in Paris
Paris, France · vertdabsinthe.com
386
TWEETS
19
ABONNEMENTS
320
ABONNÉS
Abonné

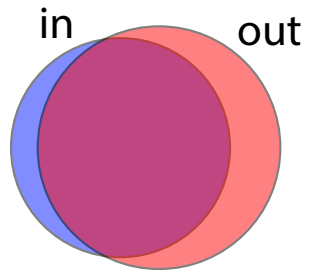
Notion de capitalisme social [GVK+12]

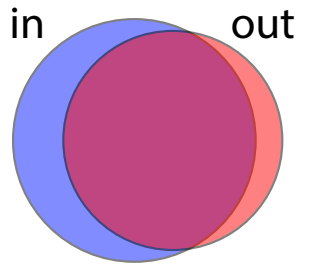
Introduite par Gosh et al.[GVK+12] : Utilisateurs qui suivent le plus les spammeurs.
Liste de 100.000 utilisateurs.

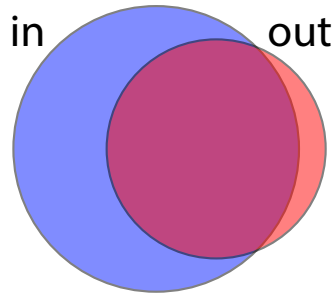
- Obtenir rapidement un maximum de visibilité
- Pourquoi les étudier ?
  - Comprendre leur influence sur le réseau
  - Améliorer la qualité de service
  - Appliquer leurs méthodes à d'autres domaines (marketing)

Stratégies des capitalistes sociaux

- I Follow You, Follow Me (IFYFM)
- Follow Me, I Follow You (FMIFY)
- État passif







Exemples de capitalistes sociaux

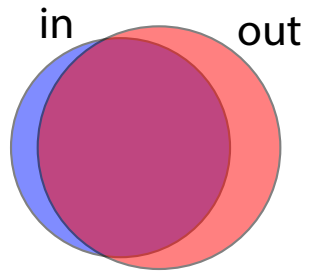
screen name	name	followers	friends
IFOLLOWBACKJP	TFBJP	$1.2 \cdot 10^5$	$1.1 \cdot 10^5$
itsrealchris	iFollowBack	$1.7 \cdot 10^5$	$1.6 \cdot 10^5$
AllFollowMax	TFBJP	$4.2 \cdot 10^4$	$4.3 \cdot 10^4$
BarackObama	Barack Obama	$2.5 \cdot 10^7$	$6.7 \cdot 10^5$
britneyspears	Britney Spears	$2.2 \cdot 10^7$	$4.1 \cdot 10^5$
JetBlue	JetBlue Airways	$1.7 \cdot 10^6$	$1.1 \cdot 10^5$
Starbucks	Starbucks Coffee	$3.2 \cdot 10^6$	$7.9 \cdot 10^4$

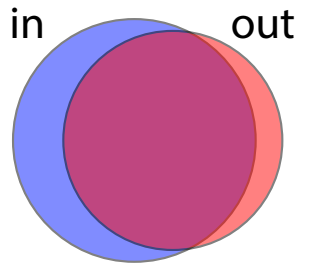
Table: Followers and friends numbers are rounded.

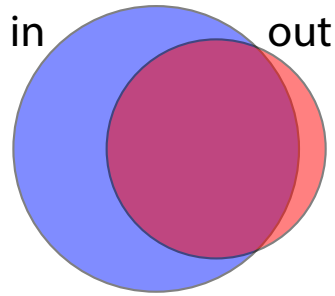
Plan

Rappel des stratégies

- I Follow You, Follow Me (IFYFM)
- Follow Me, I Follow You (FMIFY)
- État passif







Comment les détecter ?

Rappel

- FMIFY : Follow Me and I Follow You;
- IFYFM : I Follow you, Follow me;
- Former social capitalists;

Mesures de Similarités

$$Overlapindex = \frac{|N^+(v) \cap N^-(v)|}{\min(|N^+(v)|, |N^-(v)|)}$$

$$Ratio = \frac{|N^+(v)|}{|N^-(v)|}$$

Seuil de détection

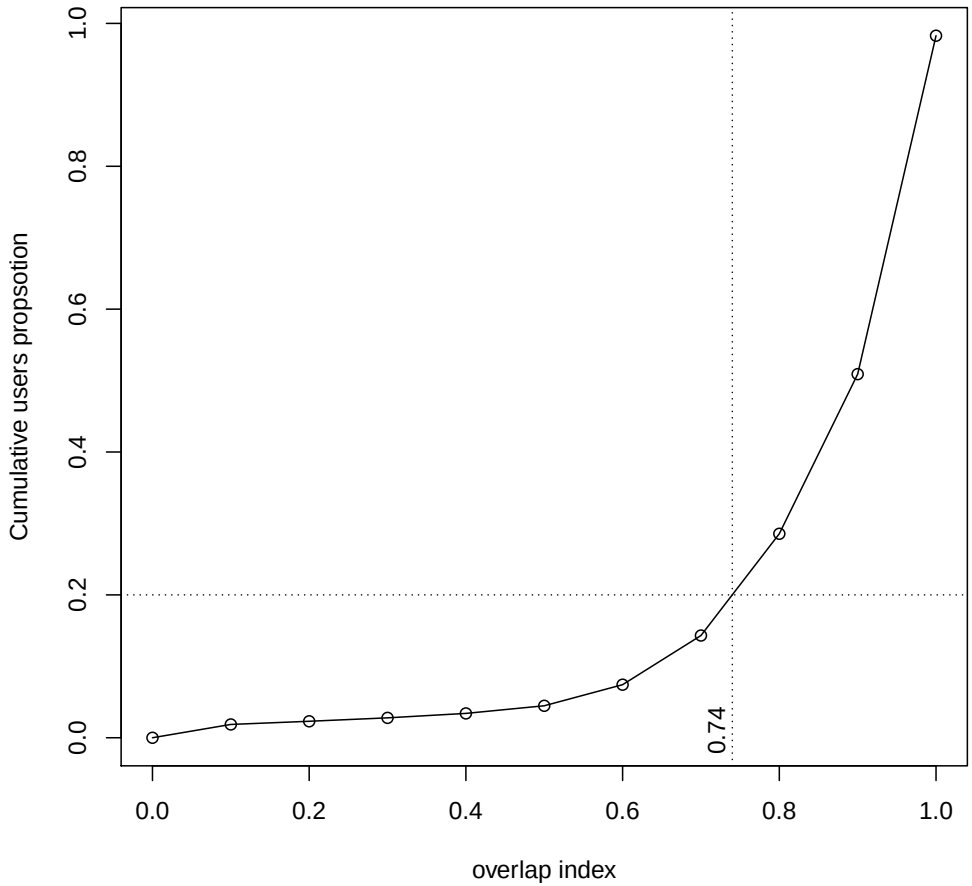


Figure: Overlap index des 100000 capitalistes sociaux de Ghosh et al.

Quelques exemples...

screen name	name	followers	friends	overlap	ratio
IFOLLOWBACKJP	TFBJP	$1.2 \cdot 10^5$	$1.1 \cdot 10^5$	0.97	0.92
itsrealchris	iFollowBack	$1.7 \cdot 10^5$	$1.6 \cdot 10^5$	0.81	0.94
AllFollowMax	TFBJP	$4.2 \cdot 10^4$	$4.3 \cdot 10^4$	0.99	1.04
BarackObama	Barack Obama	$2.5 \cdot 10^7$	$6.7 \cdot 10^5$	0.77	0.03
britneyspears	Britney Spears	$2.2 \cdot 10^7$	$4.1 \cdot 10^5$	0.82	0.02
JetBlue	JetBlue Airways	$1.7 \cdot 10^6$	$1.1 \cdot 10^5$	0.74	0.06
Starbucks	Starbucks Coffee	$3.2 \cdot 10^6$	$7.9 \cdot 10^4$	0.77	0.02

Plan

Les données

Le graphe de Twitter

Un graphe collecté par Kwak et. al en 2009 [KLPM10] :
41 millions d'utilisateurs
1,4 milliard d'arcs

Les données

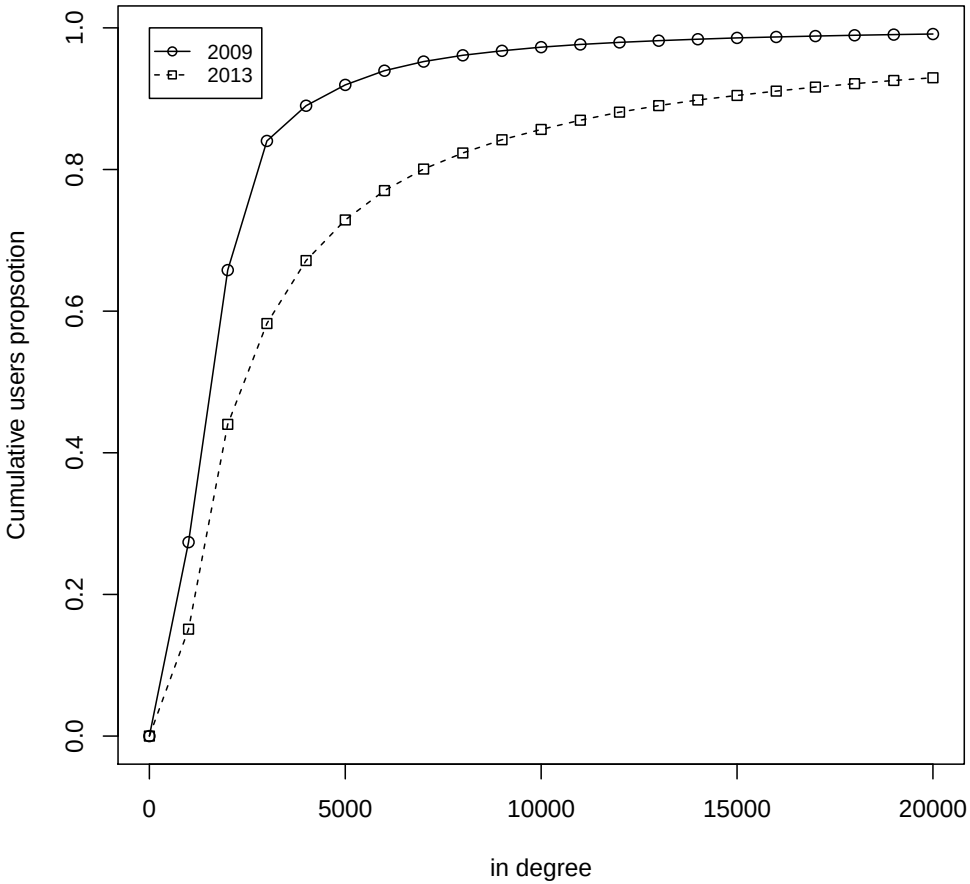
Le graphe de Twitter

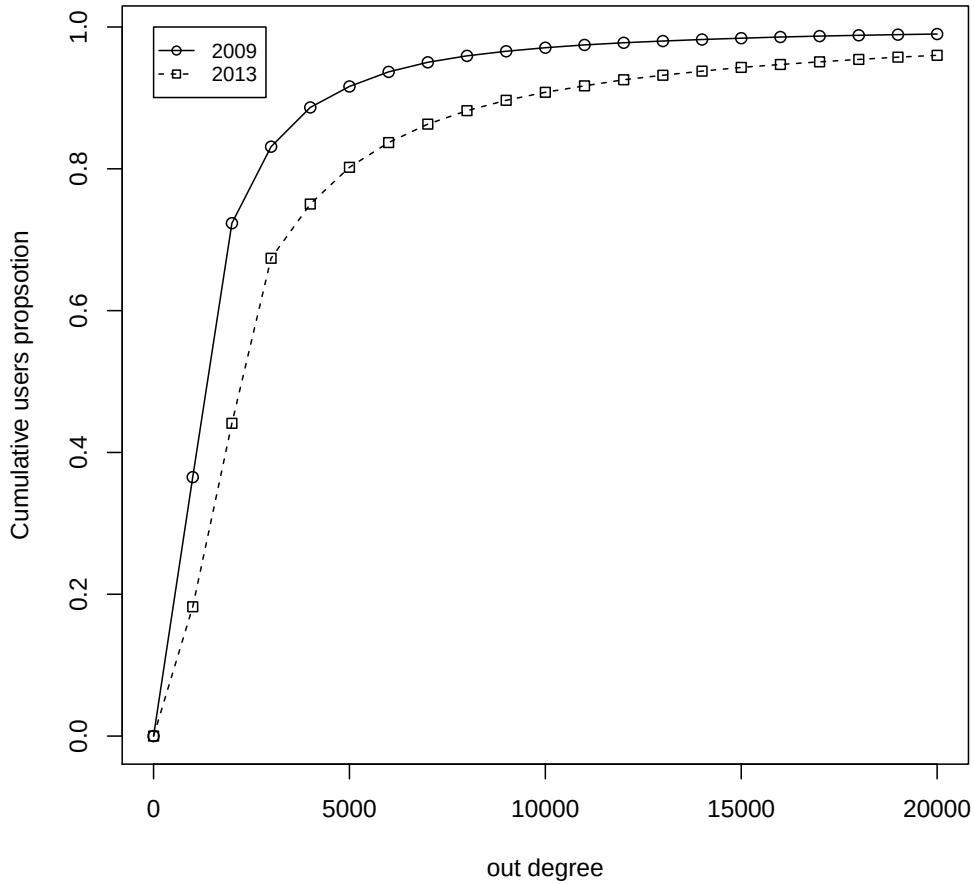
Un graphe collecté par Kwak et. al en 2009 [KLPM10] :
41 millions d'utilisateurs
1,4 milliard d'arcs

Capitalistes sociaux

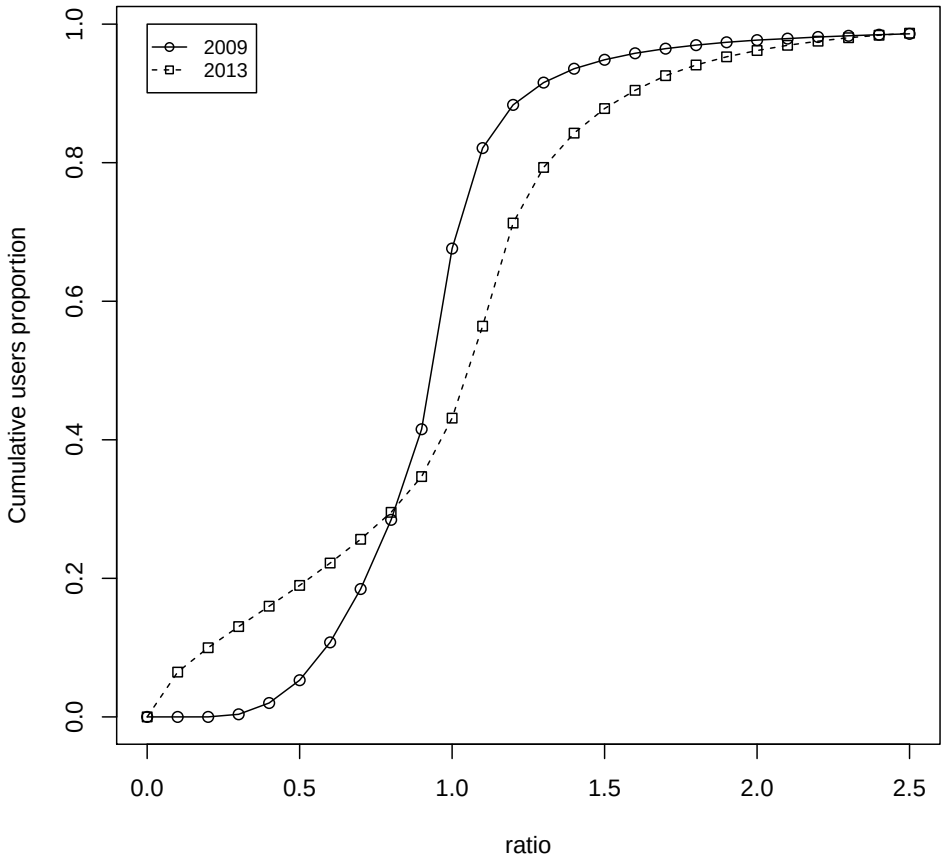
145.000 détectés
$\Rightarrow$ Echantillon de 75% en 2013

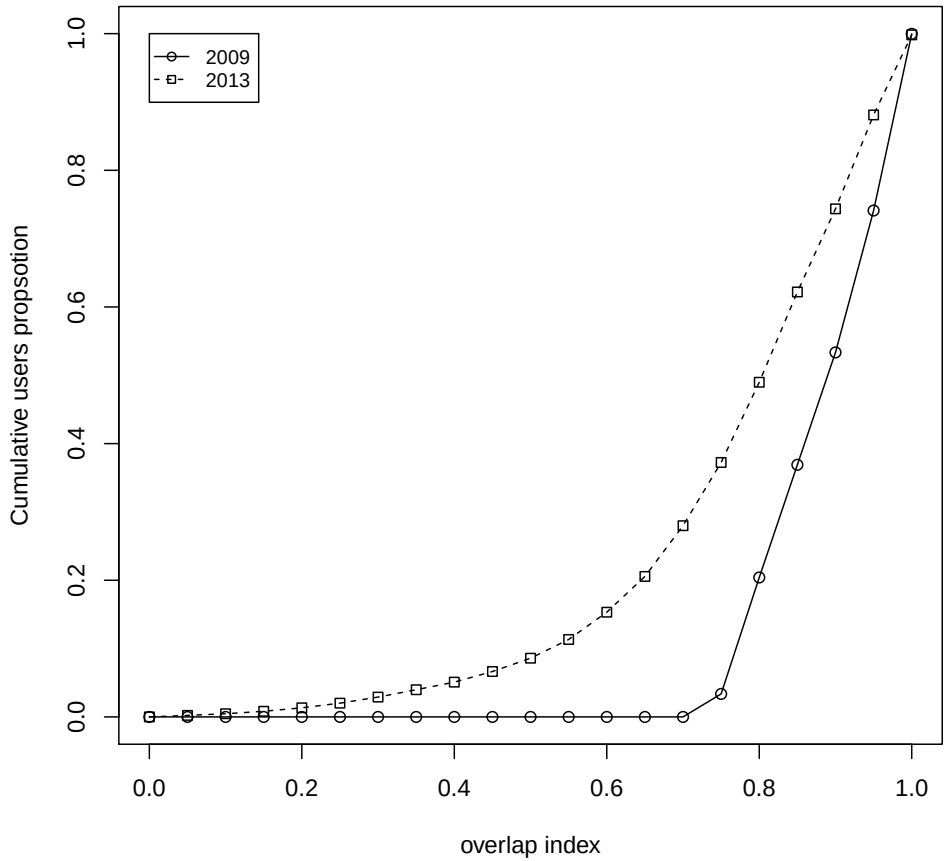
Les degrés


2009
2013
Cumulative users propsotion
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
15000
20000
in degree


2009
2013
Cumulative users propsotion
0.0
0.2
0.4
0.6
0.8
1.0
0
5000
10000
15000
20000
out degree

Ratio / Overlap index





Toujours pas lassés de capitaliser en 2013

	$\mathbf{o}^{2013} \geqslant 0.74$	$\mathbf{r}^{2013} > 1$	$\mathbf{r}^{2013} \in [0.7; 1]$	$\mathbf{r}^{2013} < 0.7$
$\mathbf{r}^{2010} > 1$	19892	40%	**25**%	**35**%
$\mathbf{r}^{2010} \in [0.7; 1]$	37789	**80**%	14%	6%
$\mathbf{r}^{2010} < 0.7$	10435	**89**%	7%	4%

Table: r^i et o^i : respectivement le ratio et l'overlap index pour $i \in \{2010, 2013\}$.

Las en 2013

	$\mathbf{o}^{2013} < 0.74$	$\mathbf{r}^{2013} > 1$	$\mathbf{r}^{2013} \in [0.7; 1]$	$\mathbf{r}^{2013} < 0.7$
$\mathbf{r}^{2010} > 1$	14103	16%	25%	**59**%
$\mathbf{r}^{2010} \in [0.7; 1]$	15377	36%	23%	40%
$\mathbf{r}^{2010} < 0.7$	8894	**61**%	14%	25%

Table: r^i et o^i : respectivement le ratio et l'overlap index pour $i \in \{2010, 2013\}$.

Username	out-	in-	ratio	ovp
@ladygaga	636929	73274	**8.69**	0.97
@BarackObama	1882889	770155	**2.44**	0.91
@BritneySpears	2674874	406238	**6.58**	0.95
@paulocoelho	75423	48446	**1.56**	0.98
@paulpierce34	815197	524	**1555.72**	0.95

Username	out-	in-	ratio	ovp
@ladygaga	136386	37485540	**0.00**	0.82
@BarackObama	680428	30836226	**0.02**	0.77
@BritneySpears	412703	27763836	**0.01**	0.81
@paulocoelho	98	7721670	**0.00**	0.86
@paulpierce34	85	2804060	**0.00**	0.78

Plan

Les données

Ashley Paternoster @Rain_Bow_Ash 20m
(◕∀◕)/☆♡+˚*。Ⓛ Ⓤ Ⓒ Ⓚ Ⓨ #TFBJP ☀ R E T W E E T
☀ONLYIF☀YOU☀WANT☀NEW☀FOLLOWERS☀ #HitFollowsTeam
#Teamfollowback 18.02
Expand

Fa$t Life™ @LilMarcus_DMV 7 Sept
#90sBabyFollowTrain
1. ♻RT this
2. I follow you
3. You followback
4. Fav for a S/O
5. IG : _lilmarcus
#TeamFollowBack

Retweeté par Ashley Paternoster

Les données

#TeamFollowBack

- 725.000 tweets récoltés en février 2013
- 125.000 utilisateurs : 12% dont on connait les voisinages

Les données

#TeamFollowBack

- 725.000 tweets récoltés en février 2013
- 125.000 utilisateurs : 12% dont on connait les voisinages

- Quelles sont les sources des tweets ?
- Quels sont les autres hashtags utilisés ?
- Ces utilisateurs sont il détectés par notre méthode ?
- Peut on automatiser cette méthode ?

Les données

Type	Number
Tweets	726470
Hashtags	4227703
Distinct hashtags	25028
Average hashtag number by tweet	5.82
Distinct users	124786
Mentions	719972
Distinct user mentioned	43199

Les hashtags

Hashtag	Occurence numbers
TeamFollowBack	766421
TFBJP	339917
sougofollow	177064
500aday	172655
OPENFOLLOW	148304
FollowBack	143174
RT	125211
instantfollowback	107989
AutoFollow	105528
HitFollowsTeam	102100

Table: Les 10 hashtags les plus utilisés et leur nombre d'apparitions (50% du dataset).

Les détecte-t-on ?

Type	Number
Users tweeting on the hashtag	124786
Sample with followers, friends crawled	15226
Sample with followers, friends $>$500	8442
Sample with followers, friends $>$500 and overlap $>$0.74	6740

Table: Overlap index des utilisateurs observés sur *#TeamFollowBack*.

Sources	Occurence numbers
Twitter for BlackBerry	196911
twitter.com	196747
Twitter for Android	66473
Twitter for iPhone	49556
twitterfeed	39010
Mobile Web (M2)	28482
BotMaker	14260
BestFollowers App.2.85	11189
dlvr.it	10443
TweetCaster for Android	9777
Twitter for iPad	9329
twittbot.net	8300
UberSocial for BlackBerry	8037
Write Longer	5155

Table: Sources utilisées pour 90% des tweets du datasets.

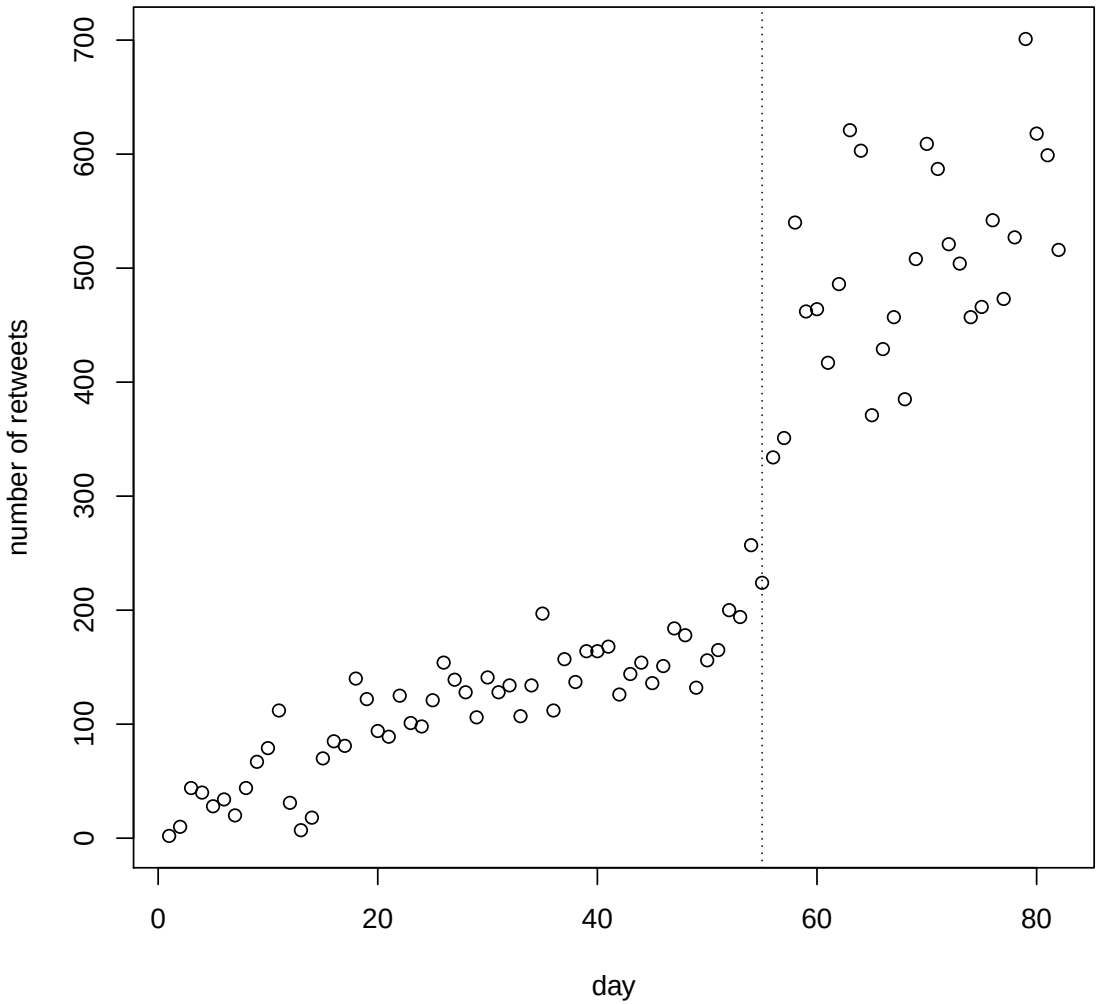
Compte automatisé

Des actions complexes et sophistiquées...

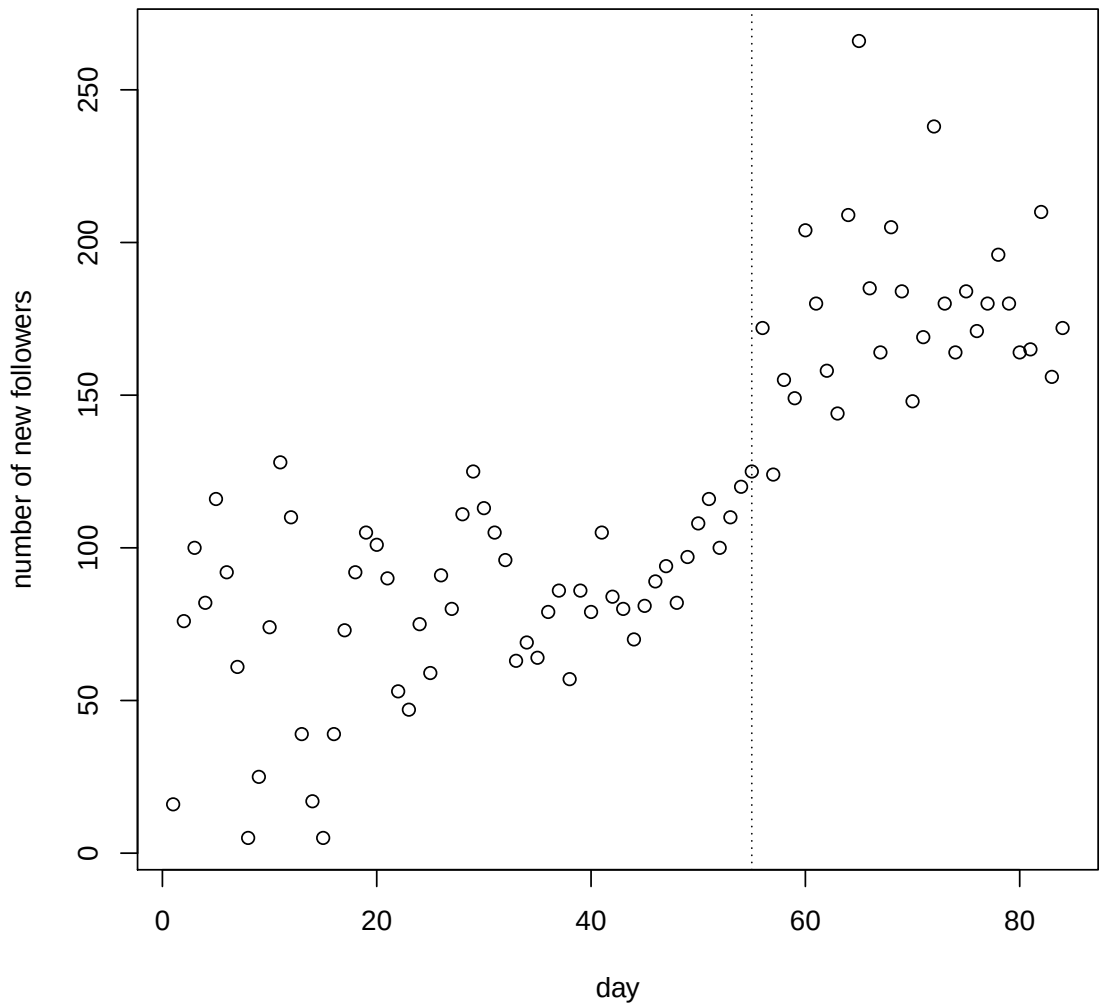
- Rejoue les tweets du dataset
- ”Follow back”
- Suit les utilisateurs qui le mentionne
- Suit les utilisateurs qui le retweete

Compte automatisé

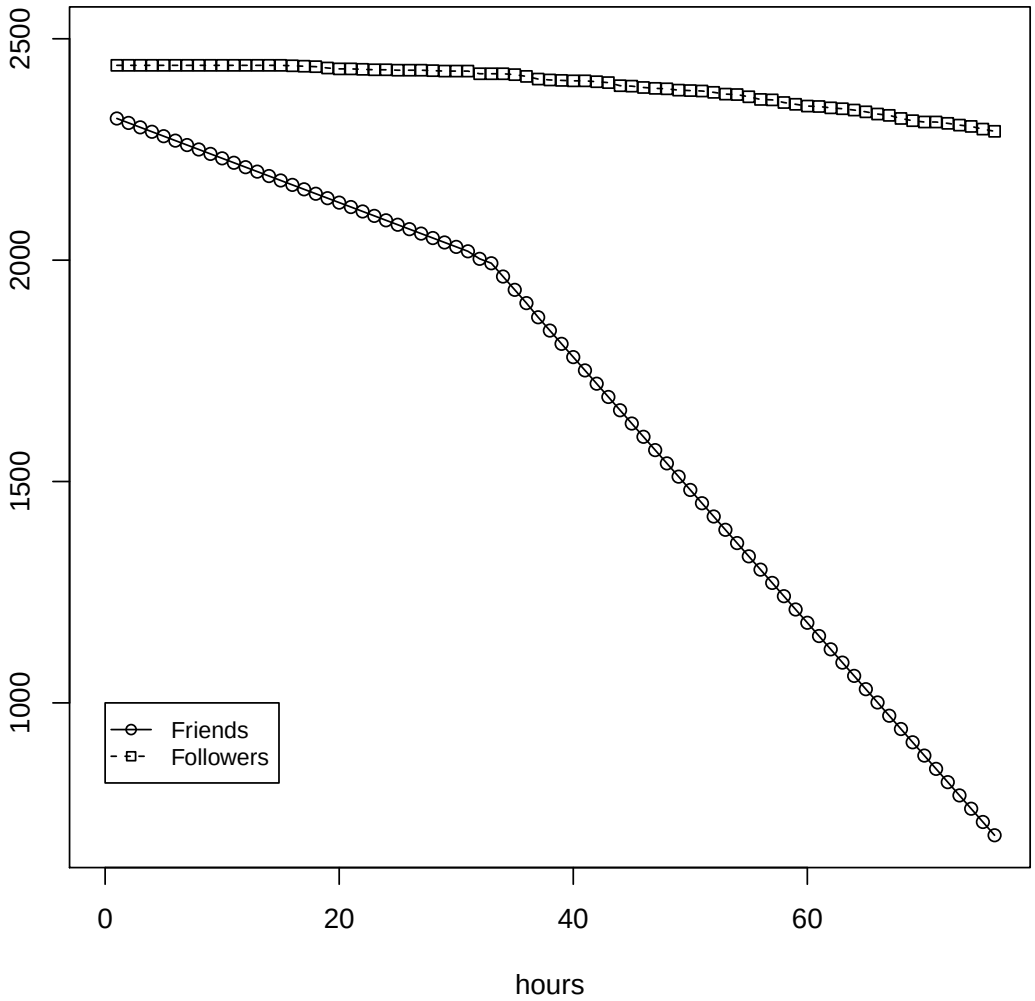
Retweets quotidiens



Nouveaux followers quotidiens



Unfollow



Plan

Les capitalistes sociaux dans le graphe ?

Intuition

Les stratégies de capitalisme social devraient mieux marcher si les cibles de ces stratégies sont des capitalistes sociaux.

Les capitalistes sociaux dans le graphe ?

Intuition

Les stratégies de capitalisme social devraient mieux marcher si les cibles de ces stratégies sont des capitalistes sociaux.

Un ensemble de sommets très connectés

- Une composante faiblement connexe
- Un coefficient de clustering supérieur à la moyenne
- Un voisinage de capitalistes sociaux

Les capitalistes sociaux dans le graphe ?

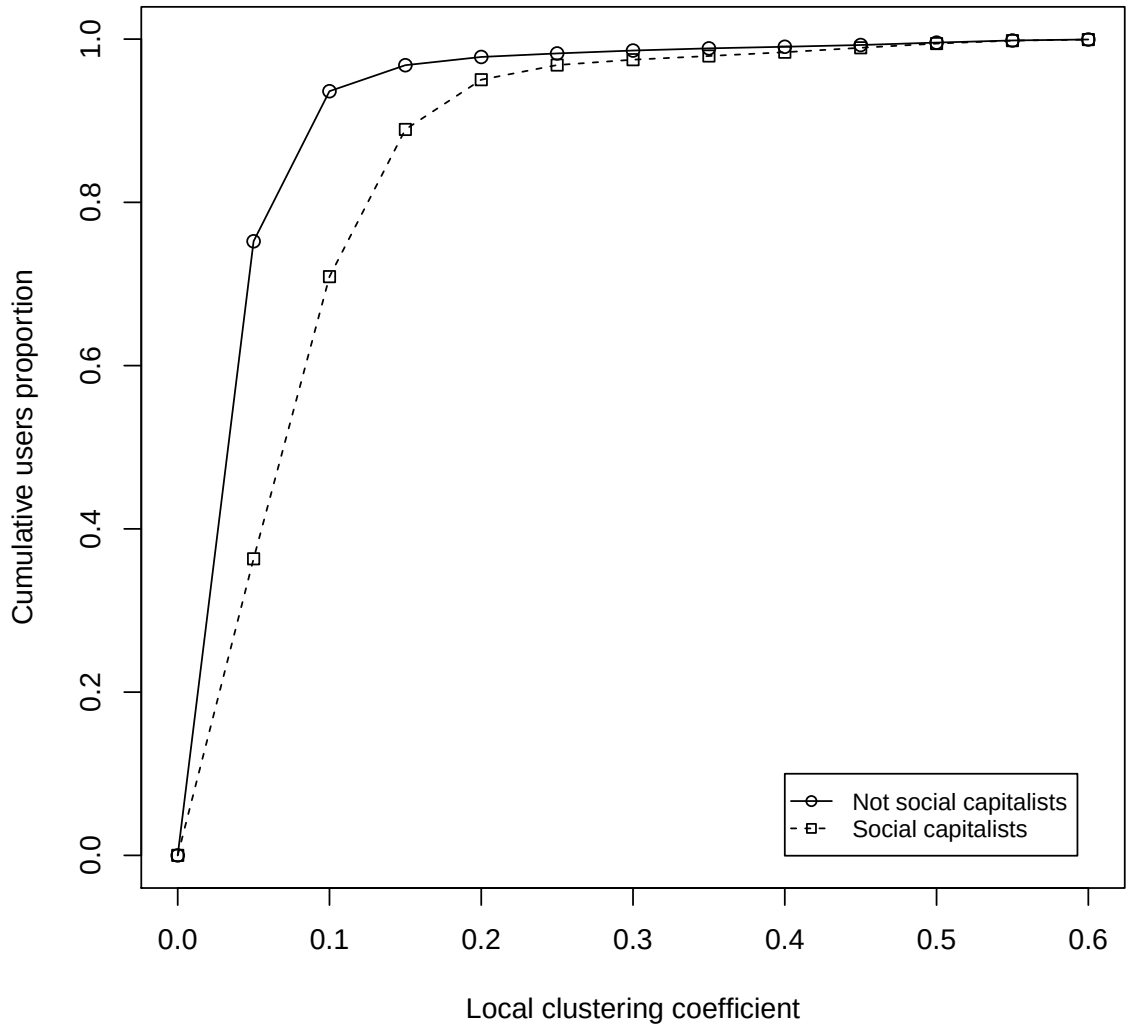
Coefficient de Clustering

Soit $G = (V, E)$ un graphe, et $v \in V$ n'importe quel de ses sommets. Le coefficient de clustering cc_v de v est défini comme :

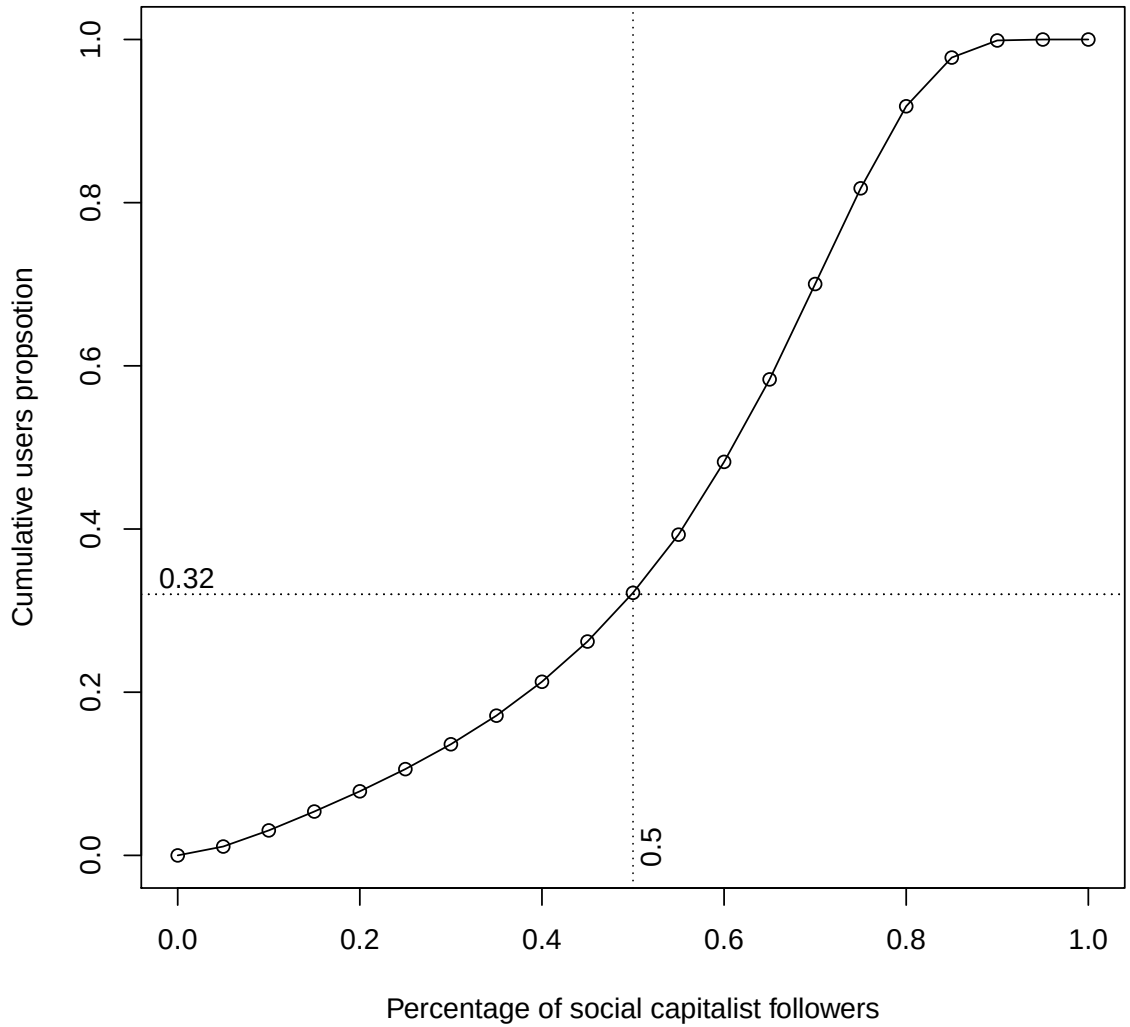
$$cc_v = 2 \cdot n_{edges}/(d_v(d_v - 1))$$

où n_{edges} est le nombre d'arêtes entre les voisins de v, et d_v le nombre de voisins de v.

Coefficient de Clustering



Voisinage entrant



Communauté

- Composant faiblement connexe
- Coefficient de clustering supérieur à la moyenne
- Voisinage de capitalistes sociaux

$\rightarrow$ Une/des communautés de capitalistes sociaux

Communauté

- Composant faiblement connexe
- Coefficient de clustering supérieur à la moyenne
- Voisinage de capitalistes sociaux

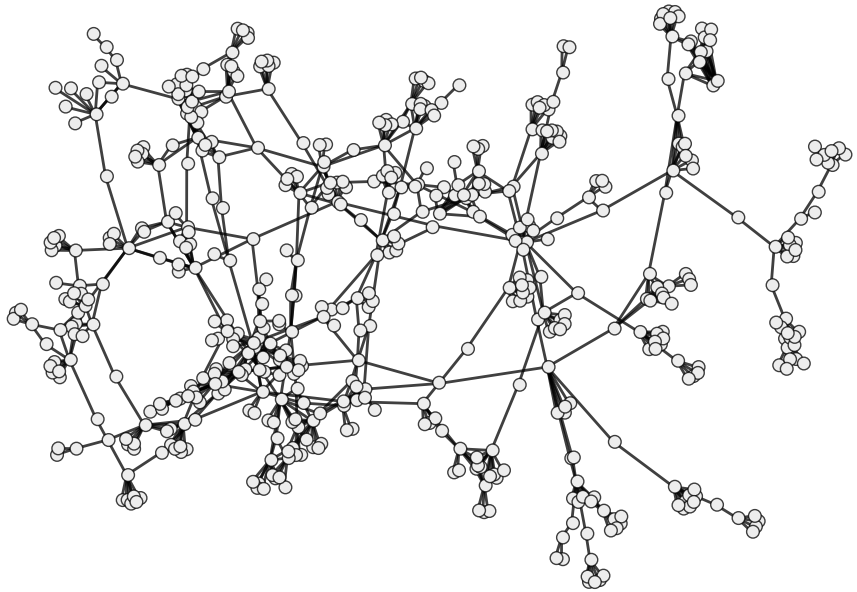
$\rightarrow$ Une/des communautés de capitalistes sociaux

Rôle communautaire

- Sont ils liés avec des communautés extérieures ?
- Sont ils hubs dans les communautés ?
- Sont ils isolés dans certaines communautés ?
- Qui suivent ils ?

Méthode de Guimerà & Amaral [GA05]

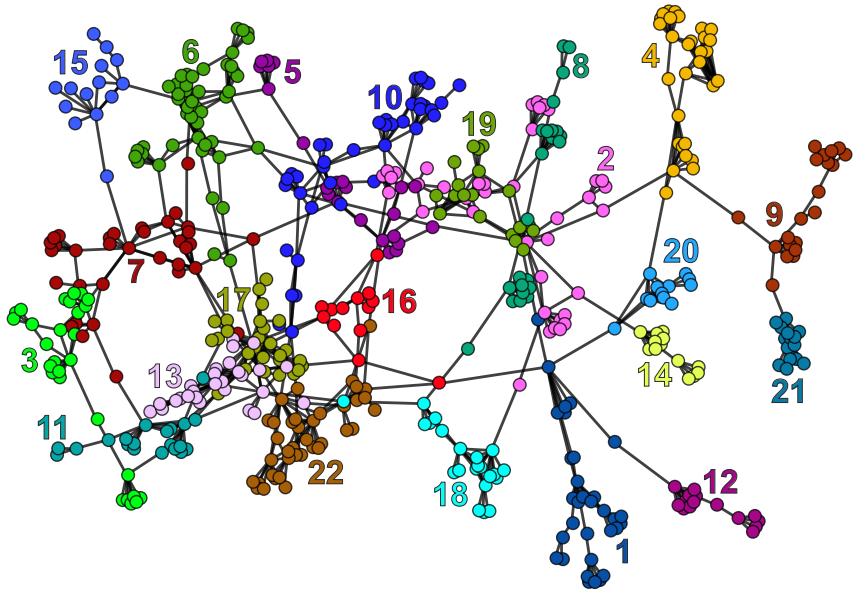
- **Principe** :
  - Caractériser la position d'un nœud en fonction de sa connectivité communautaire
  - Connectivité communautaire décrite par 2 mesures
- **Processus** :
  1. Identification des communautés
  2. Calcul des 2 mesures nodales
  3. Partition de l'espace $2D$ obtenu
  4. Mise en correspondance des rôles



[LB12]

Méthode de Guimerà & Amaral [GA05]

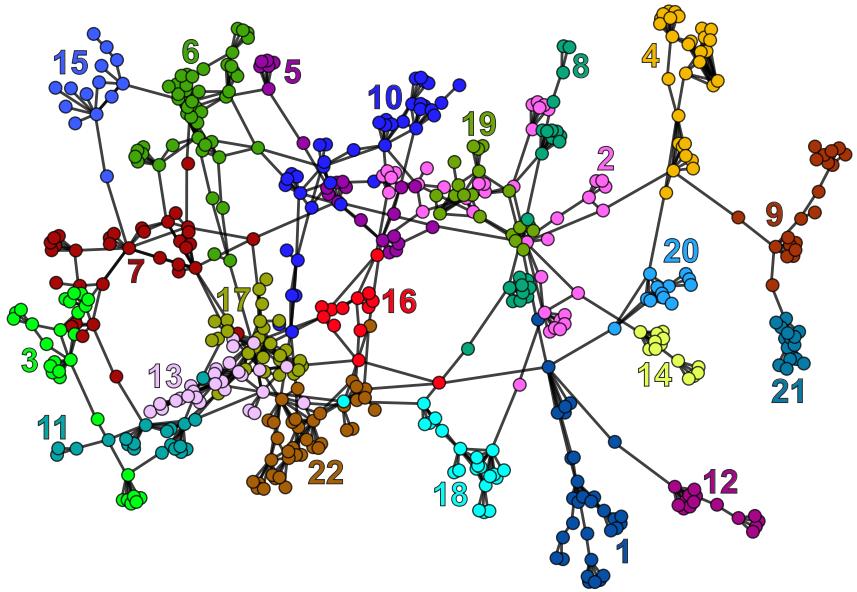
- **Principe** :
  - Caractériser la position d'un nœud en fonction de sa connectivité communautaire
  - Connectivité communautaire décrite par 2 mesures
- **Processus** :
  1. Identification des communautés
  2. Calcul des 2 mesures nodales
  3. Partition de l'espace $2D$ obtenu
  4. Mise en correspondance des rôles



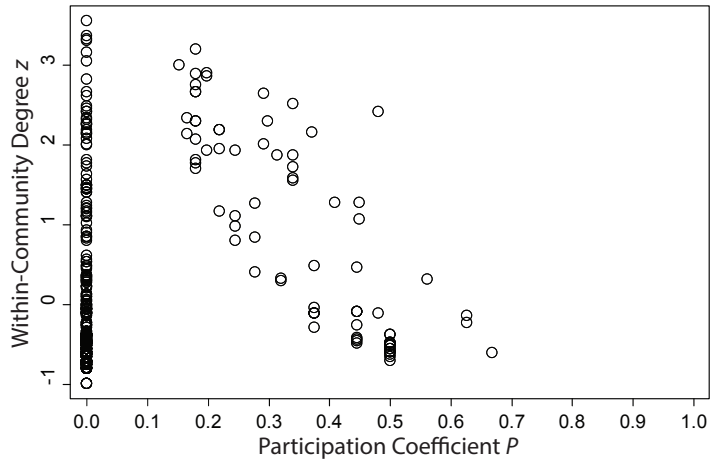
[LB12]

Méthode de Guimerà & Amaral [GA05]

- **Principe** :
  - Caractériser la position d'un nœud en fonction de sa connectivité communautaire
  - Connectivité communautaire décrite par 2 mesures
- **Processus** :
  1. Identification des communautés
  2. Calcul des 2 mesures nodales
  3. Partition de l'espace $2D$ obtenu
  4. Mise en correspondance des rôles

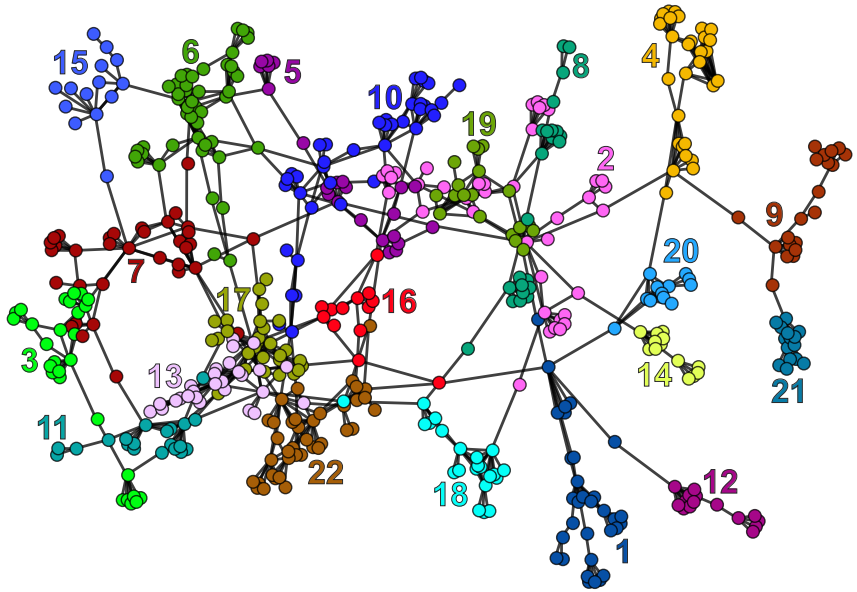


[LB12]

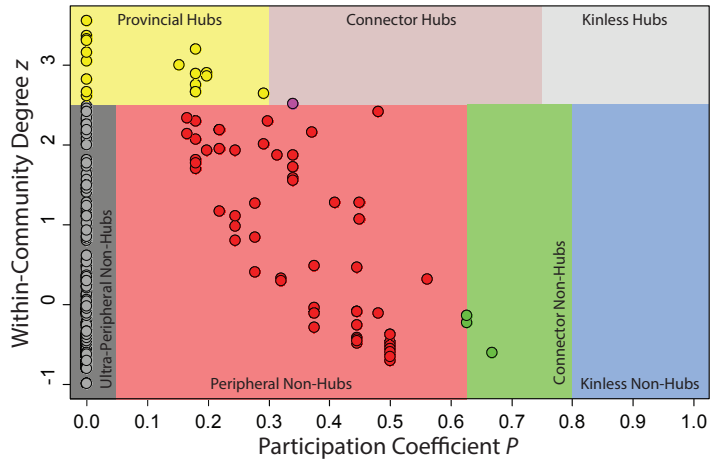


Méthode de Guimerà & Amaral [GA05]

- **Principe** :
  - Caractériser la position d'un nœud en fonction de sa connectivité communautaire
  - Connectivité communautaire décrite par 2 mesures
- **Processus** :
  1. Identification des communautés
  2. Calcul des 2 mesures nodales
  3. Partition de l'espace $2D$ obtenu
  4. Mise en correspondance des rôles

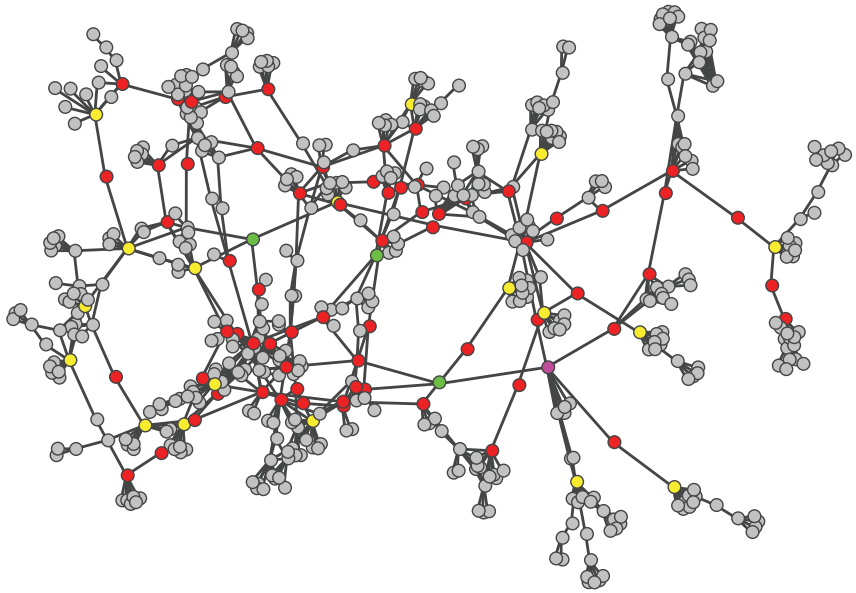


[LB12]

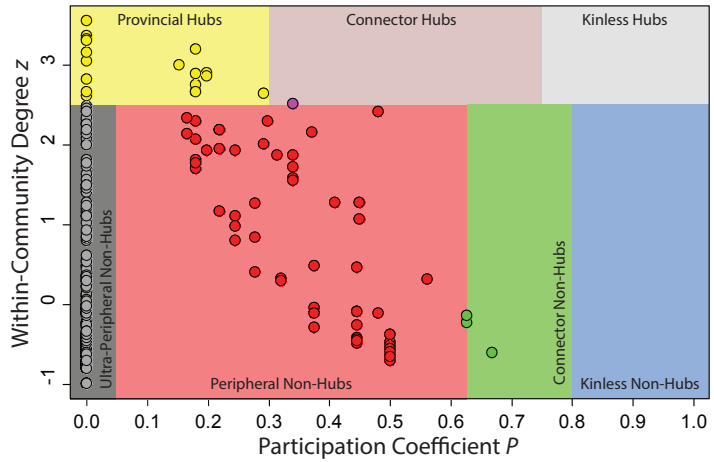


Méthode de Guimerà & Amaral [GA05]

- **Principe** :
  - Caractériser la position d'un nœud en fonction de sa connectivité communautaire
  - Connectivité communautaire décrite par 2 mesures
- **Processus** :
  1. Identification des communautés
  2. Calcul des 2 mesures nodales
  3. Partition de l'espace $2D$ obtenu
  4. Mise en correspondance des rôles

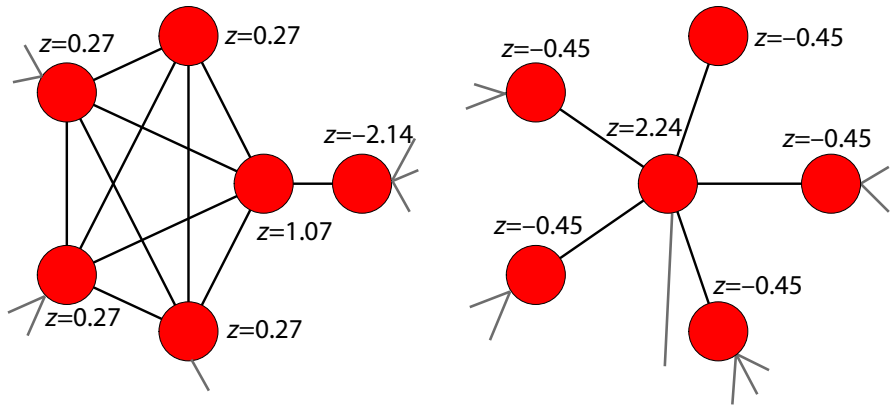


[LB12]



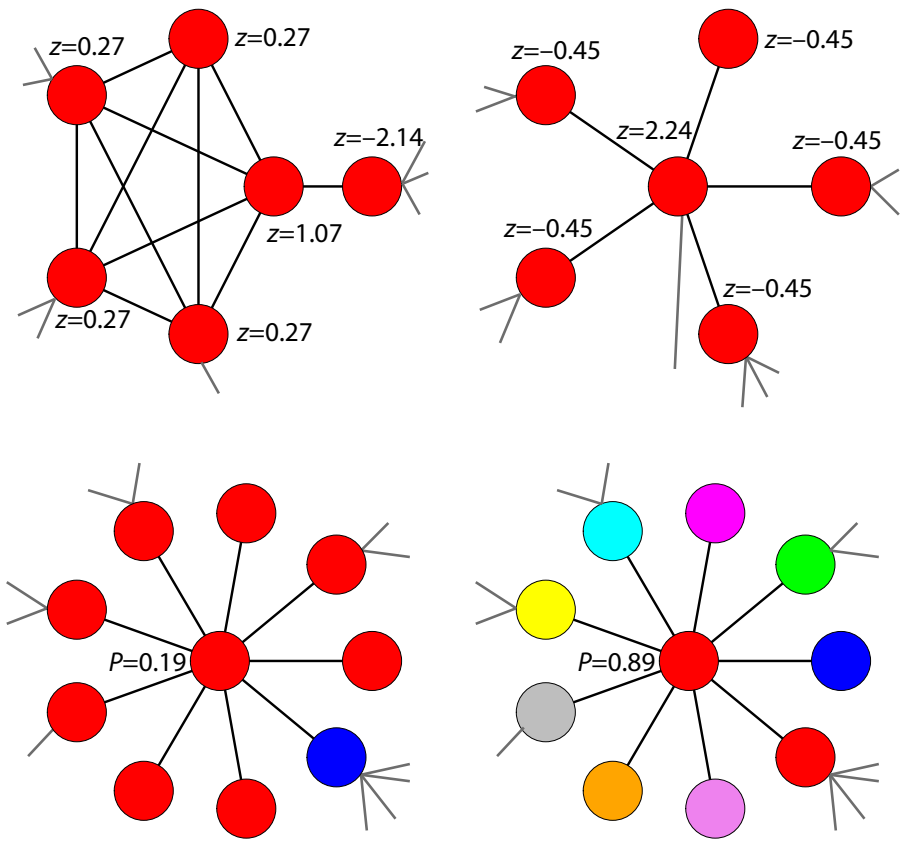
Mesures de rôle

- **Degré interne normalisé**
  - Connectivité *interne*
    $z(u) = \frac{k_{int}(u) - \mu_i(k_{int})}{\sigma_i(k_{int})}$, $u \in C_i$
  - z-score du degré interne k_{int}
  - Bornes pas fixées
- **Coefficient de participation**
  - Connectivité *externe*
    $P(u) = 1 - \sum_i \left(\frac{k_i(u)}{k(u)}\right)^2$
  - k_i : degré pour C_i
  - $P(u) = 0$:
    - Une seule communauté
  - $P(u) \approx 1$:
    - Nombreuses communautés
    - Même nombre de liens



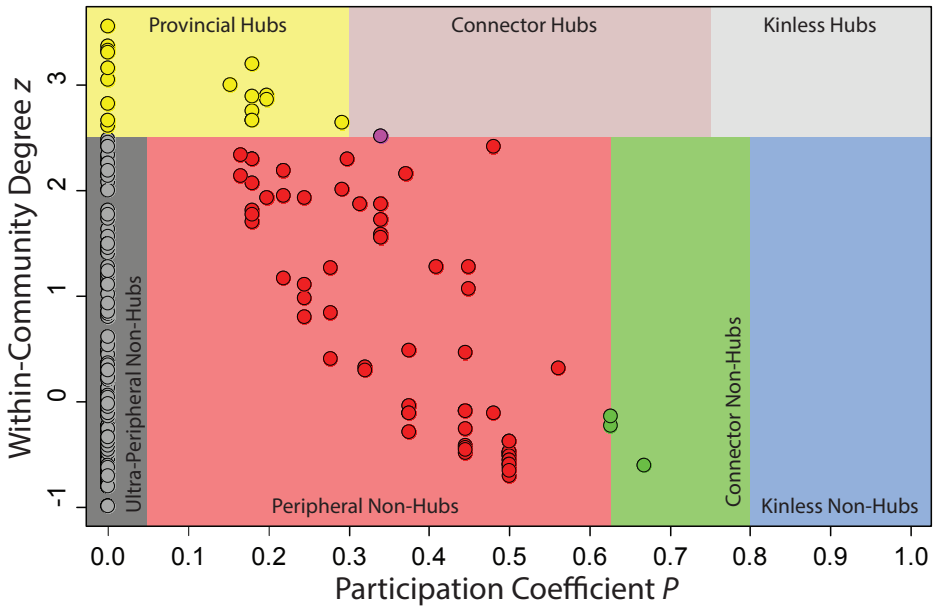
Mesures de rôle

- **Degré interne normalisé**
  - Connectivité *interne*
    $z(u) = \frac{k_{int}(u) - \mu_i(k_{int})}{\sigma_i(k_{int})}$, $u \in C_i$
  - z-score du degré interne k_{int}
  - Bornes pas fixées
- **Coefficient de participation**
  - Connectivité *externe*
    $P(u) = 1 - \sum_i \left(\frac{k_i(u)}{k(u)}\right)^2$
  - k_i : degré pour C_i
  - $P(u) = 0$:
    - Une seule communauté
  - $P(u) \approx 1$:
    - Nombreuses communautés
    - Même nombre de liens



Mesures de rôle

- **Degré interne normalisé**
  - Connectivité *interne*
    $z(u) = \frac{k_{int}(u) - \mu_i(k_{int})}{\sigma_i(k_{int})}$, $u \in C_i$
  - z-score du degré interne k_{int}
  - Bornes pas fixées
- **Coefficient de participation**
  - Connectivité *externe*
    $P(u) = 1 - \sum_i \left(\frac{k_i(u)}{k(u)}\right)^2$
  - k_i : degré pour C_i
  - $P(u) = 0$:
    - Une seule communauté
  - $P(u) \approx 1$:
    - Nombreuses communautés
    - Même nombre de liens

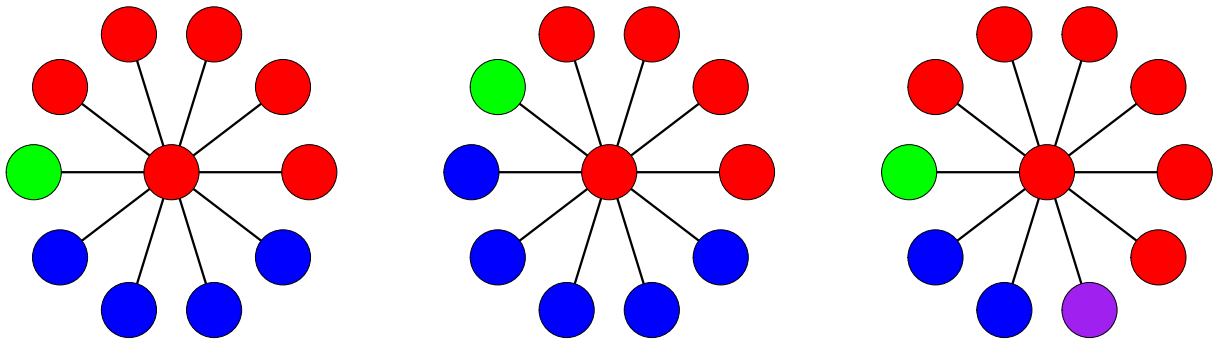


Limitations de l'approche

- Orientation des liens ignorée
  - Systèmes à relations asymétriques
  - Twitter : followers / followees
  - Seuils différents pour 4 mesures (orientées)
- Imprécision du coefficient de participation
  - Degré, nombre de communautés, distribution des liens
  - Liens externes, mais aussi internes

Limitations de l'approche

- Orientation des liens ignorée
  - Systèmes à relations asymétriques
  - Twitter : followers / followees
  - Seuils différents pour 4 mesures (orientées)
- Imprécision du coefficient de participation
  - Degré, nombre de communautés, distribution des liens
  - Liens externes, mais aussi internes



$P = 0.58$

Connectivité externe

- Restriction aux communautés externes
- 3 aspects distincts considérés :
  - **Diversité** D
    - $\epsilon(u)$: nombre de communautés externes
    - $D(u)$: z-score de ϵ
  - **Intensité externe** I_{ext}
    - k_{ext} : nombre de liens externes
    - $I_{ext}(u)$: z-score de k_{ext}
  - **Hétérogénéité** H
    - Dispersion des liens externes
    - $\lambda(u)$: écart type de k_i
    - $H(u)$: z-score de λ

Connectivité externe

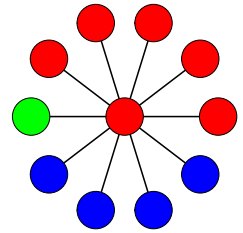
- Restriction aux communautés externes
- 3 aspects distincts considérés :
  - **Diversité** D
    - $\epsilon(u)$: nombre de communautés externes
    - $D(u)$: z-score de ϵ
  - **Intensité externe** I_{ext}
    - k_{ext} : nombre de liens externes
    - $I_{ext}(u)$: z-score de k_{ext}
  - **Hétérogénéité** H
    - Dispersion des liens externes
    - $\lambda(u)$: écart type de k_i
    - $H(u)$: z-score de λ

Connectivité externe

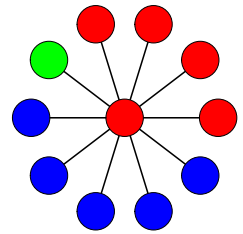
- Restriction aux communautés externes
- 3 aspects distincts considérés :
  - **Diversité** D
    - $\epsilon(u)$: nombre de communautés externes
    - $D(u)$: z-score de ϵ
  - **Intensité externe** I_{ext}
    - k_{ext} : nombre de liens externes
    - $I_{ext}(u)$: z-score de k_{ext}
  - **Hétérogénéité** H
    - Dispersion des liens externes
    - $\lambda(u)$: écart type de k_i
    - $H(u)$: z-score de λ

Connectivité externe

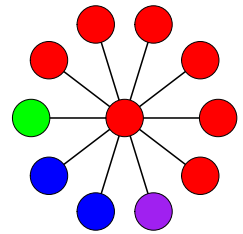
- Restriction aux communautés externes
- 3 aspects distincts considérés :
  - **Diversité** D
    - $\epsilon(u)$: nombre de communautés externes
    - $D(u)$: z-score de ϵ
  - **Intensité externe** I_{ext}
    - k_{ext} : nombre de liens externes
    - $I_{ext}(u)$: z-score de k_{ext}
  - **Hétérogénéité** H
    - Dispersion des liens externes
    - $\lambda(u)$: écart type de k_i
    - $H(u)$: z-score de λ



$\epsilon = 2, k_{ext} = 5, \lambda = 1.5$



$\epsilon = 2, k_{ext} = 6, \lambda = 2$



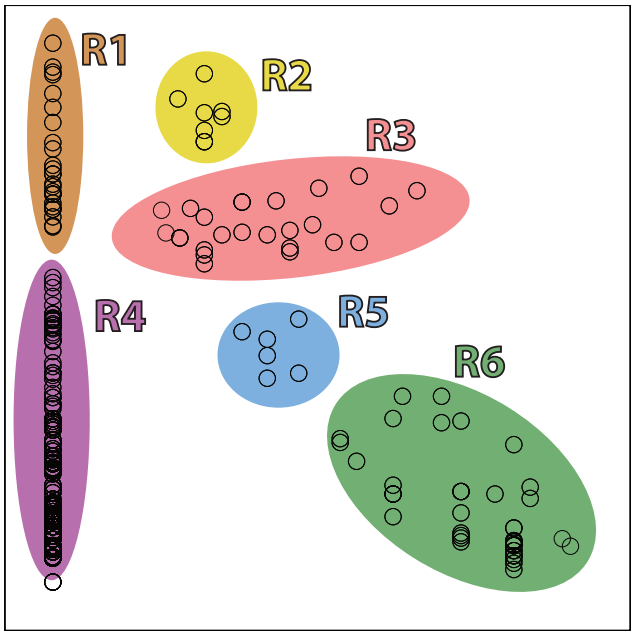
$\epsilon = 3, k_{ext} = 4, \lambda = 0.5$

Identification non-supervisée des rôles

- Difficultés à traiter
  - Traitement des nombreuses mesures
  - Mesures sans bornes fixées
  - Variabilité des données
  - Certains rôles non remplis
- Analyse de regroupement (clustering)
  - Appliquée à toutes les mesures simultanément
  - Chaque groupe obtenu correspond à un rôle

Identification non-supervisée des rôles

- Difficultés à traiter
  - Traitement des nombreuses mesures
  - Mesures sans bornes fixées
  - Variabilité des données
  - Certains rôles non remplis
- Analyse de regroupement (clustering)
  - Appliquée à toutes les mesures simultanément
  - Chaque groupe obtenu correspond à un rôle



Méthodologie

- Détection de communautés : Louvain
- Calcul des 8 mesures
- Analyse de regroupement : k-moyennes distribué [Lia09]
- Sélection des groupes : indice Davies-Bouldin $\rightarrow$ méthode du coude

Propriétés des groupes

Groupe	Taille	Proportion	Rôle
1	24543667	$46,68\%$	Hub ultra-périphérique
2	304	$< 0,01\%$	Hub très connecteur (entrant)
3	303674	$0,58\%$	Hub connecteur
4	11929722	$22,69\%$	Non-Hub périphérique (entrant)
5	10828599	$20,59\%$	Non-Hub périphérique (sortant)
6	4973717	$9,46\%$	Non-Hub connecteur

Taille des groupes

Répartition dans les groupes

Ratio	G1 NU	G2 HTC	G3 HC	G4 NP(E)	G5 NP(S)	G6 NC
< 1	$0,03\%$	$0,00\%$	$\mathbf{14,64\%}$	$11,53\%$	$\mathbf{13,65\%}$	$\mathbf{60,15\%}$
	$< 0,01\%$	$0,00\%$	$\mathbf{4,29\%}$	$0,09\%$	$0,11\%$	$1,07\%$
> 1	$0,03\%$	$0,00\%$	$\mathbf{19,38\%}$	$0,48\%$	$\mathbf{14,07\%}$	$\mathbf{66,05\%}$
	$< 0,01\%$	$0,00\%$	$\mathbf{7,31\%}$	$< 0,01\%$	$0,14\%$	$1,52\%$

Capitalistes de faible degré

Ratio	G1	G2	G3	G4	G5	G6
$< 0,7$	$0,00\%$	$10,43\%$	$\mathbf{81,67\%}$	$0,00\%$	$0,00\%$	$7,90\%$
	$0,00\%$	$\mathbf{31,25\%}$	$0,24\%$	$0,00\%$	$0,00\%$	$< 0,01\%$
$> 0,7$ et < 1	$0,00\%$	$1,52\%$	$\mathbf{95,72\%}$	$0,00\%$	$0,00\%$	$2,76\%$
	$0,00\%$	$7,24\%$	$0,46\%$	$0,00\%$	$0,00\%$	$< 0,01\%$
> 1	$0,00\%$	$0,03\%$	$\mathbf{98,02\%}$	$0,00\%$	$0,00\%$	$1,96\%$
	$0,00\%$	$0,33\%$	$1,24\%$	$0,00\%$	$0,00\%$	$< 0,01\%$

Capitalistes de degré élevé

Répartition dans les groupes

Ratio	G1 NU	G2 HTC	G3 HC	G4 NP(E)	G5 NP(S)	G6 NC
< 1	$0,03\%$	$0,00\%$	**$14,64\%$**	$11,53\%$	**$13,65\%$**	**$60,15\%$**
	$< 0,01\%$	$0,00\%$	**$4,29\%$**	$0,09\%$	$0,11\%$	$1,07\%$
> 1	$0,03\%$	$0,00\%$	**$19,38\%$**	$0,48\%$	**$14,07\%$**	**$66,05\%$**
	$< 0,01\%$	$0,00\%$	**$7,31\%$**	$< 0,01\%$	$0,14\%$	$1,52\%$

Capitalistes de faible degré

Ratio	G1	G2	G3	G4	G5	G6
$< 0,7$	$0,00\%$	$10,43\%$	**$81,67\%$**	$0,00\%$	$0,00\%$	$7,90\%$
	$0,00\%$	**$31,25\%$**	$0,24\%$	$0,00\%$	$0,00\%$	$< 0,01\%$
$> 0,7$ et < 1	$0,00\%$	$1,52\%$	**$95,72\%$**	$0,00\%$	$0,00\%$	$2,76\%$
	$0,00\%$	$7,24\%$	$0,46\%$	$0,00\%$	$0,00\%$	$< 0,01\%$
> 1	$0,00\%$	$0,03\%$	**$98,02\%$**	$0,00\%$	$0,00\%$	$1,96\%$
	$0,00\%$	$0,33\%$	$1,24\%$	$0,00\%$	$0,00\%$	$< 0,01\%$

Capitalistes de degré élevé

Observation sur le positionnement des capitalistes sociaux

Présence dans des groupes bien spécifiques

- Hubs : G2 et G3
- Connecteurs : G3 et G6
- Très connecteurs : G2 $\rightarrow$ capitalistes sociaux passifs

- Confirmation mode passif : 31% de G2 (mesures entrantes élevées)
- Connecteurs et très connecteurs : Connectés au reste du graphe
- Diversité sortante élevée (G3, G6) $\rightarrow$ suivre de nombreuses communautés

Conclusion

- Contributions
  - Détection des capitalistes sociaux
  - Etude de leur évolution : Efficacité de *IFYFM*
  - Etude d'un hashtag dédié
  - Mise en place d'un compte automatisé
  - Position des capitalistes sociaux au sein des communautés
- Perspectives
  - Visibilité $\rightarrow$ Influence ?
  - Méthode de détection moins arbitraire : machine learning
  - Structure communautaire des capitalistes sociaux ?
  - Cycle de vie d'un capitaliste social

Références

[DB79] David Davies and Donald Bouldin.

A cluster separation measure.

IEEE Transactions on Pattern Analysis and Machine Intelligence, 1(2):224–227, 1979.

[GA05] R. Guimerà and L. Amaral.

Functional cartography of complex metabolic networks.

Nature, 433:895–900, 2005.

[GVK+12] Saptarshi Ghosh, Bimal Viswanath, Farshad Kooti, Naveen Sharma, Gautam Korlam, Fabricio Benevenuto, Niloy Ganguly, and Krishna Gummadi.

Understanding and combating link farming in the twitter social network.

In *21st International Conference on WWW*, pages 61–70, 2012.

Références

[KLPM10] Haewoon Kwak, Changhyun Lee, Hosung Park, and Sue Moon.
What is Twitter, a social network or a news media?
In *Proc. of the 19th int. conference on World wide web*, WWW '10, pages 591–600, 2010.

[LB12] Vincent Labatut and Jean-Michel Balasque.
Detection and interpretation of communities in complex networks: Methods and practical application.
In Ajith Abraham and Aboul-Ella Hassanien, editors, *Computational Social Networks: Tools, Perspectives and Applications*, chapter 4, pages 81–113. Springer, 2012.

[Lia09] Wei-Keng Liao.
Parallel k-means data clustering, Oct 2009.