

## Socio-semantic dynamics in a blog network

**Jean-Philippe Cointet**

*CREA (CNRS/EP, France)*

AND

**Camille Roth**

*CAMS (CNRS/EHESS, France)*

IEEE SOCIALCOM09, VANCOUVER, BC – AUG 29–31, 2009

# A SOCIAL network

## Three kinds of links for each blog...

- **citation**: post citation links
- **interaction**: comment links
- **affiliation**: blogroll links

## ...where contents circulate

- in terms of topics ( $\mathcal{W}$ )
- in terms of cultural items ( $\mathcal{U}$ )

## Dataset: US blogosphere

- **scope**: 4 months of '08 campaign
- **network**: citations
- **nodes**: 1, 066 blogs (RTGI)

The screenshot shows a blog post on **versac.net** dated 29 June 2007. The main text discusses the denunciation of Nicolas Sarkozy by Nicolas Sarkozy. Annotations include:

- citation link**: A dashed line points from the text "Mouahaha" to the word "Mouahaha" in the post content.
- comment link**: A dashed line points from the text "Mouahaha" to a comment on the page that says "Mouahaha" and "le 29 Juin 2007 à 11:02".
- blogroll link**: A dashed line points from the text "Blog Roll" in the sidebar to a list of other blogs including "W Lambert", "Bernard Galabré", "Antipathique", and "Big Bang Blog".

The page layout includes a header with the site name, a main content area with a post, a sidebar with "LES NOTES RÉCENTES", "LES COMMENTAIRES RÉCENTS", and "PDR INTÉRIEUR", and a footer with "ET SUR PUBLIUS.FR".



# A socio-SEMANTIC network

## Three kinds of links for each blog...

- **citation:** post citation links
- **interaction:** comment links
- **affiliation:** blogroll links

## ...where contents circulate

- in terms of topics ( $\mathcal{W}$ )
- in terms of cultural items ( $\mathcal{U}$ )

## Dataset: US blogosphere

- **scope:** 4 months of '08 campaign
- **network:** citations
- **nodes:** 1,066 blogs (RTGI)

## Setting the Record Straight: More from our Canadian Cousins

Written by Robert Justin Lipkin on August 28th, 2009

Listen to some more. Finally, Canadians are fighting the slanderous characterizations of the Canadian health care system by U.S. insurance companies and other ideologues. As stated, medical care is a human right, not an economic product to be evaluated simply in terms of the bottom line. We, Americans, would do well to heed the advice of our northern cousins. But we also need more help from Canada in setting the record straight about its own system and pointing out the

systematic attempt on the part of those Americans opposing health insurance reform to distort, obscure, and simply lie about the Canadian system. Click [here](#) for more.

$\mathcal{U}$

$\mathcal{W}$

## semantic characterization

- “relevant” syntagms  
(“health insurance”, “climate change”, “national security”, “super Tuesday”, “human rights”...)
- urls: “www.youtube.com/x1hqwkeac”, etc.

# A DYNAMIC socio-semantic network

## Three kinds of links for each blog...

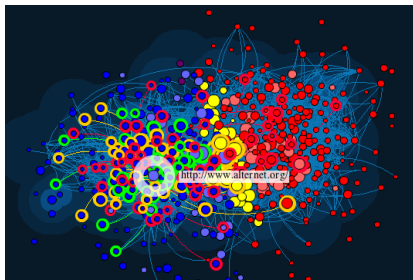
- **citation:** post citation links
- **interaction:** comment links
- **affiliation:** blogroll links

## ...where contents circulate

- in terms of topics ( $\mathcal{W}$ )
- in terms of cultural items ( $\mathcal{U}$ )

## Dataset: US blogosphere

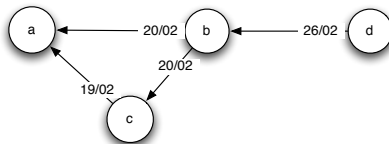
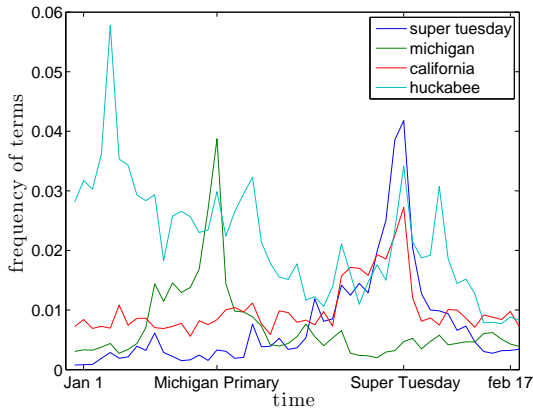
- **scope:** 4 months of '08 campaign
- **network:** citations
- **nodes:** 1,066 blogs (RTGI)



<http://presidentialwatch08.com/>



# Socio-semantic configuration



# Socio-semantic configuration

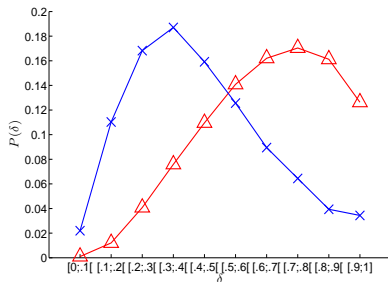
semantic profile of a blog  $i$ :

$$\hat{\mathbf{W}}_i(w) := \frac{\mathbf{W}_i(w)}{\sum_{w=1}^{|\mathcal{W}|} \mathbf{W}_i(w)} \cdot \log \frac{|\mathcal{B}|}{|\{j, \mathbf{W}_j(w) > 0\}|}$$

## semantic distance

between blogs  $i$  and  $j$ :

$$\delta(i, j) = 1 - \frac{\hat{\mathbf{W}}_i \cdot \hat{\mathbf{W}}_j}{\|\hat{\mathbf{W}}_i\| \|\hat{\mathbf{W}}_j\|}$$



**Semantic distance distributions.** *Triangles:* computed over the whole set of possible blog pairs. *Crosses:* distribution computed on linked blogs.

# Computing link creation propensity

→ estimate the “propensity of interaction”  
...that it is more or less likely for a node (or  
a dyad) with property “ $m$ ” to receive a link  
...which may be simply estimated by:

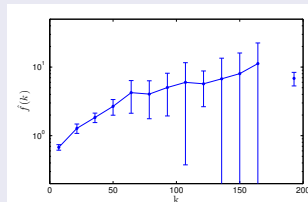
$$\hat{f}(m) = \frac{\nu(m)}{N(m)}$$

- $\nu(m)$  = number of links pointing towards an agent of type  $m$  (resp. number of new dyads of type  $m$ ) during a time period,
- $N(m)$  = number of agents (resp. of dyads) of type  $m$ .

# Computing link creation propensity

→ estimate the “propensity of interaction”  
 ...that it is more or less likely for a node (or  
 a dyad) with property “ $m$ ” to receive a link  
 ...which may be simply estimated by:

$$\hat{f}(m) = \frac{\nu(m)}{N(m)}$$



- $\nu(m)$  = number of links pointing towards an agent of type  $m$  (resp. number of new dyads of type  $m$ ) during a time period,
- $N(m)$  = number of agents (resp. of dyads) of type  $m$ .



# Dynamics of the social network

## in-degree effects

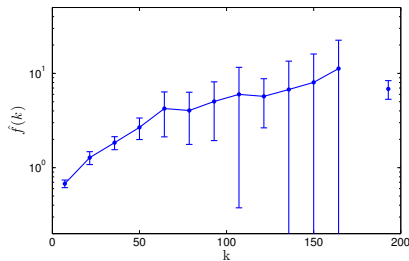
→ increasing, plateauing

## topological distance effects

→ strong trend to repetition and local interaction

## semantic distance

→ strong trend to homophily



# Dynamics of the social network

## in-degree effects

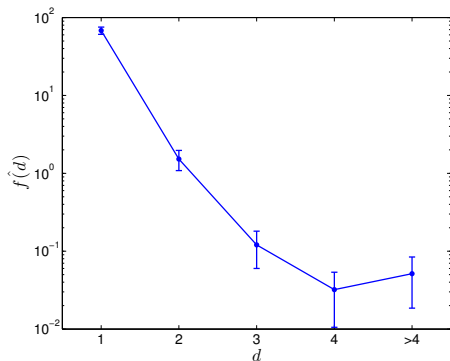
→ increasing, plateauing

## topological distance effects

→ strong trend to repetition and local interaction

## semantic distance

→ strong trend to homophily



# Dynamics of the social network

## in-degree effects

→ increasing, plateauing

## topological distance effects

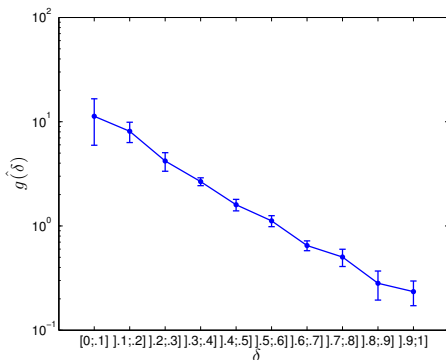
→ strong trend to repetition and local interaction

## semantic distance

→ strong trend to homophily

## primarily "social"?

⊗ social distance and degree



# Dynamics of the social network

## in-degree effects

→ increasing, plateauing

## topological distance effects

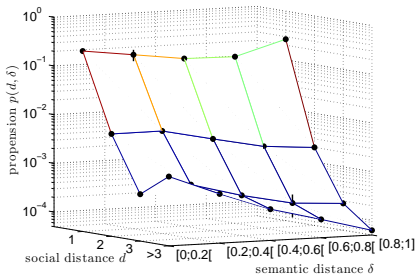
→ strong trend to repetition and local interaction

## semantic distance

→ strong trend to homophily

## primarily "social"?

● social distance and degree



# Dynamics of the social network

## in-degree effects

→ increasing, plateauing

## topological distance effects

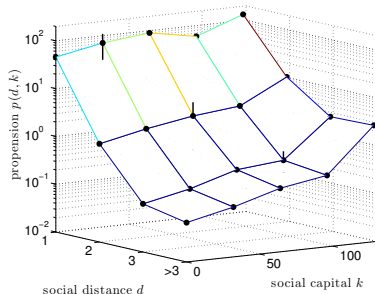
→ strong trend to repetition and local interaction

## semantic distance

→ strong trend to homophily

## primarily “social”?

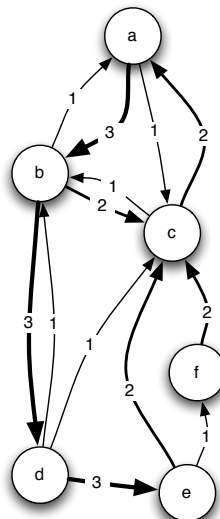
- social distance and degree



# Information flows: measures on the post network

## Dyadic measures:

- raw, weighted network, aggregated on 4 months
- attentional matrix a...  
→ and total attention  
 $\alpha_a = 5/6$
- detachment matrix



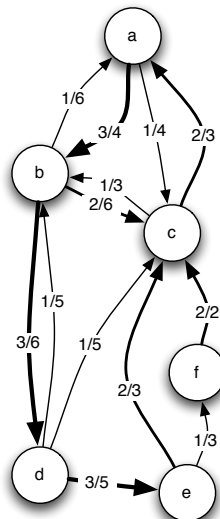
# Information flows: measures on the post network

## Dyadic measures:

- raw, weighted network, aggregated on 4 months
- attentional matrix  $\mathbf{a}$ ...  
→ and total attention  
 $\alpha_a = 5/6$
- detachment matrix

## “edge range”:

- quantifying shortcuts



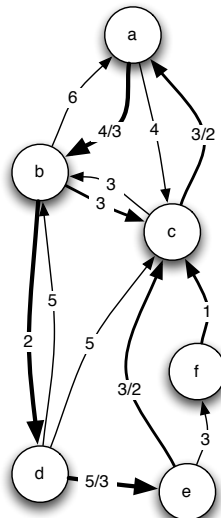
# Information flows: measures on the post network

## Dyadic measures:

- raw, weighted network, aggregated on 4 months
- attentional matrix  $\mathbf{a}$ ...  
→ and total attention  
 $\alpha_a = 5/6$
- detachment matrix

## "edge range":

- quantifying shortcuts





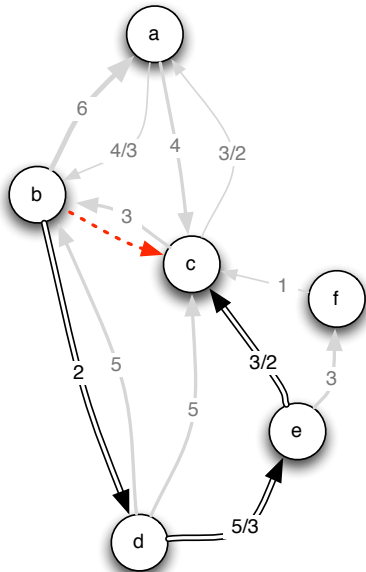
# Information flows: measures on the post network

## Dyadic measures:

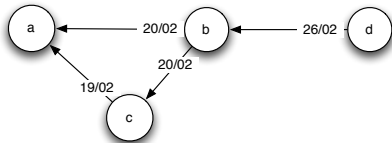
- raw, weighted network, aggregated on 4 months
- attentional matrix  $\mathbf{a}$ ...  
→ and total attention  
 $\alpha_a = 5/6$
- detachment matrix

## "edge range":

- quantifying **shortcuts**

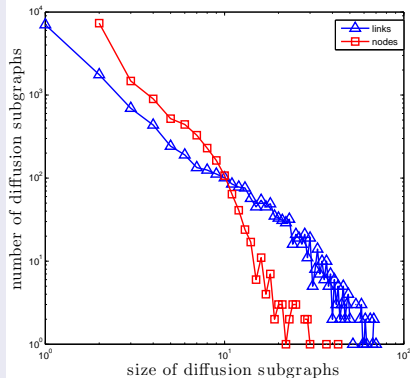


# Information cascade



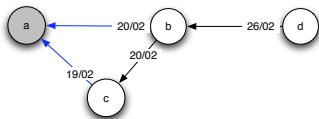
An example of diffusion subgraph, a common "resource" and a set of citation links between blogs

## Diffusion subgraphs



⇒ heterogeneous cascade sizes

# An ego-centered perspective

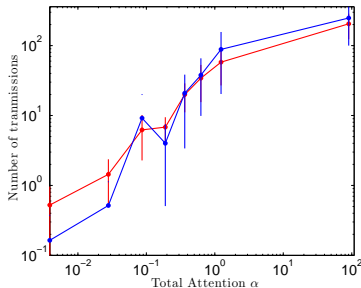


We focus on the total number of "transmissions" generated by blogs with a given total attention  $\alpha$

a bit more "global"...

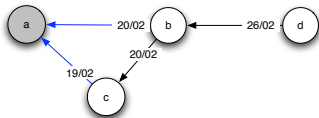
**second transmissions:** we focus on "later transmissions", i.e. after a first transmission event

role of the *total attention* on the number of *diffusion links*



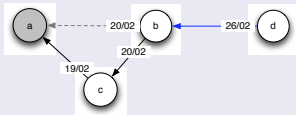
Larger *active readership*  $\Rightarrow$  larger number of diffusion links, **yet not linearly**

# An ego-centered perspective



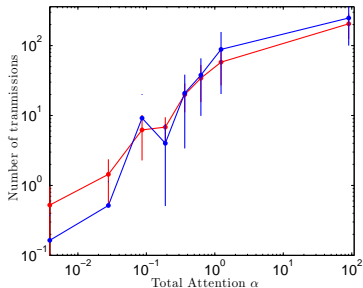
We focus on the total number of "transmissions" generated by blogs with a given total attention  $\alpha$

## a bit more "global"...



**second transmissions:** we focus on "later transmissions", i.e. after a first transmission event

## role of the *total attention* on the number of *diffusion links*

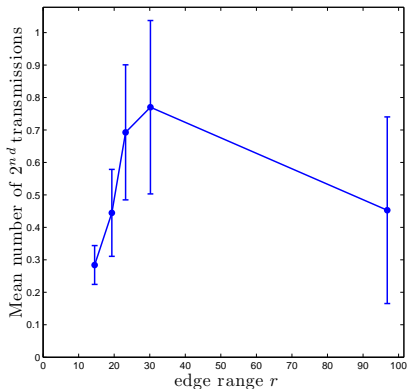
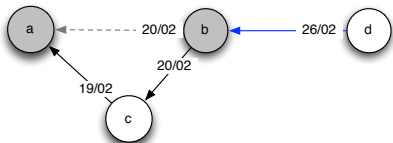


Larger *active readership*  $\Rightarrow$  larger number of diffusion links, **yet not linearly**

# A more global perspective

→ role of *edge-range* on the number of grand-children

We focus again on transmissions occurring after a first transmission event



An information which has been transmitted through a "*median*" link generates a larger number of grandchildren

## Concluding remarks

### Co-evolution of content and relationships

- Patterns *not necessarily linked to authority only*
- Patterns *not necessarily ego-centered only*
  - **divergent from the “neighbor-based-influence” perspective**

# Concluding remarks

## Co-evolution of content and relationships

- Patterns *not necessarily linked to authority only*
- Patterns *not necessarily ego-centered only*
  - divergent from the “neighbor-based-influence” perspective

## Thank you!

cointet@poly.polytechnique.fr & roth@ehess.fr