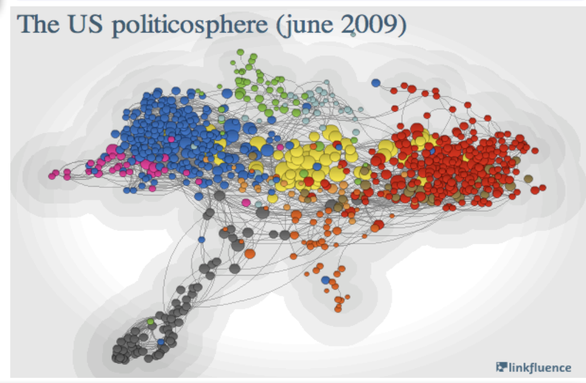
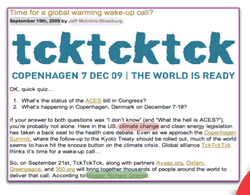
Local Networks, Local Topics: Structural and Semantic Proximity in Blogspace

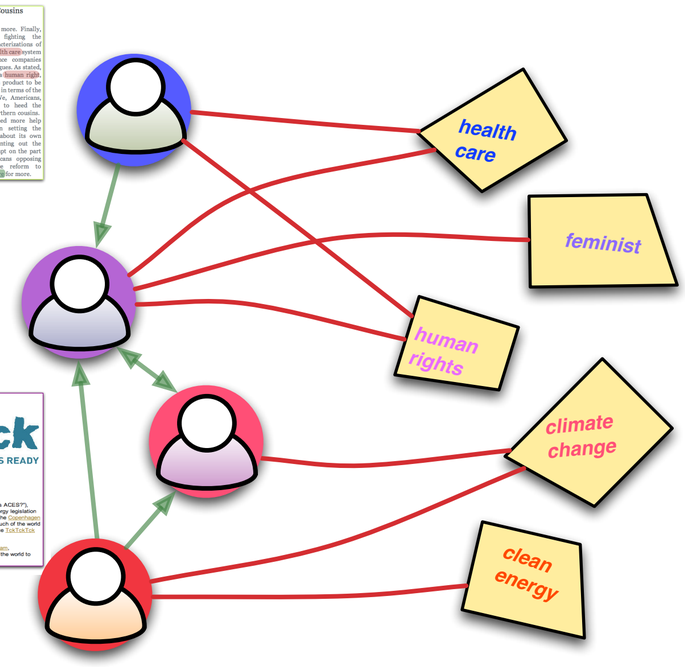
Jean-Philippe Cointet
Camille Roth

Blog networks are often described as "small world" social networks where individuals would potentially be topologically close to most of the other actors. On the other hand, links would be primarily created towards similar-minded individuals and well-connected bloggers, suggesting a "balkanization" of Internet communities.
Examining a portion of the US blogosphere on several months, we show that bloggers relate to each other essentially in a local fashion, overwhelmingly and preferentially establishing links towards a limited neighborhood rather than the whole network. Furthermore, while long-distance interactions may indeed be dominated by homophily and authority effects, we show that close neighborhoods feature significantly more horizontal and diversified interactions — thereby challenging the conjecture of a widespread balkanization of Internet communities. We shed further light on this issue by describing the dual evolution of social and semantic proximity before and after two individuals interact with each other. We discuss in particular whether interactions are preceded or followed by a structural "contraction" and/or by an increasing similarity of the surrounding local social network.

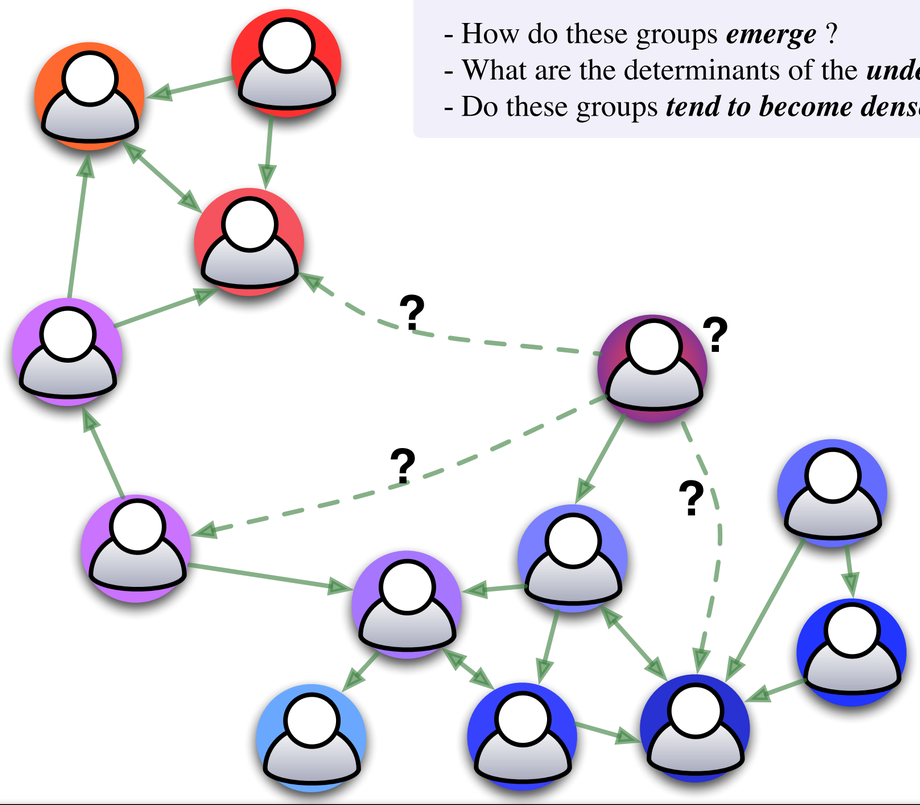


A socio-semantic perspective



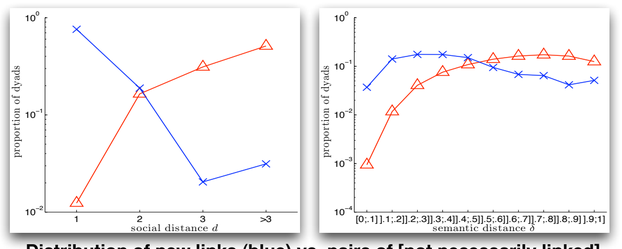


A ***socio-semantic*** description of blog networks dynamics.
Structural features matter as well as content distribution.



Citations links exhibit **tightly inter-connected groups of semantically similar blogs**.

- How do these groups ***emerge*** ?
- What are the determinants of the ***underlying link creation dynamics***?
- Do these groups ***tend to become denser and topically specialized***?

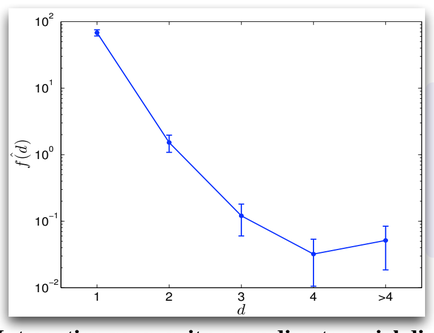


Distribution of new links (blue) vs. pairs of [not necessarily linked] blogs (red) at a given structural (left) and semantic distance (right)

Interactions are generally ***local,*** happening at distance max 2 rather than beyond (mainly ***repeated interactions***).
Moreover, ***topically similar blogs*** are overwhelmingly more likely to interact.

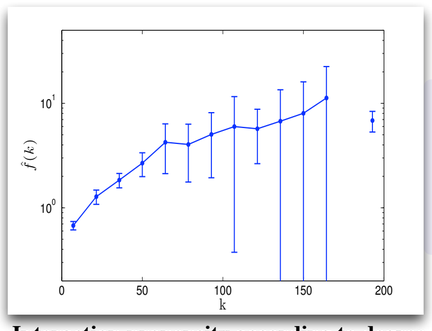
Questioning the "Web balkanization"

Determinants of new link creation



Interaction propensity according to social distance

New citations are mainly directed towards ***topologically close*** nodes

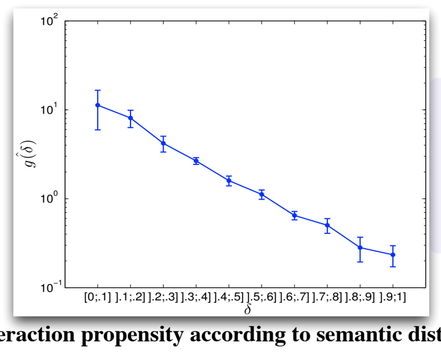


Interaction propensity according to degree

Classical degree-based preferential attachment

Propensity measures on citation networks

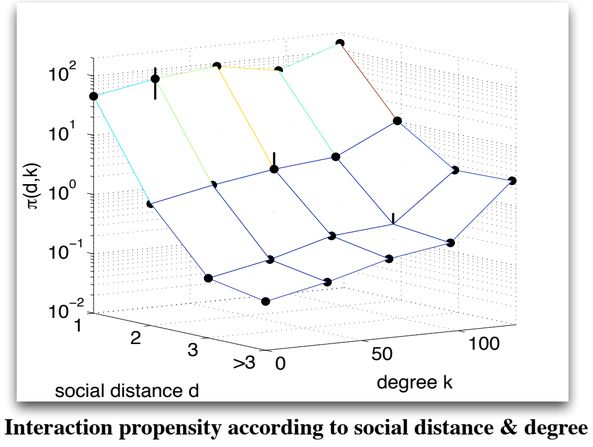
We compute the ***likeliness for two actors to interact*** according to some given social or semantic profiles.
On a blog citation network, we observe a strong tendency to ***homophily*** and ***interaction repetition***.



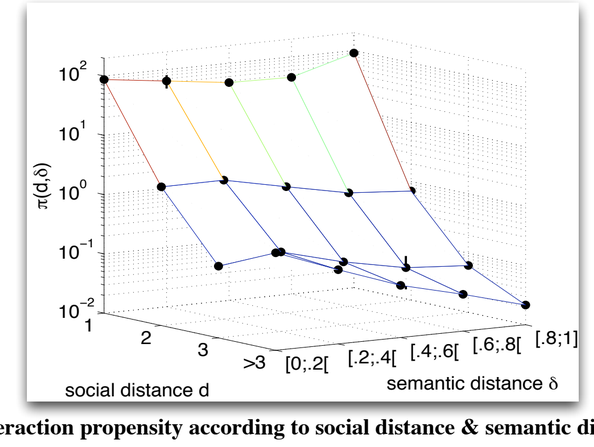
Interaction propensity according to semantic distance

Similar blogs tend to cite each other more frequently

Integrating social and semantic features:
The behavior of new link creation is however ***very different*** if we distinguish interaction between close nodes ("*friend-of-friend*" *exploration*) from those towards distant nodes ("*search engine*" *exploration*)
Within the local arena, both homophily and "rich-get-richer" effects do not hold anymore: interaction propension is not anymore biased towards more popular blogs, and, surprisingly, it also seems to be biased towards both similar and topically different blogs.



Interaction propensity according to social distance & degree



Interaction propensity according to social distance & semantic distance

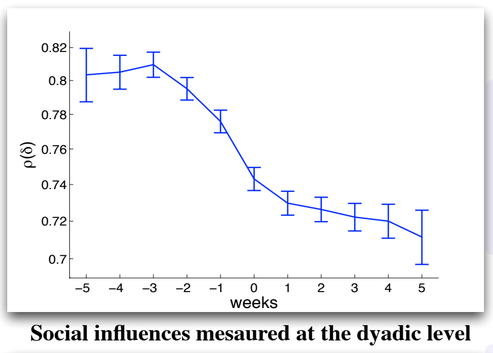
Structural and semantic contraction

Semantic contraction

To what extent does a blogger's content publication activity evolve according to her/his neighborhood ?

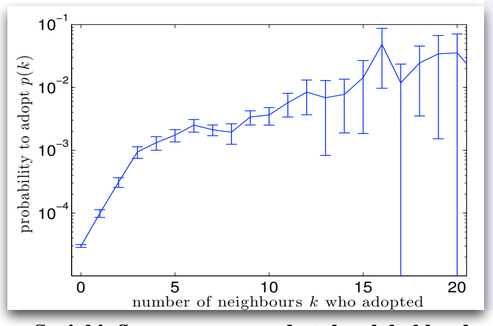
We show, at a *dyadic level*, that ***two bloggers who are going to interact*** *get* ***semantically closer both before and after*** *the very moment of* ***interaction***.

We also observe, at a more *global level,* that the whole neighborhood of a given blogger is largely influencing her content production behavior.



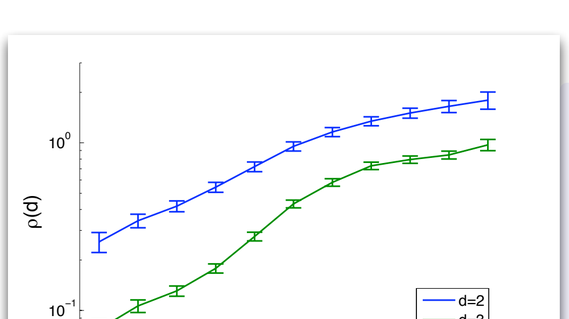
Social influences mesaured at the dyadic level

Two blogs are both getting closer *before* and *after* interaction event



Social influences measured at the global level

Probablity of adoption exponentially grows, then plateaus, with respect to the number of former adopters within its neighborhood



Evolution of the normalized number of different paths of length *d* between two blogs interacting for the first time at *t=0*

The normalized number of paths of length *2* and *3* is growing *both* before and after interaction

Structural contraction

The ***local neighborhood of a pair of bloggers*** who interact for the first time is ***subjected to a structural contraction*** phenomenon, similar to the above-mentioned semantic alignment.

webfluence
web: http://webfluence.csregistry.org
This work has been partially supported by the French ANR through grant "Webfluence" ANR-08-SYSC-009.
ICWSM-10