Dynamics on blogs network

Intertemporal topic correlations in weblogs and news websites

Jean-Philippe Cointet*, Camille Roth**, Emmanuel Faure*

*CREA, CNRS/Ecole Polytechnique, Paris, France
**CRESS, Department of Sociology, University of Surrey, Guildford, UK

July 2007 — SFI, USA

Social System in vivo

French political blogosphere

- In the context of the french presidential elections,

- stigmergic media,

- relatively autonomous system.

Social System in vivo

French political blogosphere

- In the context of the french presidential elections,
- stigmergic media,
- relatively autonomous system.

Data collection

- a socio-semantic network,
- retrieve the system dynamics,
- by adopting the blogger viewpoint.

Dataset collection

Snowball

1. from a given seed "http://versac.net",

2. we follow its blogroll links,

3. and select every active political blogs,
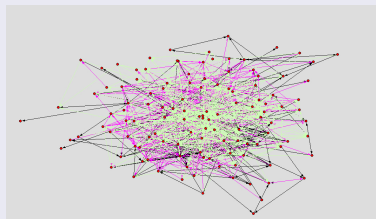
4. repeat the process.

The selection

- 123 blogs

- crawled over 5 months, from 1/01 to 31/05

The dataset

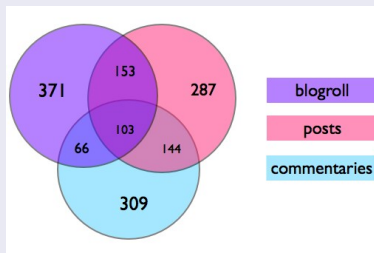## Structure

Three kinds of links:

- blogroll
- post
- comments

The dataset

Structure

Three kinds of links:
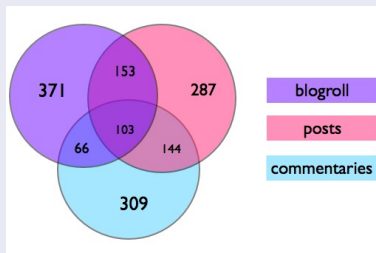
- blogroll
- post
- comments

The dataset

Structure

Three kinds of links:

- blogroll
- post
- comments

Semantics

the content of each post is collected and indexed with classical linguistic treatments.

Dynamical features

Structure

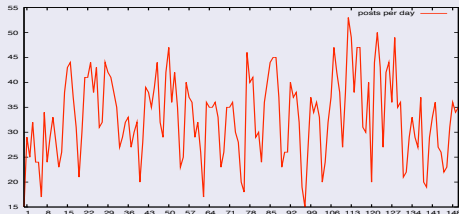posts and comments are dated, thus providing a dynamical network.

Dynamical features

Structure

posts and comments are dated, thus providing a dynamical network.

Semantics
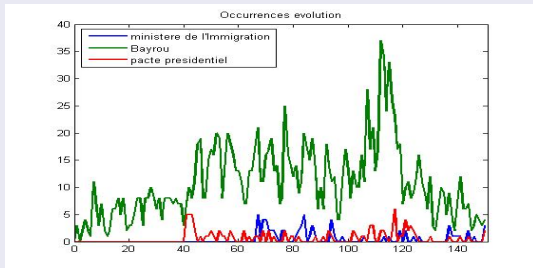
- posting activity evolution

Dynamical features

Structure

posts and comments are dated, thus providing a dynamical network.

Semantics

- posting activity evolution
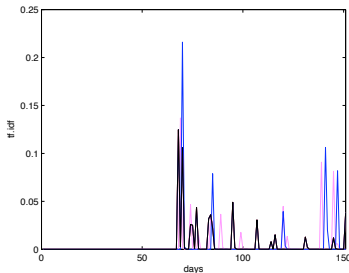- thematic occurrences evolution

Context

Intertemporal correlations

- We are interested in **generalized** patterns of **intertemporal** topic **correlations** between various **information sources**
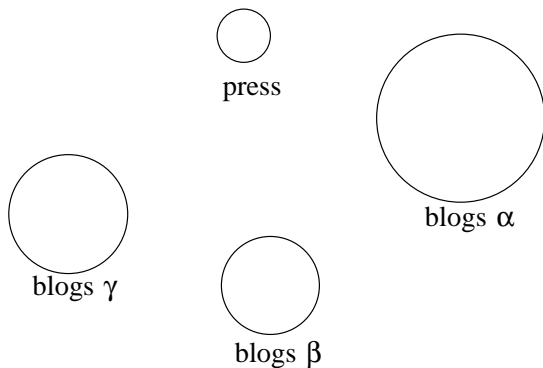
Global, macro-level viewpoint

We try to infer causal relationship between sources by creating a map of systematic topic correlations
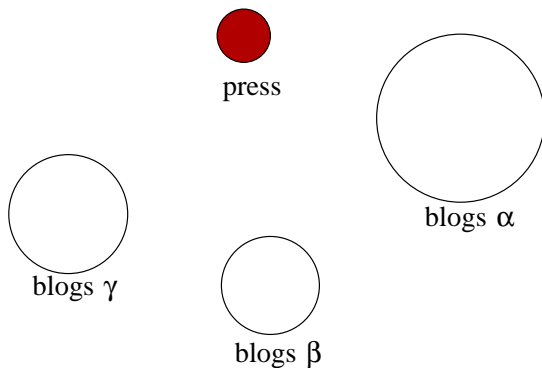

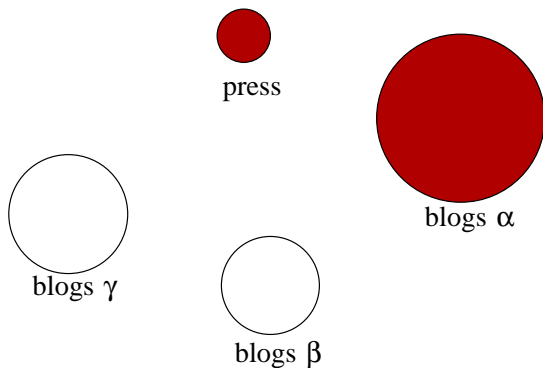occurrences of the topic "ministère de l'immigration" for UMP blogs (*blue*), PS (*pink*), UDF (*black*)
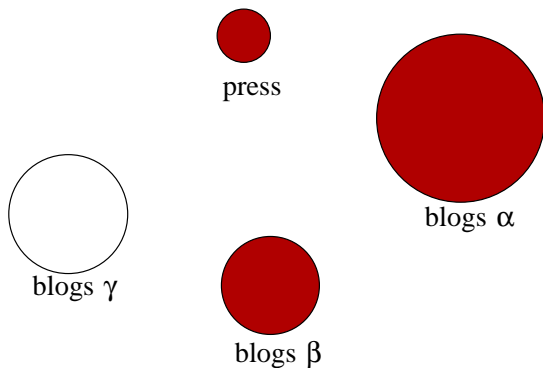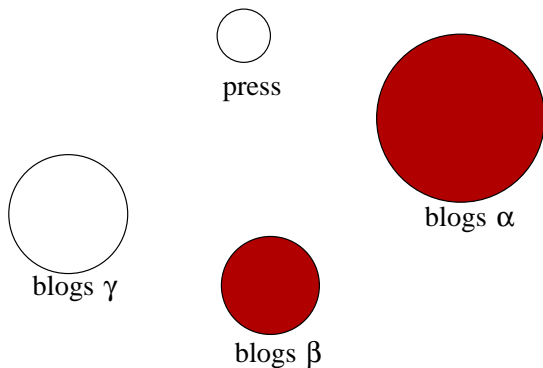
Example

Example

press

blogs α

blogs γ

blogs β

Example
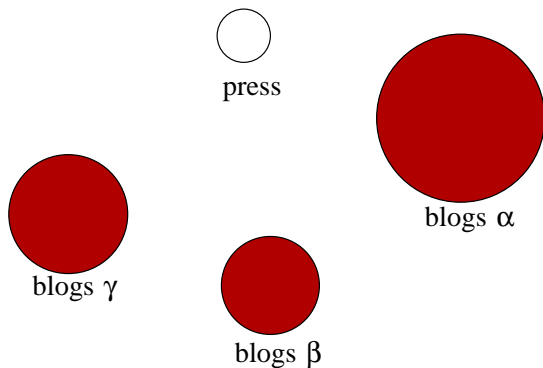
Example

Example
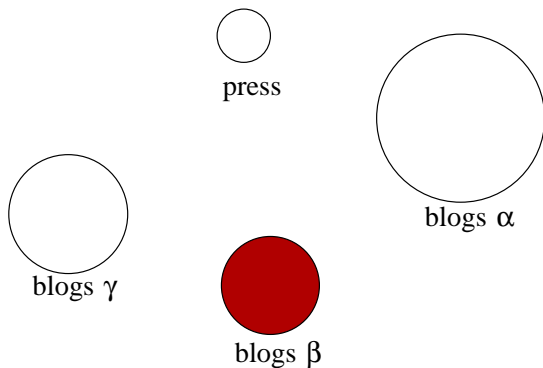
Example
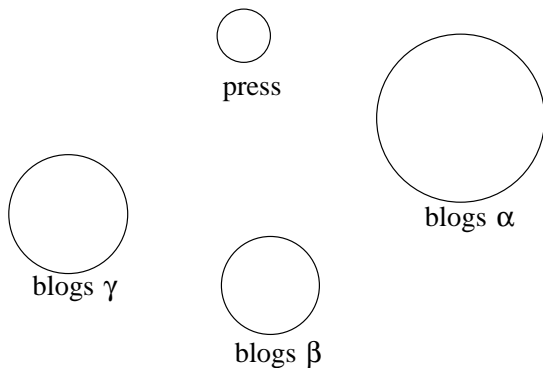
Example
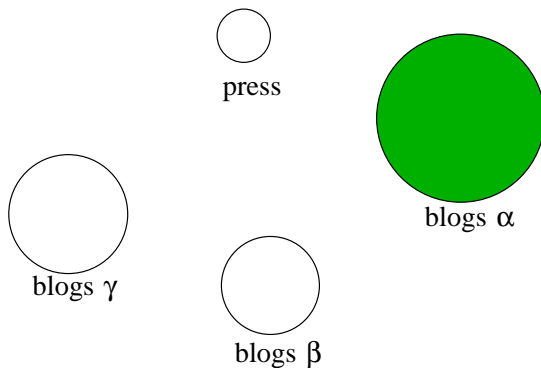
Example
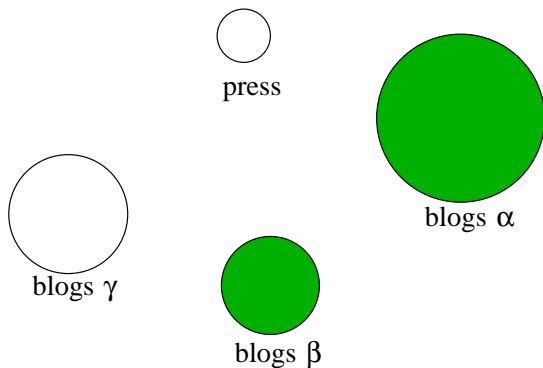
Example

Example

Example
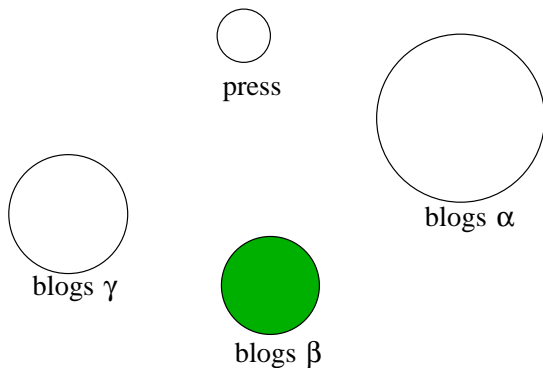
Causal-states model

A symbolic dynamics...

blogs α	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
blogs β	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
blogs γ	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
press	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1
alphabet	a	b	c	d	e	f	g	h	A	B	C	D	E	F	G	H

symbolic dynamics | ...A   H   H   f   **d   c   a**   e   F   f   b   c   b   d   e
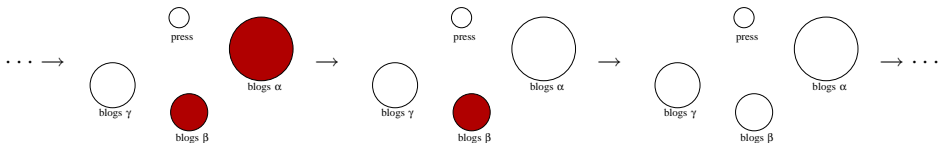
Causal-states models

Causal-state machine

(Crutchfield & Young, 1989; Shalizi, 2001)

- automatically inferring (variable-length) hidden states...
- ...made of equivalence classes of signal histories...
- ...along with transition probabilities.

Causal-states models

Causal-state machine

(Crutchfield & Young, 1989; Shalizi, 2001)

- automatically inferring (variable-length) hidden states...
- ...made of equivalence classes of signal histories...
- ...along with transition probabilities.

signal │ A H H a a a h H H a a a A H H...
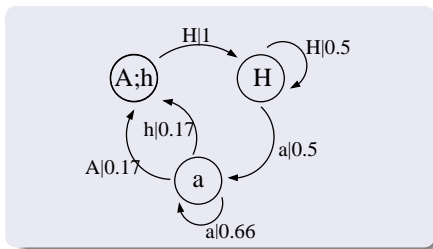
Data

Hand-made selection

- Sample of 33 very active political blogs, 6 press sources.

- Daily collection of posts during November 2006: presidential primary for the French *Parti Socialiste* (center-left).

- Selection of 75 (lemmatized) terms — this set makes our "topics".

Data

Hand-made selection

- Sample of 33 very active political blogs, 6 press sources.
- Daily collection of posts during November 2006: presidential primary for the French *Parti Socialiste* (center-left).
- Selection of 75 (lemmatized) terms — this set makes our "topics".

Practical matters

Creation of blog groups

Classical Salton (1975) categorization
Three groups: α, β, γ, plus the press
Roughly left-, right-, indep.-leaning (α, β, γ)

Data

Hand-made selection

- Sample of 33 very active political blogs, 6 press sources.
- Daily collection of posts during November 2006: presidential primary for the French *Parti Socialiste* (center-left).
- Selection of 75 (lemmatized) terms — this set makes our "topics".

Practical matters

Creation of blog groups

Classical Salton (1975) categorization
Three groups: α, β, γ, plus the press
Roughly left-, right-, indep.-leaning (α, β, γ)

Signal creation

For each term: evolution of occurrences in each blog group transformed into a signal vector.

$$(...A \quad B \quad d \quad c...)$$

Causal-state machine
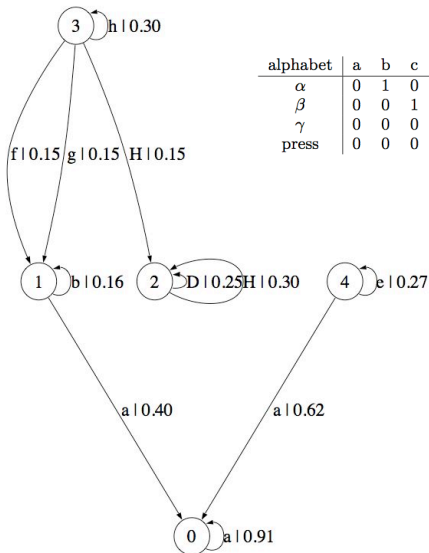
alphabet	a	b	c	d	e	f	g	h	A	B	C	D	E	F	G	H
α	0	1	0	1	0	1	0	1	0	1	0	1	0	1	0	1
β	0	0	1	1	0	0	1	1	0	0	1	1	0	0	1	1
γ	0	0	0	0	1	1	1	1	0	0	0	0	1	1	1	1
press	0	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1

S 0 : {a; G}
S 1 : {b; c; d; f; g; A; C; E; b}
S 2 : {B; D; F; H}
S 3 : {h}
S 4 : {e}

Perspectives

- Which correlation between these high-level causal relationships and the underlying networks?
- what about individual strategies?
- ...